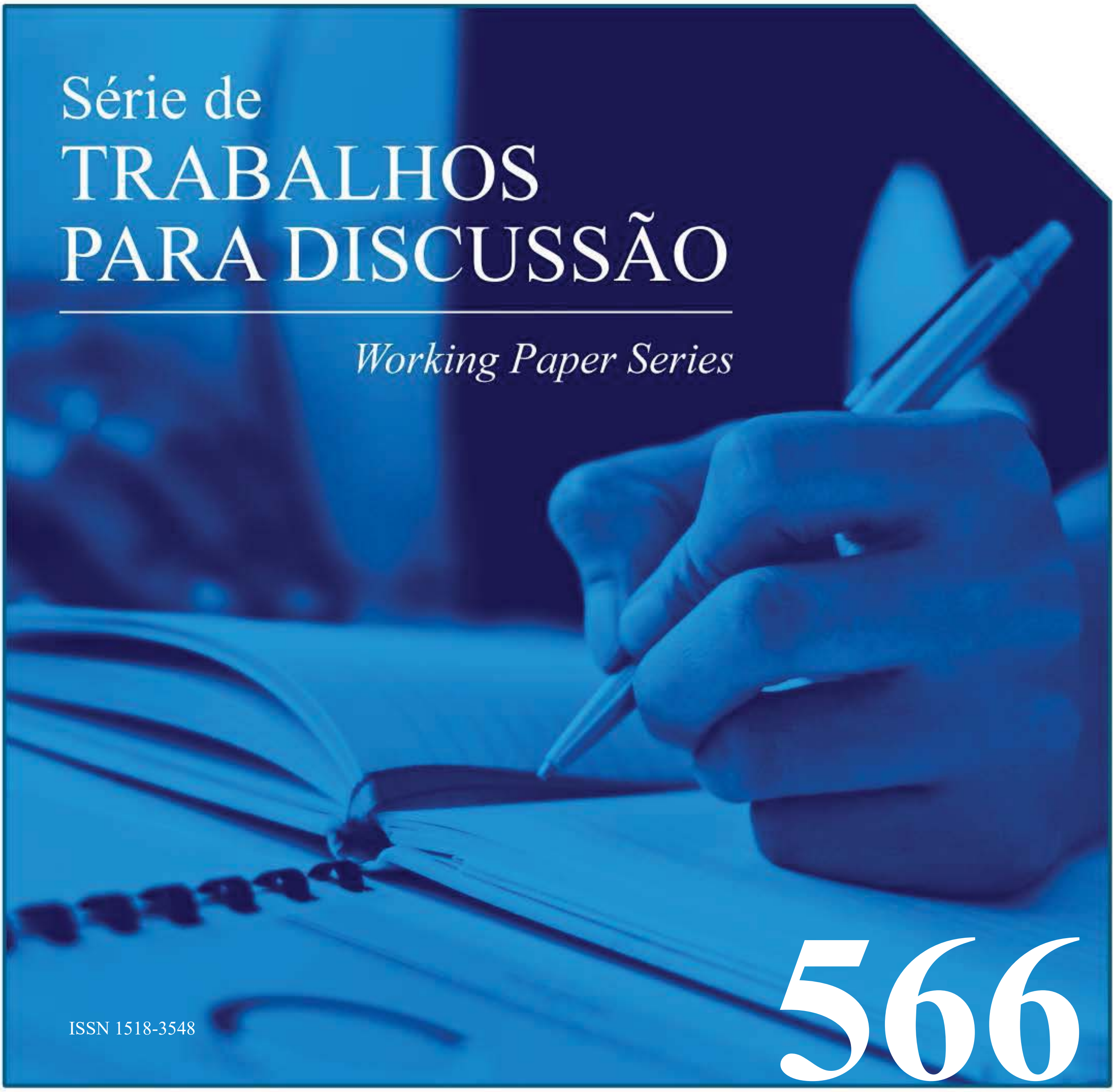

September 2022

Nestedness in the Brazilian Financial System

Michel Alexandre, Felipe Jordão Xavier, Thiago Christiano Silva, Francisco A. Rodrigues

BANCO CENTRAL DO BRASIL

ISSN 1518-3548
CGC 00.038.166/0001-05

Working Paper Series	Brasília	no. 566	Setembro	2022	p. 3-29

Working Paper Series

Edited by the Research Department (Depep) – E-mail: workingpaper@bcb.gov.br

Editor: Rodrigo Barbone Gonzalez

Co-editor: Eurilton Alves Araujo Jr

Head of the Research Department: André Minella

Deputy Governor for Economic Policy: Diogo Abry Guillen

The Banco Central do Brasil Working Papers are evaluated in double-blind referee process.

Although the Working Papers often represent preliminary work, citation of source is required when used or reproduced.

Citizen Service Division

Banco Central do Brasil

Deati/Diate

SBS – Quadra 3 – Bloco B – Edifício-Sede – 2º subsolo

70074-900 Brasília – DF – Brazil

Toll Free: 0800 9792345

Fax: +55 (61) 3414-2553

Internet: http//www.bcb.gov.br/?CONTACTUS

Non-Technical Summary

Nestedness is a feature often observed in real networks, including financial ones. The main characteristic of a nested network is the presence of *generalists* – i.e., agents that interact with many counterparties – and *specialists*, who mainly interact with generalists. In a perfectly nested network, the neighbors of a node are always a subset of – that is, they are "nested" in – the neighborhood of the nodes with a higher degree. The concept of nestedness has proved to be relevant for economic analysis, given its predictive power for many economic phenomena, such as firm survival, industry location, and economic growth. In this paper, we assess the nestedness in two Brazilian financial networks: the bank-firm credit network and the interbank network. We compute the nestedness for both networks using one of the metrics used to quantify it, the *nestedness metrics based on overlap and decreasing fill* (NODF). We compute also the individual nestedness contribution (INC) of the nodes (banks and firms). Applying machine learning techniques (random forest, XGBoost, and Shapley values) to assess the determinants of the INC, we conclude the INC of lenders is mainly correlated to their degree (i.e., their number of counterparties), while the INC of borrowers does not have a clear main driver. Moreover, we show nodes with a higher INC are also those that would cause more damage to the network if they were hit by a shock, represented here by partial/complete depletion of net worth. However, nodes with higher INC are not necessarily the most vulnerable to shocks on the network.

