

Network Resilience and Assessment of the Credit Granting Policy

Els Heinsalu^b, Francesca Pampurini^c, Marco Patriarca^b and Anna Grazia Quaranta^{a*}

*Within the credit risk context, this paper shows how the knowledge of the network's structure, implicit in the relationships between a bank's customer firms, can help to understand whether, and to what extent, the difficulties of one firm can produce negative effects on the others via a contagion mechanism. **The** aim of the paper is to propose a new approach never implemented in the literature that is useful for bank managers and Regulatory Authorities. For this purpose, we study the network's resilience implicit in an exclusive and unique data set that includes the cross-relationships, during five years, among the firms that are customers of an Italian commercial bank.*

Keywords: credit risk; complex network theory; dynamics of contagion.

Jel Classification: G21; G32; C60; C63.

1. Introduction

The analysis of the creditworthiness of the banks' customers is a timeless topic that has always played a central role in banking activity. Bankruptcy prediction and credit scoring have long been considered crucial topics for bank intermediaries and regulators (Shin 2010) interested in achieving a high level of stability to benefit from the trust of the public because the efficiency of the banking system plays an important role in the economic development of each country. Most of the crisis that stroke the banking market originated from the interaction of two elements: (i) the inability of banks to recognize promptly the gradual deterioration in the health status of some customers and (ii) the high level of interconnections that characterize the business of the customers within the same economy.

The issue of correlation between credit losses plays a crucial role in credit risk analysis. The bank must be able to include this element in the evaluation process since it has an impact on the pricing of the loan, on the level of capital requirements, on the composition of the credit portfolio and, of course, on systemic risk (Pu and Zaho 2012). Even if it is well known that credit risk exhibits this problem of

^{a*} Corresponding author, Department of Economics and Law, University of Macerata, Macerata, Italy, Email: annagrazia.quaranta@unimc.it

^b National Institute of Chemical Physics and Biophysics, Estonia, Email: els.heinsalu@kbf.ee, marco.patriarca@kbf.ee

^c Department of Economic and Social Sciences, Catholic University of the Sacred Heart of Milan, Italy, Email: francesca.pampurini@unicatt.it

Acknowledgements

This work was supported by institutional research funding IUT (IUT39-1) of the Estonian Ministry of Education and Research, by the Estonian Research Council grant PUT (PUT1356), and EU through the European Regional Development Fund (ERDF) Center of Excellence (CoE) program grant TK133.

correlation, there is not a common agreement about its determinants. This lack of understanding triggers the choices of managers, practitioners, investors and regulators since it is recognized that it was one of the factors that contributed most to the burst of the non-performing loans (NPLs) within bank portfolios after the last global crisis.

Therefore, as stated above, it is useful for banks to embed the effects of these correlations into the analysis of the credit risk of each customer to establish a correct price for the loans and to achieve a satisfactory level of diversification of the loan portfolio. This is particularly important in the case of local banks because, by definition, most of their customers belong to the same local community and the same environment, so their connections are strong.

Also, the European Central Bank and other Regulatory Authorities are interested in the possibility to employ an efficient model for the evaluation of the interconnections between subjects to monitor systemic risk and set the level of capital requirements. The analysis of the level of creditworthiness of bank customers has been mostly based on statistical techniques¹ that, taking different (accounting and managerial) variables as inputs, produce a numerical output (score) that lets the bank assess the customer's default probability. Credit institutions usually improve these models adding *hard* and *soft* information (Khandani et al. 2010). They combine public data coming from accounting documentation and from credit bureau agencies with private information regarding the past evolution of the bank-customer relationship.

The business dynamics of bank customers are a good signal of the health status of the economy hence it is expected to be useful to incorporate this information in the model to anticipate difficult cases or even defaults. Concerning this point, the bank has a big advantage in the huge quantity of *soft information* coming from the number of cross transactions between its customers. Using this proprietary data set the bank can improve the quality of a credit analysis model. By analyzing patterns in customers' expenditures and transactions, one can capture particular relationships that are impossible to detect with standard credit default models. Besides, there is the possibility to identify the transactions that hold back a high level of risk because they can represent the vehicle through which the problems of a single customer may be transferred to other subjects and, as a consequence, to the entire network of relationships giving birth to a chain of defaults.

In fact, in the credit risk context, it is extremely important to study the dynamics of contagion because it can help a bank to maintain the NPLs ratio on a low level and so to achieve an overall improvement of its credit portfolio quality. The novel point of the method suggested in this work relates exactly to this approach that aims to show how the knowledge of network structure of relationships and cross-transactions between bank customer firms can allow understanding if, and to what extent, the economic-financial difficulties of one of them can produce negative effects on the others due to contagion (like the propagation of a virus). This analysis can also highlight potential *hot areas* in the network where contagion risk between firms is higher giving the bank the possibility to take into account this element in the credit risk analysis and, as a consequence, in the pricing of loans.

