Measuring Financial Risks: The Application of Network Theory in Fintech Risk Management

Branka Hadji Misheva

A thesis submitted in fulfillment of the requirements for the award of the degree of Doctor of Philosophy in Economics and Management of Technology at the University of Pavia, Italy.

Supervisor: Prof. Paolo Giudici

2019

Contents

Executive Summary

Issuance of loans by financial intermedieries, to other firms and individuals, is always associated with major risks. The failure of loan recipients to honor their obligation at the time of maturity leaves the investor vulnerable and affects their operations. The risk associated with such transactions is referred to as credit risk. It is well known that some percentage of these non-performing loans are eventually imputed to economic losses. To minimize such risk exposures, various methods have been extensively discussed in the credit risk literature to enable credit-issuing institutions undertake thorough assessment to classify loan applicants into risky and non-risky customers. Some of these methods range from logistic and linear probability models to decision trees, neural networks and support vector machines. A conventional individual-level reduced-form approach is the credit scoring model which attributes a score of credit-worthiness to each loan applicant based on the available history of their financial characteristics.

Recent advancements, gradually transforming the traditional economic and financial system, are mainly characterized with the emergence of digital-based systems. Such systems present a paradigm shift from traditional infrastructural systems to technological (digital) systems. Financial technological (**Fintech**) companies are gradually gaining ground in major developed economies across the world. The emergence of Peer-to-Peer (P2P) platforms is a typical example of a Fintech system. The P2P platform aims at facilitating credit services by connecting individual lenders with individual borrowers without the interference of traditional banks as intermediaries. Such platform serves as a digital financial market and an alternative to the traditional physical financial market. P2P platforms significantly improve the customer experience and the speed of the service and reduce costs to both individual borrowers and lenders as well as small business owners. Despite the various advantages, P2P systems inherit some of the challenges of traditional credit risk management. In addition, they are characterized by the inability to solve for asymmetric information as efficiently as banks and by differences in risk ownership which in turn might motivate them to push volume even in view of reduced credit standards. Finally, P2P systems note a strong interconnectedness among their users which makes distinguishing healthy and risky credit applicants difficult, thus affecting credit issuers. There is, therefore, a need to explore methods that can help improve credit scoring of individual or companies that engage in P2P credit services.

We argue that P2P platforms, through the use of non-traditional data sources as well as advance modelling, can offer a new approach on credit risk evaluation in the context of P2P systems. Specifically, we suggest that the use of alternative data that summarize the interconnections that emerge between borrowers could counterbalance the inherent risks of the business model and in turn lead to higher accuracy in risk classes assignment. Namely, P2P systems can benefit from the inclusion of information on the interconnections or similarities that emerge between different participants on the platform, i.e can benefit from the application of network theory in the credit risk evaluation.

Consequently, the overall objective of this thesis is to test the predictive accuracy of traditional credit scoring models as they are employed in the context of P2P systems and investigate whether the inclusion of network parameters i.e. information on how borrowers are connected, can improve the predictive accuracy of models.

In this work, we propose several approaches on how network theory can be employed to improve the statistical-based credit scoring for P2P systems and those are: (i) correlation-based credit scoring (in the case in which time-varying financial information on borrowers is available on the platform), (ii) similarity-based credit scoring (for cross-sectional data), and (iii) factor-network-based segmentation. Furthermore, the thesis also includes an application of network theory in improving Fintech risk management, in a context beyond Fintech credit. Specifically, we also provide an application of network theory in understanding the dynamics of Bitcoin blockchain trading volumes and, how different trading groups, in different geographic areas, interact with each other. The bitcoin network is a peer-to-peer payment network that operates on a cryptographic protocol hence it represents a natural extension to the study of P2P systems.

The empirical results presented in this thesis suggest that credit risk management of SMEs engaged in P2P credit services can be improved by employing network theory. Specifically, we demonstrate the effectiveness of our approach through empirical applications analyzing the probability of default of several different samples of SMEs involved in P2P lending across Europe. In each case, we compare the results from our network-augmented model with the one obtained with standard credit score methods and throughout we find that the network-based methodologies lead to an improvement in predictive accuracy. This finding further remains valid also in the context of alternative P2P systems i.e. the Bitcoin network. We find that our proposed network-based model for understanding the dynamics of Bitcoin blockchain trading volumes overperforms a pure autoregressive model.

Keywords: Fintech, Network models, Credit Scoring, P2P Systems

1 Introduction

1.1 Motivation

The finance industry has changed extensively over the course of the last two decades mostly due to technological advancements. Within the last decade, we have witnessed the introduction of new currencies, technologies, business models and forms of transactions and all of this has occurred within an environment of global economic instability. One of the most significant changes has been the emergence of **Fintech Technology (Fintech)** (see Arner, Barberist, and Buckley, 2016, Buchak et al., 2018, Nicoletti, 2017).

The Financial Stability Board (FSB, 2017) defines Financial Technology as *"technologically enabled financial innovation that could result in new business models, applications, processes, or products with an associated material effect on financial markets and institutions and the provision of financial services"*. In other words, the term Fintech covers all innovative technologies that help the provision of financial services (He et al., 1996).

While innovation in finance is not a new concept, the focus on technological innovations and its pace have increased significantly. Fintech solutions that make use of big data analytics, artificial intelligence and blockchain technologies are currently introduced at an unprecedented rate and are changing the nature of the financial industry (EC, 2018). Such innovations can significantly disrupt existing industry structures as well as blur sector boundaries, facilitate strategic disintermediation, reinvent how existing intermediaries and financial-service providers create and offer products and services (Philippon, 2017). Put differently, the emergence of Fintech has motivated innovations that are altering many financial products, services, production processes, and organizational structures.

Examples of innovations that are central to Fintech include: cryptocurrencies and the blockchain, new digital advisory and trading systems, artificial intelligence and machine learning, peer-to-peer lending, equity crowdfunding and mobile payment systems. A report by the Association of Supervisors of Banks of the Americas (Nino et al., 2017) provides a classification of Fintech products and services into the following categories and subcategories: (i) payments (ex. P2P transfers, digital and mobile wallets, mobile point of sale, digital and cryptocurrencies), (ii) investment management services (e-trading, high-frequency trading, copy trading, robo-advice), (iii) credit and capital raising (lending marketplaces, crowdfunding, online/mobile banking), and (iv) insuretech.

In the context of the payment industry, novel, technology-driven services have been introduced, that overcome some of the restrictions of the traditional payment systems, including large transfer fees, shorter client verification period, less bureaucracy to transfers, etc (Nino et al., 2017). Furthermore, Fintechs have also offered innovative investment and foreign exchange services that include: automatized investment strategy and advice (Bayon and Vega, 2018), investment simulation platforms, crypto-trading (Mannaro, Pinna, and Marchesi, 2017), etc.

Looking at the credit services, Fintech players have had one of the most relevant impacts, as their services have led to improved financial inclusion (Jagtiani and Lemieux, 2018, Sanford, 2013, Bourreau and Valletti, 2015). Fundamental advances in the internet, mobile communications, distributed computing, and information collection and processing has enabled online markets to provide an alternative to traditional financial intermediaries particularly in the consumer and commercial credit space where they have introduced many opportunities for both borrowers and lenders (Jagtiani and Lemieux, 2018). By relying extensively on customer data and advanced methodologies, P2P platforms can offer a new approach to credit risk evaluation, valid even in a context of very short credit history of clients, one that might not satisfy traditional lending requirements. In other words, the advances in digital technology allow P2P platforms to reduce the cost of credit and increase financial inclusion.

The advantages notwithstanding, Fintech solutions leave the door open for many challenges such as cyber-attacks, underestimation of creditworthiness, potential for fraud, compliance concerns, consumer and investor protection issues and disrupted market integrity, which represent central points of interest for regulators and supervisory bodies (Giudici, 2018). Across various regulatory reports, the risks associated with Fintech credit take central stage (ECB, 2009, EC, 2018, Naoko, David, and Bihong, 2019, Tijn et al., 2018). Namely, a key point of concern for regulators is related with the potential for bad debt accumulation arising from the P2P business models. Specifically, these online marketplaces receive revenues in proportion to the loan volume originated and yet the risk is fully borne by the investors. The platforms, therefore, face financial incentives to maximize loan origination even at the expense of credit standards which in turn can lead to bad debt accumulation (Tijn et al., 2018). This is particularly relevant in view of the significant growth that P2P platforms have experienced within the last two decades. Looking at the micro-level, the report by Cambridge Centre for Alternative Finance states that in 2016, 14,521 companies have raised close to 1.13 million euros by using online alternative finance platforms which in volumes represents a 110% annual growth against the previous year's total business funding.

In this context, a key issue becomes identifying whether the ratings assigned by P2P platforms are a good predictor of default and consequently, proposing an approach for improving the predictive accuracy of P2P scoring models by leveraging the advantages of these platforms relative to traditional financial intermediaries. Namely, P2P platforms have evolved into sophisticated networks of institutional investors and lenders which through the use of alternative data sources and advanced technologies (including big data, machine learning and other complex AI algorithms) could offer a new approach to credit risk evaluation, valid even in the context of very short credit history of clients, one that might not satisfy traditional lending requirements. The use of non-traditional data sources reflecting the interconnections that emerge between borrowers on the platform, could counterbalance the stated risks and in turn lead to accurate credit risk evaluation. In this work, we claim that P2P systems can benefit from the inclusion of information on the interconnections or similarities that emerge between different participants on the platform, i.e can benefits from the application of network theory in the credit risk evaluation.

The idea of modeling credit risk by taking into account the socio-economic network was first proposed by the work of Eisenberg and Noe (2001) for a financial system. This work considers the properties of inter corporate cash flows that are assumed to have cyclical interdependence amongst the players and default rates are determined endogenously by use of a clearing vector in the network (Ntwiga, 2016). Among the main reasons for arguing that interconnections emerging between borrowers can improve credit scoring and default rates estimations, are:

- no agent exists in a vacuum. All must interact within a network in order to achieve their goals (Elisabeth et al., 2008);
- most agents considering Fintech credit share financial and social properties and those similarities can be utilized for discriminating between different risk classes;
- P2P platforms are at a disadvantage concerning the volumes of data they can use for training and testing classification models (Namvar et al., 2017). By identifying community structures of connected and non-connected entities, they are able to account for the different networks that exist within their samples without splitting the data and drawing conclusion on very small sub-samples;
- the links and similarities that exist between agents can paint a different picture of a specific agent's credit-wordiness compared to an evaluation based on historical data. Consider a company that scores well based on historical data and yet is significantly positively correlated with a company that defaulted within the last period. The existence of this link should be taken into account in the credit scoring conducted by the platform.

The nature of the P2P business model, its flexibility to employing advanced methodologies and alternative data, and the highlighted points and rationals stated above, form the main motivation undertaking this research. We strongly believe that network models are crucial for understanding the issue of credit risk in the context of P2P systems.

1.2 Background

The financial stability board uses the term Fintech credit to broadly include all credit activity facilitated by electronic (online) platforms that are not operated by commercial banks (Tijn et al., 2018). This approach is consistent across most regulators' definitions. Depending on the jurisdiction, these platforms are referred to as "peer-to-peer (P2P) lenders", "loan-based crowdfunders" or "marketplace lenders". Throughout this thesis, the term "P2P platforms", "online lending markets", "P2P systems" are used interchangeably.

The literature identifies many factors which explain the increasing role of P2P lending platforms in the global world of finance. First and foremost, these online market players are using new digital technologies and more-granular consumer data which may lead to greater convenience, lower transactional costs (Pokorna and Sponer, 2016) and in turn, higher financial inclusion. Second, these platforms do not collect deposits, which in turn allows them to operate at lower costs as they can avoid many of the intermediation costs associated with traditional financial intermediaries (Serrano-Cinca and Gutiérrez-Nieto, 2016b, De Roure, Pelizzon, and Tasca, 2016). Furthermore, in most regions, Fintech legislation is lacking thus enabling P2P platforms to work with less administrative and regulatory burden.

From a different viewpoint, the rapid growth of P2P lending platforms can pose significant risks to financial stability. One of the main risks identified within the literature is the lack of comprehensive regulation for Fintech credit, which in turn suggests that investors are at a higher risk of losing the funds invested, compared to investing in traditional financial intermedieries for which deposit insurance is set in place (Naoko, David, and Bihong, 2019). Moreover, P2P are less able to solve for asymmetric information compared to banks as in most cases, banks have a monopoly on the financial history of their clients (Akerlof, 1970, Myers and Majluf, 1984, De Roure, Pelizzon, and Tasca, 2016). Finally, a main stability concern that emerges from the P2P business model is associated with the difference in risk ownership. Namely, in the context of the traditional financial intermediation, the financial institution that provides the score - is the same institution that takes on the credit risk hence it is in the bank's best interest to have the most accurate scoring possible. In the context of the P2P lending platform – the P2P platform assigns the score on the basis of which investors make their decision, and yet the credit risk is fully board by the investors, whereas the platform wants to push volumes. To summarise, as a combined result of lacking regulation, misaligned incentives, asymmetric information, differences in the business model and in the risk ownership, the credit scores provided by a P2P platform may be inadequate to predict loan defaults.

These arguments notwithstanding, P2P platforms have other advantages for the economic system, that can counterbalance, if properly accounted, the mentioned disadvantages. Namely, as sophisticated networks of institutional investors and lenders which are flexible to use alternative data sources and advanced technologies in their analysis, P2P platforms could offer a new approach to credit risk evaluation (Jagtiani and Lemieux, 2018), one that leverages strongly on their competitive advantages. P2P platforms are based on a *universal* banking model, fully inclusive, without space and business type limitations. They automatically receive data from the participants in the platforms, that concern transactions and/or relationships of each company not just with the platform but also with each other. Provided that enough agents populate the platform, the resulting networking data is rich of information providing deeper insight into the agents' creditworthiness. In particular, P2P can use the data on companies' interactions, in terms of payments, demand and supply chains, control and governance. The latter information can be used for the purpose of creating a network model which can quantify how borrowers are interconnected with each other; a model that can be employed to improve loan default predictions even in a context in which consumers or businesses have a very short credit history.

The need for a novel approach is also evident when one reviews the scoring models that are currently being utilized by P2P platforms. Namely, despite the fact that they are flexible to use different, more elaborate data sources as well as advanced technologies in their analysis, P2P platforms have not strayed significantly from traditional scoring methodologies. Most platforms that disclose some details on the methodology they use for conducting the credit rating task, indicate conventional approaches and traditional data sources (i.e. Lending Club, Twino etc.). An online investigation conducted reveals that most of the platforms have enabled *cookies*[1] on their

[1] A cookie is a small text file (up to 4KB) created by a website that is stored in the user's computer either

websites indicating they are able to collect various type of information including the user's digital footprint, social media activity, online public profiles, etc. Whether the platforms make full use of this information is unclear.

1.3 Objectives of the Study

We aim to build a network model from the available platform data, and to achieve this goal the main issue is to quantify the information contained in networking data, often available from different perspectives: financial transactions between companies, economic similarities, common holdings, presence in common demand or supply markets and so on, giving rise to a "multi-layered" network (Allen and Gale, 2000, Leitner, 2005, Giudici and Spelta, 2016, Aldasoro and Alves, 2016, Furfine, 2003, Leitner, 2005, Poledna et al., 2016, Calabrese and Giudici, 2015). The quantification of multi-layered information requires the development of an appropriate statistical methodology.

In this thesis, we propose and test several approaches for augmenting traditional credit scoring models with information on the interconnections that emerge between borrowers. The final objective is to compare the predictive accuracy of the baseline classification (i.e without accounting for the information emerging from the interconnections of the borrowers) and the augmented specification (i.e. with the inclusion of the information emerging from the interconnections of the borrowers).

Specifically, the thesis has the following objectives:

- test whether the ratings assigned by P2P platforms are a good predictor of default. We claim that because of P2P's inability to solve asymmetric information issues as efficiently as traditional banks, and the difference in risk ownership between P2P and banks' business models, the scoring system of P2P platforms may not sufficiently reflect the probability of loan default;
- test whether the inclusion of network information, or information on how borrowers are connected with each other can improve the predictive accuracy of the credit scoring methodology. We argue that even though P2P platforms have inherent disadvantages in producing an accurate scoring, they have other advantages that can be used to address the main risk concerns. Specifically, we claim that P2P systems can benefit from the inclusion of information on the interconnections or similarities that emerge between different participants on the platform - information that they possess and are further more flexible to use (compared to traditional financial intermediaries that are burdened by legacy infrastructures and procedures).

1.4 Summary of Research Design

In this thesis, we explore a variety of approaches for integrating the networks that emerge between borrowers within the credit scoring framework. A brief overview of the individual approaches is presented below:

- at first instance, we assume a situation in which time-varying financial information on borrowers is available on the platform. In this context, we employ correlation networks to derive similarity measures among borrowers, based on their correlations on a set of observed financial properties, from which we derive centrality measures for each borrower active[2] on the platform. Once we have derived the centrality measures for the active borrowers, we specify a dependency model, based on logistic regression, that allows modelling multi-layer networks into a single model that is not just descriptive but also predictive. Researchers have also introduced graphs into the financial literature as a tool for dealing with the volume and nature of complex interactions and relationships that emerge between economic agents and

temporarily for that session or permanently on the hard disk (persistent cookie). It is a way by which websites can recognize a user and keep track of his or her preferences

[2]In this thesis, active borrowers are considered all individuals or SMEs that have asked for a loan on a P2P lending platform

industries as a whole (Mantegna and Stanley, 1999). In classical financial networks, each weight represents financial transactions between the two corresponding nodes. As transactional data are often not available, the adjacency matrix can be substituted by a correlation matrix between the observations (Giudici and Spelta, 2016). Correlation models have been used as a viable alternative to classic network models as they seem to be able to assess common exposures and complement direct linkages;

- the second approach assumes a cross-sectional data set (a situation more common for P2P platforms which work with economic agents i.e. SMEs or private consumers, that cannot access funding through traditional channels mostly due to their limited credit history). In this context, we test whether the credit risk accuracy of P2P platforms could be improved by leveraging topological information embedded into similarity networks derived from borrowers balance-sheet features at a single point in time. Relevant patterns of similarities describing institutions' importance and community structures are extracted from the networks and employed as additional explanatory variables for improving the performance of different classes of scoring models. Compared to the correlation-based approach, here we build a network that accounts for the similarity between borrowers emerging from **all** information that the platform possesses rather then on the time-varying components only. Namely, we have financial information about borrowing companies collected in a vector x_n representing the financial composition of institution n. Consequently, we define a metric that provides the relative distance between companies by applying the standardized Euclidean distance between each pair (x_i, x_j) of institutions feature vectors;

- the third approach is a factor network-based segmentation for credit score modeling. Here, we attempt to construct a network of borrowers where links emerge from the co-movement of latent factors, which allows us to segment the heterogeneous population into clusters. In addition to the general logistic regression, we also present a credit score model for each cluster via lasso-type regularization.

1.5 Contribution of the Thesis

We remark that this work is related to two main research streams. First, some authors have carried out investigations on the accuracy of credit scoring models of P2P platforms (Serrano-Cinca and Gutiérrez-Nieto, 2016a, Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios, 2015, Guo et al., 2016, Zhanga et al., 2016, Mezei, Byanjankar, and Heikkila, 2018, Lin and Viswanathan, 2015). We improve these contributions with a more formal statistical testing procedure and, furthermore, with the extension to SME lending. Second, our network models relate to a recent and fast expanding line of research which focuses on the application of network analysis tools, for the purpose of understanding flows in financial markets (as in the papers of Allen and Gale, 2000, Leitner, 2005, Giudici and Spelta, 2016, Aldasoro and Alves, 2016, Furfine, 2003, Leitner, 2005, Poledna et al., 2016). We improve these contributions, extending them to the P2P context and linking network models, that are often merely descriptive, with machine learning classifiers, thus providing a predictive framework.

Going further into details, the thesis contributions are:

- correlation network models can combine the rich structure of financial networks (see, e.g. Mantegna, 1999, Battiston et al., 2012, Lorenz, Battiston, and Schweitzer, 2009) with a parsimonious approach based on the dependence structure among market prices. Important contributions in this framework are Billio et al. (2012), Diebold and Yilmaz (2014), Hautsch (2015), Ahelegbey, Billio, and Casarin (2016), Giudici and Spelta (2016), Giudici and Parisi (2018), who propose measures of connectedness based on similarities, Granger-causality tests, variance decompositions and partial correlations between market price variables. Our model extends the above approaches, as it employs: i) correlation networks to derive similarity measures among borrowers, based on their correlations on a set of observed financial ratios, from which we derive centrality measures for each company; and ii) a dependency model, based on logistic regression, that allows modelling multi-layer networks into a single model that is not just descriptive but also predictive. Our model also extends the literature in adapting it to the context of P2P lending platforms. Correlation network models cannot be fit as such to the P2P lending application, as they are based on the correlation among (univariate) market prices, using mono-layer financial networks. In the context of P2P systems instead, due to the different data available, we need to focus on correlation between (multivariate) financial variables that describe the state of a company, using multi-layer financial networks;

- understanding the structure of the similarity network (see Mantegna and Stanley, 1999 and Newman, 2018) is key for understand the origin of companies failures and to inform policymakers on how to prepare for, and recover from, adverse shocks hitting the network. Graph-theoretical tools are, indeed, instrumental when we lack a precise mathematical description of the system since they can reveal how the nature and the evolution of financial relationships induce distinguishable patterns of structural changes in the set of economic variables; We contribute to the literature by proposing a methodology for credit risk estimation that can account for such topological information. We further show that the forecasting gain obtained by the inclusion of these variables into different families of credit scoring models is substantial and this in turn could constitute a new instrument in both policy-makers and practitioners' toolboxes.

- we also propose a latent factor-based classification technique that can improve the accuracy of credit risk models for P2P systems. In this context, the contributions are also manyfold. Firstly, we extend the ideas contained in the factor network-based classification of Ahelegbey, Giudici, and Hadji-Misheva (2019b) to a more realistic setting, characterized by a large number of observations which, when links between them are the main object of analysis, becomes extremely challenging. Secondly, we extend the network-based scoring model proposed in Giudici and Hadji-Misheva (2017) to a setting characterized by a large number of explanatory variables. The variables are selected via lasso-type regularization (Trevor, Robert, and Jerome, 2009, Tibshirani, 1996) and, then, summarized by factor scores. Thus, we contribute to network-based models for credit risk quantification. Network models have been shown to be effective in gauging the vulnerabilities among financial institutions for risk transmission (see Billio et al., 2012, Diebold and Yilmaz, 2014, Lorenz, Battiston, and Schweitzer, 2009, Ahelegbey, Billio, and Casarin, 2016), and a scheme to complement micro-prudential supervision with macro-prudential surveillance to ensure financial stability (see IMF, 2011, Moghadam and Viñals, 2010, Viñals, Tiwari, and Blanchard, 2012). Recent application of networks have been shown to improve loan default predictions and capturing information that reflects underlying common features (see Ahelegbey, Giudici, and Hadji-Misheva, 2019b, Letizia and Lillo, 2018). Thirdly, our empirical application contributes to modeling credit risk in SMEs particularly engaged in P2P lending. For related works on P2P lending via logistic regression, see Andreeva, Ansell, and Crook (2007); Barrios, Andreeva, and Ansell (2014); Emekter et al. (2015); Serrano-Cinca and Gutiérrez-Nieto (2016a);

- finally, we cross-validate the accuracy of network theory in improving the predictive accuracy of Fintech risk management models by providing an application in a context beyond Fintech credit which also contributes the literature. Specifically, the work intends to extract the network of payment relationship between Bitcoin users, owners, similarly to Tasca, Liu, and Hayes (2016). We further extend Tasca, Liu, and Hayes (2016) by investigating whether trading volumes behaviors of different groups of Bitcoin traders are interconnected and, whether lead-follower behavior can be identified. We contribute the literature by proposing an extension of Vector Autoregressive models based on network theory which improves pure autoregressive models, as they introduce a contemporaneous contagion component, that describes contagion effects between groups of traders. Our main methodological contribution consists in the introduction of partial correlations and correlation networks into VAR models. This allows to describe the correlation patterns between trading volumes and to disentangle the autoregressive component of volumes from its contemporaneous part. The introduction of VAR correlation networks also allows to build a volume predictive model, that leverages the information contained in the correlation patterns. Looking to the literature, many researchers have proposed correlation network models able to combine the rich structure of financial networks (see, e.g. Battiston et al., 2012, Lorenz, Battiston, and Schweitzer, 2009) with a more parsimonious approach that can estimate contagion effects from the dependence structure among market prices (Billio et al., 2012, Diebold and Yilmaz, 2014). More recently, Ahelegbey, Billio, and Casarin (2015), Giudici and Spelta (2016), Giudici and Parisi (2017) have extended this methodology introducing stochastic correlation networks. The proposal outlined in Section 5 extends: (a) the approach of Ahelegbey, Billio, and Casarin (2015), by enriching their VAR model using partial correlations; and (b) the approach of Giudici and Spelta (2016), by enriching their graphical Gaussian models with an autoregressive component derived through a VAR model.

1.6 Thesis Organization

The thesis is organized in the following manner: Section 2 provides an overview of the P2P lending markets, their size and growth across different regions. Furthermore, the section provides a more in-depth review of the research on the driving forces of Fintech credit as well as the advantages and disadvantages associated with the growing presence of these providers in the consumer and commercial credit space. The risk concerns outlined in Section 2 provide the main motivation for this thesis.

Section 3 outlines the mathematical preliminaries and basic tools from classical credit risk models i.e. logistic regression, estimation of high-dimensional models, lasso regularization, discriminant analysis, naive bayes, support vector machine and decision tree models.

Furthermore, Section 3 also provides an overview of the approaches used to identify the underlining network structure that emerges between P2P borrowers. Specifically, the section provides full description of the three separate methodologies developed and tested within this thesis: (i) correlation-based credit scoring model; (ii) similarity-based credit scoring model and (iii) latent factor-based classification technique. Finally, Section 3 indicates the metrices used for assessing the performance of the proposed network-based credit risk models for P2P systems.

Section 4 outlines the different data sets used as well as the empirical results from the different network-based models proposed in this thesis. The section also compares the predictive accuracy of the proposed methodologies with the baseline specifications that do not account for the undelyining network structures which emerge between borrowers.

Section 5 provides an alternative context for the application of network models in Fintech risk management. Specifically, the section investigates whether network theory can be used to understand the dynamics of Bitcoin blockchain trading volumes and, specifically, how different trading groups, in different geographic areas, interact with each other. The section describes the data used as well as the methodology proposed for improving autoregressive models, introducing a contemporaneous contagion component, that describes contagion effects between trading volumes.

Section 6 discusses the results related with each of the thesis's objectives and provides concluding remarks. Recommendations for extension of the work are also presented.

2 Literature Review

2.1 Peer to Peer

The term peer-to-peer (P2P), originally emerged in the field of computer science and refers to the concept that in a network of peers (equal participants), using the appropriate IT and communication infrastructure, two or more peers can collaborate directly and without central coordination (Schoder, Fischbach, and Schmitt, 2005, Balakrishnan et al., 2003, Rao et al., 2003, Marti and Garcia-Molina, 2006). In other words, P2P is used to describe the interaction between two parties without the need for a central intermediary. Compared to a client/server networks, P2P systems should lead to *"improved scalability, lower cost of ownership, self-organized and decentralized coordination of previously underused or limited resources, greater fault tolerance, and better support for building ad hoc networks"* (Schoder, Fischbach, and Schmitt, 2005). P2P can be understood as one of the oldest architectures in the world of telecommunication i.e the internet is itself can be classified as a P2P network. The advancements in IT as well as the fast growth of the internet has enable the rise of a great variety of P2P services. Among the first such service that become widely popular was P2P sharing of files. This enabled users which had downloaded a particular software on their computers to be able to connect directly with other users and engage in file-sharing. In terms of the main characteristics of P2P systems, Schoder, Fischbach, and Schmitt (2005) outline three main properties of these networks: (i) they enable sharing of distributed resources and services, (ii) they provide decentralization, and (iii) they are characterized with autonomy, i.e. each of the nodes (entities or vertices) that are part of the network can decide when and to what extend they share the resources with the other participants in the network.

2.2 Peer to Peer in Finance

The history of P2P services in finance can be traced back to the establishment of two companies, the UK-based Zopa in 2005 and the US-based Prosper in 2006. Both companies allowed for borrowers and lenders to match themselves by dealing directly rather than going through a central marketplace (Milne and Parboteeah, 2016b).

This idea of eliminating or reducing the role of central intermediaries and moving to a peer-to-peer decentralized system is not new neither in the context of finance. *"Banking is necessary, banks are not"* was stated by Bill Gates in 1994 and this phrase has served as the mantra for the first wave of Fintech. As a result of Silicon Valley's changed focus towards disrupting industries, a great deal of Fintechs started thinking of solutions and products that can challenge different aspects of banking and could deliver better financial services directly to consumers. These developments undeniably point to a shift in customers' behaviour and preference towards digital solutions which in turn means that traditional intermediaries must rethink their digital strategy or otherwise innovative technologies *will eat their lunch.*

As a result of these changes in customers' preferences and the fast technological advancements, a wide range of peer-to-peer lending solutions have emerged, supporting personal loans or small business lending, invoice discounting and more. As suggested in the beginning of this thesis, the Financial Stability Board (Tijn et al., 2018) uses the term Fintech credit to broadly include all credit activity facilitated by electronic (online) platforms that are not operated by commercial banks. It also includes platforms that use their own balance sheet to intermediate borrowers and lenders.

Going into more detail about the business model, P2P lending is the practice of lending money to individuals or businesses through online services (online platforms) that match lenders directly with borrowers (Rajeshwari, 2019). They, in effect, allow lenders and borrowers to match themselves and provide the service of (somewhat) eliminating asymmetric information between borrowers and lenders by reviewing borrowers' information and providing a rating that summarizes that person's creditworthiness (Klafft, 2009, Bachmann et al., 2011). P2P lending makes micro-finances possible without the need of a traditional financial intermediary (Hongke et al., 2017).

The idea behind alternative finance is to disrupt the banking industry and to democratize the

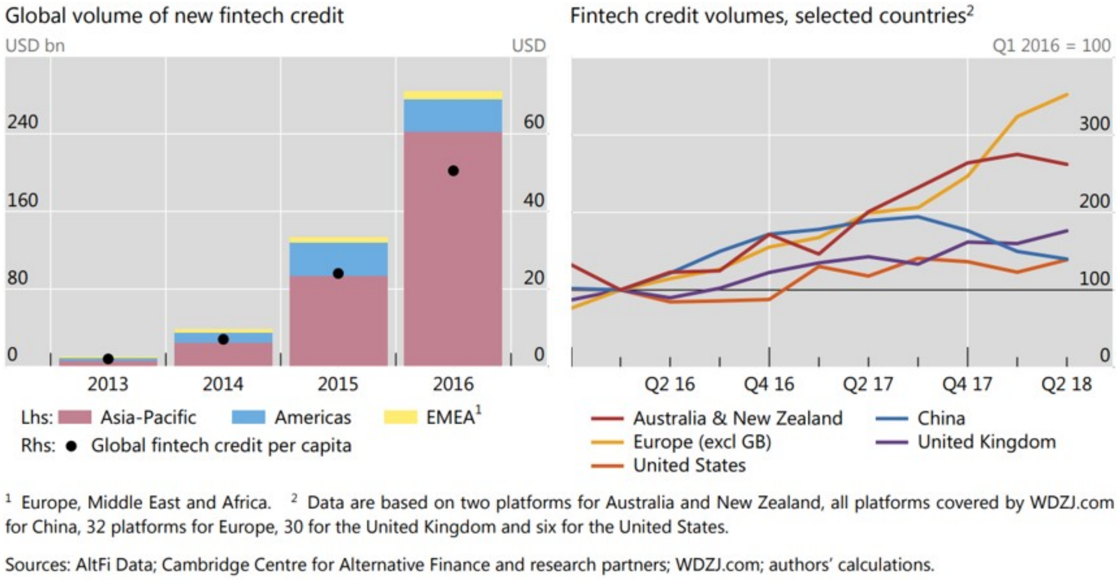


Figure 1: **Growth of Fintech (alternative) credit**

access to capital, initially in the spirit of direct lending among communities. Hence, peer to peer lending is an alternative to the traditional bank industry. Instead of borrowing from a bank, SMEs and private consumers may borrow capital from investors through an online platform. This is an attempt of *disintermediating* the financial services.

Additional to this, a key property of P2P lending platforms is that they make use of innovative digital technologies and advance methodological solutions for the purpose of addressing customers' needs and problems (Xu, Lu, and Chau, 2015). They attempt to solve the problems of lending by utilizing automated processes that reduce costs as well as credit risk models that use nontraditional data (or novel methodology). This is not to say that innovative technologies are not available to traditional financial intermediaries, but simply to stress the flexibility of Fintechs to make full use of their potential compared to the rigid nature of legacy banks where any change in existing infrastructure and processes might be very slow.

Finally, as these online marketplaces do not collect deposits, they can avoid many intermediation costs typically associated with traditional financial services. For instance, most peer-to-peer lending platforms are not required to respect bank capital requirements nor to pay fees associated with state deposit insurance practices, and this allows them to operate with lower costs. All these properties have enabled P2P lenders to offer attractive alternatives to customers and investors in the credit space.

In terms of the size of Fintech credit, measuring can be challenging in part because of its novelty, small size and diversity. Official national data are limited, as in most cases Fintech credit platforms are not subject to regulatory reporting requirements. One of the most comprehensive data sets on the topic of Fintech credit have been collected by the Cambridge Centre for Alternative Finance (CCAF), together with academic or industry partners (Ziegler et al., 2018). Estimates from the Cambridge Center for Alternative Finance (CCAF) (see Ziegler et al., 2018) indicate that \$284 billion in alternative funding was extended globally in 2016, starting from a very low base of \$11 billion in 2013 (Figure 1). Fintech credit has, however, evolved differently across different regions. In absolute terms, China was by far the largest market in 2016; the United States and the United Kingdom followed at a distance, with other large advanced economies further behind. In per capita terms, Fintech credit was relatively high in several smaller economies, including Estonia, Georgia and New Zealand.

Looking at the micro-level, the report by CCAF states that in 2016, 14,521 companies have raised close to 1.13 million euros by using online alternative finance platforms which in volumes

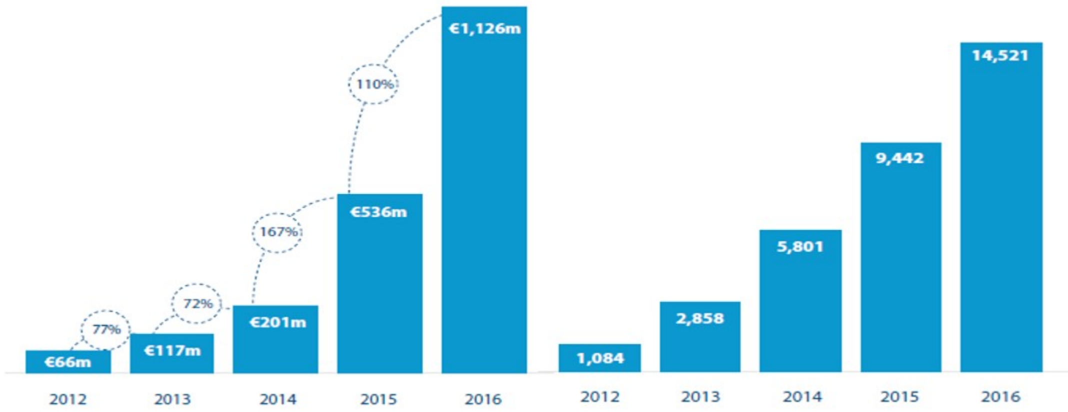


Figure 2: **Total online alternative business volumes (left) and number of European businesses raising alternative finance (right)**

represents a 110% annual growth against the previous year's total business funding (Figure 2).

One of the key priorities for Europe is promoting the health, strength and growth of the SME sector. Nevertheless, this development is often obstructed by SMEs' inability to access appropriate levels of financing (Ziegler et al., 2018). As indicated by the numbers, the emergence of Fintech credit providers has become of crucial importance in this context as they have evolved as a viable funding medium for entrepreneurs, start-ups and small and medium sized businesses across Europe. Additionally, what makes this topic of great interest is the fact that Fintech credit could become dominant in certain market segments. A report by ECB, suggests that in the United States, for example, 36% of unsecured personal lending was issued by Fintechs in 2017 (Mersch, 2019).

Furthermore, a clear example of the growing importance of P2P platforms is LendingClub which is a US peer-to-peer lending company, headquartered in San Francisco, California. This online platform is the first peer-to-peer lender to register its offerings as securities with the Securities and Exchange Commission (SEC), and to offer loan trading on a secondary market (Nowak, Ross, and Yencha, 2017). At the time of the writing of this thesis, LendingClub is the world's largest peer-to-peer lending platform. On the company's website, they claim that as of May 2019, the total loans issues amounted to 44 billion US dollars. Although this amount can be considered very small compared to the assets accumulated by a traditional financial intermediery, it can be considered relevant as it is growing significantly and in a very short period.

Looking at the reasons have supported these developments, a report by Morgan Stanly (Reid and Zwan, 2019) provides a summary of the evolution of alternative lending and states that Fintech credit gathered pace mostly due to: (i) the global credit crisis, which drove bank retrenchment from consumer and small business lending, and (ii) the introduction of new bank regulations, which in turn further increased the cost of capital for traditional financial intermediaries, imposing further pressure on the traditional banking model. Available data confirm the fast growth of Fitech credit.

The fast growth has also led to the emergence of many different business models within the context of alternative finance. Within the classification offered by the CCAF (see Ziegler et al., 2018), there are now 14 models of alternative finance with the most dominate being consumer and business lending. Namely, in 2016 which is the last report by the CCAF, P2P Consumer Lending accounted for 34% of European Alternative Finance volumes, followed by P2P Business Lending with 17%, Invoice Trading (12%), Equity-based Crowdfunding (11%) and Reward-based Crowdfunding (9%) (Table 1).

After analyzing the fact growth of P2P platforms in the last few years, we now provide an overview of the business models of these novel (dis)intermediaries. The platform may follow a variety of different methodologies but in a simple model, the online platform provides a low-cost

Alternative Finance Model	Definition	2016	Market Share
P2P Consumer Lending	Individuals or institutional funders provide a loan to a consumer borrower.	696.81m	33.8%
P2P Business Lending	Individuals or institutional funders provide a loan to a business borrower.	349.96m	17%
Invoice Trading	Individuals or institutional funders purchase invoices or receivable notes from a business at a discount.	251.87m	12.2%
Equity-based Crowdfuning	Individuals or institutional funders purchase equity issued by a company.	218.64m	10.6%
Reward-based Crowdfunding	Backers provide finance to individuals, projects or companies in exchange for non-monetary rewards or products.	190.76m	9.2%
Real Estate Crowdfunding	Individuals or institutional funders provide equity or subordinated-debt financing for real estate.	109.45m	5.3%
P2P Property Lending	Individuals or institutional funders provide a loan secured against a property to a consumer or business borrower.	95.15m	4.6%
Balance Sheet Business Lending	The platform entity provides a loan directly to a business borrower	59.13m	2.9%
Donation-based Crowdfunding	Donors provide funding to individuals, projects or companies based on philanthropic or civic motivations with no expectation of monetary or material return.	32.40m	1.6%
Debt-based Securities	Individuals or institutional funders purchase debt-based securities, typically a bond or debenture at a fixed interest rate.	22.85m	1.1%
Balance Sheet Consumer Lending	The platform entity provides a loan directly to a consumer borrower.	16.74m	0.8%
Mini-Bonds	Individuals or institutions purchase securities from companies in the form of an unsecured retail bonds.	10.16m	0.5%
Profit Sharing	Individuals or institutions purchase securities from a company, such as shares or bonds, and share in the profits or royalties of the business.	8.36m	0.4%

Table 1: **Taxonomy of P2P Lending**

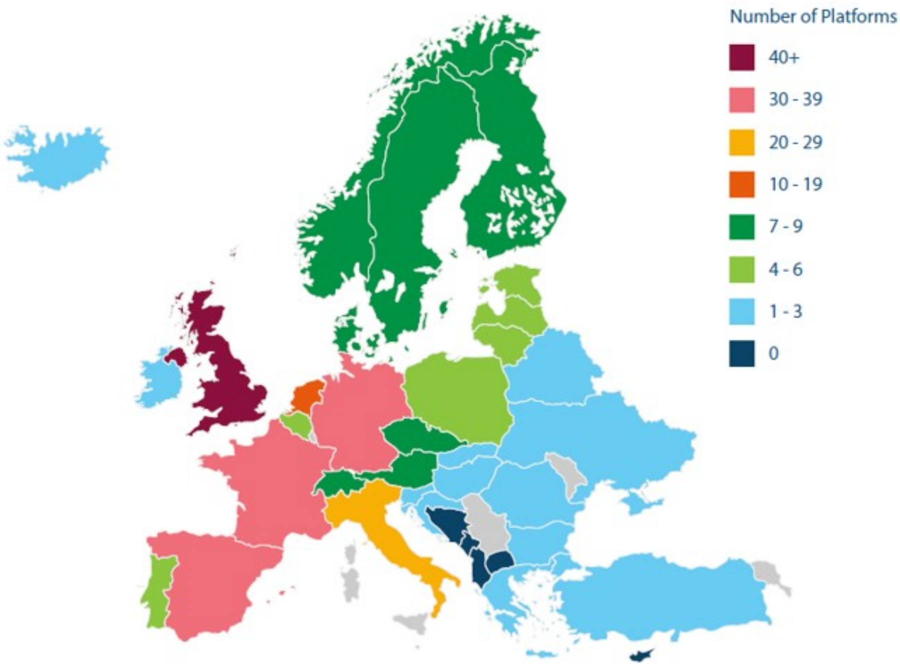


Figure 3: **Number of platforms per region**

standardized loan application process and facilitates direct matching and transacting of borrowers and investors (lenders) (Milne and Parboteeah, 2016a).