Sumário Não Técnico

O aninhamento (*nestedness*) é uma característica frequentemente observada em redes reais, inclusive financeiras. A principal característica de uma rede aninhada é a presença de *generalistas* – ou seja, agentes que interagem com muitas contrapartes – e *especialistas*, que interagem principalmente com generalistas. Em uma rede perfeitamente aninhada, os vizinhos de um nó são sempre um subconjunto do – ou seja, estão "aninhados" no – conjunto de vizinhos dos nós com maior grau. O conceito de aninhamento provou ser relevante para a análise econômica, dado seu poder preditivo para muitos fenômenos econômicos, como sobrevivência de firmas, localização de indústrias e crescimento econômico. Neste artigo, avaliamos o aninhamento em duas redes financeiras brasileiras: a rede de crédito banco-firma e a rede interbancária. Calculamos o aninhamento para ambas as redes usando uma das métricas usadas para quantificá-lo, a *nestedness metrics based on overlap and decreasing fill* (NODF). Calculamos também a *individual nestedness contribution* (contribuição de aninhamento individual – INC) dos nós (bancos e firmas). Aplicando técnicas de aprendizado de máquina (*random forest*, XGBoost e valores de Shapley) para avaliar os determinantes da INC, concluímos que a INC dos concessores de crédito é determinada principalmente pelo seu grau (ou seja, seu número de contrapartes), enquanto o INC dos tomadores de crédito não tem um determinante principal claro. Além disso, mostramos que nós com maior INC também são aqueles que causariam mais danos à rede se fossem atingidos por um choque, representado aqui por um esgotamento parcial/completo do patrimônio líquido. No entanto, nós com maior INC não são necessariamente os mais vulneráveis a choques na rede.

Nestedness in the Brazilian Financial System

Michel Alexandre*

Felipe Jordão Xavier**

Thiago Christiano Silva***

Francisco A. Rodrigues****

Abstract

In this paper, we assess the nestedness in the Brazilian financial system. We rely on data from two Brazilian financial networks: the bank-firm credit network and the interbank network. We computed the nestedness of the networks, as well as the Individual Nestedness Contribution (INC) for each node. The analysis of the determinants of the INC shows lenders – in both networks – have their INC mainly determined by the degree, while the INC of borrowers has not a clear main determinant. Moreover, we found nodes with a higher INC would cause more damage to the network if they were hit by a shock, but are not necessarily those more vulnerable to shocks on the network.

Keywords: nestedness, complex networks, systemic risk, financial networks, machine learning

*Research Department, Central Bank of Brazil. Email: michel.alexandre@bcb.gov.br

**Institute of Mathematics and Computer Science, University of São Paulo, Brazil. E-mail: felipe.xavier11@gmail.com

***Research Department, Central Bank of Brazil and Universidade Católica de Brasília. E-mail: thiago.silva@bcb.gov.br

****Institute of Mathematics and Computer Science, University of São Paulo, Brazil. E-mail: francisco@icmc.usp.br

1 Introduction

One of the most pervasive and studied patterns observed in complex networks is nestedness. It refers to a hierarchical organization of the network, where the set of neighbors of a node is a subset of the neighborhood of the nodes with a higher degree – or conversely, a superset of the neighborhood of the nodes with a lower degree. A nested network is composed of generalists – that interact with many counterparties – and specialists, who interact mostly with generalists. Therefore, specialist-specialist interactions are quite rare (Bascompte et al., 2003).

A simple illustration of a perfectly nested network is depicted in Figure 1. We portray a bank-firm credit network. Each row (column) corresponds to a firm (bank). Banks (firms) are labeled as B1,..., B6 (F1,..., F7). A colored square represents a loan extended by the bank in the corresponding column to the firm in the corresponding row. The two types of nodes – banks and firms – are ranked in descending order according to the degree (firms from top to bottom, banks from left to right). The banks connected to a given firm are also connected to firms with a higher degree. For instance, bank 3 is connected to firm 5 (the blue square in the figure). It is also connected to firms above firm 5 (with a higher degree), but it is not connected to firms below firm 5 (with a lower degree). Similarly, firms connected to a given bank are also connected to banks with a higher degree (e.g., firm 5 is connected to banks on the left of bank 3, but not to banks on the right of bank 3). The more generalist (specialist) banks correspond to the columns located in the left (right) of the figure. Similarly, the more generalist (specialist) firms correspond to the rows located in the top (bottom) of the figure.

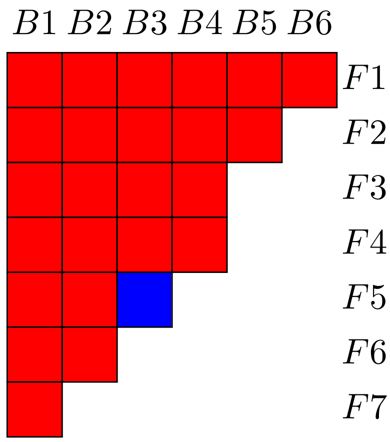

Figure 1: *Example of a perfectly nested bank-firm credit network. The connection between bank 3 and firm 5 is represented by the blue square. Bank 3 is connected only to the firms above firm 5, as they have a degree higher than that of firm 5. Similarly, firm 5 is connected only to the banks on the left of bank 3, i.e., those with a degree higher than that of bank 3.*

Similarly, we can represent a perfectly nested unipartite network. In Figure 2, we show a perfectly nested interbank network from the lenders' point of view. If the link representing the loan granted by B4 to B6 is removed, the network becomes also perfectly nested from the borrowers' point of view.