After the identification of the network structure defined by the firms and (incoming and outgoing) flows between them, it is possible to analyze the robustness and the resilience of the network itself. This approach represents an important innovation for a bank because it provides useful information to discover weak nodes and to understand the dynamics of interconnections in order to have an idea about the behavior and the evolution of the community of customers. For a bank, this information is fundamental to assess the quality of the credit granting policy and to be able to assemble a well-diversified portfolio of loans that can survive the problems that would be caused by a single firm. The information about the interconnections' dynamics allows the bank to adjust in time the credit granting policy hardening or softening the criteria and the rules to grant loans. In this way, it is easier to set the correct loan price, to obtain a satisfactory return and to contain the overall credit risk. The value of the approach proposed lies in the fact that a network can capture the strength and the intensity of each bilateral and multilateral connection. Indeed, it is possible to identify – if they exist – the main nodes (firms) that are supposed to be able to trigger a contagion effect; the characteristic of a contagion chain is that it does not depend on the credit risk of a single firm, but on the topology of the network, that is the connections between firms. Such a feature cannot be taken into account with a traditional credit risk model, so it is useful to combine the results of a traditional credit risk model with the network resilience analysis. This is a novel methodology; to the best of our knowledge, the idea to combine the results coming from traditional credit risk models with network resilience analysis has never been studied earlier.

In this paper we analyze, as an example, a real and unique data set provided by an Italian commercial bank (the name of the bank cannot be mentioned to preserve confidentiality²). A strength of this study is that the data set comprises information about the cross transactions between the customer firms of the same bank since our aim is just to highlight the potential *hot areas* where contagious risk is higher within the same bank's customer firms network. Hence, this gives the bank the possibility to take into account in the credit risk analysis and, as consequence, in the loan pricing, how much the economic-financial difficulties of one of its customers can produce negative effects on the others because of the contagion effect. Moreover, data refer to firms that are customers of a local bank; this is even more useful for our purpose because, as already stated, most of a local bank's customers belong to the same community and the same environment. Consequently, their connections are stronger than those between the customers of a bigger bank whose credit portfolio, as known, is more diversified.

We carry out an analysis based on a network model, in which the customer firms represent the nodes and the total transaction amounts between firms define the links and the corresponding weights. Of course, the results we obtain from the data provided by the local commercial bank are valid exclusively in that particular situation and cannot be generalized. However, the procedure sketched has a methodological value, since it may be useful also in other types of networks and analogous situations. The analysis done provides an example of how the approach of network analysis can be applied to a real situation. As such, it can be useful not only for a bank manager (who can combine it with the traditional credit risk model to improve bankruptcy prediction and assess the quality of the credit granting policy), but also

for Regulatory Authorities, to study the dynamics of systemic risk in case of financial crisis and distress and to set optimal capital requirements.

The structure of the paper is as follows. Section 2 offers a review of the main literature on network analysis and its applications to economic and financial topics. Section 3 describes the data we deal with in this study and the characteristics of the network of relationships between the bank's customer firms. Section 4 discusses the results and Section 5 concludes.

2. Literature Review

Risk assessment of financial intermediaries and, in particular, of banks has been one of the most studied topics for many decades both from a credit institution and from a regulatory perspective.

Traditional models rely on the use of early warning systems (Galindo and Tamayo 2000) following the famous model introduced by Altman (Altman 1981). These models, also known as credit scoring models, provide a number (score) that represents the creditworthiness of a subject based on some information regarding personal characteristics and the financial situation of the subject itself. Then, statistical discrimination is performed to fix the threshold that separates the "good" from the "bad" customers. Actually, within the same context, the implementation of a statistical method is not the only way to obtain such a division of the customers. An alternative approach can be successfully employed that is based on artificial neural networks (West 2000).

The main aim of models *à la Altman* is to predict the probability of failure (or problematic conditions) given the aforementioned explanatory variables, but without assessing the factors that produce the default (Chen and Huang 2003). Risk decomposition and aggregation models that have their roots in the CAPM theory and the Contingent Claim Analysis face the latter problem. These models try for the first time to take into account the problem of correlations among subjects, especially in case of credit losses, and try to analyze the determinants of credit risk correlation.

The very first model of this type was the one by Merton (Merton 1974) who introduced the problem of credit risk correlation by assuming that the stochastic processes followed by the assets of two companies are correlated. Within this strand of literature other models focused in particular on the strong contagion effects among firms in the state of bankruptcy (Jorion and Zhang 2007; Horst 2007; Hatchett and Kühn 2009) and tried to incorporate into the analysis new kinds of risk factors (Collin-Dufresne et al. 2003; Duffie et al. 2009; Giesecke 2004; Giesecke and Weber 2004 & 2006; Jarrow and Yu 2001; Schönbucher and Schubert 2001). Unfortunately, all these models can turn out to be useless for banks, since they rely on market data and, as it is well known, most of the bank customers are not listed on any market.