In the first step of the process, the prospective borrowers provide information on their own finances and the project or purpose for which they seek funding.

In the following step, investors are able to access the applications and review them through the online platform. It is important to mention that in most cases, the applications are available online only after the platform has verified the information provided.

Once a borrower and investor are matched, loan contracting comes into force directly between them. This ensures that the investor, rather than the platform operator, takes on the risks immediately.

Investments and loans are usually duration matched, meaning that investors are typically unable to liquidate their investments before expiration. The only option for investors to take out their investment earlier is to find another investor willing to take over the same participation. Across Europe, there are several P2P platforms that assist this process by establishing a secondary market where investors can trade with their investments or credit rights can be transferred, however in most cases, this service is not available.

Once the loan is originated, the P2P lender serves as an agent for the investors by servicing the loan in return for ongoing fees. Among the tasks that the platform takes on are: record maintenance, collection of borrowers' payments; distribution of payments to investors and management of the unmet obligations. In addition to these common services, some types of P2P platforms offer partial protection against loan defaults even though this is not the most widely used model.

The geographic distribution of participating platforms from Europe shows UK to have the highest concentration of platforms followed by Germany (35), France (33), Spain (32) and Italy (26) and the Netherlands (19). While individually the Nordic Countries (Denmark, Finland, Iceland, Norway and Sweden) had fewer than 10 platforms each, the region recorded 32 participating platforms (Figure 3).

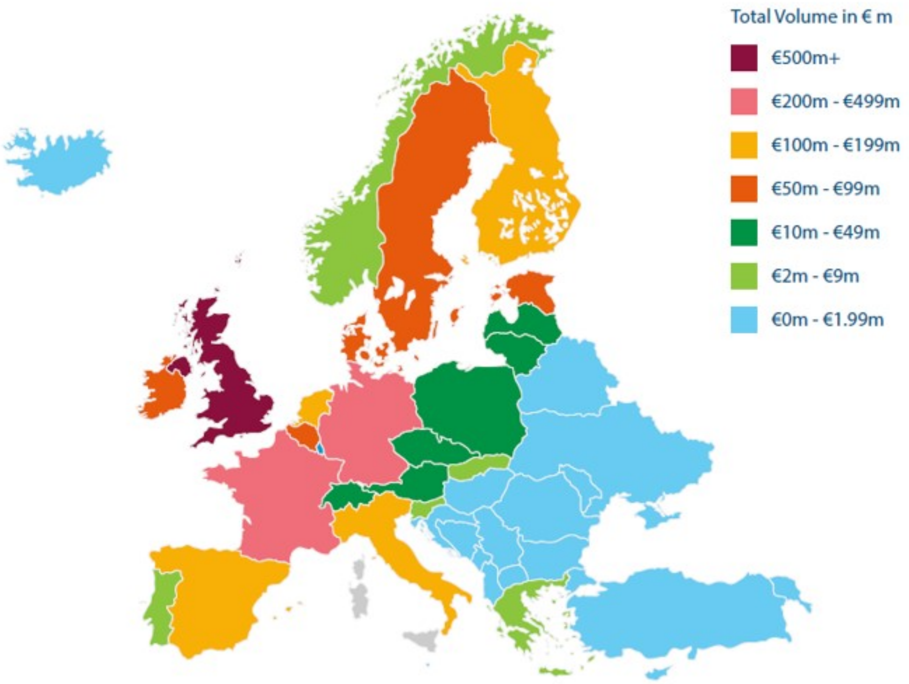


Figure 4: **Market volume per region**

After the United Kingdom, the top five volume-driving countries were France (443.98m), Germany (321.84m), the Netherlands (194.19m), Finland (142.23m) and Spain (130.90m). Interestingly, the volume generated did not necessarily correlate with platform distribution; for instance, Finland had only 8 platforms but ranked fourth in terms of volume whereas Italy, with 26 platforms, ranked in sixth place as related to volume.

2.3 P2P Lending Platforms: Drivers, Benefits and Risks

Drivers of P2P Lending Platforms. Here, we review the drivers as well as the benefits and risks that are associated with the rapid growth of these alternative financial intermediaries, which in turn form the motivation of this thesis.

As presented in Figure 1, despite the fact that technological developments have a global reach, the size of Fintech credit differs significantly across different regions. This in turn suggests that there might be country-specific factors that determine the growth of Fintech credit. Hence, if we consider what are the main factors driving this phenomenon, we have to discuss two separate set of factors. Some factors have influenced all forms of credit. Among these factors, researchers have linked: economic growth and level of economic and financial development, the quality of the country's legal institutions (Tijn et al., 2018), the public trust in national institutions, etc. Another group of factors are those that have a direct impact on the growth of Fintech credit. In this context, researchers have identified the degree of competition in credit markets and regulation as the main factors (see Group, 2019). Following Tijn et al. (2018) a less competitive banking system could mean higher margins on bank credit and thus a boost for alternative credit sources like Fintech credit. Furthermore, if Fintech credit providers are able to assess borrowers' credit-worthiness using non-traditional data sources and thus capture underserviced customers, that might also motivate higher volumes of Fintech credit. Additional to this, Tijn et al. (2018) argue that the intensity and quality of financial regulation could also directly influence the growth of Fintech credit even though, a priori, the overall effects can be ambiguous. More stringent overall regulation might lead to a higher trust by the public concerning the novel forms of financial intermediation. On the other hand, the lack of regulation of Fintech activities, services and products could stimulate their growth as they would be able to operate at lower costs. Such regulatory environment could also encourage arbitrage to the extent that similar risks are regulated more tightly in the traditional lending sector (Tijn et al., 2018).

In order to investigate the drivers of Fintech credit in more detail, we also report the results from a study conducted by the Bank for International Settlements (Tijn et al., 2018) which established a research group to study the size, drivers and policy issues associated with the growth of Fintech credit markets around the world. The regression results confirm that an economy's Fintech credit volume per capita (amount of funding collected through Fintech platforms per capita) is positively associated with GDP per capita. The results also report a negative coefficient estimate on squared GDP per capita which in turn might be indicative of the positive effect becoming less important at higher levels of development. In the context of the impact that regulation has, the study finds a statistically significant positive estimate on the Lerner index, a measurment of firm's market power. Specifically, the Lerner index describes the relationship between elasticity and price margins for a profit-maximizing firm. This result further suggest that jurisdictions that have a less competitive banking sector, are characterized with more Fintech credit activity. This is also consistent with the premise that Fintech provides, which are able to advantage of innovative technologies and alternative data sources can offers relatively lower cost services and greater convenience to customers.

A general conclusion from the study conducted by Tijn et al. (2018) is that at this stage of development, we cannot accurately evaluate the impact Fintech credit will have on the overall financial industry. Furthermore, going back to an earlier argument, overall measuring and data collection concerning Fintech's operations can be challenging because of the novelty, size and diversity of the sector. With these contrains in mind, it is possible to identify a set of benefits and risks associated with the fast growth of P2P lending platforms.

Benefits of P2P Lending Platforms. The literature identifies the following as the main benefits of Fintech credit: lower costs, greater convenience and financial inclusion. In principle, borrowers benefit because they are able to receive credits at lower interest rates, and in some cases with little or no collateral, whereas lenders can receive higher rates of return on investment, due to reduced transaction costs (see Emekter et al., 2015). Furthermore, the use of new digital technologies and more granular customer data allow Fintechs to provide greater convenience at lower costs. The literature provides some evidence in favor of this argument. Using market-wide, loan-level data on U.S. morgage applications, Andreas et al. (2018) find that Fintech lenders process mortgage applications about 20 percent faster than other lenders, even when controlling for detailed loan, borrower, and geographic observables. They further find that such faster processing does not come at the cost of higher defaults. In terms of the rational for such finding, Andreas et al. (2018) argue that Fintech lenders adjust supply more elastically than other lenders in response to exogenous mortgage demand shocks, thereby alleviating capacity constraints associated with traditional mortgage lending.

Turning to the financial inclusion argument, Jagtiani and Lemieux (2018) find that the score (assigned to borrowers by using both traditional and non-traditional data) perform well in predicting loan performance over the two years after origination. The authors further argue that the use of non traditional data sources (such as social media, digital footprint etc.) has allowed some borrowers who would have been classified as subprime by traditional criteria to be slotted into "better" loan grades, which allowed them to get lower priced credit, thus improving financial inclusion. Furthermore, the authors argue, that these borrowers that have obtain access to the credit market via Fintech credit provides, obtain more favorable loan terms. The information advantages of Fintech has also been discussed by other researchers. J. Yan, Yu, and Zhao (2015) stress that many P2P platforms rely not only on "hard" (data that is directly measurable, factual and indisputable)but also on "soft" (data that has been collected from qualitative observations and quantified) information for the purpose of evaluating a candidate's creditworthiness, a practice not typically employed by traditional banks. This in turn allows P2P platforms to extend the access to credit to unserved or underserved consumers. The literature also offers several empirical studies on the relevance of of such "soft" information (among which the applicants' pictures, descriptions concerning loans' usage as well as social networking activity) for credit scoring. For instance, Ge, J. Feng, and Gu (2016) find that two main forms of social media information serve as a signal concerning an applicant's creditworthiness: (i) the self-disclosure of social media accounts and (ii) the social media network engagement.

In a similar vein, Berg et al. (2018) investigate the accuracy of alternative data in credit scoring tasks. Specifically, they analize the applicability and usefulness of incorporating the information content of the digital footprint in the prediction of consumer default. The author finds that even simple, easily accessible variables from the digital footprint match the information content of credit bureau scores. They further ague, that these alternative information emerging from consumers' digital footprint complements rather than substitutes traditional information such credit bureau information. This provides evidence in favor of the premise that Fintech platforms widened access to credit. It is also important to mention that the European directives concerning the payment industry are also proving very favorable development for P2P lenders. For instance, with the new revised Payment Service Directive (PSD2), which is the implementation of a European guideline designed to further harmonize money transfers inside the EU, the monopoly which banks have on their clients' account information and payment transactions becomes weaker as this information can be disclosed through application payment interfaces, thus paving a way for P2P platforms to connect with banks and, thereby, improve both their "hard" (traditional) and their "soft" (non-traditional) information.

In China, Fintech credit is arguably well suited to fund small businesses, start-ups and less affluent consumers. As their access to credit through traditional channels might be constrained, these borrowers have often had to resort to informal private and more expensive lenders. For example, a study by Harald et al. (2018) finds evidence that automated credit lines to companies trading on Alibaba's e-commerce platform increase access to credit for firms with a low credit score. In a survey of retail borrowers on a large Chinese platform, more than half reported that they had no borrowing history from a financial institution (Deer and Yin, 2015).

Risk concerns. From a different viewpoint, the rapid growth of the importance of P2P lending platforms can pose significant risks to financial stability. This because P2P lenders typically produce inadequate measures of credit risk. In a nutshell, the increased volume of lending, which brings a higher commission revenue to the platform, could be associated with the risk of a deterioration in the credit risk of the counterparts. Secondly, in comparison with traditional banks, P2P platforms are less able to eliminate asymmetric information, thus increasing the risk of bad debt accumulation because they have no access to detailed information on borrowers past financial transaction, which in turn allows banks to better discriminate between credit applicants with different credit risk levels, and to better sustain borrower monitoring, once a loan has been assigned. Indeed, economic theory argues that banks represent an institutional solution to the problem of asymmetric information between borrowers and lenders in the credit market (Akerlof, 1970, Myers and Majluf, 1984, De Roure, Pelizzon, and Tasca, 2016). This happens because they are able to access detailed information on borrowers past financial transaction, which in turn allows them to better discriminate between credit applicants with different credit risk levels, and to better sustain borrower monitoring, once a loan has been assigned. In line with this, De Roure, Pelizzon, and Tasca (2016) claim that banks expertise in screening and monitoring the activities of borrowers gives them a competitive advantage over P2P lenders, as both ex ante and ex post asymmetric information are mitigated. A third point of attention concerns business models (Figure 5).

In the context of traditional banking, the "many-to-one-to-many" approach, in which the financial intermediary (the bank) collects deposits from several entities, fixes a borrowing price, and takes decisions concerning to whom to lend, has a high degree of transparency since rating and price information is typically disclosed. However, the intermediary's decision is not automatically determined by such information but, rather, the intermediary controls and governs the lending process. On the other hand, P2P lending is built on the basis of a "many-to-many" approach, in which the financial intermediary empowers each lender to decide to whom borrower to lend and

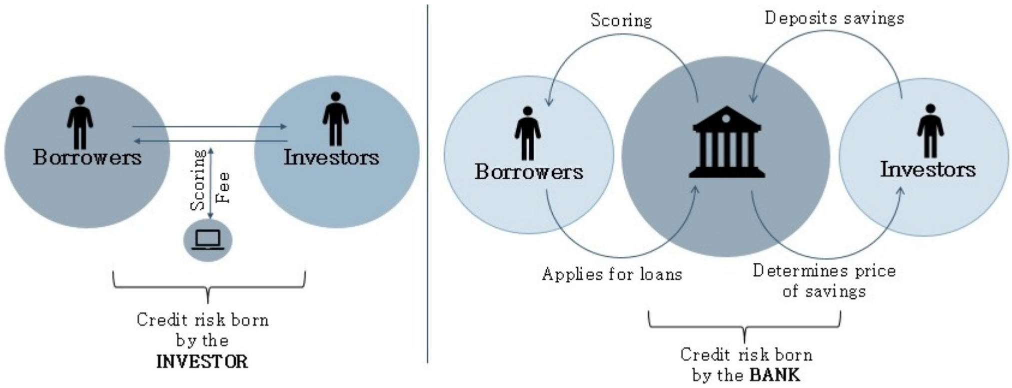


Figure 5: **Difference in risk ownership**

for what amount (Bachmann et al., 2011 and Ge, J. Feng, and Gu, 2016). To guide the process, the P2P platform provides lenders with information on the potential borrowers, their loan purpose and, most importantly, on their rating, and correlated prices. The grade of each loan represents a unifying indicator of the overall creditworthiness of each individual loan applicant, on which the decisions of the lender can be based. In other words, the intermediary does not really intermediate by making a lending decision but, rather, it provides the information on which such decision may be based. It follows that banks and P2P lenders have different approaches to credit scoring and, therefore, to credit risk measurement. For both banks and P2P lenders, a rating system has the purpose of estimating the probability of default of a loan, which is then used in the decision process concerning approval, interest rates and volume specifications. However, the incentives for model accuracy differ substantially. For banks, the grading is conducted by the financial institution itself, which is the actual entity that assumes the credit risk. A bank is thus interested to have the most accurate possible model. On the other hand, in a P2P platform, grading is determined by the platform but the risk is fully borne by the lender (see Serrano-Cinca and Gutiérrez-Nieto, 2016b). In other words, P2P lenders allow for direct matching between borrowers and lenders, without the loans being held on the intermediary's balance-sheet (see Milne and Parboteeah, 2016b). From a risk-return perspective, while in classical banking the financial institution chooses its optimal trade-off between risks and returns (subject to regulation constraints), in P2P lending, the platform maximizes its returns without taking care of the risks which are borne by the lenders. To summarise, as a combined result of misaligned incentives, asymmetric information, differences in the business model and in the risk ownership, the credit scores provided by a P2P platform may be inadequate to predict loan defaults.

2.4 Peer to Peer in Finance: Regulation

Having looked at the main benefits and risk that emerge from the P2P lending business model, this section provides a review of the emerging landscape of Fintech regulation as well as the challenges that regulators and supervisors face. As discussed in Section 2.3 Fintech solutions can bring many benefits to the overall financial system in terms of financial inclusion, lower costs and greater convenience. Nonetheless, they do leave the door open for many customer and investor protection issues which in turn represents a central point of interest for regulators and supervisors. The challenge for regulators is to identify the desired level of trade-off between innovation incentives on one hand, and consumer protection, on the other. The European regulatory framework should enable Fintech companies operating in its jurisdiction to benefit from innovations in technology and finance while at the same time ensuring both a high level of protection for consumers and investors, and resilience of the financial system. Specifically, in the context of P2P lending that we review, on one hand, regulators should ensure protection against systemic risks and maintain a fair, safe, and competitive market. At the same time, there is a need to encourage the growth of

Selected features of dedicated fintech credit policy frameworks

Jurisdiction	Tax incentives	Regulations[1]	Licensing / authorisation[1]	Investor protections[1]	Risk management requirements[1]
Australia	—	—	—	—	—
Brazil	—	✓	✓	✓	—
Canada	—	—	—	—	—
Chile	—	—	—	—	—
China	✓	✓	✓	✓	✓
Estonia	—	—	—	✓	—
Finland	—	✓	✓	—	—
France	✓	✓	✓	✓	✓
Germany	—	—	—	—	—
Japan	✓	—	—	—	—
Korea	—	—	—	—	—
Mexico	—	✓	✓	—	✓
Netherlands	—	—	—	✓	—
New Zealand	—	✓	✓	—	✓
Singapore	—	—	—	—	—
Spain	—	✓	✓	—	✓
Switzerland[2]	—	✓	✓	✓	✓
United Kingdom	✓	✓	✓	✓	✓
United States	—	—	—	—	—

[1] Specific rules for fintech credit that are separate from pre-existing rules for other financial intermediaries. [2] New rules effective from 2019.

Sources: CGFS-FSB (2017); national authorities

Table 2: **Overview of Fintech credit policy frameworks**. Table extracted from Claessens (2018), p.44

lending to realize its potential to transform small business funding and enhance economic growth (Giudici, 2018).

On the question, to what extent Fintech credit firms are regulated, practical experience shows that they can be drawn into either existing or new regulatory frameworks (Tijn et al., 2018). One of the main guidelines in this context is **neutrality** which means ensuring regulation does not favour one entity or form of activity over another provided the risks are the same. Some examples of existing regulation focused on Fintech companies are given by Australia and the Netherlands. In these countries, alternative credit provides must apply for a specific license in order to be able to engage in financial intermedation and facilitate access to credit for consumers. In the United States, P2P lenders must engage in a licensing process in each of the states in which they would like to extend operations (Naoko, David, and Bihong, 2019). Consequently, many P2P platforms enter in collaborations with banks to originate loans agreed online. A similar process is in practice in Germany as well where Fintech credit providers are forbidden from engaging in lending without a banking license and related prudential oversight (BAFIN, 2019). As Fintech credit grew, a number of countries have introduced specific new regulations and license regimes that are presented in the Table 2.

Most changes in the legislation have occurred since 2015 but some are quite recent (see Naoko, David, and Bihong, 2019). Brazil and Mexico introduced new rules and licensing practices in early 2018. Minimum capital requirements have been imposed in Spain and the United Kingdom, and have entered into force in Switzerland in January 2019.

Regulatory Feature	US	UK	People's Republic of China
Regulatory sandbox available	No	Yes	No
Operational licenses required	SEC license and licenses from state governments; platforms operating without an SEC license can seek investments from accredited investors	FCA license required but provisional licenses are common	Internet content provider license
Role of P2P platform	Facilitator of bank loans to borrowers; the platform purchases the loan using funds from investors	Facilitator of loans between investors and borrowers	Facilitator of bank loans to borrowers
Originator of the loan	Bank	P2P platform	Bank
Provision funds	Permitted but not a widespread practice	Permitted except for ISA (tax-free) investments	Not legally permitted but occur in practice

Table 3: **Regulatory features across the US, UK and China**. Table extracted from ADBI Working Paper Series (Naoko, David, and Bihong, 2019)

While the amount of legislation concerning Fintech activities is increasing, many authorities have stressed the need to ensure that regulations is not going to demotivate innovation and market entry. This is particularly relevant for regions in which access to capital is quite limited and the emergence of alternative credit sources can lead to great economic benefits. Regulatory bodies in several regions have put in place innovation facilitators, such as "regulatory sandboxes" where they implement and test different technologies in a controlled environment. Some jurisdictions have also introduced specific tax incentives for investors.

Going into more details concerning the individual regulations, Table 3 provided by the Asian Bank for Development (see Naoko, David, and Bihong, 2019, gives an overview and a comparison between different regulatory regimes in the US, the UK and China.

Clearly, the degree of regulation changes significantly across different regions. The United States is characterized with an extensive and stringent regulation in which several regulatory and supervisory bodies have oversight. In this country, P2P lenders are not a true matching platform as the legislation states that a bank must originate the loan from the platform to the borrower. Consequently, the US P2P market is very concentrated, with only a very few number of platforms actually in business. The UK Fintech market is slightly different. First, there is only the Financial Conduct Authority (FCA) that has oversight and its practice is to assess P2P lending platforms individually. The FCA, furthermore, is greatly interested in establishing a dialog with P2P platforms which in turn allows the regulator to obtain constant feedback on the regulatory burden Fintech's face as well as their scaling-up issues. Consequently, the UK counts the most P2P platforms in Europe maybe also do to the lighter regulation and friendlier environment to innovation in the financial sector.

The context of China, the P2P lending industry has grown faster than in any other country. The sector remains largely unregulated mostly because of its relevance concerning financial inclusion. It is widely thought that China's government purposefully refrains from involvement in this sector to allow it to grow quickly and thus provide access to credit to underserved parts of the economy.

Overall, regulation greatly depends on the context.

What is clear from the above regulatory overview is that many regulatory and supervisory bodies have considered the risk and benefits of innovative business models in the credit space. Furthermore, they have also tried to either integrate novel business models in existing legislation or propose new legislation addressing some of the risk. However, some issues still remain. There are problematic incentives for platforms that rate credit and originate loans without holding the risk of these loans. In addition, when investor returns are guaranteed by platforms, investors have no incentive to distinguish among risk categories (Naoko, David, and Bihong, 2019). The main question now becomes what can innovative technologies actually do and how can they help.

P2P platforms, as sophisticated networks of economic agents, possess a variety of advantages, which if accounted for properly, can mitigate some of the main risk concerns outlined by researchers and regulatory bodies. Namely, due to their flexibility in using alternative data sources, advanced technologies and novel analytical models, P2P platforms could offer a new approach to credit risk evaluation, one that leverages strongly on their competitive strengths. As they automatically receive data from the participants in the platforms, concerning transactions or relationships of economic agents with the platform and amongst themselves, P2P platforms can rely on network theory to enrich the information pool and provide deeper insight into the agents' creditworthiness. Such information can be used for the purpose of creating network models that can capture the interconnections that emerge between borrowers and these models can consequently be integrated into a standard classifier to improve loan default predictions even in the context in which consumers or business have a very short credit history.

The following section describes in detail the proposed methodology for building a network between P2P borrowers and using information provided by these network to improve P2P platforms' ability to discriminate between different risk classes.

3 Methodology and Data

As stated previously, the main purpose of this work is to test the predictive accuracy of traditional scoring models as employed by P2P lending platforms and investigate whether the inclusion of network parameters or information on how borrowers are linked or interconnected, can lead to higher predictive accuracy of P2P scoring models. In a nutshell, what we aim to active is test the predictive accuracy of classifiers with and without accounting for the underlining network structure that exists between borrowers active on the platform. In this context, the methodology chapter provides an overview of: (i) the classifiers used, (ii) the approaches for building a network between P2P borrowers active on the platform, (iii) the integration of network parameters in a predictive framework, and (iv) the approaches for assessing the predictive accuracy of the different models.

3.1 Credit Risk Models

Credit risk models are useful tools for modelling and predicting individual firm default and have been discussed extensively in the literature (Crouhy, Galai, and Mark, 2000, B.Gordy, 2000, A.Lopeza and R.Saidenberg, 2000, Carey and Hrycay, 2001). Such models are usually grounded on regression techniques or machine learning approaches (Khan and Niazi, 2017) often employed for financial analysis and decision-making tasks such as accurate forecasting, classification of risk, estimating probabilities of default, and data mining. These supervised learning techniques range from non-linear regression, generalized linear regression, discriminant analysis, to support vector machine, decision trees and neural networks (see Abellán and Castellano, 2017; Galindo and Tamayo, 2000; Khandani, Kim, and Lo, 2010; Khashman, 2011; Lean et al., 2010; Lessmann et al., 2015 to cite few).

The literature offers a wide variety of credit risk models that differ in their fundamental assumptions. Nonetheless, the joint aim of these methodologies is to predict the probability density function of losses that arise from the credit portfolio of the financial intermediary. Consider N firms having observation regarding T different variables (usually balance-sheet measures). For each institution n define a variable γ_n to indicate whether such institution has defaulted on its loans or not, i.e. $\gamma_n = 1$ if company defaults, $\gamma_n = 0$ otherwise. In a nutshell, credit risk models develop relationships between the explanatory variables embedded in T and the dependent variable γ.

The ability to measure credit risk and to successfully discriminate between different risk classes has the potential to greatly improve the risk management capabilities of financial intermedieries. In this thesis, for the purpose of investigating whether network information can improve loan default predictions and further protect lenders, in a financial stability context, we rely on various types of credit scoring models. Namely, against this background and throughout this thesis, we employ: (i) logistic regression (with and without regularization), (ii) discriminant analysis, (iii) naive Bayes classifier, (iv) support vector machines and (v) decision trees (Anderson, 2007). These models, once estimated on a training sample, can be used to predict the probability of default of a new loan, so that lenders can decide whether to invest on it, or not. This decision crucially depends on the accuracy of the prediction which, in turn, depends on the validity of the employed model. We argue that scoring models can be improved by exploiting borrowers' networking data. We believe that incorporating network information into a credit scoring model employed by P2P platforms could improve default predictive accuracy significantly and further protect investors. This requires building an appropriate network analysis model.

The following sections summarize the characteristics of the models employed in the analysis as well as the different methodologies for identifying the underlining networks structure that emerges between P2P borrowers.

3.1.1 Logistic Regression

The logistic regression model (D. W. Hosmer and Lemesbow, 1980, D. W. Hosmer, T. Hosmer, et al., 1998, Wright, 1995, Jr, Lemeshow, and Sturdivant, 2013) is one of the most widely used methods for a binary classification tasks. Namely, the model aims at classifying the dependent variable into two groups characterized by the different status (defaulted v.s. active) by the following model (Jr, Lemeshow, and Sturdivant, 2013):

$$ln(\frac{p_n}{1-p_n}) = \alpha + \sum_{t=1}^{p} \beta_t x_{nt} \tag{3.1}$$

where p_n is the probability of default for institution or client n, $x_i = (x_{i1}, ..., x_{iT})$ is the T-dimensional vector of borrower (firm or client) specific explanatory variables, the parameter α is the model intercept while β_t is the p-th regression coefficient. It follows that the probability of default can be found as:

$$p_n = (1 + exp(+\sum_{t=1}^{T} \beta_t x_{nt}))^{-1} \tag{3.2}$$

3.1.2 Estimating High-Dimensional Logistic Models

When estimating high-dimensional logistic models with a relatively large number of predictors, there is the tendency to have redundant explanatory variables (Kalina, 2014). Thus, to construct a predictable model, there is the need to select the subset of predictors that explains a large variation in the probability of defaults. Several variable selection methods have been discussed and applied for various regression models. In this thesis, variants of the lasso regularization for logistic regressions, are considered (Trevor, Robert, and Jerome, 2009).

3.1.3 Lasso

The lasso estimator (Tibshirani, 1996) solves a penalized log-likelihood function given by

$$\arg\min_{\beta} \sum_{i=1}^{n} \Big[Y_i(\beta_0 + X_i\beta) - \log\left(1 + \exp(\beta_0 + X_i\beta)\right)\Big] - \lambda \sum_{j=0}^{p} |\beta_j| \tag{3.3}$$

where n is the number of companies, p the number of predictors, and λ is the penalty term, such that large values of λ shrinks a large number of the coefficients towards zero.

3.1.4 Adaptive Lasso

The adaptive lasso estimator (Hui and Trevor, 2005) is an extension of the lasso that solves

$$\arg\min_{\beta} \sum_{i=1}^{n} \Big[Y_i(\beta_0 + X_i\beta) - \log\left(1 + \exp(\beta_0 + X_i\beta)\right)\Big] - \lambda \sum_{j=0}^{p} w_j|\beta_j| \tag{3.4}$$

where w_j is a weight penalty such that $w_j = 1/|\hat{\beta}_j|^v$, with $\hat{\beta}_j$ as the ordinary least squares (or ridge regression) estimate and $v > 0$.

3.1.5 Elastic-Net

The elastic-net estimator (Hui and Trevor, 2005) solves the following

$$\arg\min_{\beta} \sum_{i=1}^{n} \Big[Y_i(\beta_0 + X_i\beta) - \log\left(1 + \exp(\beta_0 + X_i\beta)\right)\Big] - \lambda \sum_{j=0}^{p} (\alpha|\beta_j| + (1-\alpha)\beta_j^2) \tag{3.5}$$

where $\alpha \in (0, 1)$ is an additional penalty such that when $\alpha = 1$ we a lasso estimator (L_1 penalty), and when $\alpha = 0$ a ridge estimator (L_2 penalty). For the elastic-net estimator, we set $\alpha = 0.5$ giving equal weight to the L_1 and L_2 regularization.

3.1.6 Adaptive Elastic-Net

The adaptive elastic-net estimator (Zou and H. H. Zhang, 2009) combines the additional penalties of the adaptive lasso and the elastic-net to solve the following

$$\arg\min_{\beta} \sum_{i=1}^{n} \Big[Y_i(\beta_0 + X_i\beta) - \log\left(1 + \exp(\beta_0 + X_i\beta)\right) \Big] - \lambda \sum_{j=0}^{p} (\alpha w_j |\beta_j| + (1-\alpha)\beta_j^2) \tag{3.6}$$

In one of the applications presented in this thesis, we focus on estimating the credit score using the four lasso-type regularization methods. We select the regularization parameter using ten-fold cross-validation on a grid of λ values for the penalized logistic regression problem. Two λ's are widely considered in the literature, i.e., $\lambda.min$ and $\lambda.1se$. The former is the value of the λ that minimizes the mean square cross-validated errors, while the latter is the λ value that corresponds to one standard error from the minimum mean square cross-validated errors. Our preliminary analysis shows that $\lambda.1se$ produces a larger penalty that is too restrictive in the sense that we lose almost all the regressors. Although our goal is to encourage a sparse credit scoring model for the purpose of interpretability, we do not want to impose too much sparsity that renders the majority of the features insignificant. Thus, we rather choose $\lambda.min$ over $\lambda.1se$. For the additional penalty terms, we set $\alpha = 0.5$, $v = 2$, and $\hat{\beta}_j$ as the ridge regression estimate.

3.1.7 Advantages and Disadvantages of the Logistic Regression

As suggested in the beginning of the section, the logistic regression is a widely used technique for binary classification tasks because it is very efficient. It does not require too many computational resources and it typically provides models that are very accurate (Tuffery, 1996). Furthermore, in terms of the outputs, it directly models a probability which is very useful for many different problem sets (ex. probability of default). Also, as demonstrated above, it is rather simple to regularize. On the other hand, the decision boundary of this technique is linear hence it cannot be used for identifying non-linear dependencies. Moreover, it requires the explanatory variables to be linearly independent and it is very sensitive to extreme values in the continuous variables (Tuffery, 1996).

3.1.8 Discriminant Analysis

Discriminant analysis (see Klecka and Iversen, 1980, Lachenbruch and Goldstein, 1979) assumes that different classes generate data based on different Gaussian distributions with the same variance-covariance matrix. Linear discriminant analysis (LDA) approaches the problem by assuming that the conditioal probability density functions $p(x|\gamma = 0)$ and $p(x|gamma = 1)$ are both normally distributed with mean and covariance parameters (mu_0, V_0) and (mu_1, V_0) respectively.

Under this assumption, the optimal Bayes classifier compares the a posteriori probabilities of all classes and assigns a pattern to the class with the maximal probability (Mika et al., 1999). Namely, the Bayes optimal solution is to predict points as being from the default class if the log of the likelihood ratios is bigger than some threshold τ, so that:

$$(\mathbf{x} - \mu_0)'\mathbf{V}_0^{-1}(\mathbf{x} - \mu_0)' + ln|\mathbf{V}_0| - (\mathbf{x} - \mu_1)'\mathbf{V}_1^{-1}(\mathbf{x} - \mu_1)' + ln|\mu_1| > \tau \tag{3.7}$$

3.1.9 Advantages and Disadvantages of Discriminant Analysis

The main advantages associated with discriminant analysis are (Tuffery, 1996): (i) it has a direct analytical solution which in turn makes if very easy and fast to calculate, (ii) since the coefficients are a linear combination of the input features, the results are relatively explicit, (iii) it works well with smaller dataset as well. Turning to the weaknesses, similar as it is the case with the logistic regression, the discriminant analysis detects only linear relationships. Moreover, it works only on continuous variables that do not have any missing values.

3.1.10 Naive Bayes

The naive Bayes classifier (Ng and M. Jordan, 2001 greatly simplify learning by assuming that features are independent given a class (Irina, 2001). Although this might be considered a very unrealistic assumption, empirical studies investigating the predictive accuracy of different classifiers indicate that in practice, the naive Bayes compares well with other (more) sophisticated tools.

Naive Bayes models aim at estimating the probability or probability density of features x given class γ, i.e. $P(x|T\gamma)$. The naive Bayes classifier combines this model with a decision rule. One common rule, the maximum a posteriori (MAP) decision rule, picks the hypothesis that is most probable. This means that, the corresponding classifier is a function that assigns a class label $\hat{\gamma} = C_k$ for $k = 0, 1$ as follows:

$$\hat{\gamma} = argmax_{k\in(0,1)} p(C_k) \prod_{t=1}^{T} p(\mathbf{x}_t|C_k) \tag{3.8}$$

3.1.11 Advantages and Disadvantages of Naive Bayes

The Naive Bayes algorithm is very simple and fast to train since no complex mathematics and error correction are involved (Klecka and Iversen, 1980). Furthermore, in the context of categorical data, this method typically outperforms other approaches hence it is very useful for multi-class problems. On the other hand, the Naive Bayes classifier is not able to learn interactions between features and furthermore it is not useful for large data sets as it would give poor performance.

3.1.12 Support Vector Machine

Support vector machine (SVM) classifies data by detecting the best hyperplane that separates all data points of one class from those of the other class (Gunn, 1998). Consider $(x_1, \gamma_1), ..., (x_N, \gamma_N)$ where the γ_n indicates the class to which the point x_n belongs. Each x_n is a T-dimensional real vector. SVM finds the "maximum-margin hyperplane" that separates data points x_n for which $\gamma = 1$ from the data points for which $gamma = 0$, which is defined so that the distance between the hyperplane and the nearest point x_n from either group is maximized. In formula:

$$\max_{\mathbf{w}\in R^T, b\in R} \min_{\mathbf{x}\in A\cup B} \frac{|\mathbf{w}'\mathbf{x}_{i+b}|}{||\mathbf{w}||} \tag{3.9}$$

where A and B are disjoint subsets and $wx - b = 0$ represents a hyperplane.

3.1.13 Advantages and Disadvantages of SVM

The main advantages of this technique are (Tuffery, 1996): (i) it is very useful for modelling non-linear phenomena, (ii) high predictive accuracy, and (iii) robust results (due to the fact that the optimal hyperplane is determined by the nearest points). The advantages notwithstanding, SVM models are not easily interpretable and are very sensitive to changes of the kernel parameters. Finally, depending on the size of the data set, these models can be very computationally intensive.

3.1.14 Decision Tree

A decision tree classifier is one of the possible approaches to multistage decision making where the basic idea is to break up a complex decision into a union of several simpler decisions, hoping the final solution obtained this way would resemble the intended desired solution (Safavian and Landgrebe, 1991). In a nutshell, a decision tree is a flow-chart like methodology where internal nodes represent test on an attribute and each branch represents the result of a test while each leaf node represents a response (decision taken after computing all attributes). Basically, such models create a tree for the entire data and process a single outcome at every leaf. Algorithms for building decision trees generally work top-down by extracting the variable that, at each step, best splits the set of items. Some metrics are applied to each candidate sub-set, and the resulting values are combined (e.g., averaged) to provide a measure of the quality of the split.

3.1.15 Advantages and Disadvantages of Decision Tree Models

Decision trees are used greatly in classification and regression tasks mostly because they allow for model interpretation as a sequence of if-then-else rules (Kudyba, 2014). Further advantages of this method is that it accounts for variables' interactions and it is suitable for high-dimensional data, which is in line with the exponential growth of data that we are currently facing. This been said, high-dimensionality by default means lower interpretability hence large decision trees will be harder to interpret. Probably the most relevant disadvantage of decision trees is that they use a step function that can have very large errors near the boundaries (Kudyba, 2014).

3.2 Network-based Scoring Models for P2P Lending Platforms

In this section, we provide an overview of the approaches used for identifying the underlining network structure that emerges between P2P borrowers. Namely, as stated previously, we believe that incorporating network information into credit scoring models of P2P systems, could significantly improve their predictive accuracy. This in turn requires building an appropriate network analysis model. In this thesis, we test three separate approaches and those are:

- **a correlation-based credit scoring model (Giudici, Hadji-Misheva, and Spelta, 2019a)**, based on the logistic regression, where time-varying financial features for borrower SMEs active[3] on the P2P platform are available. Specifically, we propose to augment the logistic regression with centrality measures derived from correlation networks among borrowers, deduced from the co-movement of their financial features. Centrality measures are indicators of the importance of any given node in a network. This approach requires the estimation of a correlation network emerging between the borrower companies and calculating network centralities, that are included into the model specification;
- **a similarity-based credit scoring model (Giudici, Hadji-Misheva, and Spelta, 2019b)**, where there is no information on how borrowers' financial indicators have changed over time. With this approach, we propose to enhance credit risk accuracy of peer-to-peer platforms by leveraging topological information embedded into similarity networks, derived from borrowers' financial information. Topological coefficients describing borrowers' importance and community structures are employed as additional explanatory variables in a variety of classifiers;
- **a latent factor-based classification technique (Ahelegbey, Giudici, and Hadji-Misheva, 2019b, Ahelegbey, Giudici, and Hadji-Misheva, 2019a)**, where there is no information on how borrowers' financial indicators have changed over time. With this approach, we propose a latent factor-based classification technique to divide the population into major network communities in order to estimate a more efficient logistic model. Specifically, given a number of attributes that capture firm performances in a financial system, we build a latent position model which allow us to distinguish between communities of connected and not-connected firms based on the spatial position of the latent factors. Consequently, we compare the predictive accuracy of the full specification with that of the models that are estimated separately on the two communities of connected and non-connected companies.