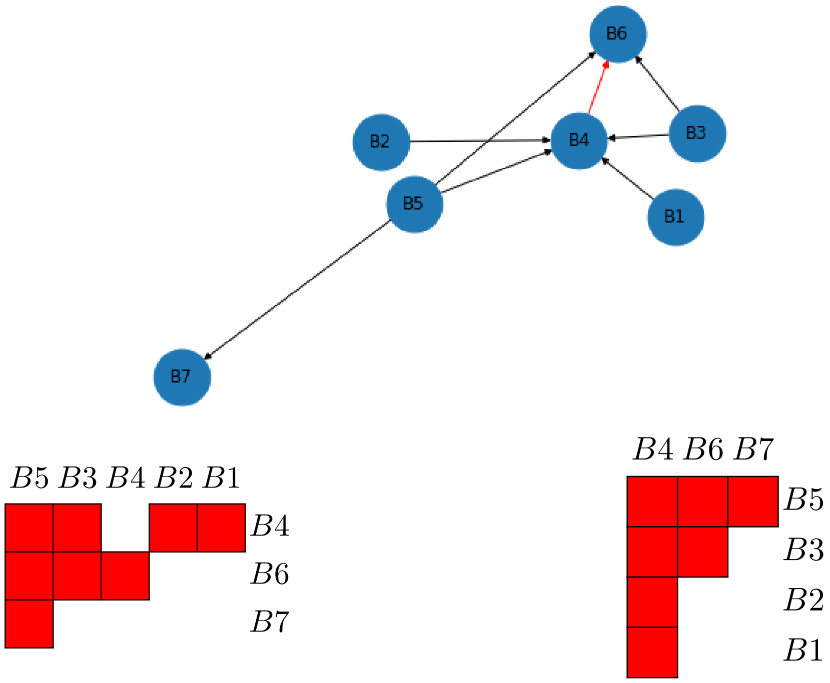

Figure 2: *Top: example of a perfectly nested interbank network from the lenders' point of view. Bottom: matrix representation of the graph with (left-hand side – LHS) and without (right-hand side – RHS) the link depicted in red in the top panel. In the LHS panel, lenders (borrowers) are displayed in columns (rows). In the RHS panel, it is the opposite. In the RHS case, the network is perfectly nested also from the borrowers' point of view.*

There is not a consensus on how nestedness should properly be quantified. For this reason, there are many metrics to measure nestedness being used simultaneously, such as the *nestedness metrics based on overlap and decreasing fill* (NODF) (Almeida-Neto et al., 2008) and the *spectral radius* (Staniczenko et al., 2013).[1] Nestedness is closely related to other network topological properties. Some studies (Abramson et al., 2011; Jonhson et al., 2013) confirmed that nestedness is significantly correlated with disassortativity. Lee et al. (2016) point out that nestedness is a generalization of the core-periphery structure. Payrató-Borras et al. (2019) propose that the most heterogeneous networks in terms of degree distribution are also the most nested ones.

Despite nested networks having been discovered (Patterson and Atmar, 1986) and mainly studied in ecology (Bascompte and Jordano, 2013), nestedness has also been reported in economic networks. Examples include country-country trading relationships (De Benedictis and Tajoli, 2011), manufacturer-contractor networks (Saavedra et al., 2009), country-product export networks (Tacchella et al., 2012), and interbank networks (König et al., 2014). The concept of nestedness has proved to be relevant for economic analysis, given its predictive power for a plethora of economic phenomena. Assessing the interactions between designer and contractor firms in the New York City garment industry, Saavedra et al. (2011) show firm's survival probability decreases as its nestedness contribution increases. Bustos et al. (2012) show nestedness in industrial ecosystems is quite stable, and hence it predicts the appearance and disappearance of individual industries in each location. The nestedness of world trade networks plays an important role in the prediction of countries' growth

[1]To more details, see, for instance, Payrató-Borràs et al. (2020) and Mariani et al. (2019), Section 3.1.

trajectory (Cristelli et al., 2017; Tacchella et al., 2012).

The purpose of this paper is to assess the nestedness of the Brazilian financial system. We compute the nestedness of two financial networks – the interbank network and the bank-firm credit network – using quarterly information from March 2012 through December 2015. Moreover, we apply the methodology developed by Saavedra et al. (2011) to compute the individual nestedness contribution (INC) of banks and firms in both networks. The INC of a given node is computed by comparing the nestedness of the network when the interactions of this node are randomized. In Figure 3, we show a simple example to explain this concept. Suppose we want to compute the INC of the node in the column highlighted in green in the original network (left-hand side of the figure). On the right-hand side of the figure, we show possible randomization of the node links. The original links are deleted, and the node is randomly connected to the nodes in the rows, keeping the same number of connections. Such randomization is performed as many times as possible. The INC of the node is given by comparing the average nestedness of the network when its links are randomized to that of the original network. If the average nestedness increases when the node links are randomized, its INC is positive, being negative otherwise (more details in Section 3.1).

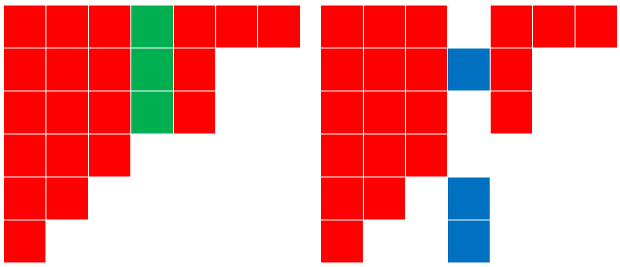

Figure 3: *Example of link randomization in order to compute the individual nestedness contribution (INC) of a given node (corresponding to the column highlighted in green in the l.h.s. chart).*

After computing the INC of the nodes, we perform two other exercises. In the first one, we apply machine learning techniques (random forest, XGBoost, and Shapley values) to identify the determinants – among a set of financial and topological variables – of the INC. In the second one, using only the interbank network, we assess the correlation between INC and two systemic risk measures presented in Alexandre et al. (2021): the systemic impact (SI) and the systemic vulnerability (SV) of the banks. While the former refers to the loss caused by a shock in the bank to the whole system, the latter measures the loss suffered by the bank in case of a shock in the system.