Another important topic faced by literature on credit risk concerns the use of soft information such as those about the historical behavior of a customer and his past relationship with the bank³. Information of this kind is useful for different purposes: in

the case of cross-selling activities, to be sure to meet real customers' needs (Dyche and Dych 2001), and in case of credit analysis. From this idea a new strand of research was born, based on the study of behavioral scoring models and techniques (Setiono et al. 1998). Such studies are highly linked to those about classification analysis based on statistical techniques (Hand 1981; Johnson and Wichern 1998), neural networks and data mining (Lancher et al. 1995). They all belong to the field of machine-learning which is a relatively new technique based on a set of algorithms built up to tackle computationally intensive pattern recognition problems in extremely large datasets (Khandani et al. 2010).

The use of machine learning in credit risk modeling has grown up in the last decades allowing the analysis of the credit quality of a subject using both quantitative and qualitative information (Petropoulos et al. 2019). The most of this literature aims to compare the performance of traditional techniques and machine learning models in the identification of default enterprises (Addo et al. 2018) showing the superior efficiency of machine learning⁴.

Network-based studies of both credit and systemic risk have been explored e.g. by Petrone and Latora (2018) and Giudici et al. (2019). The first contribution analyzes the interconnections between financial institutions to quantify systemic risk. The analysis is based on a dynamic model that combines credit risk techniques with a contagion mechanism on the network of exposures among banks. The authors illustrate how their model works on the network of the European Global Systemically Important Banks. Since most of the data are not public, the authors employ a Monte Carlo approach to simulate the values necessary to build up the network: the entity of the interconnections between the banks and the contagion mechanism.

On the contrary, the second work (Giudici et al. 2019) studies how to enhance the estimation accuracy of the credit risk of peer-to-peer lending platforms by leveraging topological information embedded into similarity networks, derived from borrowers' financial information. In particular, they use the topological coefficients describing borrowers' importance and community structures derived by a network approach as new independent variables that can be added to a traditional credit scoring model, improving its predicting performance.

The recent, very fast, development of technology, computing and machine learning algorithms has shed light on the importance of data and information, in general. Also, Central Banks, after the burst of the global crisis, launched many data statistic activities aimed at boosting their supervisory and monetary policy functions⁵. The possibility of the automated collecting of large sets of data increases the need for robust data mining processes and analysis models to exploit all possible dimensions of information. Conventional statistic and econometric methods are not able to capture the multidimensional information in these datasets, so advanced machine-learning techniques are preferred and the use of complex networks has been recognized as a useful framework.

3. Methodology

The identification of the structure of a network of (incoming and outgoing) flows

between one bank’s customer firms and the corresponding analysis of the network’s robustness can provide useful information to evaluate and assess the quality of credit granting policy.

The example considered in the present study concerns the network of payments between the customer firms of an Italian local commercial bank. Data refer to the incoming and outgoing monetary flows (and therefore to the monetary values of the commercial relationships in place) between the firms belonging to the subset of the bank’s customers firms that in the last five years (2015-2019) had a significant number of transactions. They are small-medium enterprises (SMEs); this particular kind of firm represents the most important subject of an industrial district⁶. Some general characteristics of the network under study (Reka and Barabási 2002; Evans 2004; Newman 2003) are summarized in Table 1⁷.

The original database represents a network of $N = 196$ nodes, connected to each other through $L = 24,261$ directed links (i.e. links that connect directionally pairs of nodes). Nodes represent firms and the generic directional link from a node i to a node j is defined by a quantity $l_{i,j} > 0$, representing the total amount of money transferred from the corresponding i -th firm to the j -th firm during the 5-year timeframe under consideration. The total number of different ordered pairs in the system is $N = 196$, implying that there is a link density $\rho \equiv L/L_0 \approx 63\%$ that is 63% of all the possible connections are realized.

Table 1: Some Features of the Network under Study

Symbol	Quantity	Value
N	No. of Nodes (Vertices)	196
L	No. of Links (Edges)	24261
ρ	Density (of edges)	0.63
k_{in}	Minimum in-degree	107
K_{in}	Maximum in-degree	142
k_{out}	Minimum out-degree	111
K_{out}	Maximum out-degree	134
$k_{in} = k_{out}$	Average degrees	123
σ_{in}	In-degree standard deviation	6.7
σ_{out}	Out-degree standard deviation	4

To summarize, the parameter values suggest that the network under consideration is quite homogeneous and dense. A homogeneous character means that the degrees of the nodes are very similar to each other – the standard deviations of the degree distributions are much smaller than the corresponding average degrees, $\sigma/k \ll 1$ – and the network is dense, i.e. there is a high value of its link density ρ , i.e. the ratio between the number of links and the total number of possible links.