The common thread among these approaches is graph theory which has become increasingly recognized as a powerful methodology for investigating and modeling interactions between economic agents (Minoiu and Reyes, 2013).

The studies of statistics and computer science have, in general, followed separate paths and development with each field providing useful service to the other. In recent years, their paths

[3]In this thesis, active companies are considered all SMEs that have asked for a loan on a P2P lending platform

have become increasingly intertwined; with statisticians becoming more concerned with computational issues and computer scientists becoming more concerned about the interpretability of the systems and solutions they build. One main area in which these two fileds have become massively intertwined is graphical models.

A graphical model is a family of probabilistic distributions defined in terms of a directed and undirected graph (M. I. Jordan, 2004, Whittaker, 1990, Koller and Friedman, 2009, M. Jordan, 1996). In the context of directed graphs, let $G(V, E)$ be a directed acyclic graph where V are the vertices or nodes and E are the edges or links of the graph. The nodes can represent variables or units whereas the edges represent the statistical dependence between the variables or units.

In the following sections, we elaborates on the three separate approaches used in this thesis to account for the interconnections that emerge between borrower companies active on P2P lending platforms.

3.2.1 Correlation-based Credit Scoring Model

Correlation network models, that rely on correlations between the units of analysis (borrowers, in our context), according to a given set of statistical variables, have been employed by a variety of researchers (Takayuki, Hideki, and Misako, 2006, Epskamp and Eiko, 2018, Bazzi et al., 2016, Opgen-Rhein and Strimmer, 2007). In the context of inter-bank lending, correlation networks have been proposed in Giudici, Sarlin, and Spelta (2017a). The authors compare correlation networks with "physical" networks, based on actual transactions, and show that they can achieve comparable predictive performances.

Mathematically, correlation network models are related to graphical models. A graphical model can be defined by a graph $\mathcal{G} = (V, W)$ where V is a set of vertices (nodes) and $W = V \times V$ is a set of weights (links) between all the vertices.

In a graphical Markov model (see e.g. Lauritzen, 2011) the weight set specializes to an edge set E, that describes whether any pair of vertices (i, j) is connected $(i, j) \in E$ or not $(i, j) \notin E$. A graphical Markov model can be fully specified by an adjacency matrix, A. The adjacency matrix A of a vertex set V is the $I \times I$ matrix whose entries are $a_{ij} = 1$ if $(i.j) \in E$, and 0 otherwise.

From a statistical viewpoint, each vertex $v \in V$ in a graphical Markov model can be associated with a random variable X_v. When the vector of random variables $(X_v, v \in V)$ follows a multivariate Gaussian distribution, the model becomes a graphical Gaussian model, characterised by a correlation matrix R which can be used to derive the adjacency matrix. This because the following equivalence holds:

$$(i, j) \notin E \iff (R^{-1})_{ij} = 0 \tag{3.10}$$

which states that a missing edge between vertex i and vertex j in the graph is equivalent to the partial correlation between variables X_i and X_j being equal to zero.

Building on the previous equivalence, a graphical Gaussian model is able to learn from the data the structure of a graph (the adjacency matrix) and, therefore, the dependence structure between the associated random variables. In particular, an edge can be retained in the model if the corresponding partial correlation is significantly different from zero.

In a network analysis model (see e.g. Barabasi, 2016), the set W is a set of weights, which usually connect each variable with all others. In other words, the graph is fully connected.

From a statistical viewpoint, each vertex $v \in V$ in a network analysis model is associated with a statistical unit, and each weight describes an observed relationship between a pair of units, such as a quantity of goods or a financial amount. While the adjacency matrix in a graphical Markov models is symmetric, the weight matrix does not need to be so. For instance, in interbank lending, which is one of the main application of network analysis to the financial domain, the weights are financial transactions, with w_{ij} indicating how much i lends to j and w_{ji} indicating how much j lends to i. The aim of a network analysis model is not to learn from the data the structure of a graph but, rather, to summarise a complex structure, described by a graph, in terms of summary measures, or topological properties.

A correlation network model (see e.g. Mantegna, 1999, Brunetti et al., 2015, Giudici and Hadji-

Misheva, 2017) is a network analysis model for which the weights are not directly observed, but are calculated as pairwise correlations between the values of a given random variable X_v, observed at different time instances $(1, \ldots, N)$, for each pair of statistical units.

Note that correlation network models are similar to graphical Markov models, as they are based on statistical relationships between variables. However, differently from graphical Markov models, (and similarly to network analysis models) they relate units, rather than variables, and they are based on correlations, rather than on partial correlations.

Note also that correlation networks are different from financial networks, the network analysis models typically considered in the financial literature (see e.g. Stefano et al., 2012). Financial networks are based on data that describe the actual financial flows between each pair of borrowers, in a given time period. If this information is available, we could use them as weights, directly.

However, this is an approach that we cannot follow when the transactions between borrowers are not available or, even when they are, when they lead to a sparse weight matrix. In addition, Giudici and Hadji-Misheva (2017) showed that, in the context of international banking, financial networks can be matched, or even improved in terms of predictive performance, by correlation networks.

In the P2P lending context, each vertex of a correlation network can correspond to a borrower company; while each edge can represent the correlation between the vector of values that a statistical variable takes, along time, for two different companies.

To exemplify, we can associate with each borrower $i = 1, \ldots, I$ a vector $X^i = (X_t^i, t = 1 \ldots, N)$ that contains the values of a random variable, such as the total assets of a company, in N distinct time periods. A weight w_{ij} between any two vertices can then be defined by the correlation between the time series X^i and X^j, as follows:

$$w_{ij} = \frac{N(\sum_t X_t^i X_t^j) - (\sum_t X_t^i)(\sum_t X_t^j)}{\sqrt{[N\sum_t (X_t^i)^2 - (\sum_t X_t^i)^2][N\sum_t (X_t^j)^2 - (\sum_t X_t^j)^2)]}}, \tag{3.11}$$

where $X^i = (X_1^i, \ldots, X_N^i)$ and $X^j = (X_1^j, \ldots, X_N^j)$ are the two series of observed values of the random variable, respectively for units i and j, at times $t = 1, \ldots, N$.

According to the above definition, the weight between any two vertices is a correlation coefficient, with the corresponding properties. In particular, a high positive value of w_{ij} means that the two companies are "similar": they move along time in the same direction. Conversely, a high negative value means that they move in opposite directions.

We now extend correlation network models. In analogy with graphical Markov models, we replace the weight matrix W with an adjacency matrix E, and associate the absence of an edge in E with with a zero correlation between the corresponding pair of companies. More formally, we take $\mathcal{G} = (V, E)$, and let

$$(i, j \notin E) \iff Corr(X_i, X_j) = 0 \tag{3.12}$$

Then, similarly as in graphical Markov models, an edge can be retained in the model if the corresponding correlation is significantly different from zero. If we assume that the underlying random variable is Gaussian, a reasonable assumption in finance, we can test whether the correlation is different from zero employing the t-test given by:

$$\frac{\sqrt{n-2} \times Corr(X_i, X_j)}{\sqrt{1 - Corr^2(X_i, X_j)}}, \tag{3.13}$$

which can be shown to be distributed as a student's T distribution with $N - 2$ degreees of freedom.

We remark that we could also employ partial correlations, as in graphical Markov models; however, this would make the computations and the interpretation of the results quite challenging.

Note that our proposed correlation network model is based on a random variable, X, that takes different values, for different borrowers, and in different time periods. In practice, when data on borrower companies are available, for example from their annually reported balance sheet, we may observe many of such variables, and, therefore, we can construct more than one correlation network model. This requires the construction of a multi-layer correlation network model.

A multi-layer correlation network can be mapped into a tensor $\mathcal{X} \in \mathcal{R}^{I\times I\times K}$ where I represents the number of borrowers and K the number of considered random variables. Each element of the tensor, x_{ij}^k represents the correlation between borrower i and borrower j, using variable k, as in formula (3.11). The tensor is composed by K weight matrices $X \in\mathcal{R}^{I\times I}$, each of which represents a correlation network between borrower companies, using one variable.

We remark that each weight matrix can be transformed into an adjacency matrix, according to the testing procedure in (3.13).

It is important to understand how a multi-layer network model operates on the available data. Suppose we have data on I companies, in N distinct time points, according to K measurement variables. The data can be organised into a longitudinal data array, A, with dimension $I \times K \times N$, a sequence of $I \times K$ observations in N time points. A multilayer correlation network takes as input the array A and produces as output a cross-sectional array of dimension $I \times I \times K$, a collection of pairwise correlations between all companies, derived according to K different measurements.

One drawback of the proposed multi-layer correlation networks is that it only considers linear relationships between the variables and thus ignores any non-linear dependencies. In addition, this approach gives rise to K different correlation networks which, particularly when the measurements are highly correlated with each other, may be redundant. One way to address the latter problem is to project the available data matrix in a lower dimensional space, based on the most important principal components, that are by construction uncorrelated with each other (Geron, 2017). However, this may lead to a loss of information, the greater the lower the percentage of variability explained by the chosen principal components. An alternative procedure is to embed all the information contained in the different layers into a linear model, along with other exogenous explanatory variables, which is the approach followed in this section.

It is evident that a multilayer network is a complex object, which requires, to be utilised, some form of summarisation. Centrality measures are useful network summaries, that can be extended to the multi-layer context, as shown in Avdjiev, Giudici, and Spelta (2018).

For simplicity, and ease of intepretation, here we consider, for each measurement variable k, the degree centrality, which in our corrrelation network context indicates the total number of nodes to which a node is significantly correlated. Or, equivalently, the number of edges connected to a particular node.

For a correlation network $\mathcal{G} = (V, E)$ described by the binary edge set E, the degree centrality of a node $x \in V$ is defined by:

$$d_x = \sum_{y\neq x} e_{xy} \tag{3.14}$$

From a statistical viewpoint, the degree centrality is the simplest and most interpretable centrality measure. In addition, it is quite robust to changes in the topology of a network: for instance, adding or removing one node has a very limited effect, in a large network, as the degree of each node can go up or down at most by one unit.

From an economical viewpoint, the existence of a positive significant correlation between two borrowers can indicate that they have the same buyers, or that they operate in complementary markets. It seems intuitive that, if an active company is positively correlated with several defaulted companies, its credit scoring should be negatively affected. Conversely, the existence of a negative significant correlation between two borrowers can indicate that they compete in terms of buyers and/or markets. It seems intuitive that, if an active company is negatively correlated with several defaulted companies, its credit scoring should be positively affected.

In the P2P lending context, it is very important to evaluate the extent to which defaulted companies can affect active companies. Indeed a correlation network for P2P lending contains two types of distinct nodes: (i) defaulted and (ii) active companies. Let V be a set of vertices, E a set of edges, and S a binary variable that indicates whether a company has defaulted $(S = 1)$ or not

($S = 0$). A network model can be rewritten as

$$\mathcal{G} = (VxS, ExS) \tag{3.15}$$

giving rise to two distinct types of "marked" vertices: $(v, s = 1)$, $(v, s = 0)$; and three distinct types of edges, that will be named, without loss of generality, A, B, C: $A = (v_x, s = 1), (v_y, s = 1)$, $B = (v_x, s = 0), (v_y, s = 0)$, $C = (v_x, s = 1), (v_y, s = 0)$. According to this decomposition, the degree centrality of a company can be decomposed into the sum of three distinct centralities, as follows:

$$d_x = \sum_{y \neq x} e_{xy} \in A + \sum_{y \neq x} e_{xy} \in B + \sum_{y \neq x} e_{xy} \in C = d_A + d_B + d_C \tag{3.16}$$

The decomposition of the degree centrality in the three components can give important insights. For example, if the main aim of the analysis is to predict in advance default cases, it seems natural to consider the "type C" centrality, measured by:

$$d_C = \sum_{y \neq x} e_{xy \in C} \tag{3.17}$$

as this measure informs on which active companies are in contact with many defaulted companies. We will discuss these implications in more detail in the application section.

The final part of our model specification is to embed the obtained centrality measures, one for each measurement, into a predictive model. We propose to extend Chinazzi et al. (2013), who incorporate network measures in a linear regression model, to the logistic regression context, and taking the multi-layer dimension into account through an additive linear component. More formally, our proposed network-based scoring model takes the following form:

$$ln(\frac{p_i}{1 - p_i}) = \alpha + \sum_j \beta_j x_{ij} + \sum_k \gamma_k g_{ik} \tag{3.18}$$

where p_i is the probability of default, for borrower i, $x_i = (x_{i1}, \ldots, x_{ij}, \ldots, x_{iJ})$ is a vector of borrower-specific explanatory variables, g_{ik} is the degree centrality measure for borrower i, under the measurement k, the intercept parameter α and the regression coefficients β_j and γ_k, for $j = 1, \ldots, J$ and $k = 1, \ldots, K$ are to be estimated from the available data.

It follows that the probability of default can be obtained as:

$$p_i = \frac{1}{1 + e^{\alpha + \sum_j \beta_j x_{ij} + \sum_k \gamma_k g_{ik}}} \tag{3.19}$$

We expect that by *augmenting* a logistic regression credit scoring model, by means of the proposed centrality measures, its predictive performance will improve.

3.2.2 Similarity-based Credit Scoring Model

The correlation-based credit scoring model developed in Section 3.2.1 shows how the traditional logistic classifier can be augmented to account for the interconnections that emerge between borrower companies active on a P2P lending platform. This allows us to obtain deeper insights into otherwise unobservable similarities between economic agents which in turn might improve the platform's ability to discriminate between different risk classes. In terms of the practical rational for such an approach, we argue that the existence of a positive or negative statistically significant correlations between two borrowers can be indicative of joint, unobservable forces exitsing between borrowes (i.e. same buyers, or servicing complementary markets). It seems thus intuitive that, if a well performing company is significantly (and positively) correlated with several defaulted or bad-performing companies, its credit scoring should be affected.

The main limitation in this context is related with the availability of time-varying data on borrowers. Namely, in the majority of the cases, companies applying for a loan to a P2P platform cannot access funding through traditional financial intermediaries mostly due to the lack of long credit history.

With the following approach, we address this concern precisely and propose how to infer the network structure emerging between borrowers without possessing time-varying information on their financial performance. In this case, we can rely on similarity patters between borrowers' features to extract meaningful networks revealing recurrent topological structures that provide additional information on the multivariate nature of credit risk (see Giudici, Sarlin, and Spelta, 2017b).

Let the financial information about borrowing companies be collected in a vector $\mathbf{x}_n$ representing the financial composition of institution n. We define a metric that provides the relative distance between companies by applying the standardized Euclidean distance between each pair $(\mathbf{x}_i, \mathbf{x}_j)$ of institutions feature vectors (see Mantegna, 1999, Fontes, Rodrigues, and Craig, 2005, M.Merigo and Casanovas, 2011, Berry, Guillen, and Zhou, 2010). More formally, we define the pairwise distance $d_{i,j}$ as:

$$d_{i,j} = (\mathbf{x}_i - \mathbf{x}_j)\mathbf{\Delta}^{-1}(\mathbf{x}_i - \mathbf{x}_j)' \tag{3.20}$$

where Δ is a diagonal matrix whose i-th diagonal element is S_i^2, being S the vector of standard deviation. Namely, each coordinate difference between pairs of vectors $(x_i - x_j)$ is scaled by dividing by the corresponding element with the standard deviation. The distances can be embedded into a $N \times N$ dissimilarity matrix $\mathbf{D}$ such that the closer the companies i, j features are in the Euclidean space, the lower the entry $d_{i,j}$. In other words, the stronger the similarity (i.e. the force that connects two companies' characteristic vectors), the shorter the length of the links connecting the institutions. Pairs of companies that are dissimilar receive higher weights since they are placed far away from each other, while values approaching zero are assigned to pairs with highly similar characteristics.

Although D can be informative about the distribution of the distances between the companies, the fully-connected nature of this set does not help to find out whether there exist dominant patterns of similarities between institutions. Therefore, the extraction of such patterns demands a representation of the system where sparseness replaces completeness in a suitable way. To accomplish this, we derive the Minimal Spanning Tree (MST) representation of borrowing companies' balance-sheet similarities (see Bonanno et al., 2003; Mantegna and Stanley, 1999; Spelta and Araujo, 2012, Chazelle, 2000, Graham and Hell, 1985). To find out the MST representation of the system, we perform hierarchical clustering by applying the nearest neighbor method. At the initial step, we consider N clusters corresponding to the N institutions. Then, at each subsequent step, two clusters l_i and l_j are merged into a single cluster if:

$$d\left(l_i, l_j\right) = \min\left\{d\left(l_i, l_j\right)\right\} \tag{3.21}$$

with the distance between clusters being defined as:

$$d\left(l_i, l_j\right) = \min\left\{d_{rq}\right\} \tag{3.22}$$

with $r \in l_i$ and $q \in l_j$. These operations are repeated until a single cluster emerges. This clustering process is also known as the single link method since one obtains the MST of a network. Given a connected graph, the corresponding MST is a tree of $N-1$ edges that provides the minimum value of the sum of the edge distances. More specifically, the hierarchical clustering procedure takes $N-1$ steps to be completed when the graph is composed by N nodes, and it exploits, at each step, a particular distance $d_{i,j} \in D$ to merge two clusters into a single one.

Once we obtain the final network, we extract relevant information from the topology of the network i.e. we compute different measures from complex network theory. In particular, the research in network theory has dedicated a huge effort to developing measures of interconnectedness, related to the detection of the most important player in a network. The idea of centrality was initially proposed in the context of social systems, where a relation between the location of a subject in the social network and its influence on group processes was assumed. Moreover, beside investigating the importance each institution has in the network, we are also interested in assessing whether the network is characterized by a community structure and to exploit such feature. This topological characteristic indicates the presence of sets of companies usually defined as very dense sub-graphs, with few connections between them. Individual communities can shed light on the function of the system represented by the network since communities often correspond to functional units of the system. Being able to identify these sub-structures within a network we can provide insight into how network function and topology affect each other.

Various measures of centrality have been proposed in network theory such as the count of neighbors a node has, i.e. the degree centrality, which is a local centrality measure, or measures based on the spectral properties of the graph (see Perra and Fortunato, 2008). Spectral centrality measures include the eigenvector centrality (Bonacich, 2007), Katz's centrality (Katz, 1953), PageRank (Brin and Page, 1998), hub and authority centralities (Kleinberg, 1999). These measures are feedback, also know as global, centrality measures and provide information on the position of each node relative to all other nodes.

For our purposes, we employ both families of centrality measures (Opsahl, Agneessens, and Skvoretz, 2010, X. Yan, Zhai, and Fan, 2013, Wei et al., 2012). In particular, for each node we compute the degree and strength centrality together with the PagePank centrality. The degree k_i of a vertex i with $(i = 1, ..., N)$ is the number of edges incident to it. More formally, let the binary representation of the network be $\hat{D}$ such that:

$$\hat{\mathbf{D}}_{ij} = \begin{cases} 1 & if \;\; d_{ij} > 0 \\ 0 & otherwise \end{cases} \tag{3.23}$$

Similarly, the strength centrality measures the average distance of a node with respect to its neighbours. Formally, the strength of vertex i is:

$$s_i = \sum_{j=1}^{N} \mathbf{D}_{ij}. \tag{3.24}$$

These centrality measures provide no information about the higher order similarities among institutions i.e. no information is provided about the way in which these similarities compound each other affecting the overall system.

The PageRank centrality (Halu et al., 2013, Pedroche, Romance, and Criado, 2016), on the other hand, measures the importance of a node in a network by assigning relative scores to all nodes in the network, based on the principle that connections to few high scoring nodes contribute more to the score of the node in question than equal connections to low scoring nodes. More formally, the PageRank computes the probability that a random walker will land on a given node. Suppose that each unit of input in the system moves according to a Markov process defined by an $N \times N$ transition probability matrix $p = [p]_{ij}$. Under a regularity condition (ergodicity of p), there exists a real, positive vector $\pi_{in} = [\pi_{in}]_i$, $i = 1, ..., N$ such that $\pi_{in} = p\pi_{in}$ and $\sum \pi_{in}(i) = 1$. This is the PageRank vector. PageRank computes the importance of each node in a directed graph under a random surfer model. When at a node, the random surfer can either: (i) transition to a new node from the set of out-edges, or (ii) do something else (e.g., execute a search query, use a bookmark). The probability that the surfer performs the first action is known as the damping parameter in PageRank. We use ε to denote the damping parameter. The second action is called teleporting and is modeled by the surfer picking a node at random according to a distribution called the teleportation distribution vector or personalization vector (Rossi and Gleich, 2012). When p is not ergodic, one typically assumes that with some small probabilities a unity of input moves from any i to any j, so π_{in} exists. The input PageRank is formally defined as (Giudici and Spelta, 2016):

$$\pi_{in} = \varepsilon \left(D\Phi + f d'_{out} \right) \pi_{in} + (1 - \varepsilon) f \tag{3.25}$$

The parameter $\varepsilon \in (0, 1)$ is a damping parameter that determines the relative importance of the matrix $\left(D\Phi + f d'_{out} \right)$ and the teleportation distribution f. D is the adjacency matrix of the MST representation of the network and Φ is a diagonal matrix with elements $\Phi_{ii} = \min\left(\frac{1}{k_{out,i}}, 1\right)$. The second component is fd'_{out} where d_{out} is a column vector with elements $d_{out,i} = 1$ if $k_{out,i} = 0$ and otherwise 0. The vector d_{out} identifies those individuals that have no outgoing links and avoids that the random walker "gets stuck" on a dead-end node. Furthermore, not all nodes in the network are necessarily directly connected to one another. Therefore, the PageRank is adjusted again so that with probability $(1 - \varepsilon)$ the walker is allowed to jump to any other node in the network according to f . This is the reason why the vector f is called the teleportation distribution.

Notice that, in our networks that are based on distances between object, the higher the centrality measures associated to a node, the more the node is dissimilar with respect to its peers (or with respect to all other nodes in the network).

In general, centrality measures rank vertices according to their systemic importance without paying attention to whether the network is characterized by a community structure. On the contrary, several studies have analyzed the empirical characteristics of different networks have found the presence of sets of institutions usually defined as very dense sub-graphs, with few connections between them, as a result of similar patterns at the micro-level (see Pecora, Kaltwasser, and Spelta, 2016; Spelta, Flori, and Pammolli, 2018). The Louvain Method for community detection is a method able to extract communities from large networks created by Blondel et al. (2008). The identified communities maximize system's modularity, a measure that quantifies the strength of the division of the system into communities of densely interconnected nodes that are only sparsely connected with the rest of the system (see Newman, 2006). The modularity of our system is:

$$Q = \frac{1}{2m} \sum_{i,j} [D_{i,j} - \frac{s_i s_i}{2m}] \delta(c_i, c_i) \tag{3.26}$$

where $d_{i,j}$ is the weight of the edge between nodes i and j, s_i is the sum of the weights of the edges attached to node i, c_i is the community to which node i belongs, $\delta(u, v)$ is equal to 1 when $u = v$ and zero otherwise, and $m = \frac{1}{2} \sum_{i,j} D_{i,j}$. The clustering algorithm has two steps. First, each company constitutes a single community. Second, we evaluate the modularity increase of each company when a different community is joined. The conguration that provides the maximum gain in modularity is kept and the process is repeated for all nodes until no further improvements occur.

3.2.3 Latent-Factor Classification

Both approaches presented in Section 3.2.1 and Section 3.2.2, respectively propose methodologies that would allow P2P lending platforms to account for the interconnections emerging between borrowers by including network centrality parameters into the model specification.

In this following methodology, we approach the problem of credit risk for P2P systems by presenting a latent factor-based classification technique to divide the population into major network communities (connected and non-connected nodes) in order to estimate a more efficient logistic model. By representing SMEs which have applied for a loan to a P2P lending platform as vectors in real space, expressed as linear combinations of orthogonal bases described by singular value decomposition (SVD), orthogonality becomes a metric for classifying the respective SMEs into communities (see Golub and Reinsch, 1971, Abdi, 2007).

The information contained in the eigenvectors and eigenvalues are at the center of all spectral graph partitioning approaches. In the context of this method, nodes (i.e. P2P borrowers) are partitioned in two groups such that borrowers connected to each other belong to the same community. Once the adjacency matrix based on the SVD approach is inferred, we estimate and compare the predictive accuracy of traditional scoring models for connected and not-connected nodes independently. Note that, in applying this approach, we attempt to replicate the segmentation practices which are an imperative factor of bank's service offering and main determinant of the accuracy of their scoring models.

To summarize, the latent-factor-based classification technique for improving credit scoring models in P2P systems, can be described by three consecutive steps. We first obtain the latent factors (positions) for each credit applicant (i); we then infer communities between connected and not-connected nodes (ii); last, we estimate and compare the predictive performance of logistic regression models, separately for connected and not-connected nodes (iii).

This approach contributes the literature by also proposing a new network-based scoring model, which different from the previous two methodologies, leverages the structure of network communities, obtained using all available information, rather than network centralities. By increasing the ability of P2P lenders to successfully discriminate between different risk classes, the proposed methodology indirectly helps stabilize the overall financial system. Namely, although regulators are greatly interested in managing credit risk exposure, they cannot apply such models as they are based on commercial data made available only to the P2P lending platform.

Latent Factor Model Let Z be an arbitrary $n \times m$ matrix. The singular value decomposition (SVD) of any $n \times m$ non-symmetric matrix Z can be expressed as (Gander, 2008)

$$Z = UDV' \tag{3.27}$$

where $D = \Lambda^{1/2}$ is a diagonal matrix, with Λ interpreted as the diagonal matrix of nonzero eigenvalues of $Z'Z$ and ZZ', U and V are matrices whose columns are the orthonormal eigenvectors of ZZ' and $Z'Z$ respectively. More specifically, U is referred to as the matrix of the left singular vectors that span the column space of Z and the columns of V span its row space and are referred to as the right singular vectors.

Let $X = (x_1, \ldots, x_n)'$ be a stacked collection of the institutional features. Following Hoff (2007), we relate the observations in X to the following model

$$X = Z + E = UDV' + E \tag{3.28}$$

where Z is the expectation of X and E is the errors assumed to be normally distributed with mean zero and covariance matrix Σ.

We assume the observed institutional attributes are driven by some unobserved underlying factors that signal the financial conditions of the institution. We consider a lower dimensional number of factors (i.e, $k < m$). Thus, following (3.28), we express each x_i as:

$$x_i = u_i DV' + \varepsilon_i = f_i V' + \varepsilon_i \tag{3.29}$$

where $u_i = (u_{i,1}, \ldots, u_{i,k})'$ is a k dimensional vector representing the i-th row of U, D is a $k \times k$ diagonal matrix, V is the $m \times k$ matrix of factor loadings and $f_i = u_i D$ is the vector of factor scores.

Network Inference Following the literature on latent space models, we consider a class of network models commonly referred to as inner-product models Durante and Dunson (see 2014) and Hoff (2008). In this framework, we project the latent factors onto a "social space" and nodes that are "close" to each other are more likely to be connected. Let A be an $n \times n$ binary adjacency matrix where $A_{ij} = A_{ji} = 1$ indicates a link between nodes i and j, and zero otherwise. We parameterize the ij-th entry of A via a probit mapping function given by

$$P(A_{ij} = 1 | f_i, f_j, \theta) = \pi_{ij} = \Phi(\theta + f_i' f_j) \tag{3.30}$$

where π_{ij} is the probability of a link between nodes i and j, Φ is the cumulative density function of the standard normal distribution, and θ is a constant. In estimating a network for a large number of nodes, a common approach is to approximate the network by a sparse structure. Following the literature on sparse graphical models, we set $\theta = \Phi^{-1}(\frac{2}{n-1})$. From (3.30), we define a link between nodes i and j by

$$A_{ij} = (\pi_{ij} > \pi_0) \tag{3.31}$$

Thus, node i is connected to node j if and only if the probability of a link between the two nodes exceeds π_0. The choice of the threshold π_0 may be put in correspondence with the proportions of the event of interest within the total sample.

Network scoring At the final step, once the communities of connected and non-connected companies are identified, a logistic regression model is trained and tested separately for each of the found communities. The predictive accuracy of the models accounting for the emerging communities in consecutively compared with a full-specification model that does not account for the different communities within the sample.

3.3 Assessing the Predictive Accuracy of the Models

For assessing whether accounting for the underlining network that emerges between P2P borrowers, has an ex-ante forecasting capability for predicting default events, we rely on standard measures from classification and machine learning literature.
For evaluating the performance of different model, we employ, as a reference measure, the indicator $\gamma \in \{0, 1\}$ that is a binary variable which takes value one whenever the institutions has defaulted on its loans and value zero otherwise. For detecting default events represented in γ, we need a continuous measurement $p \in [0, 1]$ to be turned into a binary prediction B assuming value one if p exceeds a specified threshold $\tau \in [0, 1]$ and value zero otherwise. The correspondence between the prediction B and the ideal leading indicator γ can then be summarized in a so-called contingency matrix, as described in Figure 6.

		Actual Class	
		Defaulted	Not Defaulted
Predicted Class	Defaulted	True positive (TP)	False Positive (FP)
	Not Defaulted	False Negative (FN)	True Negative (TN)

Figure 6: **Contingency matrix.** The figure reports the four possible cases for default signaling. The rows of the contingency matrix correspond to the true class and the columns correspond to the predicted class. Diagonal and off-diagonal cells correspond to correctly and incorrectly classified observations, respectively.

From the contingency matrix, we can easy illustrate the performance capabilities of a binary classifier system. To this aim, we can compute the receiver operating characteristic (ROC) curve and the associated area under the curve (AUC) value which is one of the main measurements used for the evaluation of predictive classifiers. The ROC curve plots the false positive rate (FPR) against the true positive rate (TPR). To be more explicit:

$$FPR = \frac{FP}{FP + TN} \qquad \text{and} \qquad TPR = \frac{TP}{TP + FN} \tag{3.32}$$

Furthermore, the AUC depicts the true positive rate (TPR) against the false positive rate (FPR) depending on some threshold. As it is made clear by equation 3.32 TPR is the number of correct positive predictions divided by the total number of positives whereas FPR is the ratio of false positives predictions overall negatives.

One drawback of the AUC measure is that it is highly dependent upon the cut-off points or the unique probabilities of default estimated by the classifier. One way to overcome this limitation is by applying the Somers' D measure (Somers, 1952, Orth, 2012). Somers' D represents a conditional version of Kendell's coefficient which maps each combination between one observed value and one predicted value into simple binary measure: -1 in case of discordant pair and +1 in case of a concordant pair (Agosto, Giudici, and Raffinetti, 2019, Agosto and Raffinetti, 2019). Somers' D is often used in binary classification tasks and if there is no ties on independent variable, Somers'D is related to the AUC value in the following manner:

$$AUC = \frac{SD_{xy} + 1}{2} \tag{3.33}$$

$$SD_{xy} = \frac{P_c - P_d}{P_c + P_d + P_t} \tag{3.34}$$

where, SD_{xy} is Somers' D, P_c is the number of concordant pairs, P_d is the number of discordant pairs, and P_t is the number of neither concordant nor discordant pairs. This measure too has some drawbacks. Namely, it can be highly computationally intensive and it employs a rather crude binary summary which does not take into account the actual distance between each combination pair (Agosto and Raffinetti, 2019). In this context, the work by Agosto and Raffinetti (2019) proposes a new predictive classification accuracy measure that attempts to overcome the drawbacks associated with the Somers' D statistic. Specifically, the authors propose a measure that is based on the Lorenz curve and is obtained by reordering the observed response variable values in non-decreasing order and on the concordance curve, obtained by reordering the observed response variable values re-arranged with respect to the corresponding predicted value. Compared to the Somers' D statistic, the proposed RGA index is also a comparison between actual and predicted variable ranks but instead of relying exclusively on the ranks, the new index uses the actual value of the response variable corresponding to those ranks (for more detail, see Agosto and Raffinetti, 2019 and Agosto, Giudici, and Raffinetti, 2019). This measure can be very effective in real credit scoring tasks as it overcomes some of the main drawbacks identified with the frequently used predictive metrices.

Going back to the contingency matrix, another measure that can be easily calculated is the precision recall (PR) curve. The PR curve plots precision (P) versus the recall (R), or, more explicitly:

$$P = \frac{TP}{TP + FP} \quad \text{and} \quad R = \frac{TP}{TP + FN} \tag{3.35}$$

The PR curve complements the ROC curve since the former is directly influenced by class (im)balance because of the false positive measure. The PR curve is thus better to highlight differences between models for highly imbalanced data sets where the cardinality of default events is much lower with respect to the number of active institutions.
Moreover, we also compute other measures for assessing models performance such as the accuracy and the F1-score. The accuracy of each model can be computed as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.36}$$

and it characterizes the proportion of true results (both true positives and true negatives) among the total number of cases under examination.

The F1-score is also a measure of a classification accuracy. It considers both the precision and the recall of a model classification to compute the score. The F1-score is computed as the harmonic average of the precision and recall, where an F1-score reaches its best at 1 (standing for perfect precision and recall) and worst at 0. It can be computed as:

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{3.37}$$

Finally, since the main scope of this work is to provide evidence of the usefulness of network measures for discriminating between defaulted and active institutions, in some of the use cases, we also compute the Net Reclassification Improvement (NRI) as in Pencina et al. (2008). Briefly, consider a situation in which defaulted probabilities are estimated using two models that share only a sub-set of explanatory variables. Define upward movement (up) as a change into higher category (defaulted) based on the the model feeded with the new variables and downward movement (down) as a change in the opposite direction. The NRI focuses on reclassification tables constructed separately for companies facing or not the defaulted event, and quantifies the correct movement in categories — upwards for events (defaults) and downwards for non-events (actives).

For backtesting, while assessing the performance of each model, available information must be exploited in a realistic manner. To this end, for each of the use cases, we perform out-of-sample testing by employing a k-fold cross validation approach. In a nutshell, a model is usually given a data set of known data on which training is run (training data set), and a data set of unknown data against which the model is tested (validation data set or testing set). The goal of cross validation is to test the ability of the model in forecasting new data that was not used for estimating it. One round of cross validation involves partitioning a data set into complementary subsets, performing the analysis on one subset (called the training set), and validating the analysis on the other subset (called the validation set or testing set). To reduce variability, we apply 10 rounds of cross validation using different partitions, and the validation results are averaged over the rounds to give an estimate of the model's predictive performance.

4 Empirical Results: Traditional Scoring vs Network-Based Scoring

4.1 Datasets Used

For the purpose of achieving the main aim of this thesis, several different data sets were used. Each is explained in detail in the following sections.

4.1.1 Consumer Lending

For the purpose of testing whether the ratings assigned by P2P platforms are a good predictor of default, data was collected from LendingClub (**DATA-SET A**), which is the largest online marketplace connecting borrowers and investors. The analysis relied on loans' data covering the period 2007– 2016, obtained from the platform's official web page. Specifically, the data set contained all loans approved by the platform within the period 2007-2016. The full list of variables used for the analysis is presented in Table 4. The important variables of interest are the ratings ("grades") assigned to each loan applicant and the status of the loan, which allows identification of the proportion of those that have defaulted over the period of analysis. Loans that at the time of the analysis were still ongoing, that is, had a *current* loan status, were removed from the analysis.

4.1.2 SME Lending

For the purpose of testing whether the inclusion of network information, or information on how borrowers are connected with each other, can improve the predictive accuracy of the credit scoring models employed by P2P platforms, we used several different data sets. Specifically, a collaboration was established with a European Credit Assessment Institution (ECAI) that supplies credit scoring to P2P platforms specialised in business lending. The ECAI provided us with different data sets on SMEs that have applied for a loan at a P2P lending platform. All obtained data sets are explained in detail in the following paragraphs.

DATA-SET B. DATA-SET B is composed of financial information on 727 companies which have asked for a loan to a specific P2P lending platform in 2017. In order to provide certain amount of anonymity, the ECAI did not provide information on the specific platform on which the SMEs have applied for a loan. Table 9 provides the summary statistics of the variables included in this data set. Going into more detail, most of the SMEs included in the sample are businesses with operations in Southern Europe and specializing in manufacturing. The specific country of operation was not provided. he proportion of observed defaults in the sample is equal to 6.01%. The data-set contains only variables that are provided to the platform and necessary for evaluating the companies' creditworthiness.

The available data include the status of the companies, classified as [1 = Defaulted] and [0 = Active] which in turn can be used to estimate a credit scoring model, aimed at predicting the default status, on the basis of the observed values of a set collection of financial variables, derived as ratios from the yearly balance sheet of each company.

What is noticeable from Table 9, is that, as in most real-world data sets (and particularly those reflecting the operations of start-ups and small and medium enterprises), for most variables, there is a noticeable presence of unusually large or small values when compared to the mean. The literature recognizes many methods for dealing with outliers however in most cases the correct application of these methods is based on very strong assumptions concerning the size and distribution of the data set as well as the randomness of the outliers. In this context, we do not substitute or cancel outliers because we believe they can provide important insights concerning the companies included in the sample. Indeed, what we do is carry out a row-standardization of the data frame before training various classifiers.

Variable	Explanations
Grade	Assigned loan grade
Annual income	The self-reported annual income provided by the borrower
Loan amount	The listed amount of the loan applied for by the borrower
Loan over income ratio	Loan amount divided by annual income
Ownership	Binary variable taking value [1] if the borrower owns/morgages his home
Total number of accounts	The total number of credit lines currently in the borrower's credit file
Fico score	The upper boundary range the borrower's FICO belongs to.
Inquiries	The number of inquiries in past 6 months
Address_group1-3	Locations were classified into four groups based on default odds ratios
Months since last delinquency	The number of months since the borrower's last delinquency.
Revolving balance	Total credit revolving balance
Revolving line utilization rate	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit.
Verification status	Indicates if income was verified by Lending club
Debt ratio	A ratio calculated using the borrower's total monthly debt payments
on the total debt obligations	
Purpose1-3	Debt purposes were classified into four groups based on default odds ratios
income_group1-3	Interaction between annual income and purpose group
loan_group1-3	Interaction between loan over income ratio and purpose group
balance_group1-3	Interaction between revolving balance and purpose group
util_group1-3	Interaction between revolving line utilization rate and purpose group
total_group1-3	Interaction between total number of accounts and purpose group
issue_year07-16	Dummies referring to the year
fico_group1-3	Interaction between fico score and purpose group

Table 4: **Explanations of variables (DATA-SET A)**

Finally, from the financial indicators obtained, only a portion of the variables were observable over time. Specifically, only for three of the ratios obtained, we had information for the period 2007-2015 and those ratios are: (i) activity ratio, (ii) solvency and (iii) ROE.

DATA-SET C. A second dataset obtained from the ECAI is composed of official financial information from 813 SMEs that have applied for a loan through a P2P lending platform in 2017. Similarly as it is the case with DATA-SET B, in order for the ECAI to ensure some level of anonymity, no information was provided on the specific P2P platforms on which the SMEs have applied for a loan. The difference between DATA-SET B and DATA-SET C is related with the amount of available financial data for the units of analysis. Namely, DATA-SET C has a total of 44 balance sheet and income statement variable of the 813 companies (from the year 2015) hence this data-set is augmented with additional information collected from Orbis – Bureau van Dijk database. This action was conducted by the ECAI itself hence the P2P platform not necessarily would opt-in for a service that augments the amount of information it can use for the scoring. Similarly as it is the case with DATA-SET B, the available data include the status of the companies, classified as [1 = Defaulted] and [0 = Active] which in turn can be used to estimate a credit scoring model. On a different note, this data set does not contain any information on the changes of the financial ratios over time. Another point of difference between DATA-SET B and C is related with non-available (NA) data. Namely, as it is the case with most studies, when collecting balance sheet data on companies, the required count or value for some variables are not available. Although this is a problem that arise in most empirical research relying on financial and non-financial data about economic agents, very rarely do authors stray from the complete cases analysis. This approach can be problematic as it considered inefficient and possibly leading to bias conclusions Briggs (2003). In line with this, with DATA-SET C, we adapt the imputation method for generating replacements values for missing data which in turn allows for the use of the full number of observations.