This study is related to the literature on the role of topological features in identifying systemically important banks (Alexandre et al., 2021; Ghanbari et al., 2018; Kuzubas et al., 2014; Martinez-Jaramillo et al., 2014). Our main conclusions are the following: i) the degree is the main determinant of the INC of lenders in both networks – banks in the bank-firm network and lending banks in the interbank market. On the other hand, there is not a clear main driver of the INC of borrowers. The importance of the features in determining the INC of borrowers is distributed much more evenly and depends on the network; and ii) in the interbank network, INC correlates positively to SI. Thus, nodes that contribute to the nestedness of the network are those that would cause more damage to the network if they were hit by a shock. Also, nodes with higher INC are not necessarily the most vulnerable to shocks on the network. A positive correlation between INC and vulnerability is observed only when the lenders' INC is considered.

This paper proceeds as follows. Sections 2 and 3 discuss, respectively, the data set and methodological issues. In Section 4, we bring the general results concerning the computation of nestedness and INC, as well as the analysis concerning our two exercises: the determinants of the INC and the correlation between INC and systemic risk. Finally, final considerations are presented in Section 5.

2 The data set

Using several unique Brazilian databases which comprises supervisory and accounting data, we extract quarterly information from March 2012 through December 2015 (16 periods) and build two networks: the bank-bank (interbank) network and bank-firm bipartite network.

The interbank network comprises all types of unsecured financial instruments registered in the Central Bank of Brazil (BCB). Credit, capital, foreign exchange operations, and money markets are among the main types of financial instruments. Different custodian institutions register and control these operations: Cetip[2] (private securities), the BCB's Credit Risk Bureau System – SCR[3] (credit-based operations), and the BM&FBOVESPA[4] (swaps and options operations).

[2]Cetip is a depositary of mainly private fixed income, state and city public securities, and other securities. As a central securities depositary, Cetip processes the issue, redemption, and custody of securities, as well as, when applicable, the payment of interest and other events related to them. The institutions eligible to participate in Cetip include commercial banks, multiple banks, savings banks, investment banks, development banks, brokerage companies, securities distribution companies, goods and future contracts brokerage companies, leasing companies, institutional investors, non-financial companies (including investment funds and private pension companies) and foreign investors.

[3]SCR is a very thorough data set that records every single credit operation within the Brazilian financial system worth 200BRL or above. Up to June 30th, 2016, this lower limit was 1,000BRL. Therefore, all the data we are assessing have been retrieved under this rule. SCR details, among other things, the identification of the bank, the client, the loan's time to maturity and the parcel that is overdue, modality of loan, credit origin (earmarked or non-earmarked), interest rate, and risk classification of the operation and the client.

[4]BM&FBOVESPA is a privately-owned company that was created in 2008 through the integration of the Sao Paulo Stock Exchange (Bolsa de Valores de Sao Paulo) and the Brazilian Mercantile & Futures Exchange (Bolsa de Mercadorias e Futuros). As Brazil's main intermediary for capital market transactions the company develops, implements and provides systems for trading equities, equity derivatives, fixed income securities, federal government bonds, financial derivatives, spot FX, and agricultural commodities. On March 30th, 2017, BM&FBOVESPA and Cetip merged into a new company named B3.

We compute the net financial exposures taking into account financial conglomerates or individual financial institutions that do not belong to conglomerates (classified as "b1", "b2", or "b4" in the BCB's classification system[5]), removing intra-conglomerate exposures. We exclude institutions with negative equity. Financial institutions' equity was retrieved from `https://www3.bcb.gov.br/ifdata`.

The bank-firm network is bipartite, i.e., considers only loans granted by banks to non-financial firms listed on the Brazilian stock exchange (BM&FBOVESPA). Information on firms' equity was retrieved from the *Economatica* database. Using the SCR information, we identified the loans granted by financial institutions for each of these firms. As in the interbank network, we include in the bank-firm network financial institutions with positive equity and classified as "b1", "b2", or "b4". Some statistics of both networks are presented in Tables 1 and 2.

Table 1: *Summary statistics of the bank-firm network.*

Quarter-year	N. of banks	N. of firms	Density	Avg. weighted degree*	Avg. net worth*
01-2012	72	304	0.0709	830.1	4079.5
02-2012	73	305	0.0700	854.5	4152.5
03-2012	73	302	0.0685	899.2	4244.0
04-2012	73	312	0.0685	943.6	4305.4
01-2013	72	315	0.0692	970.8	4090.7
02-2013	70	315	0.0697	1038.0	4288.6
03-2013	66	315	0.0723	1052.0	4408.4
04-2013	65	317	0.0731	1175.6	4418.6
01-2014	71	320	0.0695	1250.6	4505.5
02-2014	73	316	0.0681	1297.1	4608.2
03-2014	70	316	0.0706	1349.5	4588.4
04-2014	67	322	0.0753	1417.9	4409.9
01-2015	70	319	0.0711	1461.3	4482.1
02-2015	70	318	0.0714	1481.1	4618.7
03-2015	70	310	0.0725	1558.2	4718.0
04-2015	71	310	0.0737	1598.1	4465.7

*: in BRL million.

[5] See `https://www.bcb.gov.br/content/estabilidadefinanceira/scr/scr.data/metodologia.pdf`.

Table 2: Summary statistics of the interbank network.

Quarter-year	N. of banks	Density	Avg. weighted degree*	Avg. net worth*
01-2012	128	0.0843	2747.6	3516.2
02-2012	128	0.0850	2940.7	3598.1
03-2012	130	0.0825	3142.7	3620.7
04-2012	130	0.0802	3257.2	3690.7
01-2013	130	0.0823	3604.7	3609.1
02-2013	128	0.0837	3401.2	3610.6
03-2013	127	0.0796	3474.4	3728.4
04-2013	127	0.0777	3557.1	3840.9
01-2014	130	0.0773	3551.3	3724.5
02-2014	130	0.0773	3433.9	3830.6
03-2014	130	0.0781	3756.4	3908.8
04-2014	129	0.0732	3970.7	3878.8
01-2015	129	0.0757	3966.3	3943.7
02-2015	130	0.0743	3819.8	4071.2
03-2015	128	0.0781	4023.9	4127.6
04-2015	126	0.0792	4111.5	4181.9

*: in BRL million.