All these features make the network under study similar to a *random network*, i.e. a network that can be constructed from a set of disconnected nodes by assigning to each node the same probability to form links with different nodes. Random networks

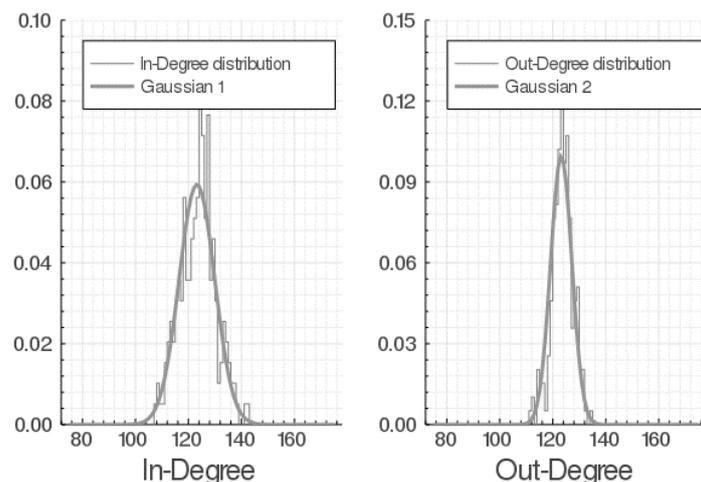
represent prototypical networks that were the first subject of the studies of graph theory.

This type of network *looks the same* if observed from any node and is to be contrasted with other types of networks that have broader degree distributions. In particular, many networks found in natural and social sciences belong to the category usually referred to as *scale-free networks*, characterized by degree distributions with a long tail. For example, the network associated with the internet has a degree distribution $f(k) \sim 1/k^\alpha$ with $\alpha \approx 2.5$ (so that even its second moment diverges).

The high average degree and the large link density of the network considered here may have a twofold origin: (i) a natural tendency of the firms to cluster together, since they are firms selected inside the same economically connected industrial district; (ii) a partial artifact turning the actual network into one closer to a random network, since the data are averaged over a relatively long time window of five years, in which many links may have been removed and/or rewired and many other new links may have formed. That is, a temporal network averaged in time is expected to be more homogeneous than the network observed instantaneously or in a smaller time window.

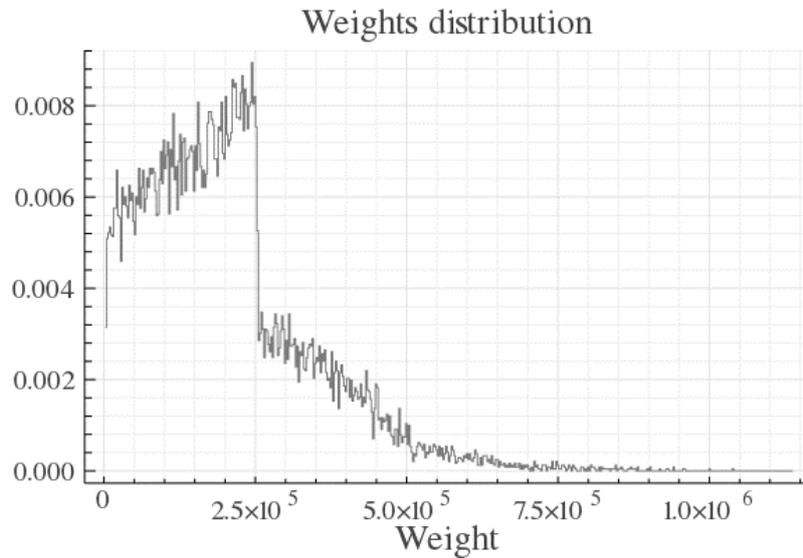
In Figure 1 the internal connectivity of the network is represented through the in-and out-degree distributions $f(k_{in})$ and $f(k_{out})$ (histograms). The continuous curves depict the corresponding fits made using Gaussian functions. The Gaussian fits only serve as a guide to the eye to demonstrate the homogeneous character of the distribution, i.e. that the degree distributions do not have a long (e.g. power-law) tail. The parameters of the fits (standard deviation and mean directly obtained from the data through basic statistics) are not used in the following.

Figure 1: In-Degree (Left) and Out-Degree (Right) Distributions (Histograms) and Respective Gaussian Fits (Continuous Lines). The Fitting Parameters for the Curves Gaussian 1 (In-Degrees) and Gaussian 2 (Out-Degrees), Extracted From the Network Data, Are: Mean $M_1 = M_2 \approx 123.0$ and Standard Deviations $\Sigma_1 \approx 6.7$ and $\Sigma_2 = 4.0$, Respectively