In the context of DATA-SET C, missing variables are present for four main variables indicating the leverage and profitability condition of the SMEs. Figure 7 provides the histogram of missing data. Looking at the histogram in Figure 7 we observe that the four variables for which most NAs are present are: (i) the leverage ratio, (ii) the return on investment ratio, (iii) the return on equity ratio and (iv) the return on sales ratio. These represent some of the most crucial financial ratios determining the companies' overall financial health which in turn impacts the probability of loan default.

The most used imputation methods take a set of complete predictors and returns a single imputation for each variable. With DATA-SET C, we follow a multiple approach, in which imputations are created by repeated calls to the elementary imputation function. Considering that all four variables which contain NAs are continuous variables, the imputation method, we employ is predictive mean matching A. Burton, Billingham, and Bryan (2007).

In the literature, there has been an extensive discussion on when it is appropriate to use an imputation function rather than following the complete case approach. In most cases, the choice depends on two main elements: (i) whether the missing data mechanism is ignorable and (ii) whether the imputations contain information coming from outside the model used for predictions. In the context of this work, some negative implications from the use of the imputation function could arise. Specifically, the variance of the variables subjected to the imputation function may be reduced once the missing values are included which in turn could lead to bias estimates. However, we justify our choice by two main arguments. First, it is the view of the authors that the missing data mechanism is not negligible as the variables with the highest NA count are crucial determinants of the probability of default hence the use of the complete case approach, could lead to bias estimates. Second, the imputations are based on information from outside the model which are relevant to the outcome and at the same time are not predicted by the other covariates. Table 5 present a description and summary statistics of the variables in our sample.

DATA-SET D. The final data set provided by ECAI (DATA-SET D) is composed of information on 15045 companies across 25 financial indicators. In this case, borrowing information are not raw

		Active		Defaulted	
	Variables	Mean	SDev	Mean	SDev
1	Fixed assets	349.06	593.54	381.68	1453.28
2	Intangible fixed assets	38.62	95.7	32.72	125.31
3	Tangible fixed assets	283.53	530.26	291.86	1130.79
4	Other fixed assets	26.91	161.16	57.1	399.81
5	Current assets	663.44	437.48	466.62	1077.46
6	Stock	161.54	268.85	80.22	258.76
7	Debtors	256.97	249.02	131.94	337.05
8	Other current assets	244.93	259.3	254.47	818.28
9	Cash	87.62	140.98	118.15	601.56
10	Total assets	1012.5	818.47	848.31	2296.79
11	Shareholder funds	275.85	512.5	-504.46	2631.02
12	Capital	44.85	94.33	138.96	689.08
13	Other shareholder funds	231	495.42	-643.43	3300.18
14	Non current liabilities	232.2	360.65	471.97	1993.71
15	Long term debt	69.67	203.91	187.05	1129.91
16	Other non current liab	162.53	260.8	284.92	1048.35
17	Provisions	13.18	56.29	92.51	426.27
18	Current liabilities	504.45	355.38	880.81	2821.72
19	Loans	94.18	154.91	191.19	1053.4
20	Creditors	181.1	197.19	248.18	1089.63
21	Other current liab	229.17	291.15	441.44	1342.65
22	Total shareholder funds & liab	1012.5	818.47	848.31	2296.79
23	Turnover	946.36	305.56	407.57	892.36
24	Sales	923.46	310.38	419.86	887
25	Profit loss	33.91	118.43	-181.85	496.42
26	Financial revenues	2.03	20.44	8.42	69.93
27	Financial expenses	13.24	21.36	23.93	103.57
28	Financial profit loss	-11.21	29.58	-15.51	64.34
29	Profit loss before tax	22.7	122.08	-197.36	530.35
30	Taxation	13.63	25.62	1.21	19.84
31	Profit loss after tax	9.07	105.39	-198.57	527.76
32	Other revenues	3.48	16.53	26.91	148.37
33	Other expenses	3.12	10.97	17.81	97.01
34	Leverage ratio	6.61	7.73	0.39	7.57
35	Total asset total liab	1.58	1.46	1.45	3.16
36	Current ratio	1.67	1.37	1.75	3.64
37	Quick ratio	1.35	1.26	1.59	3.62
38	ROI	1.83	4.5	-2.62	4.59
39	ROE	13.73	23.46	16.4	33.17
40	Asset turnover	1.32	0.83	1.17	3.89
41	ROS	0.04	0.06	-0.01	0.05
42	Debt conversion ratio	462.03	8630.04	-2712.46	27698.69
43	Debt ratio	0.78	0.33	5.67	37.38
44	ROCE	0.67	2.96	-1.92	4.82

Table 5: **Summary statistics of the variables in DATA-SET C**

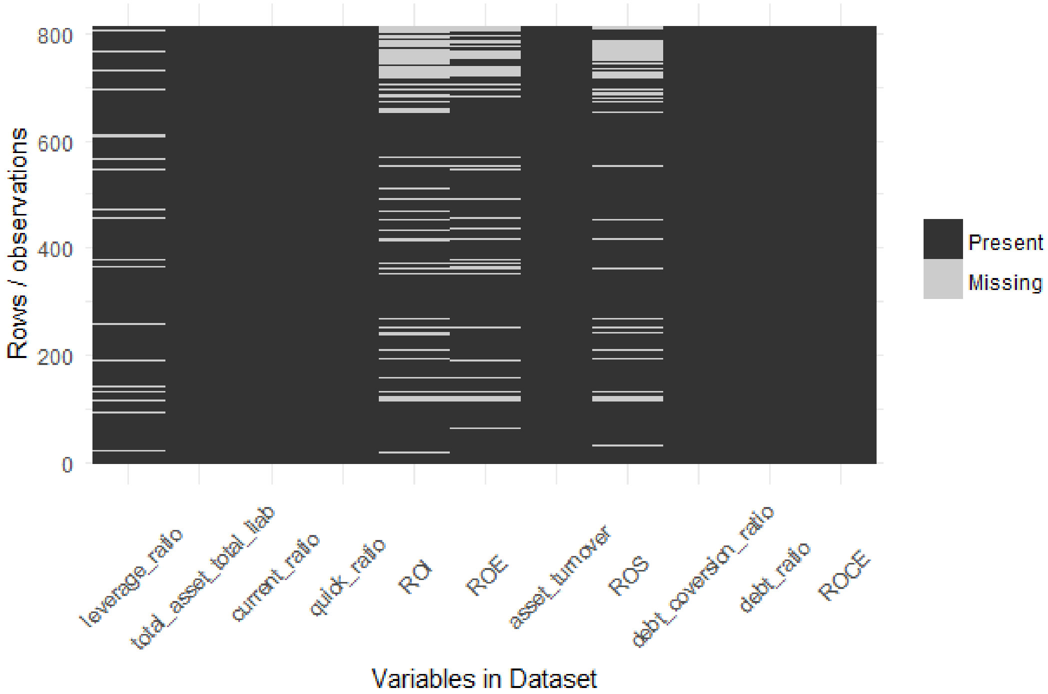


Figure 7: **Histogram of missing data (DATA-SET C)**

balance-sheet data but ratios of some of those statements. Table 6 provides a description of the formulas used to compute such ratios. Table 7 and 8, instead, provide the summary statistics of the numeric and nominal variables included in this data set. It is important to note that none of the variables included in DATA-SET D contain missing values and the proportion of defaulted companies is 10.85%. Similarly as in the context of DATA-SET B, we standardize the data frame before training various classifiers.

ID	FORMULA	Type	ID	FORMULA	Type
RATIO001	(Total assets - Shareholders Funds)/Shareholders Funds	Continuous	RATIO019	Interest paid/(Profit before taxes + Interest paid)	Continuous
RATIO002	(Long term debt + Loans)/Shareholders Funds	Continuous	RATIO027	EBITDA/interest paid	Continuous
RATIO003	Total assets/Total liabilties	Continuous	RATIO029	EBITDA/Operating revenues	Continuous
RATIO004	Current assets/Current liabilties	Continuous	RATIO030	EBITDA/Sales	Continuous
RATIO005	(Current assets - Current assets: stocks)/Current liabilties	Continuous	RATIO036	Constraint EBIT	Dichotomous
RATIO006	(Shareholders Funds + Non current liabilities)/Fixed assets	Continuous	RATIO037	Constraint PL before tax	Dichotomous
RATIO008	EBIT/interest paid	Continuous	RATIO039	Constraint Financial PL	Dichotomous
RATIO011	(Profit (loss) before tax + Interest paid)/Total assets	Continuous	RATIO040	Constraint P/L for period th EUR	Dichotomous
RATIO012	P/L after tax/Shareholders Funds	Continuous	DPO	Trade Payables/Operating revenues	Continuous
RATIO013	GROSS PROFIT/Operating revenues	Continuous	DSO	Trade Receivables/Operating revenues	Continuous
RATIO017	Operating revenues/Total assets	Continuous	DIO	Inventories/Operating revenues	Continuous
RATIO018	Sales/Total assets	Continuous	NACE	Industry classification on NACE code, 4 digits precision	Dichotomous

Table 6: **Description of variables (DATA-SET D)**

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
ratio001	15,045	8.895	19.382	−67.150	1.250	9.850	207.090
ratio002	15,045	1.258	3.341	−9.590	0.000	1.170	33.380
ratio003	15,045	1.462	0.799	0.100	1.070	1.540	8.270
ratio004	15,045	1.541	1.212	0.000	0.970	1.720	15.890
ratio005	15,045	1.189	1.007	−0.020	0.620	1.400	10.910
ratio006	15,045	7.841	23.883	−33.140	0.940	4.790	300.770
ratio008	15,045	23.228	72.012	−285.860	1.250	16.830	571.220
ratio011	15,045	0.028	0.147	−1.340	0.010	0.070	0.510
ratio012	15,045	−0.065	0.792	−8.930	0.000	0.210	1.280
ratio017	15,045	1.369	1.060	0.000	0.690	1.740	8.500
ratio018	15,045	1.331	1.056	0.000	0.660	1.700	8.500
ratio019	15,045	0.195	0.495	−3.490	0.010	0.380	3.950
ratio027	15,045	37.094	93.111	−208.860	2.530	28.410	750.360
ratio029	15,045	0.063	0.196	−2.080	0.020	0.110	0.940
ratio030	15,045	0.069	0.220	−2.660	0.020	0.120	1.410
DIO	15,045	104.540	351.178	0	0	80	5,569
DPO	15,045	76.215	114.619	0	0	100	1,493
DSO	15,045	95.753	132.277	0	0	135	1,531
turnover	15,045	3,397.433	7,532.013	2	602	2,759	79,454

Table 7: **Summary Statistics of DATA-SET D (numeric variables).** Summary statistics of the reference balance-sheet quantities. For each measure we report the average (Mean) along with the standard deviation (St. Dev.), the minimum (Min), the 25-th and 75-th percentiles (Pctl), the maximum (Max)

Variable	**Levels**	**n**	%	$\sum$ %
status	0	13413	89.2	89.2
	1	1632	10.8	100.0
	all	15045	100.0	
ratio036	0	12345	82.0	82.0
	1	2700	17.9	100.0
	all	15045	100.0	
ratio037	0	11956	79.5	79.5
	1	3089	20.5	100.0
	all	15045	100.0	
ratio039	0	969	6.4	6.4
	1	14076	93.6	100.0
	all	15045	100.0	
ratio040	0	11412	75.8	75.8
	1	3633	24.1	100.0
	all	15045	100.0	

Table 8: **Summary Statistics of DATA-SET D (Nominal variables)** [Levels = Levels of the variable; n = Number of observations in each level; % = % of total observations; $\sum$ % = Cumulative]

4.2 Empirical Results

In this section, we provide the empirical results from our analysis. Specifically, this section provides the results from the two objectives of the work i.e:

- Test whether the ratings assigned by P2P platforms are a good predictor of default;
- Test whether the inclusion of network information, or information on how borrowers are connected with each other can improve the predictive accuracy of the credit scoring models employed by P2P platforms.

Statistic	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max	Active	Defaulted
Turnover	2,432.079	13,590.780	24.279	557.056	689.907	230,236.000	2539.74	749.31
Profit-loss	87.250	696.575	−1,268.420	2.372	36.562	11,861.400	101.62	-137.42
Net Income	59.190	554.283	−1,597.930	0.468	24.573	9,953.360	70.37	-115.57
Total Assets	2,040.371	7,278.405	36.803	428.343	1,004.350	77,803.600	2082.15	1387.36
Shareholders' funds	556.100	2,131.850	0.493	35.175	322.306	28,282.000	563.60	438.90
Current ratio	1.618	1.495	0.032	0.949	1.701	12.828	1.57	2.39
Profit margin	2.256	12.622	−92.187	0.368	4.963	86.413	3.39	-15.49
ROE	12.168	96.703	−970.574	1.359	43.385	546.701	15.37	-37.85
ROCE	15.049	30.451	−219.233	3.066	24.257	145.057	16.72	-11.04
Solvency ratio	24.755	21.867	0.170	7.470	34.945	97.011	24.02	36.32
ROA	3.626	10.256	−80.348	0.242	6.364	54.470	4.31	-6.99
EBITDA	8.646	12.059	−60.161	3.871	12.936	92.573	9.45	-3.87
Long term debt	83.549	244.298	0	0	53.7	2,127	84.18	73.65
Loans	84.777	159.690	0	0	123.8	1,879	78.64	180.77
Current liabilities	404.553	455.344	10.446	196.856	458.060	4,940.398	386.55	686.00
Other current liabilities	189.081	335.987	1.462	41.845	214.764	3,726.131	179.66	336.38
Cash	56.814	93.070	0.003	3.872	65.566	707.161	55.87	71.61
Financial Expenses	12.741	22.306	−1.449	1.651	16.913	240.184	11.92	25.60
short_term_debt	593.634	753.368	20.892	277.971	660.184	8,666.529	566.20	1022.38
Debt to equity ratio	13.915	37.212	0.021	1.334	12.824	463.909	14.39	6.55
Cash over total assets	0.092	0.137	0.00000	0.004	0.136	0.849	0.09	0.14
Coverage	900.676	16,077.610	−2,237.300	0.033	8.319	340,031.000	965.40	-110.96
Activity	1.159	0.753	0.047	0.698	1.467	8.681	1.19	0.75

Table 9: **Summary Statistics of DATA-SET B.** Summary statistics of the reference balance-sheet quantities available in DATA-SET B. For each measure we report the average (Mean) along with the standard deviation (St. Dev.), the minimum (Min), the 25-th and 75-th percentiles (Pctl), the maximum (Max) and the values of the measures for Active and Defaulted firms.

Having in mind that we propose several approaches for building a high-performing credit scoring model, each subsection provides a brief overview of the specific objectives it aims to achieve as well as the data and methodology used. Finally, the subsections conclude with a discussion on the results and further steps.

4.2.1 Traditional P2P Lending Scoring Models: Do They Predict Default?

Objectives We claim that because of P2P's inability to solve asymmetric information issues as efficiently as traditional banks; and the difference in risk ownership between P2P and banks' business models, in the context of the P2P platforms, the grading system may not sufficiently reflect the probability of loan default. In this context, at first instance, we investigate whether there is a basis for such an argument i.e. we test the following hypothesis:

[H4.2.1] Ratings assigned by P2P lending platforms are not a good predictor of default.

Summary of Data and Methodology To test the specified hypotheses, DATA-SET A is used. In terms of the methodology, statistical theory offers a great variety of models for building and estimating the probability of default of lenders. All different approaches can be grouped in two broad categories: (i) parametric and (ii) non-parametric (Genriha and Voronova, 2012). For the purpose of reproducing the P2P grade-decision process and evaluating its performance in predicting loans' default, we employ the logistic regression explained in Section 3.1.1. In the context of P2P lending, logistic regression has been used in several the studies of Andreeva, Ansell, and Crook (2007), Barrios, Andreeva, and Ansell (2014), Emekter et al. (2015) and Serrano-Cinca and Gutiérrez-Nieto (2016a).

Status	Active [=0]	Defaulted [=1]
Turnover	2539.74	749.31
Profit-loss	101.62	-137.42
Net income	70.37	-115.57
Total assets	2082.15	1387.36
Shareholders' funds	563.60	438.90
Current ratio	1.57	2.39
Profit margin	3.39	-15.49
ROE	15.37	-37.85
ROCE	16.72	-11.04
Solvency ratio	24.02	36.32
ROA	4.31	-6.99
EBITDA	9.45	-3.87
Long term debt	84.18	73.65
Loans	78.64	180.77
Current liabilities	386.55	686.00
Other current liabilities	179.66	336.38
Cash	55.87	71.61
Financial expenses	11.92	25.60
Short term debt	566.20	1022.38
Debt to equity ratio	14.39	6.55
Cash to total assets	0.09	0.14
Coverage	965.40	-110.96
Activity	1.19	0.75

Table 10: **Average value of variables across status (DATA-SET B)**

Empirical results (Giudici and Misheva, 2018) For the purpose of investigating whether there is a basis for our premise that the grading system employed by P2P platforms may not sufficiently reflect the probability of loan default, we first present descriptive statistics. Table 11 provides an exploratory analysis of the continuous variables which Lending club collects from loan applicants. We investigate the average value of the indicators across different grades and what becomes clear is that for some of the variables there is not a significant variability between the highest and lowest grade (ex. total number of accounts, revolving balance). Looking at the individual indicators, the highest variability is noticed with the revolving accuracy and loan amount over income variables.

In order to see whether the platform is taking into consideration the right information when assigning the grades, we also consider the variability of the indicators with respect to loan status (Table 12). Overall, the averages are not significantly different which in turn can be an indicator that the platform should expand the scope of information necessary to accurately predict default.

Table 13 in turn, provides a cross tabulation with respect to the categorical variables. Considering the grade assigned, it is clear that there exists a relationship between the grade and the loan status. The table shows that 93.9% of the loans graded A did not default and the percentage decreases as the grades become lower. This can be considered evidence of the fact that the P2P lending platform does improve allocative efficiency as it supplies credits to consumers who are considered not creditworthy by traditional financial institutions. Similar arguments are also offered in other studies (see Serrano-Cinca, Gutiérrez-Nieto, and López-Palacios, 2015). However, cause for concern does exist when one considers that the majority of defaulted loans were ranked “C”. This can be considered an indicator of the upward bias discussed previously. Furthermore, the data presented in Table 13 indicates that a very small proportions of the applicants' self-reported infor-

Grade								
A	**B**	**C**	**D**	**E**	**F**	**G**	**Non-par test**	
Loan Amount	13196.8	12868.7	13770.4	14784.2	17837.2	19244.8	20924.9	0.000***
Annual Income	84481.48	73537.10	71158.83	69801.91	73464.61	74569.49	79923.55	0.000***
Debt Ratio	14.71	16.76	18.23	18.93	19.74	19.76	19.53	0.000***
Fico Score	736	703	692	687	686	683	681	0.000***
Revolving balance	16827	15639	15452	15156	16398	16175	16787	0.000***
Revolving line utilization rate	38.51%	52.61%	57.33%	60.15%	60.91%	61.82%	61.76%	0.000***
Total Number of Accounts	27	25	25	25	26	26	26	0.000***
Loan Amount over Income	.1786	.1982	.2149	.2333	.2675	.2814	.2934	0.000***

Table 11: **Descriptive analysis of categorical variables across grades (DATA-SET A).** The Ho hypothesis of the non-parametric test is that the distribution of the independent variable is the same across categories of grade.*** - significant at 1% l.s. Software used: R. Code available

	Loan Status	
	Not-Defaulted	**Defaulted**
Loan Amount	13979.3	15053.5
Annual Income	75478.61	67319.54
Debt Ratio	17.16	19.48
Fico Score	702	692
Credit revolving balance	15877	15337
Revolving line utilization rate	53.16%	58.14%
Total Number of Accounts	25	25
Loan Amount over Income	.2080	.2465

Table 12: **Descriptive analysis of continuous variables (DATA-SET A)**

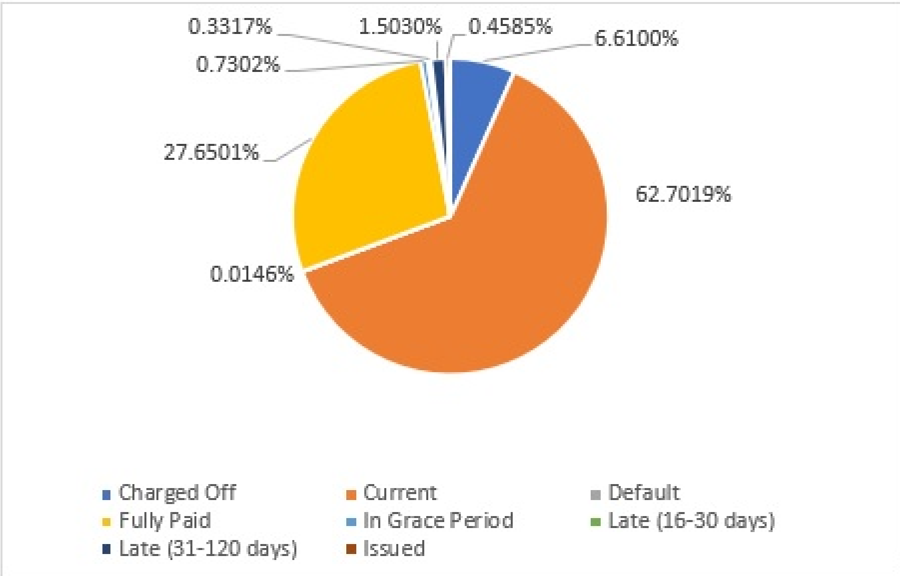


Figure 8: **Loan status variable (DATA-SET A)**

mation has been verified which is unexpected. Finally, there is some variability in the percentages of defaulted loans among different loan purposes indicating there might exists a need for a cluster- or network-dependent grading model.

Before testing 4.2.1, the Figure 8 represents the proportion of defaulted loans in the context of Lending club, for the period 2007 - 2016.

From Figure 8, defaulted loans together with those classified as charged-off comprise 6.6246% of total loans intermediated by Lending club. Although the proportion is not high, we believe that the frequency of the event is sufficient for developing and testing statistically predictive models. We remark that, in the case of lower default frequencies, and for robustness purposes, logistic regression could be extended with the Generalized Extreme regression scoring model proposed by Calabrese and Giudici (2015).

Predictive Performance of Lending Club's Scoring Model Before testing the predictive performance of Lending Club's grading system, we want to identify which information among those publicly available are most relevant in determining an applicant's creditworthiness. In order to

		Loan Status					
		Not Defaulted		Defaulted		Total	
		Count	Row N %	Count	Row N %	Count	Row N %
Grade	A	70723	93.90%	4574	6.10%	75297	100.00%
	B	124339	88.30%	16528	11.70%	140867	100.00%
	C	107626	81.10%	25158	18.90%	132784	100.00%
	D	61134	74.80%	20618	25.20%	81752	100.00%
	E	28411	68.40%	13132	31.60%	41543	100.00%
	F	9927	63.40%	5731	36.60%	15658	100.00%
	G	2452	59.90%	1644	40.10%	4096	100.00%
Term	36 months	316832	85.50%	53859	14.50%	370691	100.00%
	60 months	87780	72.40%	33526	27.60%	121306	100.00%
Home Ownership	any	8	100.00%	0	0.00%	8	100.00%
	mortage	205686	84.40%	37931	15.60%	243617	100.00%
	none	36	83.70%	7	16.30%	43	100.00%
	other	114	80.90%	27	19.10%	141	100.00%
	own	38663	81.80%	8630	18.20%	47293	100.00%
	rent	160105	79.70%	40790	20.30%	200895	100.00%
Verification Status	not verified	132397	86.30%	20945	13.70%	153342	100.00%
	source verified	132806	81.00%	31091	19.00%	163897	100.00%
	verified	139409	79.80%	35349	20.20%	174758	100.00%
Purpose of Loan	car	5126	87.70%	717	12.30%	5843	100.00%
	credit_card	85837	84.70%	15565	15.30%	101402	100.00%
	debt_consolidation	237721	81.30%	54604	18.70%	292325	100.00%
	educational	270	82.80%	56	17.20%	326	100.00%
	home_improvement	25295	84.80%	4529	15.20%	29824	100.00%
	house	2296	82.30%	494	17.70%	2790	100.00%
	major_purchase	9449	85.40%	1618	14.60%	11067	100.00%
	medical	4191	80.10%	1043	19.90%	5234	100.00%
	moving	2850	79.10%	754	20.90%	3604	100.00%
	other	21478	80.80%	5118	19.20%	26596	100.00%
	renewable_energy	338	79.50%	87	20.50%	425	100.00%
	small_business	5329	72.80%	1988	27.20%	7317	100.00%
	vacation	2475	82.20%	536	17.80%	3011	100.00%
	wedding	1957	87.60%	276	12.40%	2233	100.00%

Table 13: **Descriptive analysis of categorical variables across loan status**

achieve this, an attempt is made to reproduce Lending Club's grading process. Table 14 reports the findings from logistic regression.

In Table 14, results from three models are presented. Out of the 13 main variables included, twelve (among which ownership, total number of accounts, the FICO score, loan purpose, number of inquiries in the past 6 months, debt-to-income ratio, location, the number of months since borrower's last delinquency, revolving balance, and revolving line utilization rate) were found to have a statistically significant impact on the assigned grade. The results further suggest that annual income has no impact on the assigned grade which is somewhat surprising. Empirical research on the determinants of credit ranking in the context of traditional financial institutions have repeatedly found borrowers' income to be a significant determinant of the assigned rank or grade (Adams, B. Burton, and Hardwick, 2003, Jin and Zhu, 2015). The fact that there is not enough evidence to reject the hypothesis that annual income does not influence credit ranking in the context of Lending Club could be an indicator of the biased scoring model employed by this intermediary. Further evidence in support of this argument can be found in the estimated coefficient concerning the verification status. Common economic logic would dictate that the verification of the information provided by borrowers is of crucial importance to the credit ranking. Our empirical findings show that although the variable is found statistically significant, its sign is ambiguous.

In the next step, we proceed towards evaluating the predictive performance of the grades assigned by Lending Club with respect to loan default, in a second regression model. For this purpose, the estimated AUC is used. The AUC, which theoretically ranges between 0 and 1, is equal to 0.856, for the considered full sample model. Although the value of AUC equal to 0.856 can be considered sufficient, a need to increase the accuracy of the model is always encouraged. Even small increases in accuracy can lead to significant savings due to a superior predictive performance, as argued by West (2000). The results are presented in Table 15.

From Table 15, the grade variable is a statistically significant predictor of loan default, but its overall predictive power is limited. Namely, if we consider the estimated area under the ROC curve as a measure of predictive performance, the results suggest that the assigned grades do not have high predictive accuracy - the AUC value for the model using only the grade as a predictor of loan default is equal to 0.618. Furthermore, in order to investigate whether there is evidence if favor of the argument that grades are bias upwards we conduct additional diagnostic tests. The error matrix as well as additional statistics are presented in Table 16.

	Model 1			Model 2			Model 3		
	Estimate	**Std. Error**	**Sig.**	**Estimate**	**Std. Error**	**Sig.**	**Estimate**	**Std. Error**	**Sig.**
(Intercept)	-2.02E+01	1.22E-01		-2.09E+01	1.25E-01	***	-2.34E+01	1.34E-01	***
Annual income	-6.39E-08	7.02E-08							
Loan amount									
Loan over income				-3.72E+00	3.79E-02	***	-4.28E+00	3.78E-02	***
Ownership	9.33E-02	8.07E-03	***	9.12E-02	8.18E-03	***	2.75E-02	8.22E-03	***
Total number of accounts	1.57E-02	3.61E-04	***	1.30E-02	3.61E-04	***	1.17E-02	3.61E-04	***
Fico score	3.04E-02	1.69E-04	***	3.21E-02	1.74E-04	***	3.33E-02	1.76E-04	***
Inquiries	-4.52E-01	4.18E-03	***	-4.89E-01	4.26E-03	***	-5.04E-01	4.28E-03	***
Address Group 1	6.95E-02	8.29E-03	***	7.82E-02	8.42E-03	***	8.28E-02	8.45E-03	***
Address Group 2	-3.85E-01	1.11E-01	***	-3.94E-01	1.12E-01	***	-4.61E-01	1.13E-01	***
Address Group 3	-5.83E-01	1.56E-01	***	-5.33E-01	1.57E-01	**	-5.68E-01	1.58E-01	***
Months since the borrower's last delinquency	2.17E-03	1.71E-04	***	1.48E-03	1.73E-04	***	1.55E-03	1.74E-04	***
Revolving line utilization rate	-5.31E-01	1.89E-02	***	-5.22E-01	1.91E-02	***	-5.57E-01	1.91E-02	***
Credit revolving balance	2.97E-06	2.24E-07	***	2.47E-06	2.21E-07	***	7.79E-07	2.11E-07	***
Verification status	-7.74E-01	8.25E-03	***	-6.20E-01	8.45E-03	***			
Debt ratio	-3.81E-02	5.24E-04	***	-2.65E-02	5.21E-04	***	-2.62E-02	5.23E-04	***
Purpose1							2.00E+00	3.82E-02	***
Purpose2							1.39E+00	3.78E-02	***
Purpose3							4.90E-02	6.18E-02	
AUC	**0.7848775**			**0.7983443**			**0.8014684**		

Table 14: **Results - I.** Models (1), (2) and (3) are logistic regression models with grade as a dependent variable. Codes = *** $\alpha = 0.01$; * $\alpha = 0.1$

	Model 4			Model 5			Model 6		
	Estimate	Std. Error	Sig.	Estimate	Std. Error	Sig.	Estimate	Std. Error	Sig.
(Intercept)	-1.14726	0.01	***	4.72E+00	1.41E-01	***	3.54E+00	7.75E-01	***
Grade	-1.07895	0.01	***						
Annual income							1.14E-06	5.56E-07	*
Loan amount							1.04E-05	3.91E-06	**
Loan over income ratio				2.58E+00	4.12E-02	***	2.33E+00	7.03E-02	***
Ownership				-1.94E-01	9.23E-03	***	-1.78E-01	9.39E-03	***
Total number of accounts				-6.21E-03	4.20E-04	***	-8.10E-03	2.93E-03	**
Fico score				-1.05E-02	1.93E-04	***	-8.03E-03	1.05E-03	***
Inquiries				1.66E-01	4.17E-03	***	1.58E-01	4.21E-03	***
Address_group1				-1.70E-01	9.85E-03	***	-1.73E-01	9.91E-03	***
Address_group2				-5.09E-01	1.33E-01	***	-3.38E-01	1.36E-01	*
Address_group3				-1.65E+00	3.01E-01	***	-1.29E+00	3.01E-01	***
Months since the borrower's last delinquency				-1.89E-03	2.01E-04	***	-1.97E-03	2.02E-04	***
Revolving line utilization rate				2.15E-01	2.24E-02	***	1.41E-01	1.40E-01	
Revolving line utilization rate				-2.58E-06	3.03E-07	***	3.75E-07	1.25E-06	
Verification status				2.37E-01	1.05E-02	***	2.26E-01	1.08E-02	***
Debt ratio				2.30E-02	5.80E-04	***	2.48E-02	6.23E-04	***
Purpose1							1.49E+00	8.17E-01	.
Purpose2							1.44E+00	7.92E-01	.
Purpose3							6.63E-01	1.49E+00	
income_group1							-2.27E-06	6.03E-07	***
income_group2							-1.24E-06	5.67E-07	*
income_group3							-5.05E-06	1.51E-06	***
loan_group1							5.45E-07	4.03E-06	
loan_group2							-8.23E-06	3.92E-06	*
loan_group3							1.00E-05	8.76E-06	
fico_group1							-2.98E-03	1.11E-03	**
fico_group2							-2.74E-03	1.07E-03	*
fico_group3							-8.41E-04	2.04E-03	
balance_group1							-3.06E-06	1.38E-06	*
balance_group2							-3.10E-06	1.31E-06	*
balance_group3							1.51E-06	3.26E-06	
util_group1							-3.42E-02	1.46E-01	
util_group2							7.94E-02	1.42E-01	
util_group3							5.68E-02	2.44E-01	
total_group1							1.38E-04	3.03E-03	
total_group2							1.02E-03	2.96E-03	
total_group3							-3.84E-03	5.21E-03	
issue_year07							2.78E-01	2.32E-01	
issue_year08							1.57E-01	1.19E-01	
issue_year09							-7.06E-04	6.33E-02	
issue_year10							1.55E-02	3.92E-02	
issue_year11							1.24E-01	2.81E-02	***
issue_year12							-4.37E-02	1.85E-02	*
issue_year13							-1.02E-01	1.37E-02	***
issue_year14							2.61E-01	1.26E-02	***
issue_year15							1.08E-01	1.31E-02	***
issue_year16							-1.81E+00	4.15E-02	***
AUC	**0.6184062**			**0.6595419**			**0.6787838**		

Table 15: **Results - II.** Models (4), (5) and (6) are logistic regression models with loan status as a dependent variable. Codes = *** $\alpha = 0.01$; ** $\alpha = 0.05$; * $\alpha = 0.1$.

The confusion matrix (Table 16 and 17) provides evidence of the upward bias inherit in the P2P grading process due to both its inability to solve for asymmetric information and the different risk-ownership compared to traditional financial institutions. What we notice from Table 16 is that the model does not predict defaults. In this context, the result is associated with the default threshold which is 0.5. This might not be applicable to our context as the proportion of defaults in

		Reference	
		Not Defaulted	**Defaulted**
Prediction	**Not defaulted**	101247	21752
	Defaulted	0	0

Table 16: **Confusion Matrix**

the sample is significantly smaller hence the AUC is a more reliable measure of predictive accuracy as it measures the model's performance on different thresholds. The predictive performance of the default model does not change significantly even if we consider more advanced statistical models. In this respect, Table 15 also presents the results from two additional models aimed at capturing the determinants of loan default in the context of P2P platforms. An important finding from the conducted estimations is that the predictive performance of the scoring model improves by several percentage points once terms capturing the interaction between purpose and other control variables are included. Table 15 shows that several interaction terms were found statistically significant thus suggesting that same control variables can differently affect the probability of default dependent on the purpose for which the loan is taken. Although the predictive power of the scoring model increases as time and space predictors as well as interaction terms are included in the estimation, the improvements can be considered small as AUC values vary within the range 0.618-0.678. This is not to say that such improvements are irrelevant as even a small improvement in accuracy can lead to significant future savings (West, 2000). Still predictive performance below 70% represents a concern as a review of the literature suggest that values for the AUC above 70% are considered acceptable (Deloitte, 2016). Since the estimated models offer a predictive power below the acceptable limit, there is a clear need to increase the scoring accuracy of the credit decisions. What these preliminary insights suggest is that loan default in the context of P2P platforms is impacted by factors other than those observed and requested by Lending Club. In order to pursue improvements in the credit scoring models, it is thus necessary to explore other approaches beyond the traditional scoring models.

4.2.2 Correlation-Based Credit Scoring Model

Having obtained significant evidence that the scoring conducted by P2P platforms has the potential to perform poorly in discriminating different risk classes, in the following step, for improving the predictive accuracy of scoring models, we propose a network-based approach based on correlation networks.

Objectives. In this section, we empirically test whether the predictive performance of P2P credit scoring models can be improved using correlation networks. In other words, this section provides the results from the following two objectives:

- test the predictive performance of traditional scoring models employed by P2P lending platforms;
- test whether the inclusion of network parameters obtained from correlation networks that emerges between P2P borrowers, can improve the predictive accuracy of the scoring.

Summary of Data and Methodology We propose to augment traditional credit scoring methods with centrality measures derived from correlation networks among borrowers, deduced from the co-movement of their financial variables (DATA-SET B). As financial indicators we have chosen, without loss of generality, those reported in Table 18, which are among the most frequently reported in the literature.

Accuracy	0.8232
95% CI	(0.821, 0.8253)
Kappa	0
Mcnemar's Test P-value	0
Sensitivity	
Specificity	0
Pos Pred Value	0.8232
Neg Pred Value	NaN
Prevalence	0.8232
Detection Rate	0.8232
Detection Prevalence	1
Balanced Accuracy	0.5
Positive Class	0

Table 17: **Confusion Matrix: Additional Statistics**

Variable	Description
Activity ratio	Sales amount over total assets
Cash over total assets	Cash and cash equivalents over total assets
Coverage	Net income over financial expenses
Current ratio	Current liabilities over current assets
Return on assets	Net income over total assets
Return on equity	Net income over shareholders equity
Solvency ratio	Net income over total debt obligations
Total assets	Logarithm of total assets

Table 18: **The explanatory variablesn from DATASET B used for the analysis and their description**

From the set of variables kept, only a portion of the variables were available over the years. Specifically, only for three out of the eight financial indicators we were able to observe variations over time, and those are: (i) activity, (ii) solvency and (iii) ROE.

The methodology applied in this section can be summarized in the following steps:

- Train a logistic regression classifier using traditional financial indicators and test for predictive accuracy ;
- Using the three time-varying ratios, build correlation networks; If we consider each company to be a node in the network and we associate different time series with different nodes of the network, each pair of nodes can be thought to be connected by an edge with a corresponding weight which will be equal to the estimated correlation coefficient;
- Classify the emerging statistically significant links that emerge between borrowers and for each class calculate network centrality parameters;

- Train a logistic regression classifier using the augmented specification (including the network centrality parameters from the different classes of links emerging between borrowers).

- Compare the predictive accuracy of the traditional vs the network augmented classifier.

Empirical results (Giudici, Hadji-Misheva, and Spelta, 2019b) We first present the application of correlation networks to describe and summarise the relationships between the borrowers in a P2P platform.

As previously discussed, our proposed network models aim to infer the networking properties among the borrowers in a P2P platform from the time co-movement among the values that a given set of random variables take, when applied to their yearly financial statements.

As mentioned previously, for base variables, we take three well-known financial ratios: (1) the activity ratio, expressed as the ratio between sales and total assets; (2) the solvency ratio, expressed as the ratio between the net income and the total debt and (3) the return on equity ratio. Following this choice of variables, three time series of data can be extracted, for each of the 727 considered companies. Consequently, three 727 x 727 correlation matrices (weights) are obtained. Instead of using a fully connected correlation network, with all edges present, which would be as many as $(727 \times 726)/2 = 263901$, we consider a more parsimonious network, in which an edge between two companies is present whether the corresponding test in (3.13) is significant at a level of $\alpha = 0.01$. The application of the test gives rise to three 727 x 727 adjacency matrices, in which an edge is either present or absent, depending on whether the corresponding correlation is significant or not significant.

Figure 9 shows the network obtained using the the activity indicator to calculate correlations. In the Figure 9 nodes are colored based on their status, with red indicating companies that have defaulted in the considered period, and green - still active companies. The nodes are not equal but, rather, have a size proportional to their degree centrality, with bigger nodes indicating more connected ones. Edges are instead colored according to the sign of the found correlation: green for a positive correlation, and red for a negative correlation.

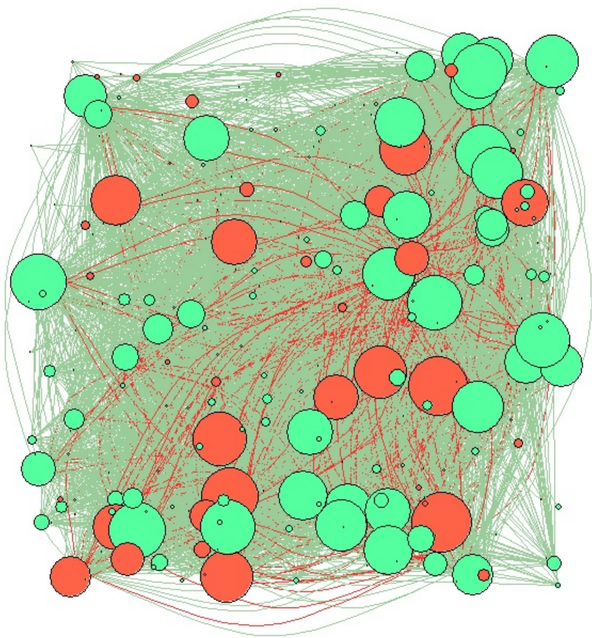


Figure 9: **Correlation network based on the activity indicator.** Number of nodes= 386.