3 Methodology

3.1 Measuring nestedness and INC

In this paper, we quantify nestedness using the NODF (Almeida-Neto et al., 2008). The nestedness of the network N is defined by the following equation:

$$N = \frac{\sum_{i<j}^{C} M_{ij} + \sum_{i<j}^{R} M_{ij}}{\left[\frac{C(C-1)}{2}\right] + \left[\frac{R(R-1)}{2}\right]}. \tag{1}$$

In Eq. 1 above, C (R) is the number of nodes of the type displayed in columns (rows). Note that these numbers can be different in bivariate networks, but will necessarily be equal in univariate networks. For every pair of nodes i and j, $M_{ij} = 0$ if $k_i = k_j$, and $M_{ij} = n_{ij}/min(k_i, k_j)$ otherwise, where k_i is the number of interactions of node i, and n_{ij} is the number of interactions in common between i and j. N varies between 0 and 1, where 1 designates a perfectly nested network.

The INC is quantified following the methodology developed by Saavedra et al. (2011). The INC of node i is given by the following equation:

$$c_i = \frac{(N - \langle N_i^* \rangle)}{\sigma_{N_i^*}}, \tag{2}$$

where N is the network's observed nestedness, $\langle N_i^* \rangle$ is the average nestedness across an ensemble of random replicates within which the interactions of node i have been randomized, and $\sigma_{N_i^*}$ is the standard deviation of N_i^*. We randomize the interactions of a node following the null model specified in Bascompte et al. (2003), generating 1,000 random replicates. The randomization of the interactions of a given node i works as follows: we cancel some link between i and another node, and then we connect i with another node with which i does not have a connection.

We innovate in the computation of the INC by considering the different roles a node can play in a network. In bivariate networks, nodes play only one role. For instance, in a bank-firm credit network, banks are always lenders and firms, borrowers. However, in univariate, directed networks, our innovation can be quite useful. For example, in interbank networks, all nodes are of the same type (banks), but a given node i can be lender and borrower at the same time. If this is the case, we will compute its lending INC INC_L and its borrowing INC INC_B. The former is obtained by randomizing only its outgoing links, which represent loans granted by i, and keeping its incoming links – loans received by i – fixed. The latter is computed in a similar fashion, through the opposite operation.

3.2 Machine learning techniques

After computing the INC for our data instances using the methodology presented in Section 3.1, we assess its determinants employing three machine learning techniques: random forest (RF), XGBoost (XB), and Shapley values. While RF and XB are ensemble learning methods that can be used for both classification and regression, Shapley values are employed to provide a better interpretability of the results.

Random forest and XGBoost. In this study, both RF (Breiman, 2001) and XB (Friedman et al., 2000) are used for regression tasks. The aim is to estimate a predicted output $\hat{y}_i$ from an observed output y_i and a vector of explanatory variables X_i. RF reports the average prediction of several decision trees.[6] The main advantage of using an RF rather than an individual decision tree is avoiding over-fitting problems. When used to solve regression tasks, decision trees are called *regression trees*. Departing from the *root node*, which encompasses the whole data set, the algorithm runs by answering true/false questions till it reaches the *leaf node*, with the output. The intermediate nodes are called *interior nodes*. Figure 4 depicts a simple regression tree. There are only three explanatory variables: x_1, x_2, and x_3. The predicted output can assume four values: y_1, y_2, y_3, and y_4. The value of the predicted output for a given data instance depends on whether or not the value of the explanatory variables is above a specific critical value.

[6]On decision trees, see, e.g., Breiman et al. (1984).

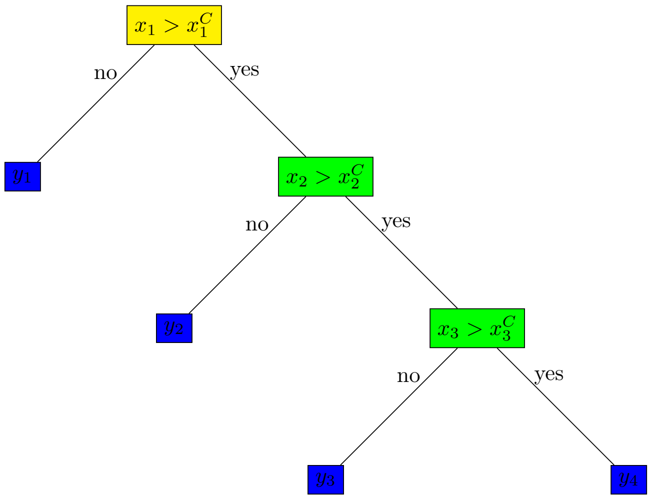

Figure 4: *A simple example of a regression tree, with the root node (yellow), the interior nodes (green), and the leaf nodes (blue).*

XB is an optimization algorithm that creates a more efficient predictor model from an ensemble of weak predictors (usually, decision trees). This is done by increasing the performance of the predecessor predictor through the inclusion of a new estimator – i.e., a new combination of explanatory variables and corresponding weights. The algorithm assigns weights to all explanatory variables, which are used as input by the first weak predictor. If the output is predicted wrong by the weak predictor, the weight of the corresponding explanatory variables is increased. In the next boosting stage, these variables are fed to the second weak predictor, and so on. The final predictor is the result of an ensemble of these weak predictors.

The process through which the models are trained and validated is known as *repeated k-fold cross-validation*. The data set – the output to be predicted and a set of potential explanatory variables – is split into k different parts (folds). $k-1$ folds are used in the training of the model. The remaining fold is used to test the efficiency of the model. Score measures (e.g., the root mean squared error – RMSE – and the R^2) are computed from the predicted output $\hat{y}_i$ and the observed output y_i of the remaining fold. The process is repeated until each fold is used as the testing data set. Therefore, the number of regressions run is equal to the number of repetitions times k.