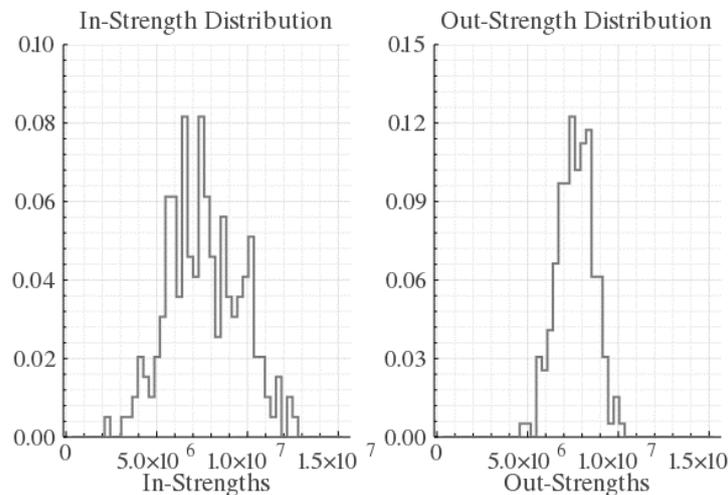
The histograms of the link weights, i.e. of the total amounts $\{l_{i,j}\}$ of the fluxes occurred between pairs of nodes i and j in the time-span considered is shown in Figure 2. The distribution suggests that medium-sized transactions are most common, which is typical of SMEs.

Figure 2: Histogram of the Link Weights (Total Amount of the Money Transfer between Pair of Nodes, Arbitrary Units)



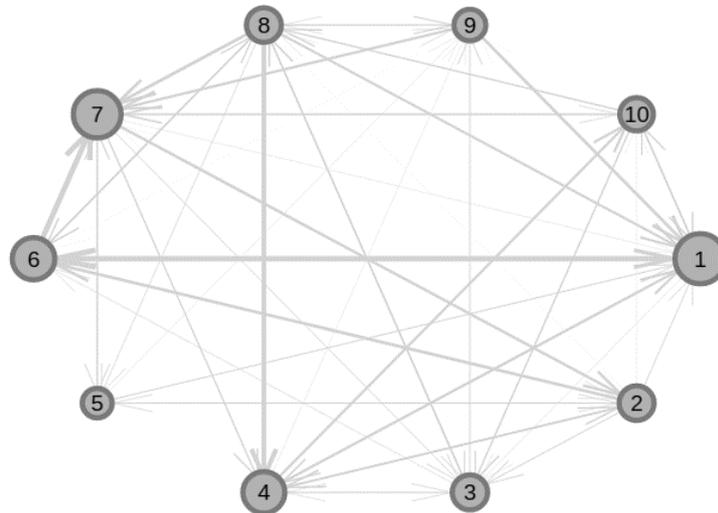
The activity of the nodes is represented in Figure 3 by the distributions of the in-and out-strengths $\{S_i\}$. While the weights $\{l_{i,j}\}$ characterize the links, the strength S_i is a property of the i th node, representing its overall degree of activity: S_i is defined as the sum of the weights of all the links starting from (out-strengths) or reaching (in-strength) node i , in other words, the histograms in Figure 3 represent the total in-and out-fluxes of the nodes. These histograms confirm that the network is quite homogeneous. The nodes' strengths are homogeneous too.

Figure 3: Histogram of the in-and Out-Strengths of Nodes (Magnitude of the Total in-and Out-Fluxes of Firms)



To give an idea of the structure of the network, a graphical representation of a small subset, made up of 10 nodes of the network chosen randomly, is shown in Figure 4.

Figure 4: The Subset of the Network Composed of Ten Randomly Chosen Nodes (Node Sizes are Proportional to Their Strengths and Link Widths to Their Weights).



The resilience of a network (or its fragility) represents its capability of survival after the failure of some of its nodes and it is recognized as an important chapter in the theory of complex networks (Albert and Barabási 2002). In many problems, it is necessary to test how a network reacts and changes under the effects of external perturbations (Cohen et al. 2000). A standard type of resilience test consists in studying the effects of some failures, resulting in the removal of some nodes or links. Depending on the internal dynamics of the network and the external loads, the removal of some nodes (or some links) can cause the failure of additional nodes (or links), eventually resulting in a *cascade of failures*.

When testing the robustness of a network, one has to beware of the fact that testing the same network in different ways against different types of perturbations may result in very different levels of resilience. In turn, how failure can take place depends on the problem considered. A prototypical example of cascading failure is that occurring in electric power grids. A power grid can be represented as a network, in which nodes represent power plants or transmission substations, connected by links, representing the power transmission lines. A cascading failure can start when the electricity fluxes crossing a certain node exceed a critical threshold value determined by the capacity of the node. When this happens, the node is removed from the network and the electricity fluxes have to be redirected towards other nodes, which can cause additional failures, possibly resulting in a cascading failure.

In the case of an economic network, characterized by nodes representing firms that exchange money and goods with each other, and by links connecting such nodes representing the actual transfers taking place between firms, the main reason of failure is, on the contrary, related to the disappearance or the excessive reduction of

the fluxes below a critical threshold. However, the way how defaults propagate and generate a cascading failure is similar.