Figures 9 indicates that 386 out of 727 companies (more than 50%) are significantly central

Edge type	N. of positive edges	N. of negative edges	N. of edges
A	381	35	416
B	1333	527	1860
C	1869	372	2241

Table 19: **Number of significant edges in the correlation network based on the activity indicator.**

and that, among them, there are both bad and good companies. The graph also indicates that both negative and positive correlations arise. To better interpret Figure 9 and, in particular, the significant correlations included in it, we can count, from the corresponding adjacency matrix, the number of positive and negative edges of the different types: "A" (between two bad companies); "B" (between two good companies); "C" (between a good and a bad company). The results are summarised in Table 19.

Table 19 shows that there are 3583 significant positive edges against 934 negative ones. This indicates that the activity ratio emphasizes mainly "similarities" between companies, expressed by positive edges, rather than dissimilarities, expressed by negative edges. From an economic viewpoint, a positive edge between two companies indicates that their two relative sales volumes move together i.e. they are complementary to each other so that when one fails the other is damaged too; a negative one indicates instead that they are competing on the market so that, when one fails, the other gets the corresponding market share. Table 19 indicates that, in the considered data, complementarity prevails.

Table 19 also shows that there are 2241 edges that indicate a significant positive correlation between a good and a bad company. When an active company is positively correlated with a defaulted one, a reduction in the sales of the latter, through complementarity, may cause a reduction in the sales of the good company, suggesting an increase in credit risk. When, conversely, the correlation is negative, competitiveness make the reduction of sales in the default company a reducing factor of the credit risk of a good company. Table 19 shows that the former effect prevails, with 1869 edges against 372, and this indicates the presence of a strong risk contagion effect, measured by correlation between sales or, more precisely, between the activity ratios. Indeed, a visual inspection of Figure 9 confirms this finding, as the proportion of defaulted companies that are central is larger than the observed proportion of defaults in the sample (6.01%).

Figures 10 shows the network obtained using the the solvency indicator to calculate correlations. The Figure 10 is based on the same assumptions used for the activity ratio, in terms of the significance level, and about the coloring and the dimensions of nodes and edges.

Figures 10 indicates that the central companies are less than before (286) and that most of them are good companies. Both negative and positive correlations arise, as before. To better interpret Figure 10 we calculate the number of positive and negative edges of the different types, as in Table 19, and report them in Table 20.

Edge type	N. of positive edges	N. of negative edges	N. of edges
A	1	2	3
B	377	81	458
C	51	17	68

Table 20: **Number of significant edges in the correlation network based on the solvency indicator.**

Comparing Figure 10 with Figure 9, we observe a lower presence of companies and, in particular, of defaulted companies. In addition, the few that can be visualized appear to have a small centrality. This can be an indication of the idiosyncratic nature of the solvency ratio, which leads to low correlations between defaulted and active companies. Conversely, what observed in Figure 9 for the activity ratio are high correlations between bad and good companies, which suggest the existence

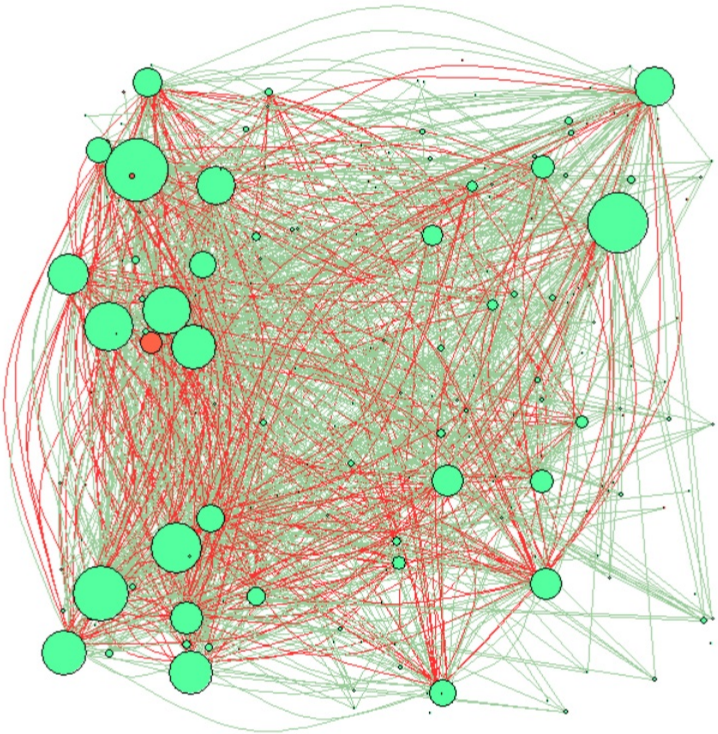


Figure 10: **Correlation network based on the solvency indicator.** Number of nodes= 288.

Edge type	N. of positive edges	N. of negative edges	N. of edges
A	0	0	0
B	780	633	1413
C	73	76	149

Table 21: **Number of significant edges in the correlation network based on the return on equity ratio.**

of a common systematic driver (such as the economic cycle), whose behaviour induces correlations between all companies.

Table 20 shows that there are only 529 significant edges against the 2241 in Figure 9. This reinforces the finding that the solvency ratio points out to a limited number of correlations between companies, among which similarities prevail. From an economic viewpoint, such similarities indicate that when one company increases debts, so does the other.

In particular, Table 20 shows that there are 51 edges that indicate a significant positive correlation between a good and a bad company. These indicate contagion risk as a defaulted company, with a high leverage, may induce a high leverage also for a good company.

Finally, we consider the network model that emerges using the correlations between companies calculated in terms of the return on equity indicator over the considered period. Figure 11 present the corresponding representation, maintaining the same assumptions as before.

Looking at Figure 11, the correlation network obtained using the return on equity indicator shows a low number of central nodes (226) and a limited presence of defaulted companies. These findings point towards the idiosyncratic nature of the return on equity indicator, which appears company specific, rather than driven by a systematic driver, consistently with the economic intuition.

To better interpret Figure 11 we report the number of positive and negative edges of the different types, in Table 21.

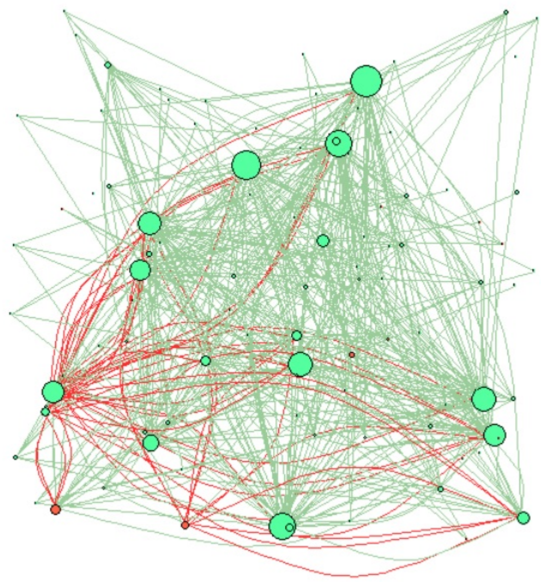


Figure 11: **Correlation network based on the return on equity ratio.** Number of nodes= 226.

Network	Statistic	Type A	Type B	Type C
Activity	Mean	16.64	23.25	21.6522
Network	Standard Deviation	16.58	29.2581	28.9158
ROE	Mean	1.2	10.6512	3.885714
Network	Standard Deviation	0.45	14.1827	4.10001
Solvency	Mean	0	18.47059	4.966667
Network	Standard Deviation	0	27.2817	6.86175

Table 22: **Degree centralities of the considered correlation networks.**

Table 21 shows that there are 1562 significant edges, less than in Table 1 but more than in Table 20. Edges are almost balanced between positive and negative ones. This indicates that there are both complementarity and competition effects among the considered companies, in terms of operational efficiency.

In particular, Table 21 shows that there are 73 edges that indicate a significant positive correlation between a good and a bad company. These indicate a contagion risk as a defaulted company, with a low ROE, may induce a lower ROE also for a good company.

To summarise this part of the analysis, Table 22 provides summary statistics about the degree centrality variable, along different edge types, for each correlation network.

From Table 22 note that the mean centrality measures reaches a high value for the activity ratio network: each company is connected, on average, to other 62. Among the latter, there are 22 of type C, and Table 2 indicates that most of them are between good and bad companies. The solvency and ROE correlation networks show much lower mean centralities.

Predictive modelling. Table 23 summarizes the results from the application of a logistic regression model to the available data and, more precisely, to the 80% of the data that has been randomly sampled to train the model.

From Table 23 note that two variables are found significant and those are: the activity ratio

	Estimate	P Value	Significance
Intercept	18.040	0.000	***
Activity ratio	-3.897	0.000	***
Cash over total assets	-0.636	0.772	
Coverage	0.000	0.672	
Current ratio	0.100	0.557	
Return on assets	-0.072	0.277	
Return on equity	-0.023	0.206	
Solvency ratio	0.004	0.774	
Total assets	-2.630	0.000	***
Area Under the Curve			0.622

Table 23: **The estimated baseline logistic regression model.**

and the total assets. The significance of the former is in line with the observation that, in the considered period, the companies in southern Europe, to which the companies in our sample belong, have suffered from a considerable decrease of the GDP. However, companies have reacted differently to recession: some, and especially those more oriented toward the internal markets, have shrunk their sales; others, and especially those more export-oriented, have maintained or increased their sales, thus explaining the significativity of the activity ratio: companies which are better able to use their assets to generate sales are less likely to default.

From Table 23 note that the estimated coefficient for the total asset variable also has the expected negative sign, suggesting that larger companies are typically less likely to default, compared to companies with smaller assets.

In the predictive analysis context, alternative models compete in terms of their predictive accuracy. As discussed previously, a widely used predictive accuracy measure for binary responses, that has the advantage of being independent of the chosen threshold, is the ROC curve and, in particular, the associated AUC, which summarises predictive accuracy into a single statistics (see e.g. Giudici (2003)). Table 23 reports also the AUC of the considered logistic regression model (corresponding to the 20% of the available observations). The AUC turns out to be equal to 0.622, which indicates the need to increase the accuracy of the model (see e.g. West, 2000).

To improve model accuracy, we employ correlation networks. To this aim, we augment the available data matrix with the degree centrality measures, calculated for each node on the basis of the correlation network models derived in the previous section. More specifically, we have added to the baseline logistic regression model the three variables corresponding to the degree centrality of each company in the Activity, Solvency and Return on Equity ratio. To avoid double counting, we have removed the original three variables from the logistic regression. Table 24 reports the results from the network based logistic regression model.

Table 24 shows that the degree centrality based on the activity ratio is significant, whereas those based on solvency and ROE are not. The sign of the significant centrality is positive. This means that the higher the centrality degree of a particular company, the higher the probability that it would be connected with a defaulted company, and this may negatively impact its overall probability of default. On the other hand, note that the negative sign of the Total assets variable is confirmed, although with a lower magnitude. The variable Return on Assets becomes significant, with the expected sign (higher values leading to a lower probability of default).

In terms of predictive accuracy, the model in Table 24 leads to an AUC of 0.836, which suggests that the inclusion of network centrality parameters does improve predictive accuracy.

We now consider, as an alternative network model specification, centrality parameters expressed rather than by the total degree of each node, by the type C degree. Table 25 contains the results of the corresponding network based logistic regression model.

From Table 25, we find that the conclusions from Table 24 are confirmed and reinforced. Both Total assets and Return on equity have a significant negative effect on the probability of default;

	Estimate	P Value	Significance
Intercept	-0.241	0.814	
Cash over total assets	0.600	0.612	
Coverage	0.001	0.672	
Current ratio	0.051	0.632	
Return on assets	-0.143	0.001	***
Total assets	-0.285	0.062	*
Degree centrality (Activity)	0.011	0.008	***
Degree centrality (ROE)	-0.038	0.412	
Degree centrality (Solvency)	-0.017	0.355	
Area Under the Curve			0.8357143

Table 24: **The estimated network based logistic regression model, with all type centralities.**

	Estimate	P Value	Significance
Intercept	0.024	0.983	
Cash over total assets	0.541	0.706	
Coverage	0.001	0.883	
Current ratio	0.001	0.983	
Return on assets	-0.150	0.001	***
Total assets	-0.428	0.018	**
Degree centrality (Activity)	0.018	0.001	***
Degree centrality (ROE)	0.043	0.067	*
Degree centrality (Solvency)	0.041	0.004	**
Area Under the Curve			0.8697479

Table 25: **The estimated network based logistic regression model, with type C centralities.**

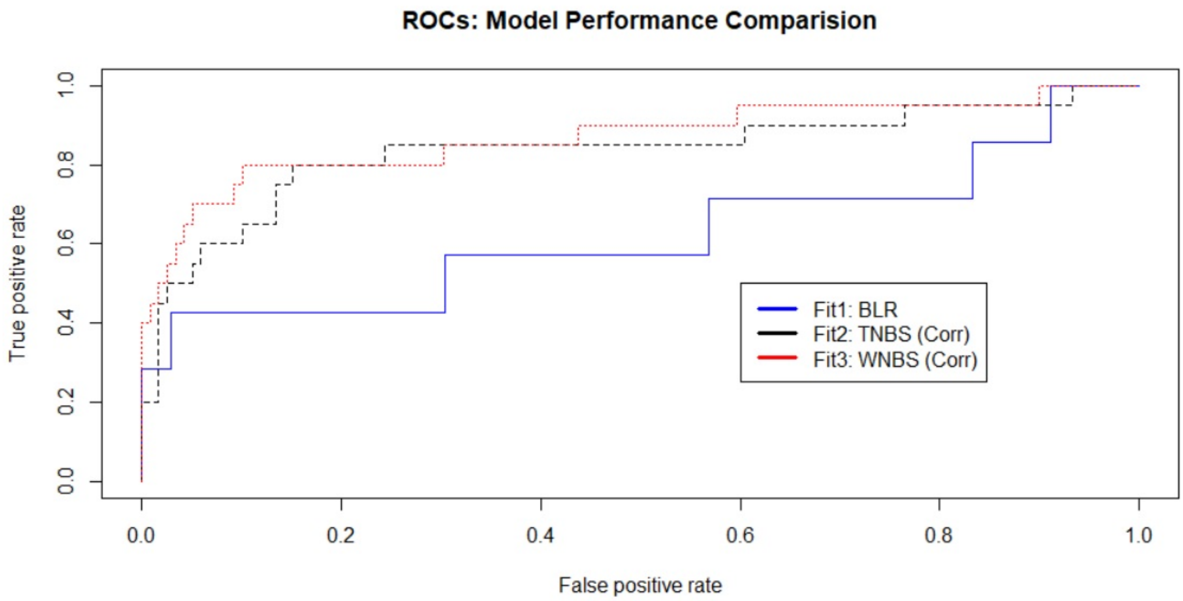


Figure 12: **Model Performance Comparison.** BLR indicated the baseline regression model; TNBS the network based model, with all edge types; WNBS the network based model, with only Type C edges.

while the activity ratio degree centrality maintains its significant positive effect, as before. In addition, both the solvency and the ROE degree centrality become significant, albeit with lower significance levels, and with a positive sign. This reflects the effect of the strong presence of positive edges of type C, which indicates a prevalence of complementarity and, hence, of a contagion effect, from bad to good companies.

From Table 25, note also that the AUC of the models improves of about three points from the previous Table 24. The predictive performance of the considered models can be compared in a more comprehensive way, contrasting the corresponding ROC curves, as in Figure 12.

Figure 12 shows that both network based models overperform the baseline regression model (BLR) and that, on the other hand, the network that includes only type C edges (WNBS) is slightly superior than the one with all edges (TNBS).

It may be worth to understand whether the difference in the out-of-sample performance of the three considered models, observed graphically comparing the corresponding ROC curves, is also statistically significant.

To this aim, we have applied the De Long test (see E. R. DeLong, D. M. DeLong, and Clarke-Pearson, 1988), which consists of a nonparametric statistical procedure to compare areas under two or more ROC curves based on the same data. The null hypotheses is that, for a given set of cut-off points on which they are calculated, two ROC curves are the same. The alternative hypotheses specifies that they are instead different. E. R. DeLong, D. M. DeLong, and Clarke-Pearson (1988) propose a test statistics which can be shown to be distributed, asymptotically, and under the null hypotheses, as a chi-squared distribution.

The application of the De Long test to our context shows that the difference between the ROC of TNBS and that of BLR is almost significant (p-value= 0.13); the difference between the ROC of BLR and that of WNBS is significant (p-value = 0.011); the difference between the ROC of TNBS and that of WNBS (p-value = 0.52) is not significant. These results suggest that the predictive accuracy improvement obtained moving from a baseline regression to a network-based regression is significant. Whereas the refinement of the network centrality measure to type C edges is not significant.

The improvement in predictive performance determined by network models can also be appreciated looking at the predictions of the alternative models for 10 randomly selected companies in the test dataset, 5 known to be bad (i.e. defaulted) and 5 known to be good (i.e. active). The results are reported in Table 26.

From Table 26 note how the estimated probability of default of active companies decreases, moving from the baseline to the network based model and, furthermore, to the network based model with type C centralities. Conversely, the probability of default of defaulted companies increases moving along the same direction. This confirms, on real out-of-sample cases, the better predictive accuracy of the proposed network based models.

4.2.3 Similarity-Based Credit Scoring

As shown in Section 4.2.2 augmenting P2P scoring models with centrality parameters obtained from correlation networks emerging between borrowers can significantly improve the predictive accuracy of the scoring models. The main concern in this context is that availability of time-varying data on borrowers. As stated previously, in the majority of the cases, companies applying for a loan to a P2P platform cannot access funding through traditional financial intermediaries mostly due to the lack of financial history.

Objectives. With the following approach, we address the concern of data availability (time-varying information) and propose how to infer the network structure emerging between borrowers without having time-varying information on their financial performance. Specifically, the objective is to show how topological information embedded into similarity networks can be exploited to increase the predictive performance of credit scoring models. Furthermore, in this approach, we cross-validate the necessity of accounting for the interconnections between borrowers for the task of credit scoring by testing the predictive accuracy of several classifiers.

Summary of Data and Methodology For this approach DATA-SET B and DATA-SET D are used. With the proposed similarity-based credit scoring approach, we show that the credit risk accuracy of peer-to-peer platforms could be improved by leveraging topological information embedded into similarity networks derived from borrowers balance-sheet features. Specifically, the methodology can be summarized in the following steps:

- For each data set used (i.e DATA-SET B and DATA-SET D), we have financial information about borrowing companies representing their financial composition. From this information, we define a metric that provides the relative distance between companies by applying the standardized Euclidean distance between each pair of institutions feature vectors;
- The adjacency matrix is used to build a graph from which we derive the minimum spanning tree representation of the borrower network and obtain our final representation of the intreconnectedness that emerges between borrowers;
- Relevant patterns of similarities describing institutions' importance and community structures are extracted from the networks and employed as additional explanatory variables for improving the performance of different classes of scoring models
- We compare the predictive performance of the classifiers with and without the augmented network centralities and community structures.

Company	PD from BLR	PD from TNBS	PD from WNBS	Status
Company A	0.325995937	0.203178979	0.102347865	Active
Company B	0.207411016	0.220493154	0.121339165	Active
Company C	0.198157788	0.101808476	0.047102588	Active
Company D	0.107315436	0.103854312	0.081768604	Active
Company E	0.006017395	0.000907596	0.000364361	Active
Company F	0.127879968	0.248898102	0.373049367	Default
Company G	0.002658514	0.125033683	0.149287131	Default
Company H	0.045663684	0.177499378	0.510471885	Default
Company I	0.000419074	0.040453597	0.06446989	Default
Company J	0.016839456	0.07174686	0.091508018	Default

Table 26: Comparison of PD Estimates across different models. BLR indicated the baseline regression model; TNBS the network based model, with all types; WNBS the network based model, with only Type C edges.

Empirical results. [4] First we report, for both data sets (DATA-SET B and DATA-SET C), the MST representation of the similarity network obtained from companies' feature distances. We report two types of network visualization, the first one shows nodes colored according to their financial soundness, red nodes represent defaulted institutions while green nodes represent sound and active companies, see Figure 13. Notice how, for both data sets, defaulted institutions occupy precise portion of the network, namely, such companies belong to the leafs of the tree and form clusters. This, in other words, suggests those companies share particular features and that unique positioning of the defaulted companies within the MST representation, indicates that the topological information obtained from the similarity network, are relevant in discriminating between different risk classes.

To highlight the retrieved clusters, we also report the networks in which nodes are colored according to the community they belong.

Figure 14 shows the log-log plot of the cumulative distribution function and maximum likelihood power-law fit for the centrality measures employed in the analysis. In the figure, we separate the cumulative distributions of such measures for the defaulted and non defaulted institutions. For all the measures and for both the two types of data sets, we observe different scaling of the power-law exponents for institutions belonging to the defaulted set and for the sound ones, this suggests that, potentially, the centrality measures that account for nodes' importance are useful variables for discriminating between companies.

In this approach, the information concerning the community structure of the networks and the centrality measures are used to provide synthetic topological variables at the node level. Such variables are embedded into the credit scoring models to assess whether they contain relevant information useful for forecasting institutions default.

Figure 15 reports the predictive accuracy of the different classifiers with and without the network augmentation, across the two different data sets. Basically, the upper left panel shows the results for the logistic regression applied to DATA-SET B, the upper right panel encompasses the same information for the support vector machine model applied to DATA-SET B, the bottom left panel refers to the performance curves of the discriminant analysis applied to DATA-SET B and finally the bottom right panel shows the results for the support vector machine classifier applied on DATA-SET D.

For sake of comparison, we have reported both the ROC and the RP curves to show that, overall, the inclusion of topological information regarding similarity patterns among companies feature, increases the forecasting performance of various credit scoring models even when the data sets are imbalanced between the two classes (defaulted v.s. active). Notice how, for most of the cases, red lines representing the performance of the models feeded with network measures lie above the blue lines representing baseline classifiers.

What is particularly noticeable from the figures presented is that the increase in the predictive accuracy (i.e. the difference between the red and blue lines), is highest for the classifiers trained and tested on the smaller data set (DATA-SET B). What this can be indicative of, is that the accuracy of the network parameters are more useful in the context in which one has limited information on the financial performance of the companies. This could be greatly relevant for P2P platforms which are often at a disadvantage concerning data availability compared to traditional financial intermediaries.

Performance improvements for all the tested models and for both data sets are also reported in Table 27. The table summarizes the values of the measure employed to assess the predictive gain of the network-augmented credit scoring models. We report, the area under the ROC curve (AUC), the area under the PR curve (AUPR), the Somers' D test (SomerD), the RGA index (RG) and Net Reclassification Improvement (NRI).

From the results collected in Table 27, it is quite clear that the inclusion of topological variables describing institutions centrality in the similarity networks and the community structure composing such networks, increases the predictive performance of most of the credit scoring mod-

[4]Paper: Giudici, P., Hadji Misheva, B. and Spelta, A. (2019). Network based scoring models to improve credit risk management in peer to peer lending platforms. Frontiers in Artificial Intelligence, 2

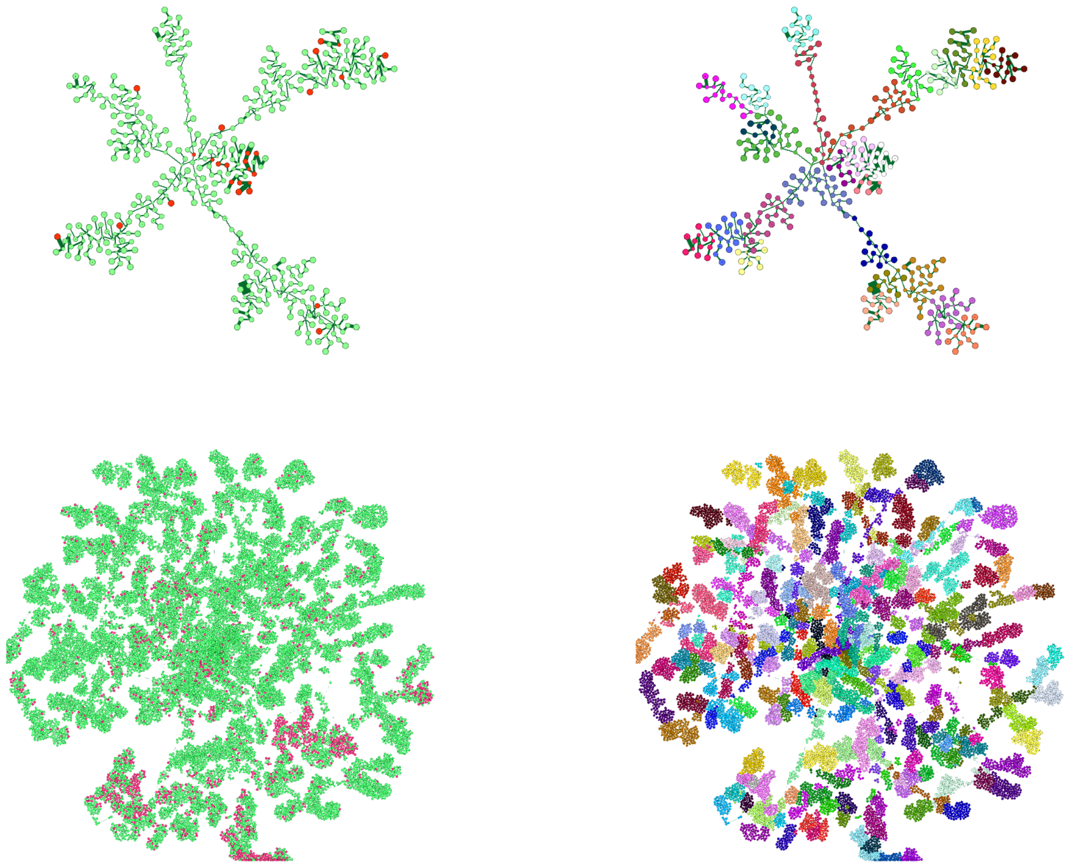


Figure 13: **Minimal spanning tree representation of the borrowing companies networks.** The tree has been obtained by using the standardized Euclidean distance between institutions features and the Kruskal algorithm. In the left panels, nodes are colored according to their financial soundness, red nodes represent defaulted institutions while green nodes are associated with active companies. The right panels show the same networks but nodes are colored according to the community they belong. Notice how defaulted institutions strongly occupy certain specific communities not being equally distributed among the networks. Moreover, the upper panels refer to the DATA-SET B while the bottom panels refer to DATA-SET D.

els. Comparing the performance of the RGA index to the conventional AUC measure, we notice that the former reports significantly lower values. This is expected considering that the RGA index is a more rigorous performance measure. This is due to the fact that it is calculated using all observed response values whereas the AUC is calculated on a selection of cut-off points. It is important to note, that the calculation of the RGA index is done using the original source material (open-source scripts) provided by the authors (Agosto and Raffinetti, 2019 and Agosto, Giudici, and Raffinetti, 2019).

4.2.4 Latent Factor Models For Credit Scoring

Both approaches presented in Section 4.2.2 and Section 4.2.3 respectively provide empirical evidence that the inclusion of network information can improve loan default predictions as it captures information that reflects underlining common features that cannot be observed otherwise. Both ap-

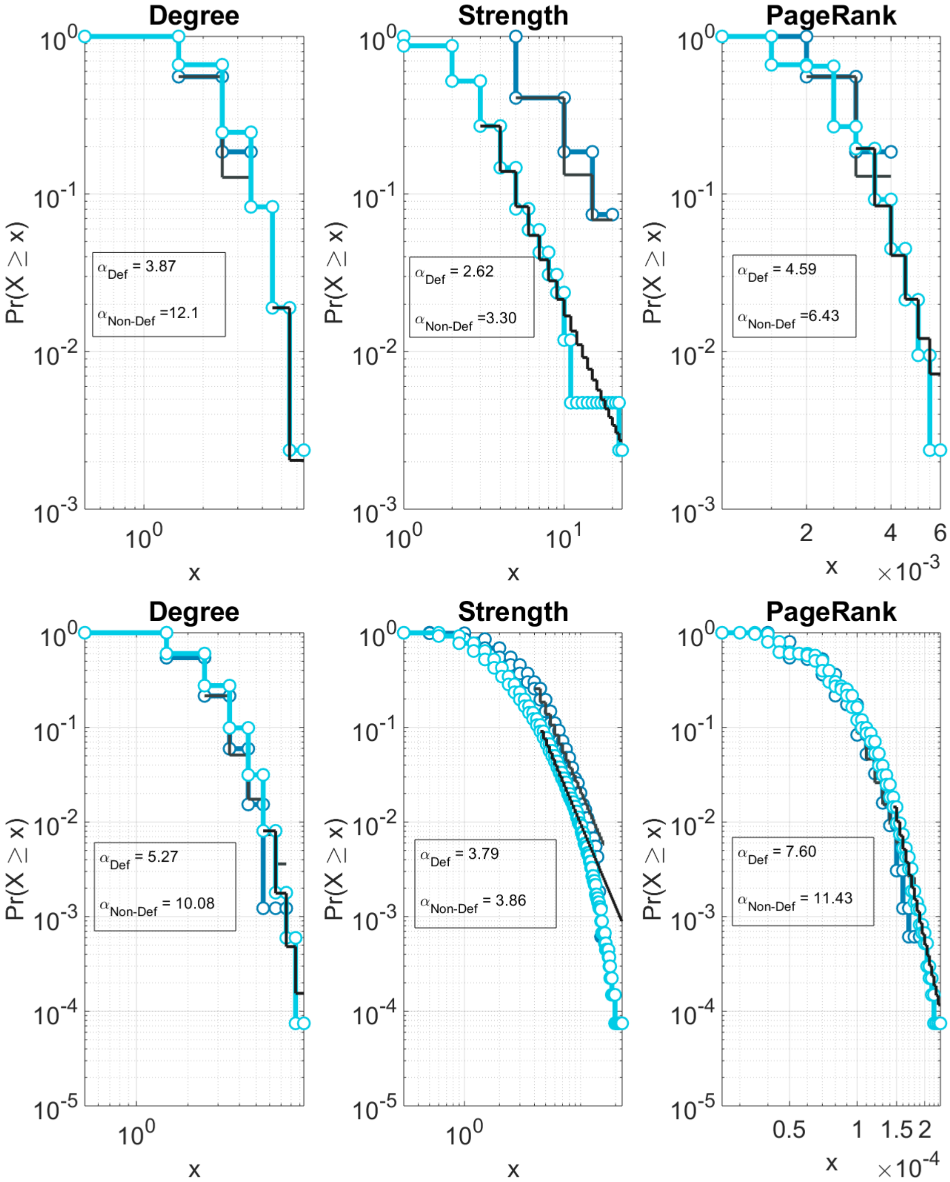


Figure 14: **Centrality measure distributions.** The panels represent the distribution of the centrality measures separated according to the defaulted indicator γ, together with the corresponding power-law coefficient estimate. In the left panels we represent the degree distributions, the central panels refer to the strength distributions while the right panels encompass the PageRank distributions. Moreover, the upper panels refer to the DATA-SET B while the bottom panels refer to DATA-SET D. The different values of the scaling coefficients related to the distributions of defaulted and active institutions suggest their potential value for discriminating between such companies.

proaches achieve this through the inclusion of network summary parameters (network centralities) into the credit scoring specification.

In this following attempt, we approach the problem of credit risk for P2P systems by presenting a latent factor-based classification technique to divide the population into major network communities in order to estimate a more efficient logistic model. Given a number of attributes

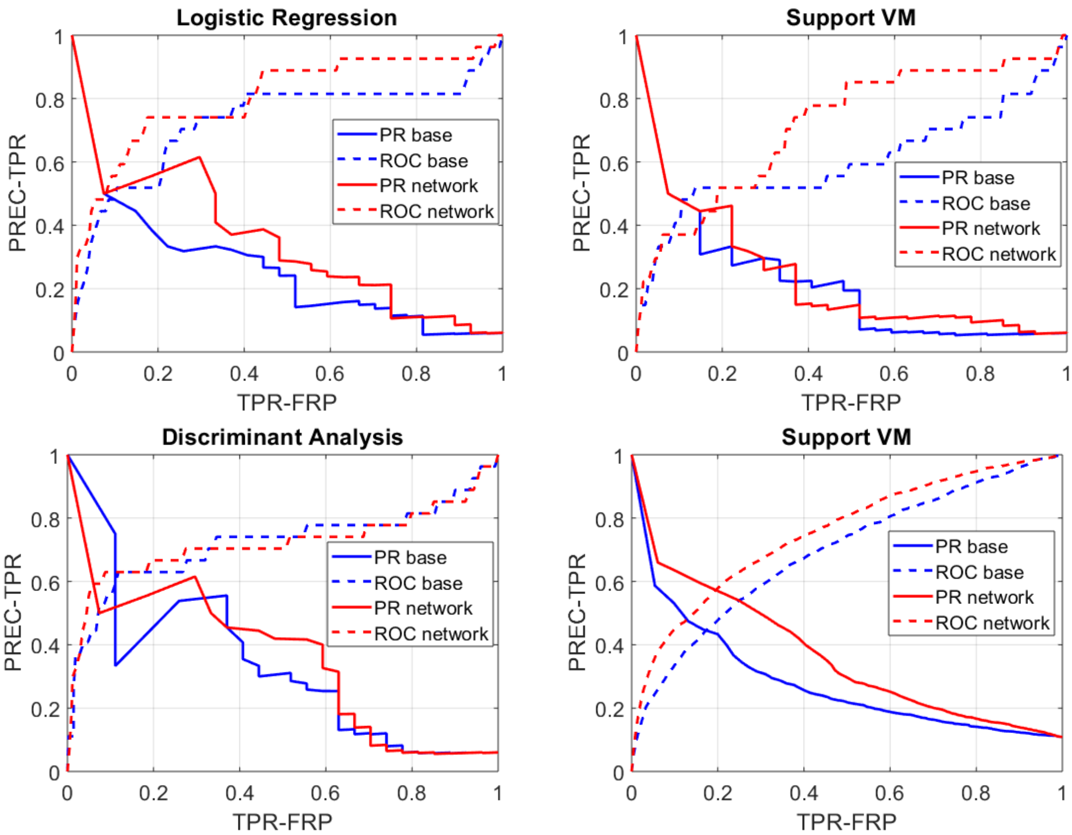


Figure 15: **Precision Recall (PR) and Receiver Operating Characteristic (ROC) curves for the baseline credit risk models and for the network-augmented models.** In each panel, dotted lines represent the ROC curves while solid lines refer to PR curves. In blue, we show the results related to the baseline models while in red we show the results related to the network-augmented models. The upper left panel refers to the logisitc Regression applied to DATA-SET A, the upper right panel shows the results related to the support vector machine model applied to DATA-SET A, the bottom left panel encompasses the performance curves of the discriminant analysis applied to DATA-SET A and, finally, the bottom right panels refer to the support vector machine applied on DATA-SET B.

that capture firm performances in a financial system, we adopt a latent position model which allow us to distinguish between communities of connected and not-connected firms based on the spatial position of the latent factors. Compared to the previous two methodologies, this approach leverages the structure of network communities, obtained using all available information.

Objectives. The main objective in this section is to investigate how factor-network-based segmentation can be employed to improve the statistical-based credit score for small and medium enterprises (SMEs) involved in P2P lending.

Summary of Data and Methodology In terms of the methodology, the following steps summarize the approach:

- construct a network of SMEs where links emerge from co-movement of the latent factors that drive the observed financial characteristics;
- segment the heterogeneous population into two subgroups of connected and non-connected clusters;

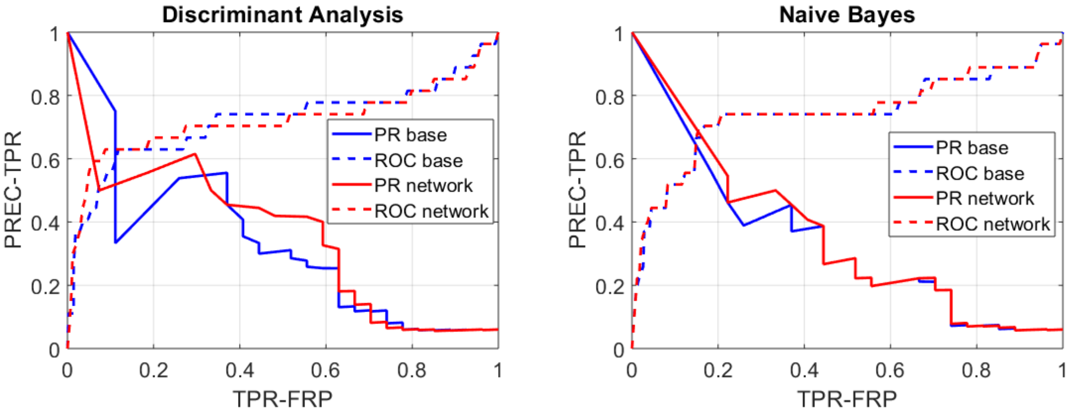


Figure 16: **Precision-Recall (PR) and Receiver Operating Characteristic (ROC) curves for the baseline credit risk models and for the network-augmented models. DATA SET A.** In each panel, dotted lines represent the ROC curves while solid lines refer to PR curves. In blue we show the results related to the baseline models while in red we show the results related to the network-augmented models.

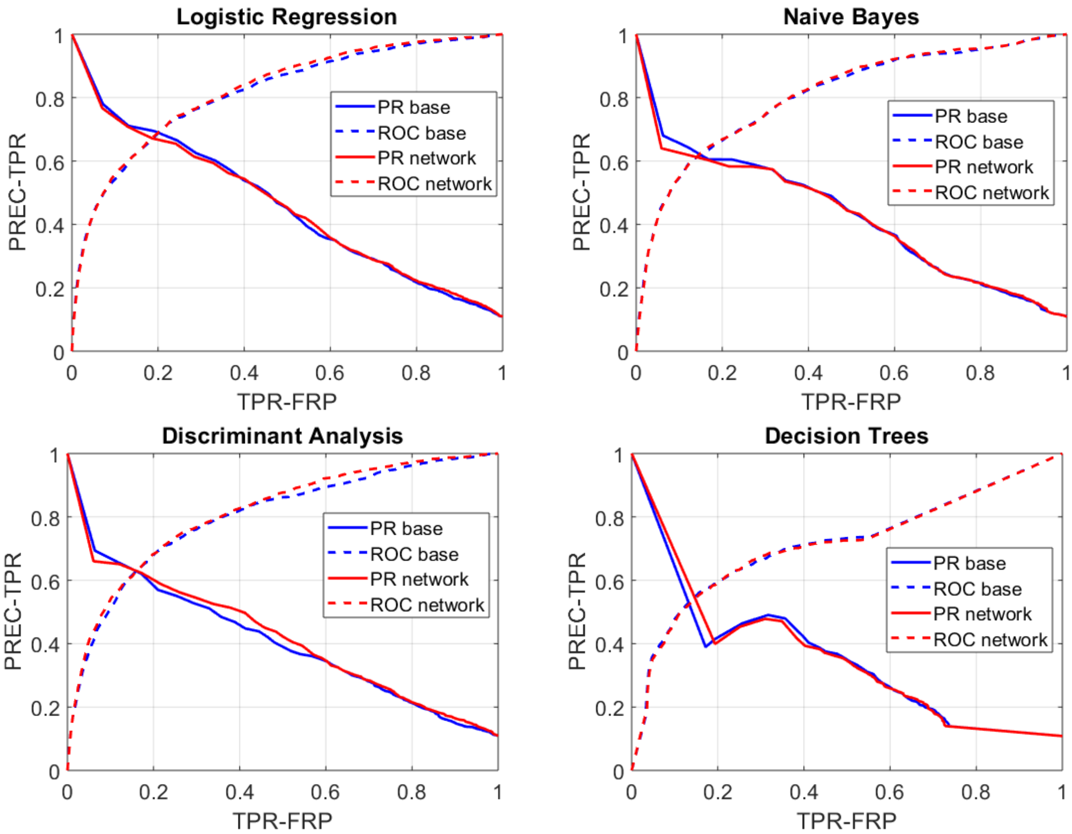


Figure 17: **Precision-Recall (PR) and Receiver Operating Characteristic (ROC) curves for the baseline credit risk models and for the network-augmented models. DATA SET B.** In each panel dotted lines represent the ROC curves while solid lines refer to PR curves. In blue we show the results related to the baseline models while in red we show the results related to the network-augmented models.