We tune the number of estimators of both methods – the number of decision trees in each forest in RF and the number of boosting stages to be performed in XB. We vary the number of estimators within a grid of ascending values and compute the average score across the regressions for each of these values. In this paper, this grid is 30, 50, 70, 100, 300, and 500. The number of estimators is set when the increase in the value does not improve the performance of the method, as measured by the score measure. In this study, the number of estimators was set as 50 in both methods. The other main parameters are reported in Table 3.

Table 3: *Main parameters: Random forest and XGBoost.*

Parameter	Value
Random forest:	
Minimum number of samples required to split	2
Minimum number of samples of a leaf node	1
Fraction of of features to consider when looking for the best split	1
XGBoost:	
Maximum depth	6
Learning rate	0.3
Fraction of columns sampled for each tree	1
Fraction of observations sampled for each tree	1

Shapley values. Shapley values, an approach that originated from the coalition games theory (Shapley, 1953; Shoham and Leyton-Brown, 2008), allows for better interpretability of machine learning predictors. It provides information on the magnitude, as well as the sign (positive or negative), of the features' importance to the output. In this study, we use the SHAP (SHapley Additive exPlanation) framework (Lundberg and Lee, 2017) to compute Shapley values. In this framework, there is an explainer model g using a set of M features as inputs aiming at predicting an output. The predicted value for a given data instance is given by

$$g(z') = \phi_0 + \sum_{i=1}^{M} \phi_i z'_i, \tag{3}$$

where z'_i is a binary variable indicating whether feature i was included in the model or not. The SHAP value ϕ_i indicates the extent in which the feature i shifts the predicted value up or down from a given mean output ϕ_0. Lundberg and Lee (2017) showed, under certain properties (local accuracy, missingness, and consistency), ϕ_i corresponds to the Shapley value of the game theory. The SHAP value of feature i is represented by the following equation:

$$\phi_i = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|!(|M|-|S|-1)!}{M!} [F(S \cup \{i\}) - F(S)]. \tag{4}$$

Therefore, the SHAP value of feature i for a given data instance computes the difference between the predicted value of the instance using all features including i, $F(S \cup \{i\})$, and the prediction excluding feature i, $F(S)$. A weight is applied to these values, which are summed over all possible feature vector combinations of all possible subsets S.[7]

[7] For details on the calculation of SHAP values, see, e.g., Lundberg and Lee (2017) and Kalair and Connaughton (2021).

3.3 Topological variables

Among the potential determinants of the INC, there are some node specific topological variables. Here, we will give a definition of each one of them.

Degree. The *degree* is one of the simplest topological measures. This is the number of links of a given node. For instance, in Figure 5, node E has degree 3, as it is connected to other three nodes (D, G, and W). In directed networks, such as the interbank network, the *in-degree* is the number of incoming links (i.e., the number of lending partners) of a given node. Similarly, the *out-degree* is the number of outgoing links – the number of partners that node grants loans to. The weight used in the *weighted degree* is the value of the loan. Therefore, the *weighted in(out)-degree* corresponds to the total amount borrowed from (granted to) other nodes.

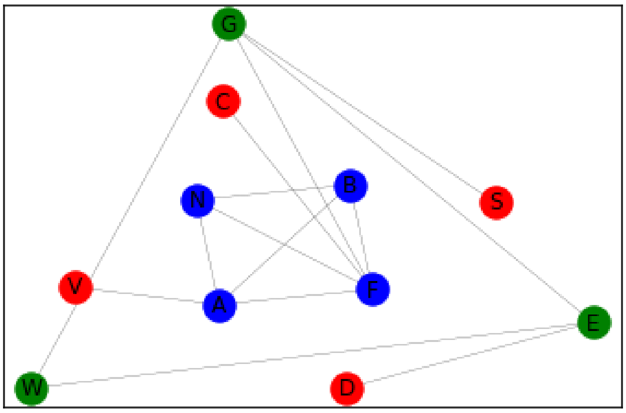

Figure 5: *Stylized undirected network, in which the nodes are colored according to their core number: blue (3), green (2), and red (1).*

Core number. A subgraph (part of a graph) H is said to be a k-core subgraph of graph G if this is the subgraph with the maximal number of nodes in which all nodes have a degree of at least k. Consider all subgraphs a given node belongs to. The *core number* of this node will be the maximum k. A simple example is depicted in Figure 5. The blue nodes compose a k-core with $k = 3$, as each blue node is connected to at least three other blue nodes. Therefore, the core number of the blue nodes is 3. Similarly, the green nodes have a core number equal to 2 and the red ones, to one.

Closeness centrality. The shortest path distance between two nodes is the one in which the sum of the weights of the links is minimized. Supposing all links have weight one, if two nodes are connected, the shortest path distance between them is equal to one. The *closeness centrality* of a given node is the average shortest path distance between this node and every other node in the network. For instance, a simple computation would show the closeness centrality of node F in Figure 5 is 0.625.

Eigenvector centrality. Differently from the closeness centrality, the eigenvector centrality (Bonacich, 1972) is not related to shortest paths. It is given by the components of the main eigenvector of the adjacency matrix representing the network. Let $A = (a_{i,j})$ be the adjacency matrix of a graph. The eigenvector centrality ec_i of node i is given by

$$ec_i = \frac{1}{\lambda} \sum_k a_{k,i}\, ec_k, \tag{5}$$

where $\lambda \neq 0$ is a constant.

PageRank. The PageRank centrality (Gleich, 2015) is an extension of the eigenvector centrality, specially designed for directed graphs. The PageRank of a node is positively impacted by its in-degree, but also by the in-degree of its direct and indirect neighbors, weighted by a damping factor. The PageRank centrality pr_i of node i is given by

$$pr_i = \alpha \sum_k \frac{a_{k,i}}{d_k} pr_k + \beta, \tag{6}$$

where $a_{k,i}$ is the component of the adjacency matrix of the graph, α and β are constants and d_k is equal to the out-degree of node k if such degree is positive, or 1 otherwise.