When there is some node failure, the network can break into separated sub-networks. Such an event may signal a starting chain of defaults. Therefore, a general approach to the estimation of the resilience of a network is based on the measurement of how the size of the remaining largest cluster of the network varies *versus* the total fraction f of the network nodes that have been removed. In the example analyzed here, we can consider the nodes disappearing from the networks as representing the firms undergoing a default. The relative size of the largest cluster is then given by the fraction of nodes that remain all connected to each other through a network of money transactions. Different criteria can be used to choose which nodes or links to remove, i.e. different *attack strategies* can be employed. The type of strategy used is very important in determining the evolution of the network. For example, one can consider a set of *random failures* of nodes or links, that is, the nodes (or links) removed are selected randomly; or one can consider an *attack*, meaning e.g. the removal of specific nodes (e.g. the most connected ones) or links (e.g. those with higher weights).

Several studies (Motter and Yang 2017; Albert and Barabási 2002) have shown that when comparing networks that have comparable average degrees k , random networks are less resilient than scale-free networks to random attacks, while the opposite is true for preferential attacks towards the most connected nodes. This is in a way to be expected since scale-free networks are characterized by the presence of highly connected nodes, the hubs. In a random network, all nodes are similarly connected to each other, so that they are all equally susceptible of failure on average, whereas in a scale-free network there is a high heterogeneity of links and most of the nodes can count on the existence of some hubs to remain connected to each other. In a scale-free network, most of the remaining nodes may still be connected to each other even after removing many nodes; a random network may decompose in sub-networks and disappear abruptly at a critical fraction f^* of nodes removed. However, this only holds for networks with the same degrees, while the degree in itself is a relevant parameter influencing the network resilience (Guillame et al. 2005).

In this respect, the time-averaged network of the example considered has a high degree, which makes it very connected internally and therefore very robust against removals of nodes. This holds in general for any network similar to a random network that will be more resilient even with respect to a scale-free network with a sufficiently low average degree. A quantitative estimate of the network resilience is provided by the theory of complex networks: a general network will break into sub networks when the remaining fraction of nodes after a series of random removals reaches the critical threshold f_0 given by (Cohen et al. 2000; Guillaume Latapy and Magnien 2005).

$$f_0 = \frac{\bar{k}}{(\bar{k}^2 - \bar{k})} \quad (1)$$

Notice that this expression depends on the first and second moments of the degree distribution only. If the average degree is high and the standard deviation of the degrees is much smaller than the average degree (i.e. the network is homogeneous, similarly to a random network), then Eq. (1) provides $f_0 \sim 1/k$, representing a small critical fraction considering the large value of \bar{k} . In other words, a homogeneous but closely connected network, as that considered in the example, is very resilient, despite its homogeneous character, to random attacks (in the example considered f_0 is less than 1%).

This conclusion holds for this particular example but can be different for other types of networks – each case requires a separate analysis. It is the natural continuation of the present work to analyze real and simulated examples of networks with different features and corresponding levels of resilience.

4. Results and Analysis

Complex Network theory is by now a standard toolbox in the study of social systems (Reka and Barabási 2002; Evans 2004; Newman 2003) and, in particular, it is a most useful tool for the analysis and description of economic systems too. It should be useful, when considering the level of credit risk within a group of connected firms that are customers of the same bank. The structure of the network is fully written in the data relative to the transactions between firms: links, degrees, the existence of communities, and the magnitudes of the transactions (representing weights of the links), strengths of the nodes (total in – and out-fluxes), etc., can be directly extracted from the data available to a bank and provide precious information on the role of every single firm inside the economic network.

For example, firms inside communities are connected to each other more strongly and therefore should be expected to influence each other more than other firms, in case of damage to the network due to the failure of some nodes. On the other hand, as shown in the example considered in this study, to know that a network is homogeneous allows one to expect that it will respond better to relatively small perturbations, even if for high levels of failures it is known to break before a scale-free network with a comparable average degree. This type of information, complemented by the more traditional information on the economic attributes of firms, should provide a more reliable picture of the system, with higher ability to predict the reaction of firms to external perturbations. In the example considered above, the network looks as homogenous and dense as a random network, in which nodes are strongly connected to each other, and therefore very resilient to external attacks or internal failures.

To assess the quality of the credit granting policy adopted by the bank, the particular network structure that emerged from the available data and the related resilience analysis would seem to indicate a good practice in selecting the credit portfolio. Any difficulty of a customer firm would seem to be easily absorbed into the system without causing particular damage to other customer firms belonging to the network. To observe a break in this network of relationships, many defaults should occur simultaneously for a very large number of nodes. This appears somewhat unlikely

and therefore the bank does not need to intervene promptly by tightening the criteria used to select its customers or by changing the pricing of loans to obtain better hedging against risk.