DATA-SET A	AUC		AUPR		SomerD		RG		NRI	
	Base	Network	Base	Network	Base	Network	Base	Network	Nri	P-val
Logistic Regression	0.7252	0.8021	0.1827	0.2653	0.4601	0.6101	0.2112	0.2911	0.5805	(0.09)
Discriminant Analysis	0.7197	0.7197	0.2590	0.2766	0.4399	0.4399	0.3445	0.3446	-0.0579	(0.54)
Naive Bayes	0.7404	0.7447	0.2358	0.2472	0.4881	0.4901	0.2311	0.2320	-0.039	(0.10)
Support VM	0.6014	0.7160	0.1361	0.1556	0.2022	0.4302	-	-	0.202	(0.55)
Decision Trees	0.7160	0.7178	0.2340	0.2416	0.4301	0.4302	0.1278	0.1288	-0.039	(0.20)
DATA-SET B	**AUC**		**AUPR**		**SomerD**		**RG**		**NRI**	
	Base	Network	Base	Network	Base	Network	Base	Network	Nri	P-val
Logistic Regression	0.8155	0.8229	0.3434	0.3418	0.6331	0.6455	0.3011	0.3018	0.004	(0.79)
Discriminant Analysis	0.8011	0.8126	0.2942	0.3038	0.6027	0.6225	0.4911	0.5001	0.011	(0.08)
Naive Bayes	0.8064	0.8090	0.3097	0.3038	0.6112	0.618	0.1921	0.1944	-0.015	(0.00)
Support VM	0.6997	0.7543	0.1615	0.2470	0.4001	0.5088	-	-	0.353	(0.00)
Decision Trees	0.7124	0.7097	0.1924	0.1899	0.4315	0.4202	0.1278	0.1288	-0.026	(0.00)

Table 27: **Summary Statistics of non-parametric analysis.** Summary statistics of the non-parametric analysis. From the left to the right: area under the ROC curve (AUC), area under the PR curve (AUPR), the Somers' D test (SomerD), the RGA index (RG) and Net Reclassification Improvement (NRI). For each measure and for all the tested models we report the results obtained by the baseline scenario and for the network-augmented configurations. Moreover, for the NRI we report the p-value of the statistic. Notice how the inclusion of topological variable increases the performance of the models, especially for the logistic regression and support vector machine classifiers.

- build a credit score model for each sub-population;
- compare the predictive performance of the full specification with that of the individual credit scoring models for each sub-population.

For this work, we use the information provided in DATA-SET C, which contains 44 balance sheet and income statement variables on 813 SMEs that have applied for a loan to a P2P platform. As stated previously, the difference between DATA-SET B and DATA-SET C is related with the amount of available financial data for the units of analysis. Namely, DATA-SET C is augmented with additional information collected from Orbis–Bureau van Dijk database.

Empirical results. [5]

We start by elaborating on the community formation identified in the context of SME-focused P2P Lending systems and investigate the systemic importance of SMEs and their potential influence on other companies within the network.

Following Jushan and Serena (2002), we estimate the number of factors, k, using the information criterion (IC) and the Bayesian information criterion (BIC). Specifically

$$IC(k) = \ln(V(k, \hat{f}^k)) + k\left(\frac{m+n}{mn}\right)\ln\left(\frac{mn}{m+n}\right) \tag{4.1}$$

$$BIC(k) = V(k, \hat{f}^k) + k\hat{\sigma}^2\left(\frac{m+n-k}{mn}\right)\ln(mn) \tag{4.2}$$

$$\text{where} \quad V(k, \hat{f}^k) = \frac{1}{m}\sum_{i=1}^{m}\hat{\sigma}_i^2, \qquad \hat{\sigma}_i^2 = \frac{1}{n}\hat{\varepsilon}_i'\hat{\varepsilon}_i, \qquad \hat{\sigma}^2 = \frac{1}{mn}\sum_{i=1}^{m}\sum_{t=1}^{n}E(\hat{\varepsilon}_i)^2$$

Given that our sample consists of 813 SME's each with 44 variables, we set $n = 813$ and $m = 44$. Table 28 shows the IC and BIC for the different values of $k \in \{1, 2, \ldots, 44\}$. For convenience, only

[5] Paper: Ahelegbey, D.F., Giudici, P. and Hadji Misheva, B. Latent factor models for credit scoring in P2P systems. Physica A 522, pp.112-121

the results of a selected few is shown in the Table. Figure 18 presents the plot of the IC and the BIC, which clearly indicates that both criteria favor higher dimensions of k. Thus, $k = 44$ is selected based on the IC and BIC.

	IC	BIC
k=1	-0.3199	0.6641
k=2	-0.3860	0.5685
k=3	-0.4636	0.4811
k=4	-0.5104	0.4198
k=5	-0.5532	0.3678
k=10	-0.6888	0.2054
k=15	-0.9095	0.1054
k=20	-1.3197	0.0447
k=25	-1.9428	0.0153
k=30	-3.2834	0.0026
k=35	-53.7105	0.0000
k=40	-53.8910	0.0000
k=44	**-53.5963**	**0.0000**

Table 28: **Comparing the information criterion (IC) and the Bayesian information criterion (BIC) for the selection of the number of factors, k.** Boldface values indicate the best choice for each metric.

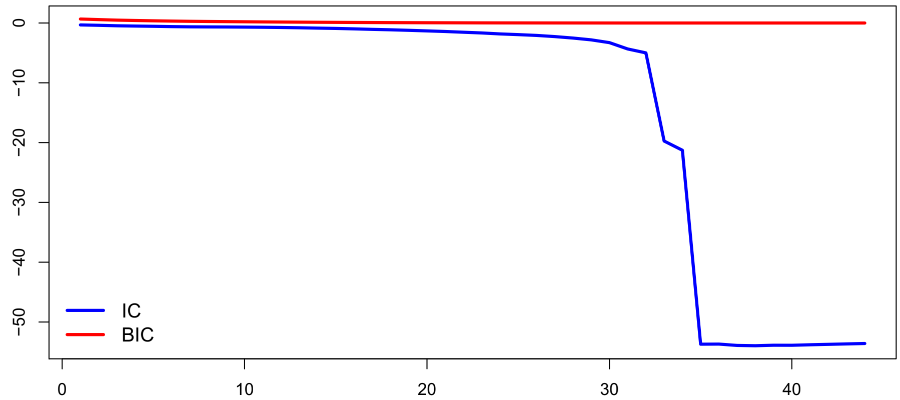
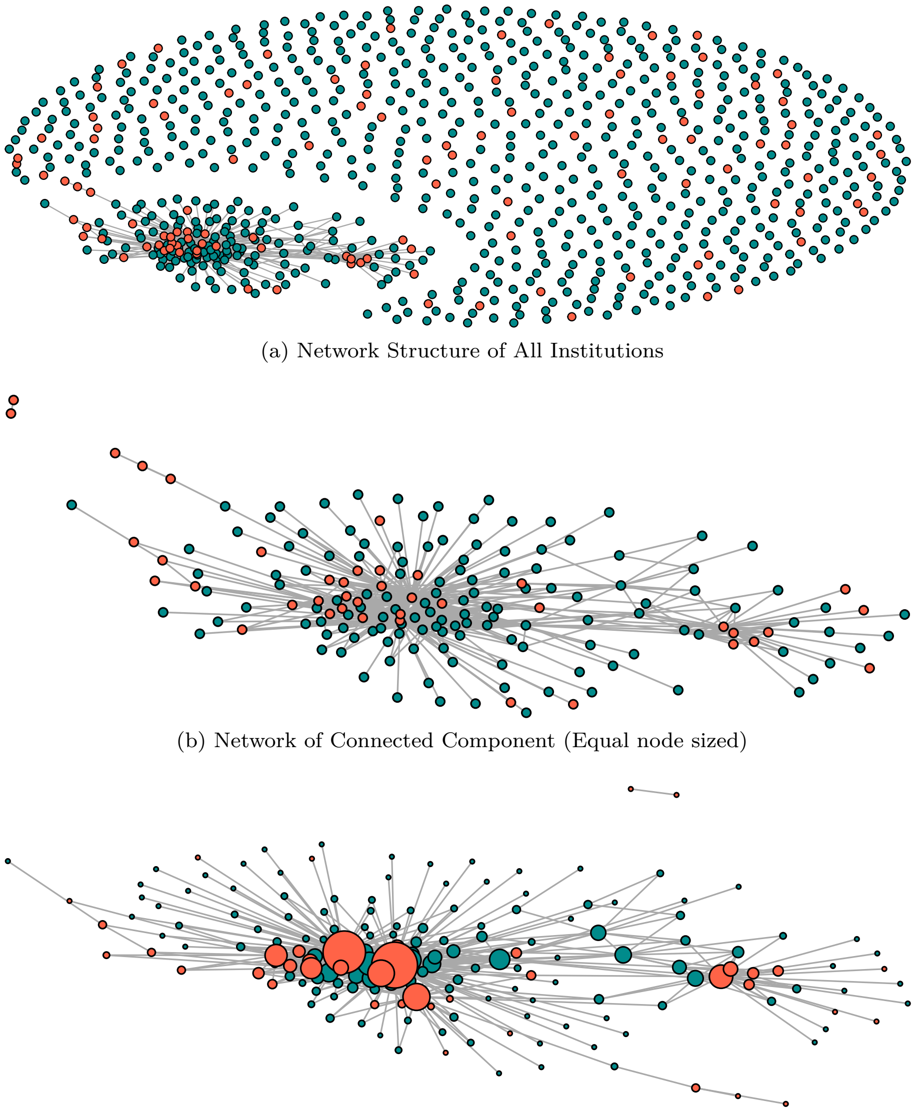


Figure 18: **Plot of information criterion (IC) and the Bayesian information criterion (BIC) for the selection of the number of factors, k.**

Using the SVD approach, we depict the network structure and the interconnections that emerge between companies participating in the P2P lending platform. In Figure 20, nodes are colored based on their status, with red indicating defaulted companies, and green active companies. The Figure contains 20 three graphs: in Figure 19a we have all nodes (companies); in Figure 19b we look only at the connected group and the nodes are equal sized; finally, in in Figure 19c nodes are proportional to companies' degree centrality. The choice of π_0 is set at 0.1 in correspondence with the proportions of defaulted companies in the overall sample. Although 0.1 is at a variance from 0.138, which is the precise proportion of defaulted companies in the sample, we use a round threshold as we found no significant difference between the number of links resulting from applying the two thresholds. Furthermore, we acknowledge that the proportion of defaulted companies within the sample is relatively high hence a more lenient approach to the process of link-inference could result in a more precise inference of the true network between the companies.

From the above graph the high interconnection that exists between active and defaulted SMEs is evident. The emergence of such linkages between companies provides evidence of the existence of joint unobservable forces linking P2P participants. From a credit risk viewpoint, if an active company is linked with a defaulted one, its credit scoring should decrease (contagion effect). Overall, network contagion seems to positively affect default, as the proportion of defaulted companies in Figure 20 is much larger than the observed proportion of defaults in the sample. Specifically, using the latent factor approach, in the community of connected nodes, 25% are companies which have defaulted which is significantly higher than the proportion of defaulted companies in the full sample which is 13.8%.



(a) Network Structure of All Institutions

(b) Network of Connected Component (Equal node sized)

(c) Network of Connected Component (Degree sized nodes)

Figure 19: **Latent factor graph of participants in a P2P lending platform.** (19a) shows both connected and non-connected participants in a P2P system, (19b) shows the connected component with equal sized nodes, and (19c) shows connected component where the size of nodes is proportional to their degrees. Nodes in green are active whiles defaulted participants are represented in red.

We remark that the identification of a network structure within the P2P lending systems could lead to the estimating on a more efficient scoring model. Namely, unlike traditional financial institutions which over the years have segmented their reference markets and in turn increased the accuracy of their scoring models, P2P platforms are based on a global and universal banking model hence developing a scoring model based on full inclusiveness of the data might lead to misleading results as it will capture the behavior and patterns of greatly varying units. In line with this, identifying underlining network structures and dividing the population into connected and not-connected nodes (based on the inferred latent variables), there is the potential of building a more efficient scoring model.

Predictive modeling We now present the application of the latent factor approach for the purpose of improving the predictive performance of the credit scoring employed by P2P lending platforms. From the many available variables, we select a small number of them, based on the research literature as well as by means of a preliminary statistical significance analysis. Specifically, we rely on step-wise regression informed via the Akaike information criterion (AIC), to decide on the variables included in the final model. The variables included in the analysis are presented in Table 29.

Variables	How they are constructed / Description
Leverage ratio	Total current liabilities over shareholders funds
Debt to assets ratio	Total liabilities over total assets
Current ratio	Current liabilities over current assets
Quick ratio	Cash and cash equivalents over total assets
Return on investment	Profit/loss for period over total assets
Return on equity	Net income over shareholders' equity
Asset turnover	Operating revenue/total assets
Return on assets	Net income over total assets
Return on sales	EBIT/operating revenues
Return on capital employed	EBIT/sum of shareholders' equity & debt liab.
Debt coverage	Net income over total loans
Debt ratio	Total current liabilities over total assets

Table 29: **Explanatory variables included in the scoring models and how they are constructed.**

Using the variables in Table 29, we can specify and estimate the benchmark model employed in the credit scoring literature: the logistic regression model. Specifically, we estimate and compare three logit models: (i) full-sample scoring, (ii) scoring for connected nodes and (iii) scoring for not-connected nodes. We argue that because of P2P platforms' universal and global banking model, higher predictive accuracy can be achieved by applying a latent factor-based classification technique which divides the population into major communities (connected vs. not-connected companies).

Table 30 presents the results from the full-sample model which does not control for the network structures that emerge on the bases of the latent variables. We remark that the model in Table 30 has been derived after a thorough activity of model selection, aimed at obtaining the best fit statistical model.

Variable	Estimate	P-value	Significance
Intercept	-2.61	0.000	***
Leverage ratio	-0.09	0.000	***
Current ratio	-1.96	0.005	**
Quick ratio	2.19	0.002	**
ROI	-0.10	0.050	
ROE	0.02	0.000	***
ROS	-6.68	0.146	
Debt conversion ratio	-0.04	0.054	
Debt ratio	0.85	0.002	**
ROCE	-0.23	0.004	**
AUC			0.856

Table 30: **Full-sample scoring – Sample size: 813 companies.** Software used: R, Code available online (fintech-h02020.eu) and upon request

From Table 30 we note that all variables are statistically significant (at 10%) expect the return on sales variable. Although identifying causality is not the prime focus of the thesis, we note that most of the coefficients have the expected sign except for the two ratios capturing liquidity and profitability performance of companies. Namely, the quick ratio and the ROE indicator report a positive and a negative sign, respectively which is contrary to our expectations. In the context of the quick ratio, the reported positive sign appears counter intuitive as it indicates that higher liquidity (measured via the quick ratio) increases the probability of default. Potential explanation for the positive sign can be in the fact that worst-performing companies tend to have higher liquidity due to their inability to invest in profitable prospects. Moreover, in the context of ROE, the results suggest that increasing ability of companies to use equity for generating profits increase the probability of default. This may be due to the fact that the companies investigated within this analysis are small businesses which do not rely extensively on equity financing.

As the main purpose of the analysis is to investigate the predictive accuracy of scoring models, we fit the models on a training set to make predications on a data that was not trained, and we consider the classification errors resulting from the models, the corresponding ROC curve and, finally, the AUC, the most widely-used measure of predictive performance for credit scoring models.

As previously commented, to improve the credit scoring accuracy, we introduce a network-based scoring. Specifically, we use the logistic estimator for connected and not-connected communities separately. Table 31 summarizes the results obtained from the scoring model for connected nodes.

Variable	Estimate	P-value	Significance
Intercept	-2.26	0.002	**
Leverage ratio	-0.05	0.187	
Current ratio	-3.29	0.024	*
Quick ratio	3.50	0.015	*
ROE	0.02	0.020	*
Debt ratio	0.86	0.059	
ROCE	-0.44	0.107	
AUC			0.949

Table 31: **Scoring for connected nodes – Sample size: 176 companies**. Software used: R, Code available online (fintech-h02020.eu) and upon request

Looking at the results for the connected nodes, similarly as it is the case with the full sample scoring, most of the variables are statistically significant and report the expected sign. ROCE remains insignificant potentially confirming the low-equity dependency of small businesses. Both the quick ratio and ROE keep the counterintuitive signs. We also note that in the context of connected companies, the liquidity ratios are among the main drivers which could be an indication of their exposure to contagion and systemic risk that in turn motivates them to keep higher liquidity.

In terms of predictive accuracy , the model, estimated only for the connected nodes, leads to an AUC of 0.949 which is a significant improvement in scoring accuracy with respect to the model previously fit on all nodes. We remark that the difference between the two models are only in the node selection. Specifically, in the latter case, the model is estimated using the same methodology but only considering those companies which are connected (have at least one link with another company in the sample). Hence variable selection is consistent throughout the individual models. As mentioned previously, we employ stepwise regression including all the key financial ratios available in the dataset and consequently select the models that best fit the data according to the AIC criteria.

Finally, we investigate the scoring model for not-connected companies. Table 32 summarizes the results. Looking at the results for the non-connected nodes, we see that most variables are found statistically significant except for the quick ratio and the debt conversion ratio. The AUC reported is 0.945 which is like the predictive accuracy of the scoring model for the connected nodes. In the attempt to validate the results, robustness checking has been carrying out by changing the π_0 threshold. We have confirmed that the community detection approach results in higher predictive accuracy relative to the full-sample scoring if the π_0 threshold is changed to 0.138, 0.2 and 0.3.

Variable	Estimate	P-value	Significance
Intercept	-0.94	0.110	
Leverage ratio	-0.06	0.052	
Quick ratio	0.28	0.298	
ROE	0.02	0.018	*
Asset turnover	-2.70	0.000	***
ROS	-18.34	0.001	**
Debt conversion ratio	-0.01	0.537	
Debt ratio	1.44	0.001	**
ROCE	-0.22	0.012	*
AUC			0.945

Table 32: **Scoring for not-connected nodes – Sample size: 637 companies.** Software used: R, Code available online (fintech-h02020.eu) and upon request

Furthermore, another robustness check performed is that of random sub-sampling. Specifically, to confirm the validity of the results, we ran the same analysis considering many different random subsets. Specifically, we divided the sample of 813 companies on two random subsets (keeping the same proportion between connected and not-connected companies) and re-ran the analysis with several different seeds. The average AUC for the random subsets with sample sizes equal to the connected and not connected communities are: 0.8618256 and 0.8555784, respectively. These predictive utilities are significantly lower compared to those obtained by splitting the sample based on the SVD approach.

We believe that these results are a clear indicator that a latent factor-based classification technique that divides the population into major communities based on the SVD approach does lead to the estimation of a more efficient logistic model, which in turn is crucial to improving credit risk measurement. The approach of identifying communities of connected and not-connected nodes, in a way, mimics the segmentation strategies employed by traditional financial institutions which in turn allows P2P platforms to obtain high predictive accuracy of the scoring model without necessarily possessing all the financial and non-financial information about individual participants in the platform. Furthermore, the proposed methodology can also help traditional financial providers to improve their scoring by further "segmenting the segments". Finally, our results can be put in context of a similar research carried out by Giudici and Hadji-Misheva (2017) which aim is to investigate whether the inclusion of network parameters (obtained from correlation networks) into a traditional credit scoring approach can improve the model's overall predictive accuracy . The study confirms this premise as the inclusion of the network-derived parameters improves the predictive accuracy of the scoring by several percentage points. However, the approach presented in this section enables a more precise differentiation of risk classes as the network is inferred based on all known financial parameters which in turn lead to a much higher predictive accuracy .

4.2.5 Factorial Network Models To Improve P2P Credit Risk Management

For the purpose of carrying out robustness test to our factor-based credit scoring application, we extend the analysis from Section 4.2.4 to a different data set (DATA-SET D). The intention is to check whether the results hold valid on a different, larger data set with a different set of predictors. In terms of the methodology employed, for the purpose of identifying the communities, we estimate a network following the approach explained in sub-section 3.2.3. Differently from subs-section 4.2.4 we test a credit score model for each cluster via lasso-type regularization logistic regression (see 3.1.3, 3.1.4, 3.1.5, 3.1.6).

Summary of Data and Methodology To estimate the underlying factors that drive the observed data matrix, we decompose the matrix of observed financial characteristics via a singular value decomposition given by

$$X = UDV = FW + \varepsilon \tag{4.3}$$

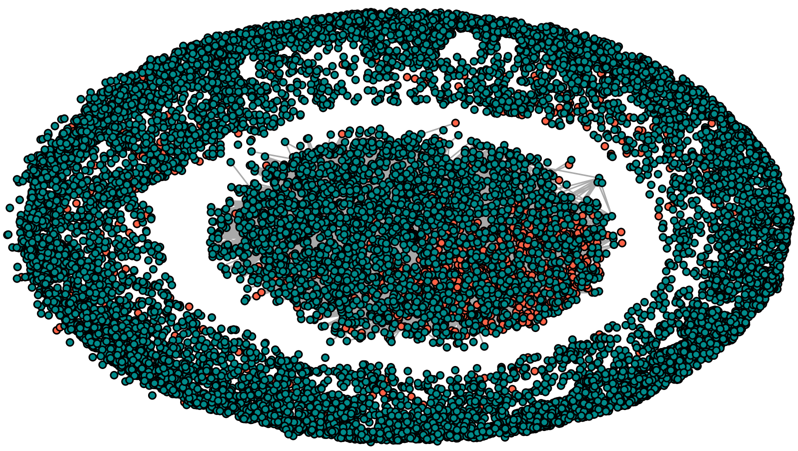
where U and V are orthonormal, and $D = \Lambda^{1/2}$ is a diagonal matrix of non-negative and decreasing singular values, with Λ as the diagonal matrix of the non-zero eigenvalues of $X'X$ and XX'. U is $n \times p$, D is $p \times p$ and V is $p \times p$. Following the error approximation criteria, we obtain the factor matrix by $F = U_{n,k}\ D_{k,k}$ and $W = V_{k,p}$, where $U_{n,k}$ is $n \times k$ matrix composed of the first k columns of U, $k < p$, $D_{k,k}$ is $k \times k$ matrix comprising the first k columns and rows of D, and $V_{k,p}$ is $k \times p$ matrix of factor loadings. The matrix F can therefore be interpreted as a projection of X onto the eigenspace spanned by $U_{n,k}$. We determine k by observing the number of eigenvalues associated with the largest variance matrix. Table 33 shows the eigenvalues of the singular value decomposition to determine the factors to retain. The eigenvalues reported are the normalized squared diagonal terms of D. From the table, we set $k = 17$ since the first 17 eigenvalues explain about 95% of the total variation in X.

No.	Eigenvalue	Variance Explained (%)	Cumulative (%)
1	5.18	21.60	21.60
2	2.58	10.73	32.33
3	2.50	10.41	42.74
4	1.60	6.69	49.42
5	1.42	5.92	55.34
6	1.30	5.40	60.74
7	1.16	4.82	65.55
8	1.09	4.56	70.11
9	0.99	4.11	74.22
10	0.93	3.88	78.10
11	0.80	3.35	81.45
12	0.79	3.31	84.76
13	0.75	3.11	87.87
14	0.56	2.35	90.22
15	0.53	2.21	92.43
16	0.51	2.12	94.55
17	0.43	1.80	96.35
18	0.37	1.54	97.89
19	0.17	0.69	98.58
20	0.11	0.47	99.05
21	0.09	0.36	99.41
22	0.07	0.27	99.68
23	0.06	0.26	99.94
24	0.01	0.06	100.00

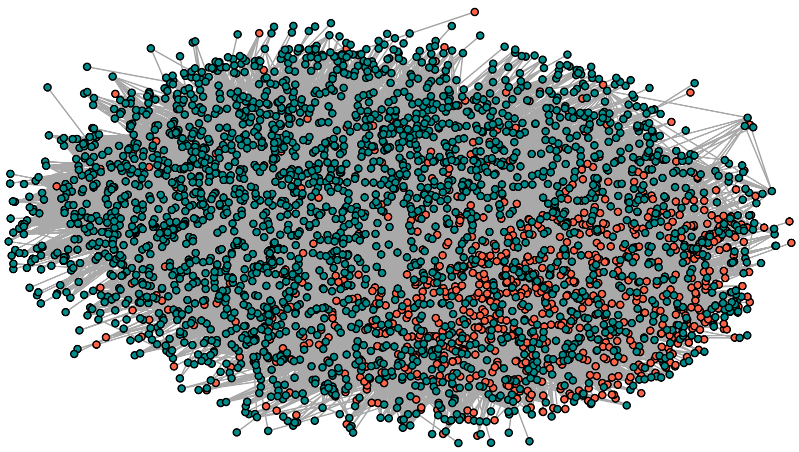
Table 33: **The eigenvalues of the singular value decomposition to determine the factors to retain.**

We use the estimated factor matrix, F, to construct the network for the segmentation of the companies. For purposes of graphical representations and to keep the companies name anonymous, we report the estimated network by representing the group of institutions with color-codes. The defaulted companies are represented in a red color code, and non-defaulted companies in the green color code (see Figure 20). Table 34 reports the summary statistics of the estimated network in terms of the default-status composition of the SMEs. The threshold is initially set at 0.1, which in turn corresponds to the proportion of defaulted loans in the sample. For robustness purposes, we compare the results obtained with a threshold value $\gamma = 0.05$ against $\gamma = 0.10$.

The result for the threshold $\gamma = 0.05$ of Table 34 shows that the connected sub-population is composed of 4305 companies which constitute 28.6% of the full sample. The non-connected sub-population is composed of 10740 (71.4%). The percentage of the defaulted class of companies are 22.4% and 6.2% among the connected- and non-connected sub-population, respectively. We notice that higher threshold values (say $\gamma = 0.1$) decrease (increase) the total number of connected



(a) Network Structure of All Institutions



(b) Network of Connected Component

Figure 20: **A graphical representation of the estimated factor network.** (20a) shows the structural representation of the factor network for threshold $\gamma = 0.05$, and (20b) depicts the connected sub-population only. The nodes in red-color are defaulted class of companies and green-color coded nodes are non-defaulted class of companies.

Threshold	Status	Conn-Sub	Non-Conn-Sub
$\gamma = 0.05$	Default	964 - 22.4%	668 - 6.2%
	Non-Default	3,341 - 77.6%	10,072 - 93.8%
	Total	4,305 - 28.6%	10,740 - 71.4%
$\gamma = 0.1$	Default	816 - 24%	816 - 7%
	Non-Default	2,580 - 76%	10,833 - 93%
	Total	3,396 - 22.6%	11,649 - 77.6%

Table 34: **Summary statistic of connected and non-connected sub-population obtained from the factor network-based segmentation for threshold values of $\gamma = \{0.05, 0.1\}$.**

(non-connected) sub-population and vice versa. Such higher threshold values also lead to a lower (higher) number of defaulted class of connected (non-connected) SMEs but (and) constituting a higher percentage of the defaulted population. Figure 20 presents the graphical representation of the estimated factor network with the sub-population of defaulted and non-defaulted companies color coded as red and green, respectively. Figure 20a shows the structural representation of both connected and non-connected sub-population while Figure 20b depicts the structure of connected sub-population only.

Empirical results. [6]

We compare the lasso, adaptive lasso, elastic-net, and adaptive elastic-net variable selection methods to model the credit score of the listed companies in our dataset. To estimate the models, we standardized each series to a zero mean and unit variance. Table 35 reports the variable selection and estimated coefficients of the four methods. The column CSM represents the benchmark credit scoring model, NS-CSM(C) - the network segmented connected sub-population credit scoring model, and NS-CSM(NC) for the network segmented non-connected sub-population credit scoring model. The top left panel represents the lasso method, the adaptive lasso is on the top right panel, elastic-net at the bottom left and adaptive elastic-net at the bottom right.

Table 36 reports the number of variables selected by each of the four competing methods for the credit score model estimation. From the table, the elastic-net is the least parsimonious, followed by the lasso, and lastly, the adaptive elastic-net and adaptive lasso are the most parsimonious. From Tables 35 and 36, we observed a significant difference in the number of selected explanatory variables for the benchmark model and the network segmented models. More precisely, the former model the credit score of a given company by using more variables while the latter on the other hand uses a significantly lower number of variables. The similar results across the four variable selection methods, given their similarities, is not terribly surprising. But they do indicate that the general approach appears to be robust in this setting, which was the main purpose of the testing. The network-based segmentation framework is therefore more parsimonious than the benchmark full population credit score model, and this helps in interpretability.

We analyzed the performance of the models by splitting the sample into 70% training and 30% testing sample. We now compare the default prediction accuracy of the models in terms of the standard AUC derived from the ROC curve. See Figure 21 for the plot of the ROC curve for the competing methods.

The comparison of the ROC curves from the competing methods shows that the CSM (in red) lies below the rest. Clearly, the curves of NS-CSM ($\gamma = 0.1$) depicted in green seems to dominate the others. The summary of the area under the ROC curve reported in Table 37 shows that NS-CSM ($\gamma = 0.1$) is ranked first, followed by NS-CSM ($\gamma = 0.05$), and the lowest AUC is obtained by the CSM. Overall, in terms of default predictive accuracy, the result of the AUC shows the NS-CSM outperforms the CSM, on average by two percentage points. This is an advantage that can be further increased considering as the cut-off the observed default percentages, which are different in the two samples.

We investigate whether the AUC of the network segmented model is significantly different from the benchmark model for the four methods. We applied the DeLong test (see E. R. DeLong, D. M. DeLong, and Clarke-Pearson, 1988) to investigate the pairwise comparison of the AUC of the benchmark model (i.e., CSM) and that of the NS-CSM for $\gamma = \{0.05, 0.1\}$. We perform these tests under the null-hypotheses that H_0: AUC (CSM) $\geq$ AUC (NS-CSM) and the alternative hypotheses, H_1: AUC (CSM) $<$ AUC (NS-CSM). Table 38 reports the one-sided statistical test of the AUC of the benchmark model relative to the network segmented models. The result of the De Long test shows that while the ROC of CSM is not statistically different from that of NS-CSM($\gamma = 0.05$), the difference between the ROC of NS-CSM($\gamma = 0.1$) and the benchmark (CSM) is statistically significant at 90% confidence level for all four methods.

[6] Paper: Ahelegbey, D.F., Giudici, P. and Hadji Misheva. (2019) Factorial Network Models To Improve P2P Credit Risk Management. Frontiers in Artificial Intelligence 2

	CSM	NS-CSM(C)	NS-CSM(NC)	CSM	NS-CSM(C)	NS-CSM(NC)
	Lasso			*Adaptive Lasso*		
V1	0.0535	.	0.0375	.	.	.
V2	.	0.0332	.	.	.	.
V3	-0.4468	-0.2818	-1.0148	-0.5298	-0.3539	-1.1990
V4	-0.3549	-0.1294	-0.5556	-0.2928	-0.1368	-0.5137
V5	.	.	.	.	.	.
V6	0.0774	.	0.1460	0.0440	.	0.0213
V7	0.2818	.	.	0.2116	.	.
V8	-0.3933	-0.3408	0.1185	-0.4356	-0.3463	.
V9	-0.0360	0.0365	-0.4690	.	.	-0.5577
V10	-0.0701	0.0287	.	.	.	.
V11	0.1291	.	0.0550	.	.	.
V12	0.0265	0.0222	0.0204	.	.	.
V13	-0.2419	.	.	-0.1759	.	.
V14	-0.0399	-0.0776	.	.	-0.113	.
V15	-0.0751	-0.0396	0.0128	-0.0520	.	.
V16	0.0520	0.2851	.	.	0.2245	.
V17	0.2213	0.1650	0.1761	0.2529	0.2092	.
V18	0.0396	0.0661	0.0143	.	0.0484	.
V19	0.2540	0.0291	0.2096	0.2755	.	0.2151
V20	0.0412	.	0.2429	.	.	0.1950
V21	0.2212	0.1620	0.2969	0.2410	0.1721	0.3185
V22	0.0930	.	0.1470	0.0541	.	0.0219
V23	-0.2262	-0.0649	-0.3452	-0.2213	-0.0650	-0.3826
V24	-0.0062	-0.0641	0.0343	.	-0.0645	.
	Elastic-Net			*Adaptive Elastic-Net*		
V1	0.0548	.	0.0568	.	.	.
V2	1.0e-04	0.0372	.	.	.	.
V3	-0.4472	-0.2692	-1.0132	-0.5293	-0.3538	-1.2208
V4	-0.3628	-0.1286	-0.6051	-0.2900	-0.1350	-0.6034
V5	0.0048	-0.0123	.	.	.	.
V6	0.0780	-0.0028	0.1862	0.0422	.	0.1528
V7	0.3003	.	.	0.1925	.	.
V8	-0.3926	-0.3310	0.2054	-0.4363	-0.3474	0.1672
V9	-0.0356	0.0435	-0.4884	.	.	-0.5195
V10	-0.1419	0.0315	.	.	.	.
V11	0.2016	0.0112	0.1025	.	.	.
V12	0.0299	0.0299	0.0545	.	.	.
V13	-0.2595	.	.	-0.1571	.	.
V14	-0.0374	-0.0785	.	.	-0.1112	.
V15	-0.0777	-0.0468	0.0597	-0.0499	.	.
V16	0.0600	0.2902	0.0669	.	0.2256	.
V17	0.2173	0.1588	0.1701	0.2527	0.2097	0.1147
V18	0.0417	0.0769	0.0439	.	0.0459	.
V19	0.2538	0.0502	0.2042	0.2747	.	0.2151
V20	0.0425	.	0.3139	.	.	0.2571
V21	0.2210	0.1634	0.3113	0.2409	0.1721	0.3036
V22	0.0933	0.0012	0.1727	0.0533	.	0.1047
V23	-0.2286	-0.0728	-0.3754	-0.2185	-0.0616	-0.4114
V24	-0.0077	-0.0724	0.0464	.	-0.0619	.

Table 35: **Estimated coefficients from lasso (top left), adaptive lasso (top right), elastic-net (bottom left) and adaptive elastic-net (bottom right).** CSM is the benchmark credit score model, NS-CSM(C) is the network segmented connected sub-population credit score model, and NS-CSM(NC) is the network segmented non-connected sub-population credit score model, estimated for threshold value $\gamma = 0.1$.

	Lasso	Adaptive Lasso	Elastic-Net	Adaptive Elastic-Net
CSM	22	12	24	12
NS-CSM(C)	16	10	20	10
NS-CSM(NC)	17	9	18	11

Table 36: **Number of selected variables of the four methods.**

	Lasso	Adaptive Lasso	Elastic-Net	Adaptive Elastic-Net
CSM	0.8089	0.8061	0.8090	0.8061
NS-CSM($\gamma = 0.05$)	0.8214	0.8204	0.8225	0.8207
NS-CSM($\gamma = 0.1$)	0.8330	0.8277	0.8342	0.8312

Table 37: **Comparing area under the ROC curve (AUC) of the four methods.** CSM is the benchmark model, NS-CSM(C) is the network segmented connected sub-population model, and NS-CSM(NC) is the network segmented non-connected sub-population model, estimated for threshold values of $\gamma = \{0.01, 0.05, 0.1\}$.

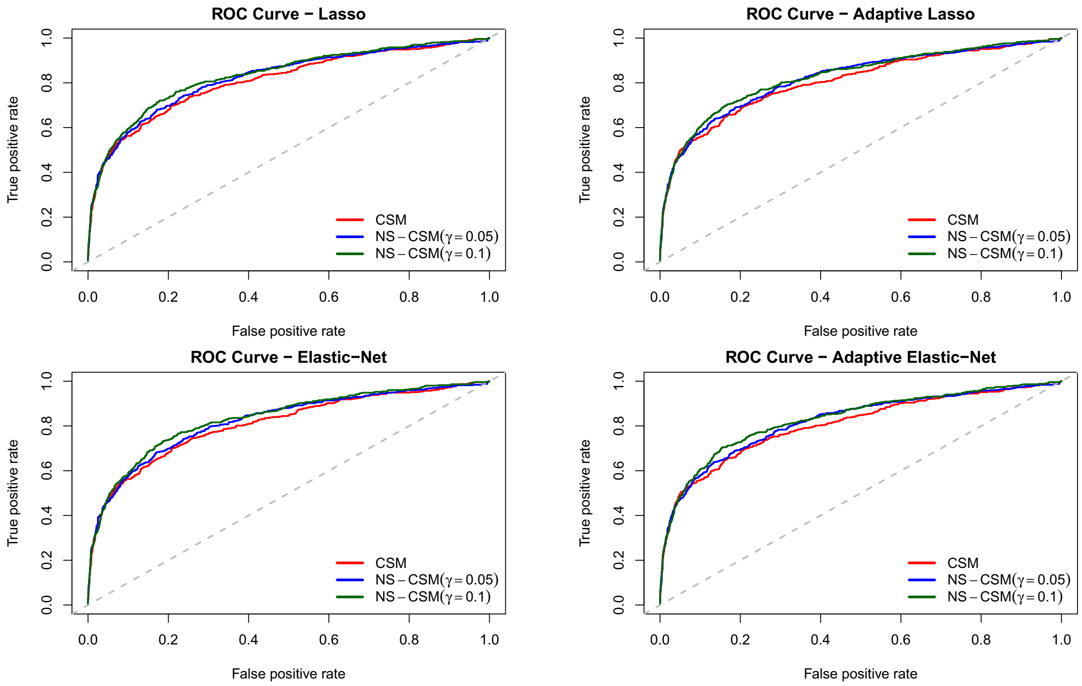


Figure 21: **ROC curves of the four methods.** CSM is the benchmark model, NS-CSM(C) is the network segmented connected sub-population model, and NS-CSM(NC) is the network segmented non-connected sub-population model, estimated for threshold values of $\gamma = \{0.05, 0.1\}$.

		Statistic	P-value	Significance	Statistic	P-value	Significance
			Lasso			*Adaptive Lasso*	
CSM	NS-CSM ($\gamma = 0.05$)	-0.7639	0.2225		-0.8598	0.1950	
	NS-CSM ($\gamma = 0.1$)	-1.4972	0.0672	*	-1.3129	0.0946	*
			Elastic-Net			*Adaptive Elastic-Net*	
CSM	NS-CSM ($\gamma = 0.05$)	-0.8241	0.2050		-0.8728	0.1914	
	NS-CSM ($\gamma = 0.1$)	-1.5770	0.0574	*	-1.5327	0.0627	*

Table 38: **AUC of the benchmark model relative to the network segmented models under the four methods.**

5 Beyond Fintech Credit: Network Models for Risk Management in P2P Crypto Markets

Throughout this thesis, a main point of interest is investigating how network models can be utilized in the task of risk management particularly in the context of improving the predictive accuracy of classification and regression models. So far, the prime context for this analysis has been Fintech credit. The reasons for this choice of context have been explained in detail in Sections 1.2 and 2.3 Namely, there is a strong need for alternative methodologies in the context of P2P scoring models mostly because these online platforms typically do not possess long and complete data sets on their clients hence they must make full use of the available data through advance modelling. Although the P2P credit space is special and, as we have shown, can benefit greatly from the application of network theory in the context of credit risk estimation, this context is not unique. Most Fintech risk management procedures can draw advantages from the inclusion of information related with the interconnections that emerge between economic agents. Yet another context is that of the fast growing crypto assets.