Clustering coefficient. The clustering coefficient measures the degree to which nodes are prone to cluster together. The clustering coefficient cc_i of a given node i is given by

$$cc_i = \frac{1}{k_i(k_i-1)} \sum_{j,k} a_{i,j} a_{j,k} a_{k,i}, \tag{7}$$

where $a_{i,j}$ is the component of the adjacency matrix of the graph and k_i is the degree of node i. Therefore, the clustering coefficient of a given node will be maximal (1) if it is connected to all neighbors of its direct neighbors.

3.4 Systemic risk

Saavedra et al. (2011) show the nodes with higher INC are those whose removal leads to a decrease in network persistence, as well as are the more vulnerable to extinction. That is, shocks in strong contributors cause more damage to the whole network, and shocks in the network affect mostly the strong contributors. We test this hypothesis using our data set. To this end, we compute the *systemic impact* and *systemic vulnerability* – SI and SV, respectively (Alexandre et al., 2021) – for the banks participating in the Brazilian interbank market, taking into consideration various levels of the initial shock. Then, we compute the correlation between these measures and the banks' INC.

Both SI and SV are computed following the *differential DebtRank* methodology (Bardoscia et al., 2015).[8] The exposure network of the interbank market is represented by $\mathbf{A} \in N \times N$, where N is the number of banks and A_{ij} is the asset invested by i in j. At period 0, we impose an exogenous shock

[8] The rest of this subsection strictly follows Alexandre et al. (2021).

on FI j, reducing its equity by a fraction of ζ. It will cause a subsequent loss $L_{ij}(1)$ to its creditors, indexed by i, equal to $\mathbf{A}_{ij}\zeta$. At period 2, j's creditors will propagate this loss to their creditors in a similar fashion, and so on. Formally, we have

$$L_{ij}(t) = min\left(A_{ij}, L_{ij}(t-1) + \mathbf{A}_{ij}\frac{[L_j(t-1) - L_j(t-2)]}{E_j}\right), \tag{8}$$

$$L_i(t) = min\left(E_i, L_i(t-1) + \sum_j \mathbf{A}_{ij}\frac{[L_j(t-1) - L_j(t-2)]}{E_j}\right), \tag{9}$$

in which $t \geq 0$ and E_j is financial institution (FI) j's equity. Thus, when an FI j suffers an additional loss equal to fraction ζ of its equity, it will impose a loss to its creditors that corresponds to ζ times their exposures towards j. Observe equity positions as well as the exposure network are time-invariant, i.e., they are taken as exogenous. The propagation considers stress differentials rather than stress absolute values (hence the methodology's name) to avoid double-counting.

Observe L_{ij} cannot be greater than A_{ij}. It means that j cannot impose to i a loss greater than i's exposures towards j. When $L_{ij} = A_{ij}$, j stops imposing losses on i. Moreover, L_i cannot be greater than E_i, i.e., i's losses cannot be greater than its equity. When $L_i = E_i$, i stops propagating losses to other FIs.

The system converges after a sufficiently large number of periods $T \gg 1$. Then we have the final matrix of losses $\mathbf{L}^{j,\zeta} \in N \times 1$, where $L_j^{i,\zeta}$ is the total loss suffered by agent j after an initial shock of size ζ on agent i. After repeating this process for the other FIs, we compute our two measures of SR. The *systemic impact* (SI) of bank i is defined as

$$SI_{i\zeta} = \frac{\sum_j \left[L_j^{i,\zeta} - L_j^{i,\zeta}(0)\right]}{\sum_j E_j}, \tag{10}$$

where $L_j^{i,\zeta}(0) = \zeta E_j$ if $j = i$ and 0 otherwise. The *systemic vulnerability* (SV) is represented by the following equation:

$$SV_{i\zeta} = \frac{1}{N}\sum_j \frac{L_i^{j,\zeta} - L_i^{j,\zeta}(0)}{E_i}. \tag{11}$$

Therefore, $SI_{i\zeta}$ measures the fraction of the aggregate FIs' equity which is lost as a consequence of an initial shock of size ζ at FI i's equity. On the other hand, $SV_{i\zeta}$ refers to the average i's equity loss when the other FIs are reduced by ζ.

As we are interested only in the losses caused by the contagion, we remove the initial shock from the computation of the SR measures. Observe we also compute $SI_{i\zeta}$ for the FI that suffered the initial shock. Due to network cyclicality, a shock propagated by a given FI can hit it back. For the same reason, we include the loss imposed by an FI on itself in the calculation of $SV_{i\zeta}$.

4 Results

4.1 General results

Figure 6 depicts the NODF of the two networks (interbank and bank-firm) for different dates. We can observe the bank-firm network always displays a higher NODF. Therefore, the hierarchical organization typical of nested networks is more noticeable in the bank-firm network than in the interbank network. The distribution of the INC through all dates is presented in Figure 7. In all cases, most of the observations are higher than zero. While the INC is roughly represented by a normal distribution in the case of the borrowers in both networks (borrower banks in the interbank network and firms in the bank-firm network), the peakedness of the distribution is higher in the case of the lenders. Moreover, the highest INC values are observed among the lenders. The exclusion of some of them would lead to a higher decrease in the nestedness of the network.