On the contrary, if a different network structure had emerged from the data and this had highlighted the presence of one or more particularly weak nodes, the bank should have remodeled the borrowers' selection criteria and, above all, those related to the loans' portfolio diversification.

5. Conclusions

The credit risk analysis is an extremely important topic both for bank intermediaries and for Regulatory Authorities for obvious reasons related to stability and profitability. Credit risk is strictly connected to systemic risk because the recent crisis has shown once again that the failure of a unit can quickly trigger others causing a default chain. In the evaluation of credit risk, each bank must be able to take into account not only the health status of the single customer but also the impact and the consequences of this systemic effect on the whole portfolio of loans.

This paper aims to provide a methodological example of how the knowledge of the network's structure implicit in the relationships between bank customer firms can help the bank to realize whether, and to what extent, the difficulties of one of these firms can produce negative effects on other customer firms through contagion mechanisms. In other words, identifying the structure of the network of flows between the firms that are customers of the bank and the subsequent analysis of the network's robustness can allow the bank to obtain useful information to evaluate/assess the quality of its policy in granting credit.

Regardless of the specific results obtained, discussed in Section 4 – exclusively referable to the bank that made the data available – the approach we propose can be useful not only for a bank's financial decision making in the field of its customers' bankruptcy prediction, credit scoring, correct quantification of credit risk and assessment of its credit granting policy, but also for regulation purposes concerning the setting of capital requirements, the evaluation of systemic risk and the forecast of financial crisis and distress.

As future development of the research presented here, various possible directions arise. First of all, improving the quality of data, which should be denser in time and concerning more customer firms. This would mean the possibility to access different types of networks, possibly characterized by the presence of communities, which are expected to provide more complex responses to the defaults of nodes and test network-based methods on new types of examples. Then one should integrate these data with other types of data concerning the status of the firms. Finally, the situation of a bank should be compared with those of other banks, possibly with different economic realities, to characterize analogies and differences.

Endnotes

¹ See Hand and Henley (1997) and Thomas (2009) for a complete review of the main statistical modeling approaches to credit scoring.

²We would like to thank the bank because, as it is well known, it is very difficult to obtain this kind of proprietary data.

³For a complete review of the models that exploit the benefits of relationship information in banking engagement, see Boot 2000 and Agarwal et al. 2009.

⁴Kandani *et al.* (2010) apply machine learning to obtain nonlinear parametric forecasting models of consumer credit risk starting from banking transactions data and credit bureau data. Butaru *et al.* (2016) use account-level credit card data from a bank, credit bureau data and macroeconomic data to predict delinquency. Galindo and Tamayo (2000) use machine learning (CART decision tree models) on mortgage loan data to detect defaults. Huang *et al.* (2004) offer a survey of corporate credit rating models and show that artificial intelligence models have a better performance than traditional statistical models.

⁵ See Petropoulos *et al.* 2019.

⁶An industrial district (ID) is a place where workers and firms (that belong to the same production chain and that buy/sell from/to each other), specialized in a particular industry and auxiliary industries, live and work. Alfred Marshall initially used this concept to describe some aspects of the industrial organization of Nations (Marshall 1919). At the end of the 1990s the industrial districts in developed or developing countries had gained recognized attention in international debates on industrialization and policies of regional development.

⁷The analysis was realized using the Julia language (<https://julialang.org>) and the specific packages for network analysis – other software freely available exists for the study of networks.

References

- Addo, P.M., Guegan, D., and Hassani, B. (2018) 'Credit Risk Analysis Using Machine and Deep Learning Models', *Risks*, 6(2), p38.
- Agarwal, S., Skiba, P. M., and Tobacman, J. (2009) 'Payday loans and credit cards: New liquidity and credit scoring puzzles?', *American Economic Review*, 99(2), p412-17.
- Albert, R., Barabási, A. L. (2002) 'Statistical Mechanics of Complex Networks', *Review of Modern Physics*, 74, p47-97.
- Altman, E. I. (1980) 'Commercial Bank Lending: Process, Credit Scoring, and Costs of Errors in Lending', *Journal of Financial and Quantitative Analysis*, 15(4), p813-832.
- Boot, A. W. A. (2009) 'Relationship Banking: What Do We Know?', *Journal of Financial Intermediation*, 9(1) p7-25.
- Chen, M. C., Huang, S. H. (2003) 'Credit scoring and rejected instances reassigning through evolutionary computation techniques', *Expert Systems with Applications*, 24(4), p433-441.
- Cohen, R., Erez, K., and Havlin, S. (2000) 'Resilience of the Internet to random breakdowns', *Physical Review Letters*, 85 (21), p4626.
- Collin-Dufresne, P., Goldstein, R. S. and Helwege, J. (2003) 'Are Jumps in Corporate Bond Yields Priced? Modeling Contagion via the Updating of Belief', *NBER Working Paper* 15733.
- Dyche, J., and Dych, J. (2001) *The CRM handbook: a business guide to customer relationship management*. Reading, MA: Addison-Wesley.