The BitCoin represents the first decentralized digital cryptocurrency which exhibited massive increase in market capitalization since its insertion. Specifically, the total market capitalization of this cryptocurrencies reached its peak of $330 billion USD in December 2017 and currently its dominance of the market is at 67.7%, making it the leading cryptocurrency (https://coinmarketcap.com/). BitCoin's immediate success instigated the emergence of new alternative cryptocurrencies with different motivations and architectures. As these *Altcoins* grow in adaptation, so does the prevalence of research and working papers focused on exploring their ecosystem and price formation processes.

In this second application, we investigate the interconnections that emerge between BitCoin users, mapping out the network of payment relationship between them. In particular, the study uses data from the BitCoin ledger that covers transactions in the time period 25.02.2012 to 17.07.2017 and groups them in approximations of traders. This then enables us to extract and plot a network of payment relationships that exists between different traders and trace the evolution of the dependence relationships between them over time thus identify leader-follower links.

The research on BitCoin and other cryptocurrencies has been substantial within the past few years with most of the research community focused on testing the applicability of traditional prediction or dynamic volatility models in the context of the crypto markets. We use the description *traditional* for the purpose of suggesting that authors are applying models found valid for traditional financial time series data. This is mostly due to the fact that many of the stylized facts that are valid for traditional financial time series, apply, to some extend, also in the context of these alternative currencies. Making a comparison with equity prices, cryptocurrencies are also characterized with time-varying and asymmetric volatility, extreme observations, etc. (Catania, Grassi, and Ravazzolo, 2018). Since one of the most commonly used methodology for modeling traditional currencies is based on the Generalized Autoregressive Conditional Hetheroskedasticity (GARCH) models, most of the research community has attempted to model the volatility of cryptocurrencies using GARCH-based models (Bouoiyour and Selmi, 2016, Katsiampa, 2017, Stavroyiannis and Babalos, 2017, Chu et al., 2017), and the results have been far from consistent. In our view, this is mostly due to the nature of the fast-growing alternative currencies. Namely, it is an indisputable fact that cryptocurrencies in general and the BitCoin in particular represent unique socio-technical ecosystems working outside any traditional market and as such, their economy is not well understood. Furthermore, they are much more volatile compared to traditional currencies, their exchange rates cannot be assumed to be independently and identically distributed (Chu et al., 2017) and their global nature limits researchers' ability to account for all latent and exogenous factors. Thus it becomes appropriate to explore the economy of the leading cryptocurrency and identify key drivers influencing BitCoin transaction volumes by looking at alternative methodologies.

This direction becomes even more important in view of the BitCoin's market concentration. Credit Swiss in January 2018 provided a study which indicates that 97% of BitCoins are held by 4% of all BitCoin addresses. Bloomberg also reported similar findings by suggesting that about

40 percent of BitCoin is held by perhaps 1,000 users meaning that the wealth in the ecosystem is very concentrated hence any movement by the few large users can ripple through the market and cause major disruptions in the price of the cryptocurrency. An example of this is the transaction that took place on 12th of November when a node moved 25,000 BitCoins worth at the time close to USD159 million to an exchange.

Having in mind this defining property, it is clear that focusing on traditional volatility models to predict price or volume movement could produce sub-optimal results, as a very important component is understanding the behaviour of individual nodes within this network. The literature on using network models for the purpose of investigating the BitCoin economy has been very limited. Notable exception to this are the papers by Tasca, Liu, and Hayes (2016) and Sean, Jonathan, and Putnins (2018). Tasca, Liu, and Hayes (2016) attempt to identify different clusters within the BitCoin economy by analyzing the transaction patterns and ascribing them to particular business categories. Using network-based methodologies, the authors have identified three marked regimes that have in turn characterized BitCoin transactions.

This work extends Tasca, Liu, and Hayes (2016), and aims to further acquire empirical evidence on whether BitCoin trading volumes behaviors of different group of traders, in different geographical regions, are interconnected. In other words, we aim to analyze interconnections among trading volumes to identify "standard" behaviours and identify anomalies.

From an econometric viewpoint, our proposed correlation network model is an extended Vector Autoregressive model, aimed at explaining the evolution of trading volumes among different groups. The extension is based on network models, which improve over pure autoregressive models, as they introduce a contemporaneous contagion component, that describes contagion effects between investment behaviours.

The validity of the model has been demonstrated in recent studies on systemic risk, in which researchers have proposed correlation network models, able to combine the rich structure of financial networks (see, e.g. Battiston et al., 2012) with a more parsimonious approach that can estimate contagion effects from the dependence structure among market prices. The first contributions in this framework are Billio et al. (2012) and Diebold and Yilmaz (2014), who derive contagion measures based on Granger-causality tests and variance decompositions. More recently, Ahelegbey, Billio, and Casarin (2015), Giudici and Spelta (2016) and Giudici and Parisi (2017) have extended this methodology introducing stochastic correlation networks.

While bivariate systemic risk models (such as Acharya, Engle, and Richardson, 2012, Adrian and Brunnermeier, 2015) explain whether the risk of an institution is affected by a market crisis event or by a set of exogenous risk factors, correlation network models explain whether the same risk depends on contagion effects, in a cross-sectional perspective.

We extend (a) the approach of Ahelegbey, Billio, and Casarin (2015), enriching their VAR model using partial correlations; (b) the approach of Giudici and Spelta (2016), Giudici and Parisi (2017), by enriching their graphical Gaussian models with an autoregressive component derived through a VAR model.

Proposal. Let y_t^i be the traded volume of Bitcoin by a specific group of traders $i\,(i = 1, \ldots, I)$, at time $t\,(t = 1, \ldots, T)$. We assume that y_t^i is a function of: (a) an autoregressive element, that captures the dependence on the past trading volumes of the same group; (b) a cross-sectional element, that captures the contemporaneous dependence on the trading volumes of other groups; (c) a stochastic residual. Mathematically, we assume that in the case of the Bitcoin traded volumes, for each volume i and time t the following equation holds:

$$y_t^i = \sum_{p=1}^{p_0} \alpha_p^i y_{t-p}^i + \sum_{j \neq i} \beta^{ij} y_t^j + \epsilon_t^i, \tag{5.1}$$

where p is a time lag (with $p_0 < t$), α_p^i and β^{ij} are the coefficients which are to be estimated, and ϵ_t^i are the residuals.

Equation (5.1) models the Bitcoin volume dynamics as a structural VAR, in which the traded volume in each group depends on its p past values, through the idiosyncratic autoregressive component $\sum_{p=1}^{p_0} \alpha_p^i y_{t-p}^i$ and, in addition, it depends on the contemporaneous values of the other groups, through the systemic component $\sum_{j \neq i} \beta^{ij} y_t^j$.

Defining B_0 as a $I \times I$ symmetric matrix with null diagonal elements containing the contemporaneous coefficients, the previous model can be expressed in a more compact matrix form, as follows:

$$Y_t = \sum_{p=1}^{p_0} A_p Y_{t-p} + B_0 Y_t + \varepsilon_t, \tag{5.2}$$

where Y_t is a I-dimensional vector containing the traded volumes of all groups at time t, Y_{t-p} is the same vector, lagged at time $t-p$, A_p is a $I \times I$ matrix that contains the autoregressive coefficients and ε_t is a vector of residuals.

In the following step, we transform the model in (5.2) into a reduced form for the purpose of facilitating the estimation process, thus becoming:

$$Y_t = \Gamma_1 Y_{t-1} + ... + \Gamma_{p0} Y_{t-p0} + U_t, \tag{5.3}$$

with

$$\begin{cases} \Gamma_1 = (\mathbb{I} - B_0)^{-1} A_1, \\ ... \\ \Gamma_{p0} = (\mathbb{I} - B_0)^{-1} A_{p0}, \\ U_t = (\mathbb{I} - B_0)^{-1} \varepsilon_t. \end{cases} \tag{5.4}$$

This reduced form allows the estimation of the vectors of modified autoregressive coefficients $\Gamma_1, ..., \Gamma_{p0}$, using time series data on the traded volumes contained in the stacked vector $\{Y_1, \ldots, Y_t, \ldots, Y_T\}$.

However, we are not interested in estimating Γ_p. In fact, the purpose of this analysis is to disentangle its autoregressive and contemporaneous components, thus separately estimating $\{A_1, ..., A_{p0}\}$ and B_0. In this sense, once B_0 is obtained, $\{A_1, ..., A_{p0}\}$ can be derived from (5.4).

To estimate B_0, note that $(\mathbb{I} - B_0)U_t = \varepsilon_t$, so that $U_t = B_0 U_t + \varepsilon_t$. This implies that, for each group i,

$$U_t^i = \sum_{j \neq i} \beta^{ij} U_t^j + \epsilon_t^i, \tag{5.5}$$

meaning that the off-diagonal elements of B_0 can be obtained regressing each modified residual, derived from the application of (5.3), on those of the other groups.

Note that the regression model in (5.5) is based on the transformation derived in equation (5.4), which makes the modified residuals correlated. The direction of such correlation is, however, unknown. In the application of (5.5) it is therefore not clear which volume residual assumes the form of a response variable, and which one of an explanatory regressor.

In order to determine the direction of such dependence, we propose to approximate each pair of regression coefficients β^{ij} and β^{ji}, with their partial correlation coefficient, which is undirected.

Mathematically, let $\Sigma = Corr(U)$ be the correlation matrix between the modified residuals, and let Σ^{-1} be its inverse, with elements σ^{ij}. The partial correlation coefficient $\rho_{ij|S}$ between the residuals U^i and U^j, conditional on the remaining residuals ($U^s, s = 1, \ldots, S$), where $S = I \setminus \{i, j\}$, can be obtained as:

$$\rho_{ij|S} = \frac{-\sigma^{ij}}{\sqrt{\sigma^{ii}\sigma^{jj}}}. \tag{5.6}$$

It can be shown that (see Epskamp and Eiko, 2018):

$$|\rho_{ij|S}| = \sqrt{\beta^{ij} \cdot \beta^{ji}}, \tag{5.7}$$

which means that the absolute value of the partial correlation coefficient between U^i and U^j, given all the other residuals, can be obtained as the geometric average between the coefficients β^{ij} and β^{ji} defined by equation (5.5) setting, respectively, i rather than j as response variables. Equation (5.7) justifies the replacement of β^{ij} and β^{ji} with their corresponding partial correlation coefficient $\rho_{ij|S}$.

From an economic viewpoint, the partial correlation coefficient expresses how the trading volume of node i is affected by the contemporaneous trading volume of node j ($j \neq i$), keeping the other volumes fixed.

An important advantage that derives from the employment of partial correlations lies in the possibility of employing correlation network models based on the conditional independence relationships described by partial correlations.

More precisely, let us assume that the vectors U_t are independently distributed according to a multivariate normal distribution $\mathcal{N}_I(0, \Sigma)$, where Σ represents the correlation matrix (that we assume to be non-singular).

A correlation network model can be represented by an undirected graph G such that $G = (V, E)$, with a set of nodes $V = \{1, ..., I\}$, and an edge set $E = V \times V$ that describes the connections between the nodes. G can be represented by a binary adjacency matrix E with elements e_{ij}, each of them providing the information of whether a pair of vertices in G is (symmetrically) linked between each other ($e_{ij} = 1$) or not ($e_{ij} = 0$). If the nodes V of G are put in correspondence with the random variables $U_1, ..., U_I$, the edge set E induces conditional independences on U via the so-called Markov properties (see e.g. Lauritzen, 2011).

Following up on (5.7), Whittaker (1990) proved that the following equivalence holds:

$$\rho_{ij|S} = 0 \Longleftrightarrow U_i \perp U_j | U_{V \setminus \{i,j\}} \Longleftrightarrow e_{ij} = 0 \tag{5.8}$$

where the symbol $\perp$ indicates conditional independence.

From a graph theoretic viewpoint, the previous equivalence means that a link between two volume residuals is present if and only if the corresponding partial correlation coefficient is significantly different from zero.

From a financial viewpoint, the previous equivalence implies that, if the partial correlation between two measures is equal to zero, the corresponding volumes residuals are conditionally independent and, therefore, the corresponding groups do not (directly) impact each other.

From a statistical viewpoint, it is also possible to test the null hypotheses that two groups of Bitcoin owners are conditionally independent by controlling whether the corresponding partial correlation coefficient is equal to zero, by means of the statistical test described in Whittaker (1990) or in Giudici (2003).

However, this poses a problem of multiple testing, and correcting for this problem might results in loss of power. Namely, Bonferroni adjustment is a widely used method for multiple comparisons but it can be very conservative especially when there are a large number of hypotheses that are being simultaneously tested (S.-Y. Chen, Z. Feng, and Yi, 2017). As an extension to this, several different adjustment methods have been proposed in the literature including Holm adjustment, Hochberg adjustment, Hommel adjustment, Benjamini-Hochberg adjustment, etc., all of which are characterized with certain advantages and shortcomings. Most relevant method in this context could be the Holm adjustment which addresses the drawback of the Bonferroni method by being much less conservative (S.-Y. Chen, Z. Feng, and Yi, 2017). An alternative, widely used method for limiting the number of spurious edges—while at the same time obtaining networks that are more interpretable, —is through the use of a regularization approach. One such prominent approach of regularization is the 'least absolute shrinkage and selection operator (LASSO) which in its essence, allows us to set estimates of exactly zero. More formally, the LASSO limits the sum of absolute partial correlation coefficients which in turn lead to overall shrinkage of estimates and inviolably some become zero. Mathematically, if $\hat{\sigma}$ represents the sample variance–covariance matrix LASSO aims to estimate the precision matrix by maximizing the penalized likelihood function (with λ_k being the penalty parameter)

$$l(\Theta) = log\ (det\Theta) - tr(\hat{\sigma}\Theta) - \lambda_k \textstyle\sum_{i,j}(|\Theta_{i,j}|)$$

For the purpose of our study, both the significance testing and the graphical LASSO serve as a robustness check for identifying the true network that emerges between Bitcoin owner groups.

Data. We consider all data from the Bitcoin blockchain, from 25 February 2012 to 17 July 2017 (1969 days with 1843 observed days), described in detail in Y. Chen, Trimborn, and Zhang (2018). Bitcoin blocks are published approximately every 10 minutes and contain information about the transaction size, the account ID (anonymous), the participating accounts and the timestamp of the transactions.

The previous information is very useful to understand the time dynamics of volume transactions, but it indicates nothing about the nature of the bitcoin owners who generate the trade. Trying to capture some kind of information on bitcoin traders, we consider the website Blockchain.info which provides information about the IP address of the relying party that provides a secure access to the originator of each transaction, and extract from it the approximate geographical provenience of the trader who generates the transaction. To avoid a too large approximation error, we decided to group geographical provenience in a few number of classes, corresponding to six continental groups: Africa (Af), Asia (As), Europe (Eu), North America (N_A), Oceania (Oc) and South America (S_A). More precisely, the continent of the bitcoin trader is identified from the data in Blockchain.info, comparing its IP address with a data set of IP address from MaxMind Inc. The approximate location of the transaction origin can be tracked by recording the first node relaying it, according to the method described in Fergal and Martin (2011). We remark that this approach works as long as the running node does not use an anonymizing technology.

We thus have a first grouping of bitcoin owners that roughly correspond to their continent of residence. To further characterize them, for each of the six continental groups we associate to each account IDs according to the absolute size of the total transaction amount they generate in the considered time period. We then further group the IDs of each continent according to the deciles of their statistical distribution. The first group, which will be labeled 1 after the continent abbreviation, has the smallest transactions, corresponding to the $0-10\%$ percentile class, while the tenth group with the largest transactions is labeled 10, corresponding to $90-100\%$ percentile class. The final result is a classification of bitcoin owenerin 60 groups: 10 groups per continent.

With this grouping we will investigate our research hypotheses, and search for the bitcoin owners who mostly impact the market. Specifically we will be able to investigate whether large-size Bitcoin owner affect the trade decisions of the others, or whether a specific continent drives the others, in terms of bitcoin trades, or both.

We remark that, although the Bitcoin is the most liquid and largest cryptocurrency, there is sometimes low liquidity in its transactions. Our data show that there are days without a single transaction in Africa, Asia, Oceania and South America, with frequency of low liquidity varying between 1% and 25%. We can overcome the liquidity problem by accumulating the 10 minute data to a daily frequency. In any case, this indicates that a further regional grouping, for example by countries, would lead to lack of data for many of them.

For each of our considered groups, our main variable of interest is the volume of transactions, in any given time point. To normalise such data, we consider the logarithm of the transaction volumes. To avoid computational problems, when no transactions in a group arise within a day, we add 1 Satoshi [7] to each transaction. Given the large numbers under consideration, the bias effect of the correction is negligible.

In Figure 22 we illustrate the daily log accumulated transaction sizes over all 10 groups in each continent. The largest transaction sizes appear in Europe and North America, whose dynamic pattern is quite steady. Asia and Oceania are evidently more volatile then Europe and North America, but less volatile than Africa and South America. The descriptive statistics, reported in Table 39, provide further evidence to these findings. Note in particular that Asia, Oceania, Africa and South America have a minimum value of zero, indicating a lack of liquidity in certain time periods.

[7] The BTC transactions are reported in Satoshi values, the smallest fraction of a BTC, where 1 BTC = $100,000,000$ Satoshi.

For deeper insights into the data features of the groups in each continent, the empirical distribution of the log transaction sizes is displayed by means of boxplots in Figure 22. For each continent, the left plot corresponds to the first group, namely the group 1 with the smallest transactions, and the right one to the group 10 with the largest transactions, respectively.

From Figure 22, the narrow box width of Europe and North America suggests that these continents are characterised by transaction sizes with low volatility and few outliers. However for Asia and Oceania the daily transaction sizes are more volatile, and lead to larger center boxes and wider whiskers. South America becomes extreme in the sense of showing even longer whiskers, with transaction sizes varying stronger between groups. Africa follows a very different picture from the other continents: it has the lowest liquidity and a much higher volatility and it shows frequent drops of the transaction volume to 0.

	Af	As	Eu	N_A	Oc	S_A
mean	142.25	193.77	232.18	230.45	186.60	155.80
sd	72.84	19.81	11.59	9.18	24.55	62.39
skewness	-1.30	-4.81	-0.86	-1.61	-4.59	-1.91
kurtosis	2.98	44.71	5.27	10.50	34.79	5.12
min	0.00	0.00	162.72	154.25	0.00	0.00
max	222.76	240.14	257.76	254.96	235.36	228.09

Table 39: **Descriptive statistics of the accumulated log transactions of the 6 regions Africa (Af), Asia (As), Europe (Eu), North America (N_A), Oceania (Oc), South America (S_A).** Eu and N_A show a related behavior in terms of the descriptive statistics, as so do As and Oc. Also Af and S_A behave related.

Empirical results (chenbranka) In this Section we present the results from the application of the proposed model. First we evaluate the model in terms of predictive accuracy, to gauge its validity in the present context; second, we interpret the model results in terms of our research hypotheses, aimed at assessing the dependency patterns among the trading behaviour of different bitcoin traders.

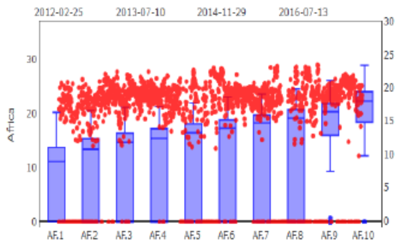
We first consider an unregularised network, whose edges are all present, even when the corresponding partial correlation is very low.

By calculating the partial correlations as specified in (5.6), we can derive the B_0 matrix and, then, the autoregressive parameters $A_1, \dots, A_{p0}$. We are thus able to disentangle the time-dependent volume of node i, separately estimating the autoregressive idiosyncratic component and the contemporaneous one, according to equation (5.2). Table 40 presents the assessment of the predictive performance of our model, to understand if the proposed approach is suitable, from a statistical viewpoint. Specifically, we want to investigate whether the inclusion of the contemporaneous component improves predictive accuracy, with respect to a much simpler pure autoregressive model. Table 40 contains the results of the predictive assessment.

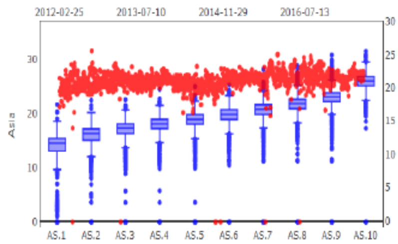
From Table 40 note that, in most cases, the proposed model overperforms a pure autoregressive model, as the corresponding root mean squared errors of the one-step ahead predictions are lower in the vast majority of cases. For robustness check, we also compute the RGA index proposed by Agosto and Raffinetti (2019). The results are presented in Table 41.

We now move towards the interpretation of the results that can be drawn from our model and, specifically, from the partial correlations (equation (5.6)). In Figure 23, each node represents one of the 60 groups of traders and each present edge indicates that two traders are dependent on each other, in terms of their transactions (conditionally on all the others). Differently, when an edge is missing, the corresponding traders behave independently of each other (conditionally on all the others). Each edge is associated to a weight, which corresponds to a partial correlation coefficient. The size of each edge in Figure 23 is proportional to such weight. On the other hand, the coloring of an edges between two nodes indicates the sign of the partial correlation coefficient:

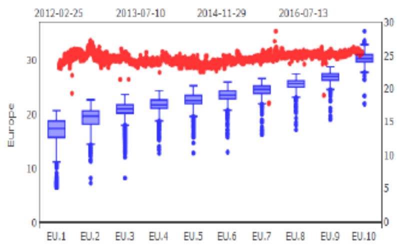
(a) Af



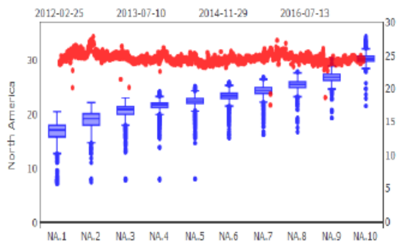
(b) As



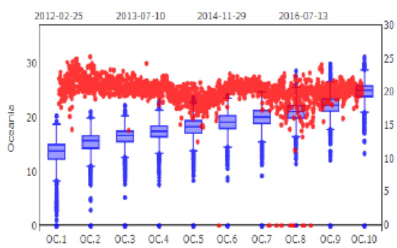
(c) Eu



(d) N_A



(e) Oc



(f) S_A

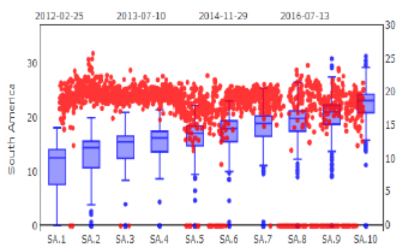


Figure 22: **Daily log transactions of the 10 groups displayed as boxplots, where the left boxplot represents the 1. group and the right one the 10. group of the respective continent.** The scatter plot displays the accumulated log transaction size of the 10 groups. The time period is 25 February 2012 until 17 July 2017 in the 6 continents.

Group	RMSE_Full	RMSE_AR	Group	RMSE_Full	RMSE_AR
Africa1	0.1945	0.2052	N_A1	0.2495	0.2500
Africa2	0.1298	0.1315	N_A2	0.4590	0.4613
Africa3	0.1600	0.1584	N_A3	0.5523	0.5596
Africa4	0.1521	0.1538	N_A4	0.3241	0.3631
Africa5	0.1492	0.1460	N_A5	0.8437	0.8530
Africa6	0.1609	0.1538	N_A6	1.2396	1.2653
Africa7	0.1385	0.1419	N_A7	0.9865	0.9951
Africa8	0.1382	0.1371	N_A8	0.8721	0.9041
Africa9	0.1276	0.1250	N_A9	0.6895	0.6962
Africa10	0.0960	0.0979	N_A10	1.2575	1.2698
Asia1	0.2258	0.2286	Oceania1	0.3182	0.3209
Asia2	0.2340	0.2264	Oceania2	0.2447	0.2477
Asia3	0.3148	0.3173	Oceania3	0.3717	0.3655
Asia4	0.3479	0.3432	Oceania4	0.4795	0.4914
Asia5	0.4328	0.4501	Oceania5	0.4909	0.5057
Asia6	0.5425	0.5493	Oceania6	0.5837	0.5782
Asia7	0.6143	0.6064	Oceania7	0.5857	0.5965
Asia8	0.6403	0.6455	Oceania8	0.8265	0.8353
Asia9	0.5294	0.6863	Oceania9	0.3350	0.3255
Asia10	0.5565	0.5623	Oceania10	0.2659	0.2733
Europe1	0.0558	0.0572	S_A1	0.2577	0.2663
Europe2	0.1414	0.1433	S_A2	0.2162	0.2183
Europe3	0.1779	0.1894	S_A3	0.2315	0.2326
Europe4	0.1405	0.1423	S_A4	0.2307	0.2302
Europe5	0.1822	0.1839	S_A5	0.2196	0.2231
Europe6	0.2241	0.2257	S_A6	0.2227	0.2234
Europe7	0.2852	0.2880	S_A7	0.2152	0.2145
Europe8	0.3673	0.3688	S_A8	0.2052	0.2061
Europe9	0.4021	0.4028	S_A9	0.1970	0.1960
Europe10	0.3460	0.3481	S_A10	0.1749	0.1757

Table 40: **Comparison between the root mean square errors obtained with our full VAR model and with a model composed by the solely autoregressive component**

Group	**RG_Full**	**RG_AR**	**Group**	**RG_Full**	**RG_AR**
Africa 1	172.67	10.14	N_America1	0.59	0.63
Africa 2	120.32	114.56	N_America2	2.54	2.29
Africa 3	59.63	77.36	N_America3	15.96	14.54
Africa 4	155.10	140.32	N_America4	36.98	27.90
Africa 5	76.51	106.17	N_America5	313.70	304.49
Africa 6	31.26	92.04	N_America6	1280.55	1226.63
Africa 7	199.74	153.74	N_America7	74.35	72.71
Africa 8	88.42	102.89	N_America8	28.79	25.41
Africa 9	68.48	104.43	N_America9	5.58	5.04
Africa 10	148.46	129.22	N_America10	18.00	16.00
Asia 1	7.88	7.14	Oceania1	51.09	41.06
Asia 2	4.11	6.52	Oceania2	21.27	17.73
Asia 3	12.10	11.30	Oceania3	32.90	44.78
Asia 4	12.91	13.92	Oceania4	429.56	313.80
Asia 5	33.70	26.81	Oceania5	10979.07	3015.27
Asia 6	79.21	76.58	Oceania6	438.52	488.45
Asia 7	132.08	136.21	Oceania7	131.47	96.04
Asia 8	236.27	233.19	Oceania8	117.85	105.48
Asia 9	374.08	152.59	Oceania9	241.78	364.92
Asia 10	104.35	99.71	Oceania10	62.23	39.85
Europe 1	0.65	0.60	S_America1	563.92	61.65
Europe 2	1.34	1.28	S_America2	1311.00	947.96
Europe 3	1.54	0.93	S_America3	974.09	682.45
Europe 4	1.64	1.59	S_America4	172.52	223.79
Europe 5	2.00	1.95	S_America5	553.35	133.24
Europe 6	1.47	1.45	S_America6	116.88	74.70
Europe 7	0.65	0.54	S_America7	20.02	35.38
Europe 8	0.49	0.39	S_America8	95.60	51.28
Europe 9	0.17	0.14	S_America9	48.20	82.92
Europe 10	1.22	1.12	S_America10	1221.34	910.02

Table 41: **Comparison between the RGA index obtained with our full VAR model and with a model composed by the solely autoregressive component**

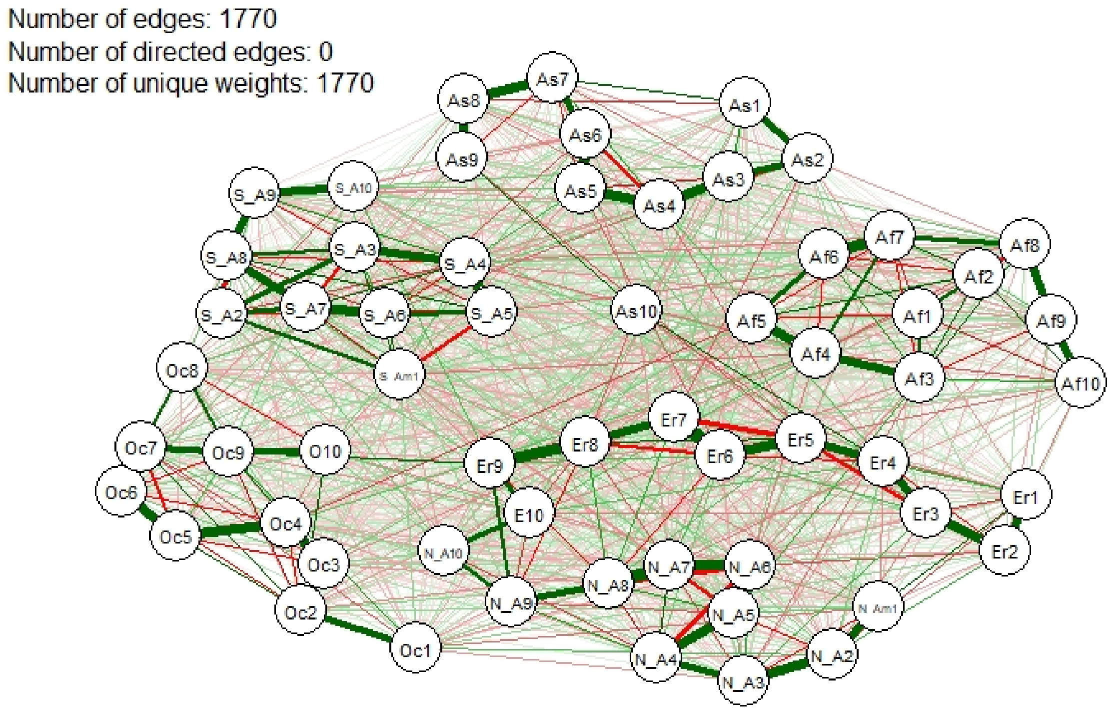


Figure 23: **Unregularized partial correlation network**

green highlights a positive partial correlation and red a negative partial correlation.

What we can observe from the network that emerges from Figure 23 is that there exist many interconnections between Bitcoin groups of users. Precisely, the summary statistics provided in the upper left corner of Figure 23 indicates that the network contains a total of 1770 non-zero links between groups. Although the graph is difficult to interpret, some clusters can be identified. We can see about five clusters which in most part correspond to the continents, with the exception of Europe and North America which are placed in the same cluster, suggesting that there exist strong dependence between the traders of the two continents. This is something that we expected to see due to the economic and political similarities among the two regions, as well as on their news sharing.

Note also that the groups representing the larger traders in Europe and North America - N_A10, N_A9, Eu10, Eu9 - show stronger positive connections than other groups. This may be explained by the fact that these groups have a comparable size of transactions, which come from a similar set of information, which induce them to behave similarly. If we match this result with that in Figure 22, which indicates the relatively larger volumes of transactions coming from these groups, we obtain a clear indication that these are the groups which can mostly impact the market. Note also that there exists a strong positive link between Oc10 and Eu9, and not between Oc9 and Eu09. This is consistent with our previous finding: the transaction volumes of Oc10 are more comparable in their size to Eu9, rather than to Eu10 (see Figure 22) and, therefore, they act similarly.

As mentioned previously, in unregularized correlation networks some edges may be present but may not be statistically significant. In the graphical representation, such situations will be visualized as very weak connections in the network. In order to prevent this and to correctly identify the significant associations between Bitcoin groups, a crucial step is to impose restrictions that will limit (or eliminate) the occurrence of spurious edges. One way to achieve this is by testing

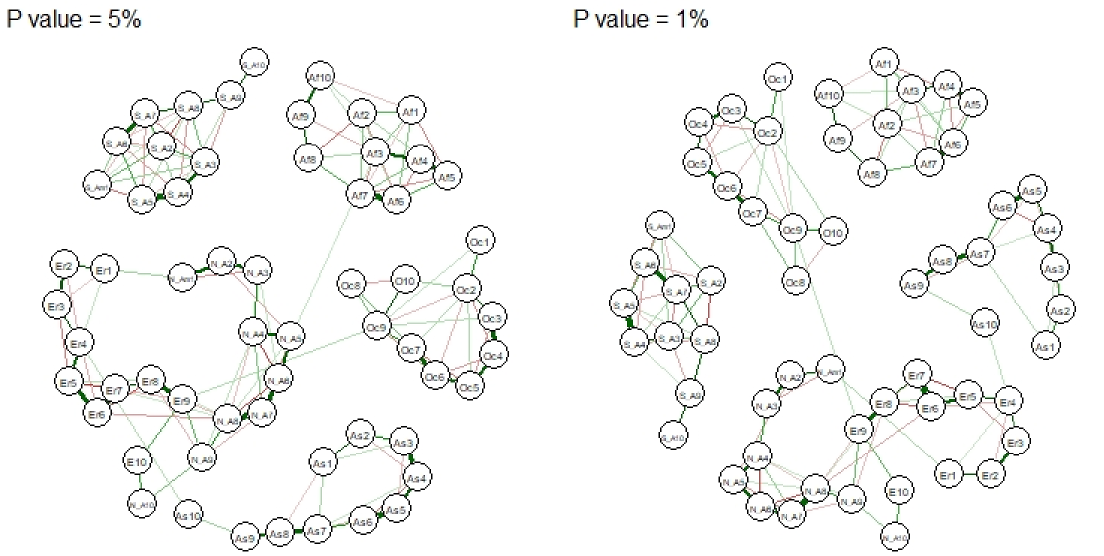


Figure 24: **Regularized partial correlation networks (without edges that are not significant)**. Software used: R, Code available upon request

the statistical significance of partial correlations.

Figure 24 presents the same network containing only links that are found statistically significant at both 5% and 1% level of significance.

Figure 24 shows that the structure of the network does not change significantly if we impose different levels of significance. What we observe from the graphs is that the majority of links that were present in the unregularized network have disappeared, reducing the total number of links from 1770 to 146 and 137, respectively. Interesting, even though a significant portion of the links have been removed, the clustering of nodes remains the same as in Figure 23. Specifically, we see the formation of clusters equivalent to the continents and we also see significant interconnection between traders in Europe and North America. Furthermore, we also see a statistically significant positive correlation Oceania's top group and Europe's and between Asia's top group and Europe's.

To further confirm our findings, we perform a further robustness check through the application of the graphical LASSO. As discussed previously, LASSO is a very popular method for eliminating spurious links. Figures 25 and 26 represent the networks that emerge by the applying graphical LASSO with two different smoothness parameters λ. We remark that, unlike the classical LASSO, in the graphical approach the choice of λ cannot be done based on cross-validation as it represents a completely unsupervised process. As a results, we attempted to identify the boundaries of λ and then use an information criterion to decide on the network that best fits the data.

From Figures 25 and 26, the changing λ does change the structure of the network, but the underlying clusters remain the same, thus confirming the close interconnection between Europe and North America, as well as those between top traders in Oceania and Europe.

Such networks can allow for powerful interpretations. Specifically, they allow for the modeling of unique interactions between users and mapping out predictive mediation. Namely, the majority of the asset allocation and finance literature is about understanding dependencies that will allow for predictive modeling. Partial correlations are closely related to coefficients obtained in multiple regression models: when an independent variable does not predict or cause, the dependent variable, we would not expect an edge in the network. Translated in this context, if the trading behavior of one node (Bitcoin user) does not help in predicting the trading behavior of another, we would not expect any linkages between the two. Hence by looking at the links that emerge, we clearly see that there is a cross dependence between Bitcoin users i.e trading volumes of individual users

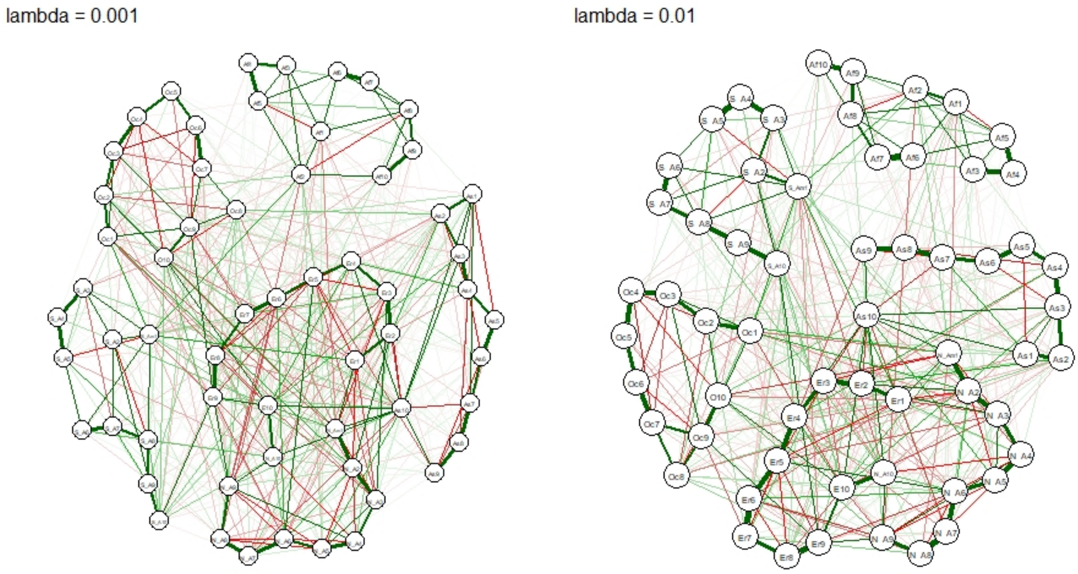


Figure 25: **GLASSO partial correlation networks [varying lambda], 1/2**

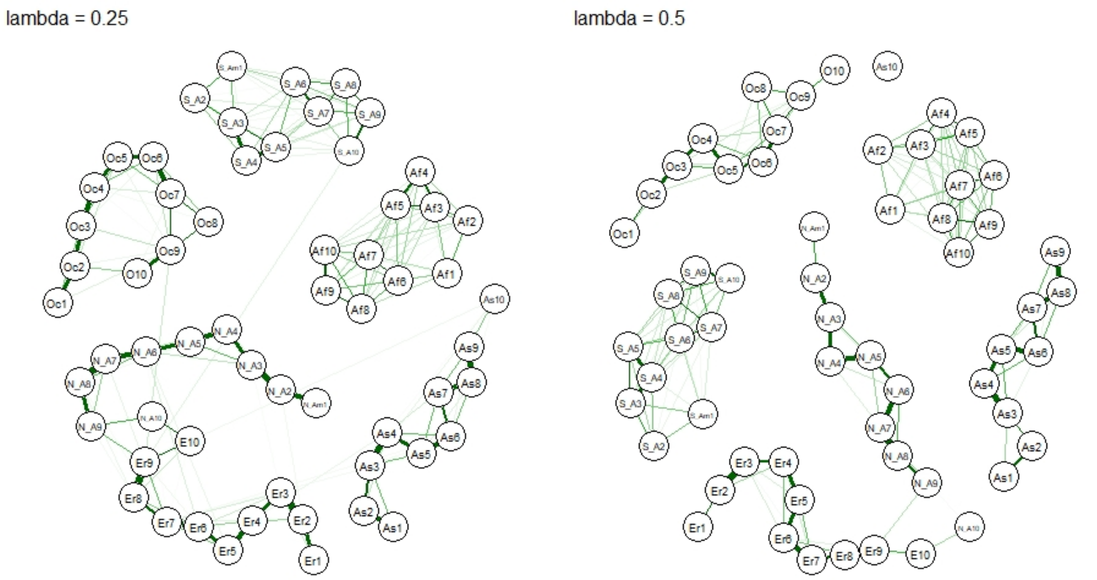


Figure 26: **GLASSO partial correlation networks [varying lambda], 2/2** Software used: R, Code available upon request

	Lambda 0.001		Lambda 0.01	
	Positive	**Negative**	**Positive**	**Negative**
Within Europe	17	14	17	13
Within North America	21	13	19	13
Between Europe and North America	48	53	45	48

Table 42: **Count of links between and within North America and Europe** Software used: R, Code available upon request

are dependent on those of other users in the network.

A closer inspection of Figure 25, reveals frequent linkages between European and North American nodes, which is in line with the previous observations. Positive linkages appear more often inside each continent, compared to negative ones. One the other hand negative and positive edges appear frequently between two continents (see Table 42).

The largest two groups in both continents share strong links with each other, confirming that that they probably share a common information set. Interestingly the largest trader group from Asia, AS10, has multiple positive edges to several groups in Europe and North America. Considering that most bitcoin mining farms are based in Asia, and especially in China, it follows that a large amount of capital is acquired and, therefore, traded, from Asia with the rest of the world. Last, note that the largest volume trading groups from Oceania and South America also share links with each other and with the larger Western-World groups. This observation leads to the conclusion that the large traders around the world are somewhat connected, possibly communicating with each other. On the other hand smaller groups, which have less information, shows less connections around the world.

Figure 26 shows what happens when we increase the penalty level to $\lambda = 0.25$. Most edges vanish, but the previously found connections persist. Still the largest trader groups from Europe and North America remain connected, while the edges from Oc9, S_A10 and As10 persist to stay connected with them. The connection goes via the largest groups in Europe, namely Eu9 and Eu10. Other persisting edges exist between the smaller groups from Asia and Europe, yet with small magnitude. Within the continents many edges are not affected by the penalty, hence emphasize the importance of the regional connectedness. Finally, when increasing the penalty parameter to $\lambda = 0.5$, most cross-continent edges are ruled out, except for the ones between the largest groups in Europe and North America. The remaining edges only appear within the continents.

To summarise, our empirical findings give an answer to our research proposition: which are the group of traders that mostly affect the bitcoin markets? These groups have been found among the top two classes of traders in North America and Europe, strongly and positively connected to each other. These traders are linked to the others, affecting their behaviours. In particular, they are especially linked with the top traders from Oceania and South America. In addition, top traders from Asia, and especially larger ones, are highly linked to the others, likely as a result of their mining activity.

Discussion. In this extensions, we have proposed a model that explains the dynamics of Bitcoin trading volumes, based on a correlation network VAR process that models the interconnections between different groups of traders. Our main methodological contribution consists in the introduction of partial correlations and correlation networks into VAR models. This allows to describe the correlation patterns between trading volumes and to disentangle the autoregressive component of volumes from its contemporaneous part. The introduction of VAR correlation networks also allows to build a volume predictive model, that leverages the information contained in the correlation patterns.

Our main financial findings show that trading volumes are highly correlated within geographical regions. Groups of traders with high transaction volumes over all continents covary in the network model, leading to the conclusion this groups share a mutual information set. The results are robust over various penalized network models. These results contribute to the identification of group of bitcoin traders that are the most likely influencers of the market. These are found to high volume traders, especially from North America, Europe and Asia.

6 Conclusion

Fintech services are becoming part of the everyday life. Such new technologies can increase financial inclusion, but they may come at the cost of increased credit risks. To cope with such risks, Fintech risk management becomes a central point of interest for regulators and supervisors, to protect consumers and preserve financial stability.

In this thesis, we develop and test several Fintech risk management tools, based on network theory, that can lead to increased predictive accuracy of statistical-based credit scoring for P2P systems. Specifically, we propose: i) correlation-based credit scoring (in the case in which time-varying financial information on borrowers is available on the platform); (ii) similarity-based credit scoring (for cross-sectional data), (iii) factor-network-based segmentation. Moreover, the thesis provides an extension beyond the world of Fintech credit. Namely, we also include an application of network theory for understanding the dynamics of Bitcoin blockchain trading volumes and, specifically, identifying how different trading groups interact with each other. We briefly summarize the main results related with the methodologies developed and tested in this work.

Predictive accuracy of traditional scoring models. The first objective of this thesis is to test the predictive performance of traditional credit scoring models as employed by one of the largest P2P platform - Lending Club and investigate whether the ratings assigned by the platform are a good predictor of default. The empirical findings from this initial analysis suggest that although there is a statistically significant relationship between the assigned rating grades and loan default, grades do not have high predictive power. Furthermore, the predictive performance does not change significantly if we apply more elaborate statistical models using the information that Lending Club collects on the borrowers. Such findings indicate a need for exploring other approaches beyond the traditional scoring models.

*** Having obtained evidence of the need to introduce alternative methodologies that would improve the predictive accuracy of P2P scoring models, in the following step, we propose network-based scoring models which will take into account the financial relationships between borrowers. Namely, a key characteristic of P2P platforms is that they are, by construction, globally interconnected. Classical banks have, over the years, segmented their reference markets into specific territorial areas thus increasing their expertise and the accuracy of their ratings. Differently, P2P platforms are based on a global, fully inclusive banking model, which in turn makes the determination of a correct rating a particularly difficult task. However, these platforms have the advantage of an improved data collection on the borrowers' network, and integrating these information into the credit scoring specification can improve the models' predictive power.

Correlation-based credit scoring model. With the correlation-based scoring model, we show how scoring models can be improved in the business P2P lending context: using information on the similarity among the borrower companies, calculated as pairwise correlations between the values of a given set of financial variables, calculated on their balance sheets. We further show how such correlations can be embedded into a multi-layer correlation network model, whose centrality measures can be extracted to improve the predictive performance of the logistic regression model, typically used in credit scoring applications. The application of the proposed correlation-based credit scoring model to a database of 727 southern European SMEs connected into a network of P2P borrowers demonstrates that the model is appealing, not only theoretically but also from an operational point of view, as it significantly increases the predictive accuracy of borrowers' default.

Similarity-based credit scoring model. One of the main limitations of the proposed correlation-based credit scoring model is related with the availability of data as it requires P2P systems to have access to time-varying information on borrowers' financial performance. Having in mind this limitation, we propose a similarity-based credit scoring model that leverages topological information embedded into similarity networks derived from borrowers' financial ratios at one point in time. We show that such topological information can be exploited to increase the predictive performance of most of the credit scoring models even if the forecasting gain obtained differ from model to model, depending also from the data set employed. This is particularly true for the models trained and tested on a smaller sample of data containing very restricted traditional information on the companies' performance for which we show that network parameters are much more relevant compared to more inclusive data sets.

Latent factor-based classification technique. We also propose a factor network-based approach that can improve credit risk management of SMEs engaged in P2P credit services. Specifically, we segment a heterogeneous population into a cluster of homogeneous sub-populations and estimating a more-efficient credit score model on the clusters using general and lasso-type regularization logistic model. We demonstrate the effectiveness of our approach through empirical applications analyzing the probability of default on two separate data sets, one of which contains information on 15,000 SMEs involved in P2P lending across Europe. In the context of both data sets, we compare the results from our model with the one obtained with standard single credit score methods and we find evidence that our factor network approach helps to obtain sub-population clusters such that the resulting models associated with these clusters are more parsimonious than the conventional full population approach, leading to better interpretability and to a modest improved default predictive performance. Our proposed factor network approach to credit score modeling presents an efficient framework to analyze the interconnections among the borrowers of a peer to peer platform and provides a way to segment a heterogeneous population into clusters with more homogeneous characteristics. The results show that the lasso logistic model for credit scoring leads to better identification of the significant set of relevant financial characteristic variables, thereby producing a more interpretable model, especially when combined with the segmentation of the population via the factor network-based approach.

Beyond Fintech credit: Crypto markets. Network theory can be applied to Fintech risk management in contexts different from Fintech credit. This thesis also presents a new model that explains the dynamics of Bitcoin trading volumes, based on a correlation network VAR process that models the interconnections between different groups of traders. Our empirical findings show that trading volumes are highly correlated within geographical regions. Also, groups of traders with high transaction volumes over all continents covary in the network model, leading to the conclusion that these groups share a mutual information set. The results are robust over various penalized network models. An important advantage of the proposed model is that it can be used to predict "regular" trading volumes and identify anomalies. Similar as in the Fintech credit application, the empirical findings from this work show that the proposed network-based model is able to predict trading volumes with an error that is lower than that of a pure autoregressive model.

Concluding remarks. In a general sense, the thesis contributes to the research literature on credit risk modelling in P2P systems, in two main directions. It first shows that classical credit scoring models may not be good predictors of default, both from a financial point of view, due to the very nature of P2P lenders in comparison with classical banks; and from an empirical and statistical perspective, as their available information is limited. Consequently, it shows how scoring models can be improved, using the information that is automatically collected in P2P systems, which typically connect economic agents, in a network of relationships and transactions. Such network structure extends beyond the classical geographical and economic sectors. Information that can be embedded into financial network models can capture non-linearities and endogenous factors that explain the spread of credit risk through the network. The application of our proposed network based scoring models to a variety of different data sets is promising as in all cases, accounting for the underlining network leads to an increased predictive accuracy of the classification and regression tasks.

Future research directions should include building network models using financial transactions between companies, rather than balance sheet variables, to get a more timely and precise network description. In addition, following this direction, research should consider building multi layered networks that take into account many levels of interactions between borrowers: transactions, common holdings, common shareholders, and etc.

We also remark how our analysis can be translated in current business and financial action. Overall, we believe that the main beneficiaries of our results may be regulators and supervisors in charge of preserving financial stability, as well as the lenders and the investors of P2P platforms who should be protected against the negative sides of Fintech innovations (related with information asymmetries) while keeping the positive sides (related with the improvement of financial services). The necessity for improving the scoring accuracy becomes even more important in view of the exceptionally fast growing P2P lending sector. Studies have repeatedly shown steady increase in the cumulative amount of loans that originate via marketplace lending platforms and the trend is expected to accelerate in the foreseeable future. Failing to ensure credit scoring models with high predictive accuracy could lead to bad debt accumulation and financial instability.

References

Abdi, H. (2007). "Singular Value Decomposition (SVD) and Generalized Singular Value Decomposition (GSVD)". In: *Encyclopedia of Measurement and Statistics*.

Abellán, J. and J. G. Castellano (2017). "A comparative study on base classifiers in ensemble methods for credit scoring". In: *Expert Systems with Applications* 73, pp. 1–10.

Acharya, V., R. Engle, and M. Richardson (2012). "Capital Shortfall: A New Approach to Ranking and Regulating Systemic Risks". In: *American Economic Review: Papers and Proceedings* 102.3, pp. 59–64.

Adams, M., B. Burton, and P. Hardwick (2003). "The Determinants of Credit Rating in the United Kingdom Insurance Industry". In: *Journal of Business Finance and Accounting* 30.3, pp. 539–572.

Adrian, T. and M. Brunnermeier (2015). "CoVar". In: *American Economic Review: Papers and Proceedings* 106.7, pp. 1705–1741.

Agosto, A., P. Giudici, and E. Raffinetti (2019). "A rank graduation accuracy measure". In: *Working Paper*.

Agosto, A. and E. Raffinetti (2019). "Validation of PARX Models for Default Count Prediction". In: *Frontiers in Artificial Intelligence* 2, p. 9. ISSN: 2624-8212. DOI: 10.3389/frai.2019.00009. URL: https://www.frontiersin.org/article/10.3389/frai.2019.00009.

Ahelegbey, D., M. Billio, and R. Casarin (2016). "Bayesian Graphical Models for Structural Vector Autoregressive Processes". In: *Journal of Applied Econometrics* 31.2, pp. 357–386.

Ahelegbey, D., P. Giudici, and B. Hadji-Misheva (2019a). "Factorial Network Models to Improve P2P Credit Risk Management". In: *Frontiers in Artificial Intelligence* 2, p. 8. ISSN: 2624-8212. DOI: 10.3389/frai.2019.00008. URL: https://www.frontiersin.org/article/10.3389/frai.2019.00008.

— (2019b). "Latent Factor Models For Credit Scoring in P2P Systems". In: *Physica A: Statistical Mechanics and its Applications* 522, pp. 112–121.

Ahelegbey, D., M. Billio, and R. Casarin (2015). "Bayesian Graphical Models for Structural Vector Autoregressive Processes". In: *Journal of Applied Econometrics*.

Akerlof, G. A. (1970). "The Market for "Lemon": Quality Uncertainty and the Market Mechanism". In: *The Quarterly Journal of Economics* 84.3, pp. 488–500.

Aldasoro, I. and I. Alves (2016). "Multiplex interbank networks and systemic importance: An application to European data. " In: *ECB*.

Allen, F. and D. Gale (2000). "Financial Contagion". In: *Journal of Political Economy*.

A.Lopeza, J. and M. R.Saidenberg (2000). "Evaluating credit risk models". In: *Journal of Banking and Finance* 24, pp. 151–165.

Anderson, R. (2007). *The credit scoring toolkit: theory and practice for retail credit risk management and decision automation*. Oxford University Press.

Andreas, F. et al. (2018). "The Role of Technology in Mortgage Lending". In: *Federal Reserve Bank of New York Staff Reports* Staff Report 836.

Andreeva, G., J. Ansell, and J. Crook (2007). "Modelling Profitability Using Survival Combination Scores". In: *European Journal of Operational Research* 183.3, pp. 1537–1549.

Arner, D., A. Barberist, and R. Buckley (2016). "The Evolution of FinTech: A New Post-Crisis Paradigm ". In: *Georgetown Journal of International Law* 1271.

Avdjiev, S., P. Giudici, and A. Spelta (2018). "Measuring Contagion Risk in International Banking". In: *Journal of Financial Stability*.

Bachmann, A. et al. (2011). "Online Peer-to-Peer Lending – A Literature Review". In: *Journal of Internet Banking and Commerce* 16.2.

BAFIN (2019). "Big data meets artificial intelligence: Challenges and implications for the supervision and regulation of financial services". In:

Balakrishnan, H. et al. (2003). *Looking Up Data in P2P Systems*. Tech. rep. COMMUNICATIONS OF THE ACM.

Barabasi, A. L. (2016). "Network science". In: *Cambridge University Press*.

Barrios, L. J. S., G. Andreeva, and J. Ansell (2014). "Monetary and Relative Scorecards to Assess Profits in Consumer Revolving Credit". In: *Journal of the Operational Research Society* 65.3, pp. 443–453.
Battiston, S. et al. (2012). "Liasons dangereuses: Increasing connectivity risk sharing, and systemic risk". In: *Journal of Economic Dynamics and Control* 36.8, pp. 1121–1141.
Bayon, P. S. and L. G. Vega (2018). "Automated Investment Advice: Legal Challenges and Regulatory Questions". In: *Banking and Financial Services Policy Report* 37.
Bazzi, M. et al. (2016). "Community Detection in Temporal Multilayer Networks, with an Application to Correlation Networks". In: *Multiscale Model. Simul* 14, pp. 1–41.
Berg, T. et al. (2018). "On the rise of fintech - credit scoring using digital footprint". In: *NBER Working Papers* 24551.
Berry, H., M. F. Guillen, and N. Zhou (2010). "Induced aggregation operators in the Euclidean distance and its application in financial decision making". In: *Journal of International Business Studies* 41.9, pp. 1460–1480.
B.Gordy, M. (2000). "A comparative anatomy of credit risk models". In: *Journal of Banking and Finance* 24, pp. 119–149.
Billio, M. et al. (2012). "Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors". In: *Journal of Financial Economics* 104.3, pp. 535–559.
Blondel, V. D. et al. (2008). "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10, P10008.
Bonacich, P. (2007). "Some unique properties of eigenvector centrality". In: *Social networks* 29.4, pp. 555–564.
Bonanno, G. et al. (2003). "Topology of correlation-based minimal spanning trees in real and model markets". In: *Physical Review E* 68.4, p. 046130.
Bouoiyour, J. and R. Selmi (2016). "Bitcoin: A beginning of a new phase?" In: *Economics Bulletin* 36, pp. 1430–40.
Bourreau, M. and T. Valletti (2015). *Enabling Digital Financial Inclusion through Improvements in Competition and Interoperability: What Works and What Doesn't?* Policy Paper 065. Center for Global Development.
Brin, S. and L. Page (1998). "The anatomy of a large-scale hypertextual web search engine". In: *Computer networks and ISDN systems* 30.1-7, pp. 107–117.
Brunetti, C. et al. (2015). "Interconnectedness in the Interbank Market". In: *FED Working Paper No.2015-090*.
Buchak, G. et al. (2018). "Fintech, regulatory arbitrage, and the rise of shadow banks". In: *Journal of Financial Economics* 130.3, pp. 453–483.
Burton, A., L. J. Billingham, and S. Bryan (2007). "Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data". In: *Clinical Trials* 4.2, p. 7.
Calabrese, R. and P. Giudici (2015). "Estimating Bank Default with Generalized Extreme Value Regression Models". In: *Journal of the Operational Research Society* 66.11, pp. 1783–1792.
Carey, M. and M. Hrycay (2001). "Parameterizing credit risk models with rating data". In: *Journal of Banking and Finance* 25, pp. 197–270.
Catania, L., S. Grassi, and F. Ravazzolo (2018). *Predicting the Volatility of Cryptocurrency Time–Series.* Working Paper. CAMP Working Paper Series No 3/2018.
Chazelle, B. (2000). "A Minimum Spanning Tree Algorithm with inverse-Ackermann Type Complexity". In: *J. ACM* 47.6, pp. 1028–1047. ISSN: 0004-5411. DOI: 10.1145/355541.355562. URL: http://doi.acm.org/10.1145/355541.355562.
Chen, S.-Y., Z. Feng, and X. Yi (2017). "A general introduction to adjustment for multiple comparisons". In: *Journal of Thoracic Disease* 9.6. ISSN: 2077-6624. URL: http://jtd.amegroups.com/article/view/13609.
Chen, Y., S. Trimborn, and Zhang (2018). "Discover Regional and Size Effects in Global Bitcoin Blockchain via Sparse-Group Network AutoRegressive Modeling". In: *working paper*.
Chinazzi, M. et al. (2013). "Post-Mortem Examination of International Financial Network". In: *Journal of Economic Dynamics and Control* 37.8, pp. 1692–1713.

Chu, J. et al. (2017). "GARCH Modelling of Cryptocurrencies". In: *J. Risk Financial Manag* 10.17.
Crouhy, M., D. Galai, and R. Mark (2000). "A comparative analysis of current credit risk models". In: *Journal of Banking and Finance* 24, pp. 59–117.
De Roure, C., L. Pelizzon, and P. Tasca (2016). *How Does P2P Lending Fit into the Consumer Credit Market?* Tech. rep. Deutsche Bundesbank, Research Centre.
Deer, M. and Y. Yin (2015). "The rise of peer-to-peer lending in China: an overview and survey case study". In: *Association of Chartered Certified Accountants*.
Deloitte (2016). *Credit Scoring: Case Study in Data Analytics*. Working Paper. arXiv:1708.00977.
DeLong, E. R., D. M. DeLong, and D. L. Clarke-Pearson (1988). "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach". In: *Biometrics* 44.3, pp. 837–845.
Diebold, F. and K. Yilmaz (2014). "On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms". In: *Journal of Econometrics* 182.1, pp. 119–134.
Durante, D. and D. B. Dunson (2014). "Nonparametric Bayes Dynamic Modelling of Relational Data". In: *Biometrika* 101.4, pp. 883–898.
EC (2018). *Fintech Action Plan: For a More Competitive and Innovative European Financial Sector*. Technical Report. ECB.
ECB (2009). *Financial Stability Review*. Technical Report. ECB.
Eisenberg, L. and H. T. Noe (2001). "Systemic risk in financial systems". In: *Management Science*.
Elisabeth, M. et al. (2008). *Learning Trust in Dynamic Multiagent Environments using HMMs*.
Emekter, R. et al. (2015). "Evaluating Credit Risk and Loan Performance in Online Peer-to-Peer (P2P) Lending". In: *Applied Economics* 47.1, pp. 54–70.
Epskamp, S. F. and I. Eiko (2018). "A tutorial on regularized partial correlation networks". In: *Psychological Methods* 23.4, pp. 617–632.
Fergal, R. and H. Martin (2011). *An Analysis of Anonymity in the Bitcoin System*. arXiv: 1107.4524 [physics.soc-ph].
Fontes, A., L. L. Rodrigues, and R. Craig (2005). "Measuring convergence of National Accounting Standards with International Financial Reporting Standards". In: *Journal of International Business Studies* 29.4, pp. 415–436.
FSB (2017). *Financial Stability Implications for Fintech: Supervisory and Regulatory Issues that Merit Authorities' Attention*. Working Paper. Financial Stability Board.
Furfine, C. (2003). "Interbank Exposures: Quantifying the Risk of Contagion. Journal of Money". In: *Journal of Money, Credit, and Banking* 35.1, pp. 111–128.
Galindo, J. and P. Tamayo (2000). "Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications". In: *Computational Economics* 15, pp. 107–143.
Gander, W. (2008). *The Singular Value Decomposition*. ETH Zurich.
Ge, R., J. Feng, and B. Gu (2016). "Borrower's Default and Self-Disclosure of Social Media Information in P2P Lending". In: *Financial Innovation* 2.1, p. 30.
Genriha, I. and I. Voronova (2012). "Borrower's Default and Self-Disclosure of Social Media Information in P2P Lending". In: *Financial Innovation* 22, pp. 42–49.
Geron, A. (2017). *Hands-On Machine Learning with Scikit-Learn TensorFlow*. USA: O'Reilly.
Giudici, P. (2003). *Applied data mining*. Chichester (UK): John Wiley and Sons.
Giudici, P. (2018). "Fintech Risk Management: A Research Challenge for Artificial Intelligence in Finance". In: *Frontiers in Artificial Intelligence* 27.
Giudici, P. and B. Hadji-Misheva (2017). *Scoring Models for P2P Lending Platforms: A Network Approach*. Working Paper. University of Pavia.
Giudici, P., B. Hadji-Misheva, and A. Spelta (2019a). "Network based credit risk models". In: *Quality Engineering* 0.0, pp. 1–13. DOI: 10.1080/08982112.2019.1655159. eprint: https://doi.org/10.1080/08982112.2019.1655159. URL: https://doi.org/10.1080/08982112.2019.1655159.

Giudici, P., B. Hadji-Misheva, and A. Spelta (2019b). "Network Based Scoring Models to Improve Credit Risk Management in Peer to Peer Lending Platforms". In: *Frontiers in Artificial Intelligence* 2, p. 3. ISSN: 2624-8212. DOI: 10.3389/frai.2019.00003. URL: https://www.frontiersin.org/article/10.3389/frai.2019.00003.
Giudici, P. and B. H. Misheva (2018). "P2P lending scoring models: Do they predict default?" In: *Journal of Digital Banking* 2.4, pp. 353–368.
Giudici, P. and L. Parisi (2017). "Sovereign risk in the Euro area: a multivariate stochastic approach". In: *Quantitative Finance* 17.12, pp. 1995–2018.
— (2018). "Credit risk contagion through correlation network models". In: *Risks* 1.7.
Giudici, P., P. Sarlin, and A. Spelta (2017a). "The interconnected nature of financial systems: direct and common exposures". In: *Journal of Banking and Finance* In press.
— (2017b). "The interconnected nature of financial systems: direct and common exposures". In: *Journal of Banking & Finance*.
Giudici, P. and A. Spelta (2016). "Graphical Network Models for International Financial Flows". In: *Journal of Business & Economic Statistics* 34.1, pp. 128–138.
Golub, G. H. and C. Reinsch (1971). "Singular Value Decomposition and Least Squares Solutions". In:
Graham, R. L. and P. Hell (1985). "On the History of the Minimum Spanning Tree Problem". In: *Annals of the History of Computing* 7.1, pp. 43–57. DOI: 10.1109/MAHC.1985.10011.
Group, F. R. A. W. (2019). *Key Aspects around Financial Technologies and Regulation Policy report*. Working Paper. CEMLA FINTECH FORUM.
Gunn, S. R. (1998). *Support Vector Machines for Classification and Regression*. Tech. rep. Faculty of Engineering, Applied Science Department of Electronics, and Computer Science.
Guo, Y. et al. (2016). "Instance-Based Credit Risk Assessment for Investment Decisions in P2P Lending". In: *European Journal of Operational Research* 249.2, pp. 417–426.
Halu, A. et al. (2013). "Multiplex PageRank". In: *PLOS ONE* 8.10, pp. 1–10. DOI: 10.1371/journal.pone.0078293. URL: https://doi.org/10.1371/journal.pone.0078293.
Harald, H. et al. (2018). "Fintech credit, financial inclusion and entrepreneurial growth". In: *Working paper*.
Hautsch, N. (2015). "Financial Network Systemic Risk Contributions". In: *Review of Finance* 19, pp. 685–738.
He, D. et al. (1996). *Fintech and Financial Services: Initial Considerations*. Oxford (UK): IMF Staff Discussion Note.
Hoff, P. D. (2007). "Model Averaging and Dimension Selection for the Singular Value Decomposition". In: *Journal of the American Statistical Association* 102.478, pp. 674–685.
— (2008). "Modeling Homophily and Stochastic Equivalence in Symmetric Relational Data". In: *Advances in Neural Information Processing Systems*, pp. 657–664.
Hongke, Z. et al. (2017). "P2P Lending Survey: Platforms, Recent Advances and Prospects". In: *ACM Transactions on Intelligent Systems and Technology* 8, pp. 1–28.
Hosmer, D. W., T. Hosmer, et al. (1998). "A Comparison of Goodness of Fit Tests for the Logistic Regression Model". In: *Statistics in Medicine* 16.9.
Hosmer, D. W. and S. Lemesbow (1980). "Goodness of fit tests for the multiple logistic regression model". In: *Communications in Statistics - Theory and Methods* 9.10.
Hui, Z. and H. Trevor (2005). "Regularization and Variable Selection via the Elastic Net". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67.2, pp. 301–320.
IMF (2011). *Global Financial Stability Report: Grappling with Crisis Legacies*. Tech. rep. World Economic and Financial Services.
Irina, R. (2001). "An Empirical Study of the Naïve Bayes Classifier". In: *T.J. Watson Research Cente*.
Jagtiani, J. and C. Lemieux (2018). "Do fintech lenders penetrate areas that are underserved by traditional banks". In: *Federal Reserve Bank of Philadelphia* Working papers 18-13.

Jin, Y. and Y. Zhu (2015). “A Data-Driven Approach to Predict Default Risk of Loan for Online Peer-to-Peer (P2P) Lending.” In: *Fifth International Conference on Communication Systems and Network Technologies*. Pp. 539–572.
Jordan, M. (1996). *Learning in Graphical Models*. Boston, USA: Kluwer Academic Publishers.
Jordan, M. I. (2004). *Graphical Models*. Computer Science Division and Department of Statistics University of California, Berkeley 94720.
Jr, D. W. H., S. Lemeshow, and R. X. Sturdivant (2013). *Applied Logistic Regression*. Canada: John Wiley and Sons;
Jushan, B. and N. Serena (2002). “Determining the Number of Factors in Approximate Factor Models”. In: *Econometrica* 70.1, pp. 191–221.
Kalina, J. (2014). “Classification methods for high-dimensional genetic data”. In: *Biocybernetics and Biomedical Engineering* 34.1, pp. 10–18.
Katsiampa, P. (2017). “Volatility estimation for Bitcoin: A comparison of GARCH models”. In: *Economics Letters* 158, pp. 3–6.
Katz, L. (1953). “A new status index derived from sociometric analysis”. In: *Psychometrika* 18.1, pp. 39–43.
Khan, B. S. and M. A. Niazi (2017). *Network Community Detection: A Review and Visual Survey*. Working Paper. arXiv:1708.00977.
Khandani, A. E., A. J. Kim, and A. W. Lo (2010). “Consumer credit-risk models via machine-learning algorithms”. In: *Journal of Banking & Finance* 34.11, pp. 2767–2787.
Khashman, A. (2011). “Credit risk evaluation using neural networks: Emotional versus conventional models”. In: *Applied Soft Computing* 11.8, pp. 5477–5484.
Klafft, M. (2009). “Peer to Peer Lending: Auctioning Microcredits over the Internet”. In: *Proceedings of the International Conference on Information Systems, Technology and Management, A. Agarwal, R. Khurana, eds., IMT, Dubai, 2008*.
Klecka, W. R. and G. R. Iversen (1980). *Discriminant Analysis*. Saga University Paper.
Kleinberg, J. M. (1999). “Authoritative sources in a hyperlinked environment”. In: *Journal of the ACM (JACM)* 46.5, pp. 604–632.
Koller, D. and N. Friedman (2009). *Probabilistic Graphical Models: Principles and Techniques*. USA: MIT.
Kudyba, S. (2014). *Big Data, Mining and Analytics: Components of Strategic Decision Making*. CRC Press.
Lachenbruch, P. A. and M. Goldstein (1979). “Discriminant Analysis”. In: *Biometrics* 35.1.
Lauritzen, S. L. (2011). *Graphical Models*. West Sussex (UK): John Wiley Sons.
Lean, Y. et al. (2010). “Support vector machine based multiagent ensemble learning for credit risk evaluation”. In: *Expert Systems with Applications* 37.2, pp. 1351–1360.
Leitner, Y. (2005). “Financial Networks: Contagion, Commitment, and Private Sector Bailouts”. In: *The Journal of Finance*.
Lessmann, S. et al. (2015). “Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research”. In: *European Journal of Operational Research* 247.1, pp. 124–136.
Letizia, E. and F. Lillo (2018). *Corporate Payments Networks and Credit Risk Rating*. Working Paper. https://arxiv.org/abs/1711.07677.
Lin, M. and S. Viswanathan (2015). “Home Bias in Online Investments: An Empirical Study of an Online Crowdfunding Market”. In: *Management Science* 62.5, pp. 1393–1414.
Lorenz, J., S. Battiston, and F. Schweitzer (2009). “Systemic risk in a unifying framework for cascading processes on networks”. In: *The European Physical Journal B - Condensed Matter and Complex Systems* 71.4, pp. 441–460.
Mannaro, K., A. Pinna, and M. Marchesi (2017). “Crypto-trading: Blockchain-oriented energy market”. In: *IEEE 2017 AEIT International Annual Conference*.
Mantegna, R. N. (1999). “Hierarchical structure in financial markets”. In: *The European Physical Journal B - Condensed Matter and Complex Systems* 11, pp. 193–197.

Mantegna, R. N. and H. E. Stanley (1999). *Introduction to econophysics: correlations and complexity in finance*. Cambridge university press.
Marti, S. and H. Garcia-Molina (2006). "Taxonomy of trust: Categorizing P2P reputation systems". In: *Computer Networks* 50.4, pp. 472–484.
Mersch, Y. (2019). *Lending and payment systems in upheaval: the fintech challenge*. Working Paper. ECB.
Mezei, J., A. Byanjankar, and M. Heikkila (2018). "Credit risk evaluation in peer-to-peer lending with linguistic data transformation and supervised learning". In: *Proceedings of the 51st Hawaii International Conference on System Sciences*.
Mika, S. et al. (1999). "Fisher Discriminant Analysis with Kernels". In: *IEEE*.
Milne, A. and P. Parboteeah (2016a). *The Business Models and Economics of Peer-to-Peer Lending*. Tech. rep. Centre for European Policy Studies.
— (2016b). "The business models and economics of peer-to-peer lending". In:
Minoiu, C. and J. A. Reyes (2013). "A Network Analysis of Global Banking: 1978–2010". In: *Journal of Financial Stability* 9.2, pp. 168–184.
M.Merigo, J. and M. Casanovas (2011). "Induced aggregation operators in the Euclidean distance and its application in financial decision making". In: *Expert Systems with Applications* 38.6, pp. 7603–7608.
Moghadam, R. and J. Viñals (2010). *Understanding Financial Interconnectedness*. Mimeo. International Monetary Fund.
Myers, S. C. and N. S. Majluf (1984). "Corporate Financing and Investment Decisions When Firms Have Information that Investors Do Not Have". In: *Journal of financial economics* 13.2, pp. 187–221.
Namvar, A. et al. (2017). *Credit risk prediction in an imbalanced social lending environment*. Working Paper. University of New South Wales.
Naoko, N., S. David, and H. Bihong (2019). "Optimal Regulation of P2P lending for small and medium-sized enterprises". In: *ADBI Working Paper Series, no.912*.
Newman, M. (2006). "Modularity and community structure in networks". In: *Proceedings of the national academy of sciences* 103.23, pp. 8577–8582.
— (2018). *Networks*. Oxford university press.
Ng, A. and M. Jordan (2001). "On discriminative vs. generative classifiers: a comparison of logistic regression and naive Bayes". In: *Proceeding NIPS'01 Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic Pages 841-848*.
Nicoletti, B. (2017). *The Future of FinTech: Integrating Finance and Technology in Financial Services*. Rome, Italy: Palgrave Studies in Financial Services Technology.
Nino, J. L. et al. (2017). *An Overview of FinTechs: Their Benefits and Risks*. Technical Report. Association of Supervisors of Banks of the Americas.
Nowak, A., A. Ross, and C. Yencha (2017). "Small Business Borrowing and Peer-to-peer Lending: Evidence from Lending Club". In: *Contemporary Economic Policy* 36.2.
Ntwiga, D. B. (2016). *Social Network Analysis for Credit Risk Modeling*. Tech. rep.
Opgen-Rhein, R. and K. Strimmer (2007). "From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data". In: *BMC Systems Biology* 37.1.
Opsahl, T., F. Agneessens, and J. Skvoretz (2010). "Node centrality in weighted networks: Generalizing degree and shortest paths". In: *Social Networks* 32.3, pp. 245–251. ISSN: 0378-8733. DOI: https://doi.org/10.1016/j.socnet.2010.03.006. URL: http://www.sciencedirect.com/science/article/pii/S0378873310000183.
Orth, W. (2012). "The predictive accuracy of credit ratings: Measurement and statistical inference". In: *International Journal of Forecasting* 1.28, pp. 288–296.
Pecora, N., P. R. Kaltwasser, and A. Spelta (2016). "Discovering SIFIs in interbank communities". In: *PloS one* 11.12, e0167781.

Pedroche, F., M. Romance, and R. Criado (2016). “A biplex approach to PageRank centrality: From classic to multiplex networks”. In: *An Interdisciplinary Journal of Nonlinear Science* 26.6.

Pencina, M. J. et al. (2008). “Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond”. In: *Statistics in medicine* 27.2, pp. 157–172.

Perra, N. and S. Fortunato (2008). “Spectral centrality measures in complex networks”. In: *Physical Review E* 78.3, p. 036107.

Philippon, T. (2017). *The FinTech Opportunity*. Technical Report. BIS Working Papers.

Pokorna, M. and M. Sponer (2016). “Social Lending and Its Risks”. In: *Procedia - Social and Behavioral Sciences* 220.31, pp. 330–337.

Poledna, S. et al. (2016). “The multi-layer network nature of systemic risk and its implications for the costs of financial crises”. In: *Journal of Financial Stability* 20, pp. 70–81.

Rajeshwari, S. M. (2019). “An Overview Study on P2P Lending”. In: *Journal of Research in Business and Management* 7.2, pp. 56–61.

Rao, A. et al. (2003). “Load Balancing in Structured P2P Systems”. In: *International Workshop on Peer-to-Peer Systems* 2, pp. 68–79.

Reid, P. and M. V. D. Zwan (2019). *An Introduction to Alternative Risk Premia*. Industry Report. Morgan Stanley.

Rossi, R. A. and D. F. Gleich (2012). *Dynamic PageRank using Evolving Teleportation*. arXiv: 1203.6098 [cs.SI].

Safavian, S. R. and D. Landgrebe (1991). “A Survey of Decision Wee Classifier Methodology”. In: *IEEE Transactions on Systems, Man and Cybernetics* 21.3.

Sanford, C. (2013). “Do agents improve financial inclusion? Evidence from a national survey in Brazil”. In: *Bankable Frontier Association*.

Schoder, D., K. Fischbach, and C. Schmitt (2005). “Core Concepts in Peer-to-Peer Networking”. In:

Sean, F., K. Jonathan, and T. J. Putnins (2018). “Sex, Drugs, and Bitcoin: How Much Illegal Activity Is Financed Through Cryptocurrencies?” In:

Serrano-Cinca, C. and B. Gutiérrez-Nieto (2016a). “The Use of Profit Scoring as an Alternative to Credit Scoring Systems in Peer-to-Peer Lending”. In: *Decision Support Systems* 89, pp. 113–122.

— (2016b). “The use of profit scoring as an alternative to credit scoring systems in peer-to-peer (P2P) lending”. In: *Decision Support Systems* 89, pp. 113–122.

Serrano-Cinca, C., B. Gutiérrez-Nieto, and L. López-Palacios (2015). “Determinants of Default in P2P Lending”. In: *PloS ONE* 10.10, e0139427.

Somers, R. (1952). “A new asymmetric measure of association for ordinal variables”. In: *American Sociological Review* 6.27, pp. 799–811.

Spelta, A. and T. Araujo (2012). “The topology of cross-border exposures: beyond the minimal spanning tree approach”. In: *Physica A: Statistical Mechanics and its Applications* 391.22, pp. 5572–5583.

Spelta, A., A. Flori, and F. Pammolli (2018). “Investment communities: Behavioral attitudes and economic dynamics”. In: *Social Networks* 55, pp. 170–188.

Stavroyiannis, S. and V. Babalos (2017). “Dynamic properties of the Bitcoin and the US market”. In: *Economics Letters*.

Stefano, B. et al. (2012). “Liaisons Dangereuses: Increasing Connectivity, Risk Sharing, and Systemic Risk”. In: *Journal of Economic Dynamics and Control* 36.8, pp. 1121–1141.

Takayuki, M., T. Hideki, and T. Misako (2006). “Correlation networks among currencies”. In: *Physica A: Statistical Mechanics and its Applications* 364.15, pp. 336–342.

Tasca, P., S. Liu, and A. Hayes (2016). “The Evolution of Bitcoin Economy: Extracting and Analyzing the Network of Payment Relationship”. In: *Working paper*.

Tibshirani, R. (1996). “Regression Shrinkage and Selection via the LASSO”. In: *Journal of the Royal Statistical Society. Series B* 58.1, pp. 267–288.

Tijn, C. et al. (2018). “Fintech credit markets around the world: size, drivers and policy issues”. In: *Bank for International Settlements*.

Trevor, H., T. Robert, and F. Jerome (2009). *The Elements of Statistical Learning: Data Mining, Inference and Prediction*.

Tuffery, S. (1996). *Data Mining and Statistics for Decision Making*. Oxford (UK): Oxford University Press.

Viñals, J., S. Tiwari, and O. Blanchard (2012). *The IMF’S Financial Surveillance Strategy*. International Monetary Fund.

Wei, D. et al. (2012). “Degree centrality based on the weighted network”. In: *2012 24th Chinese Control and Decision Conference (CCDC)*, pp. 3976–3979. DOI: 10.1109/CCDC.2012.6244633.

West, D. (2000). “Neural Network Credit Scoring Models”. In: *Computers & Operations Research* 27.11, pp. 1131–1152.

Whittaker, J. (1990). *Graphical models in applied multivariate statistics*. Chichester (UK): John Wiley and Sons.

Wright, R. E. (1995). *Logistic regression In:Reading and understanding multivariate statistics*. Washington, DC: Oxford University Press.

Xu, J. J., Y. Lu, and M. Chau (2015). “P2P Lending Fraud Detection: A Big Data Approach”. In: *Pacific-Asia Workshop on Intelligence and Security Informatics*, pp. 71–81.

Yan, J., W. Yu, and J. L. Zhao (2015). “How Signaling and Search Costs Affect Information Asymmetry in P2P Lending: The Economics of Big Data”. In: *Financial Innovation* 1.1, p. 19.

Yan, X., L. Zhai, and W. Fan (2013). “C-index: A weighted network node centrality measure for collaboration competence”. In: *Journal of Informetrics* 7.1, pp. 223–239. ISSN: 1751-1577. DOI: https://doi.org/10.1016/j.joi.2012.11.004. URL: http://www.sciencedirect.com/science/article/pii/S1751157712000946.

Zhanga, Y. et al. (2016). “Research on Credit Scoring by fusing social media information in Online Peer-to-Peer Lending”. In: *Procedia Computer Science*.

Ziegler, T. et al. (2018). *The 3rd European Alternative Finance Industry Report*. Industry Report. Cambridge Center for Alternative Finance.

Zou, H. and H. H. Zhang (2009). “On the Adaptive Elastic Net with a Diverging Number of Parameters”. In: *Annals of Statistics* 37.4, p. 1733.