Heinsalu, Pampurini, Patriarca & Quaranta

- Duffie, D., Eckner, A., Horel, G., and Saita, L. (2009) 'Frailty correlated default', *Journal of Finance*, 64, p2089-2123.
- Evans, T.S. (2004) 'Complex Networks', *Contemporary Physics*, 45(6), p455-474.
- Galindo, J., and Tamayo, P. (2000) 'Credit Risk Assessment Using Statistical and Machine Learning: Basic Methodology and Risk Modeling Applications', *Computational Economics*, 15, p107-143.
- Giesecke, K. (2004) 'Credit Risk Modeling and Valuation: An Introduction', *SSRN Working Paper*.
- Giesecke, K., and Weber S. (2004) 'Cyclical Correlations, Credit Contagion, and Portfolio Losses', *Journal of Banking & Finance*, 28(12), p3009-3036.
- Giesecke, K., and Weber S. (2006) 'Credit Contagion and Aggregated Losses', *Journal of Economic Dynamics and Control*, 30(5), p741-767.
- 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(3).
- Guillaume, J.L., Latapy, M., and Magnien, C. (2005) 'Comparison of Failures and Attacks on Random and Scale-Free Networks', in: Higashino, T. (ed.), *Principles of Distributed Systems*. OPODIS 2004. Lecture Notes in Computer Science, vol. 3544. Springer, Berlin: Heidelberg.
- Hand, D. J. (1981) *Discrimination and classification*, Chichester: John Wiley.
- Hatchett, J.P.L., and Kühn, R. (2009) 'Credit contagion and credit risk', *Quantitative Finance*, 4, p373-382.
- Horst, U. (2007) 'Stochastic cascades, credit contagion, and large portfolio losses', *Journal of Economic Behavior & Organization*, 63(1), p25-54.
- Jarrow, R., and Yu, F. (2001) 'Counterparty Risk and the Pricing of Defaultable Securities', *The Journal of Finance*, 56(5), p1765-1799.
- Johnson, R. A., and Wichern, D. W. (1998) 'Principal components', *Applied multivariate statistical analysis*, 6, p430-481.
- Jorion P., and Zhang, G. (2007) 'Good and bad credit contagion: Evidence from credit default swaps', *Journal of Financial Economics*, 84(3), p860-883.
- Khandani, A. E., Adlar, J. K., and Andrew, W. L. (2010) 'Consumer credit-risk models via machine-learning algorithms', *Journal of Banking & Finance*, 34(11), p2767-2787.
- Lancher, R. C., Coats, P. K., Shanker, C. S., and Fant, L. F. (1995) 'A neural network for classifying the financial health of a firm', *European Journal of Operational Research*, 85(1), p53-65.
- Marshall, A. (eds) (1919) *Industry and trade: A study of industrial technique and business organization*. London: Macmillan.
- Merton, R. C. (1974) 'On the pricing of corporate debt: The risk structure of interest rates', *The Journal of Finance*, 29(2), p449-470.
- Motter, A.E., and Yang, Y. (2017) 'The Unfolding and Control of Network Cascades', *Physics Today*, 70(1), p32.
- Petrone, D., and Latora, V. (2018) 'A dynamic approach merging network theory and credit risk techniques to assess systemic risk in financial networks', *Scientific Reports*, 8, p5561.

Heinsalu, Pampurini, Patriarca & Quaranta

- Petropoulos, A., Siakoulis, V., Stavroulakis, E., and Klamargias, A. (2019) 'A robust machine learning approach for credit risk analysis of large loan level datasets using deep learning and extreme gradient boosting', IFC Bulletins chapters, in: Bank for International Settlements (ed.), *Are post-crisis statistical initiatives completed?*, vol. 49.
- Pu, X., and Zhao, X. (2012) 'Correlation in credit risk changes', *Journal of Banking & Finance*, 36, p1093-1106.
- Newman, M.E.J. (2003) 'The Structure and Function of Complex Networks', *SIAM Review*, 45, p167-256.
- Réka, A., and Barabási, A.L. (2002) 'Statistical Mechanics of Complex Networks', *Reviews of Modern Physics*, 74, p47-97.
- Schönbucher, P. J., and Schubert, D. (2001) 'Copula-dependent default risk in intensity models', *Working paper*, Department of Statistics, Bonn University.
- Setiono, R., Thong, J. Y., and Yap, C. S. (1998) 'Symbolic rule extraction from neural networks: An application to identifying organizations adopting IT', *Information & Management*, 34(2), p91-101.
- Shin, H. S. (2010) 'Financial Intermediation and the Post-Crisis Financial System', *BIS Working Paper No. 304*.
- West, D. (2000) 'Neural network credit scoring models', *Computers & Operations Research*, 27(11-12), p1131-1152.
-