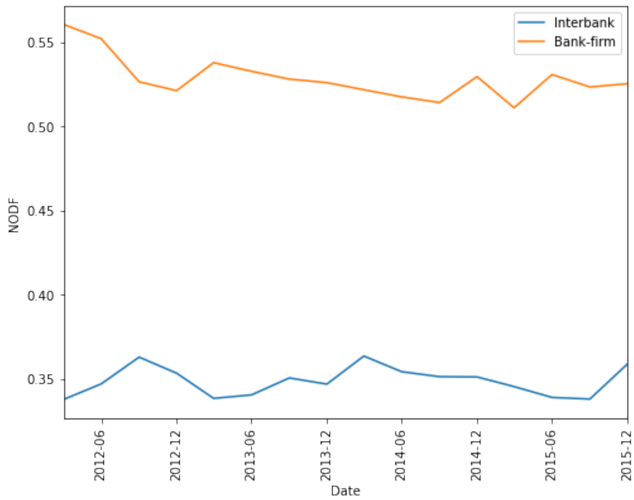

Figure 6: *NODF of the interbank and the bank-firm network.*

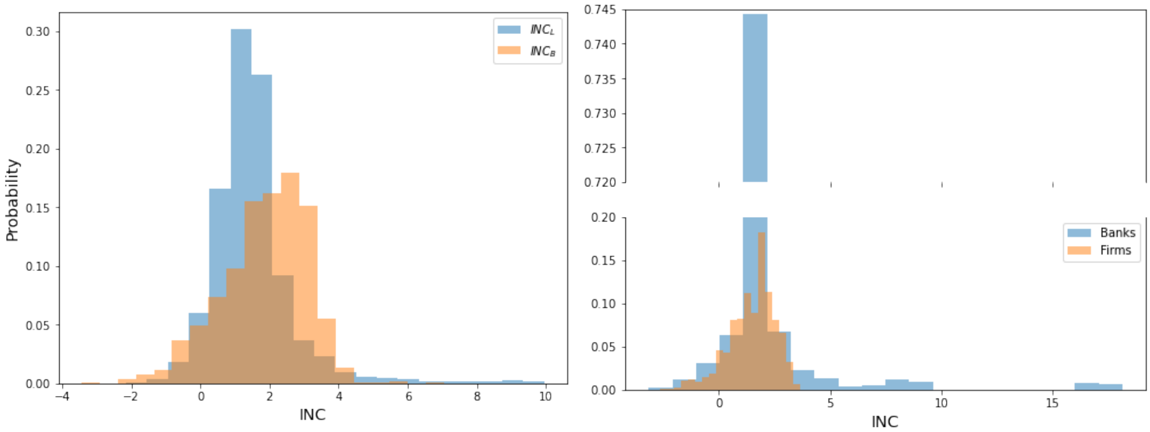

Figure 7: *Histograms of the individual nestedness contribution (INC) for the interbank (left) and bank-firm (right) network. Note the bin width is different for each chart.*

The average INC for lenders (borrowers) in the interbank network is 1.51 (1.87). In the bank-firm network, these values are 2.03 and 1.50, respectively. Saavedra et al. (2011) suggest groups with lower INC have advantages over those with higher INC, in the sense the former increases its survival probability at the expense of the latter. Therefore, lenders have an advantage over borrowers in the interbank network, but the opposite occurs in the bank-firm network.

4.2 The determinants of the INC

In this section, we assess the determinants of the INC. The variables assessed as potential determinants and their averages for each network are presented in Table 4. The weighted degree (wK) corresponds to the total amount of loan granted (in the case of the lenders) or received (in the case of the borrowers). We performed this analysis separately for banks and firms in the bank-firm network, and lenders (INC_L) and borrowers (INC_B) in the interbank network. Recalling that, in the latter case, both INC_B and INC_L will be computed for a given bank if it acts as lender and borrower. We applied the two ML techniques – RF and XB – to predict the INC. We use a repeated k-fold cross-validation with $k = 5$ and 10 repetitions, hence 50 regressions were run.

Table 4: Potential determinants of the INC assessed in the study.

Variable	Acronym	Avg. bank-firm	Avg. interbank
Degree	K[1]	4.07	–
In-degree	Kin[2]	–	1.05
Out-degree	Kout[2]	–	1.07
Weighted degree	wK[1]	6.00×10^8	–
Weighted in-degree	wKin[2]	–	3.70×10^9
Weighted out-degree	wKout[2]	–	3.76×10^9
Core number	KC	4.36	1.22
Closeness centrality	CC	1.06×10^{-2}	4.02×10^{-1}
Eigenvector centrality	EC	3.60×10^{-2}	7.15×10^{-2}
PageRank	PR	2.61×10^{-3}	8.08×10^{-3}
Clustering coefficient	C[2]	–	3.41×10^{-1}
Net worth	NW	4.40×10^9	4.02×10^9

1: Only for the bank-firm network.

2: Only for the interbank network.

Bank-firm network In the bank-firm network, there are 1,126 data instances for banks (average of 70.4 per period) and 5,016 data instances for firms (average of 313.5 per period). We have obtained a better performance – in terms of R^2 – in predicting the INC of the banks (Figure 8). The main driver of the INC of banks in the bank-firm network is the degree (Figure 9). The Shapley analysis (Figure 10) corroborates these results and shows this impact is positive.

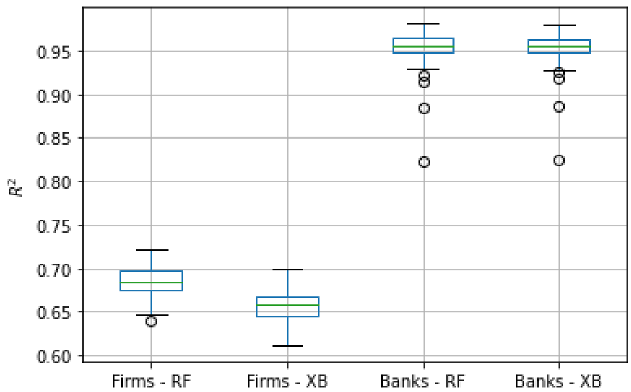

Figure 8: R^2 of the regressions for the bank-firm network.

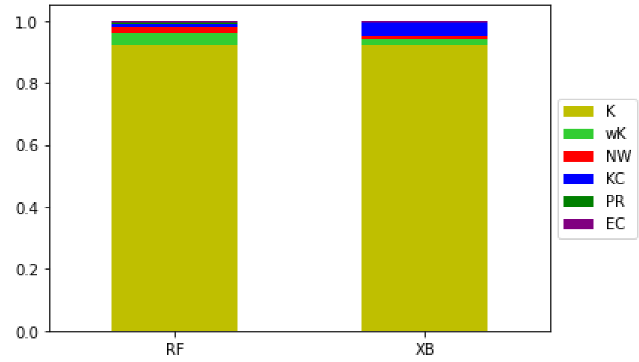

Figure 9: *Importance of the features to the prediction of the INC for the banks in the bank-firm network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) were excluded.*

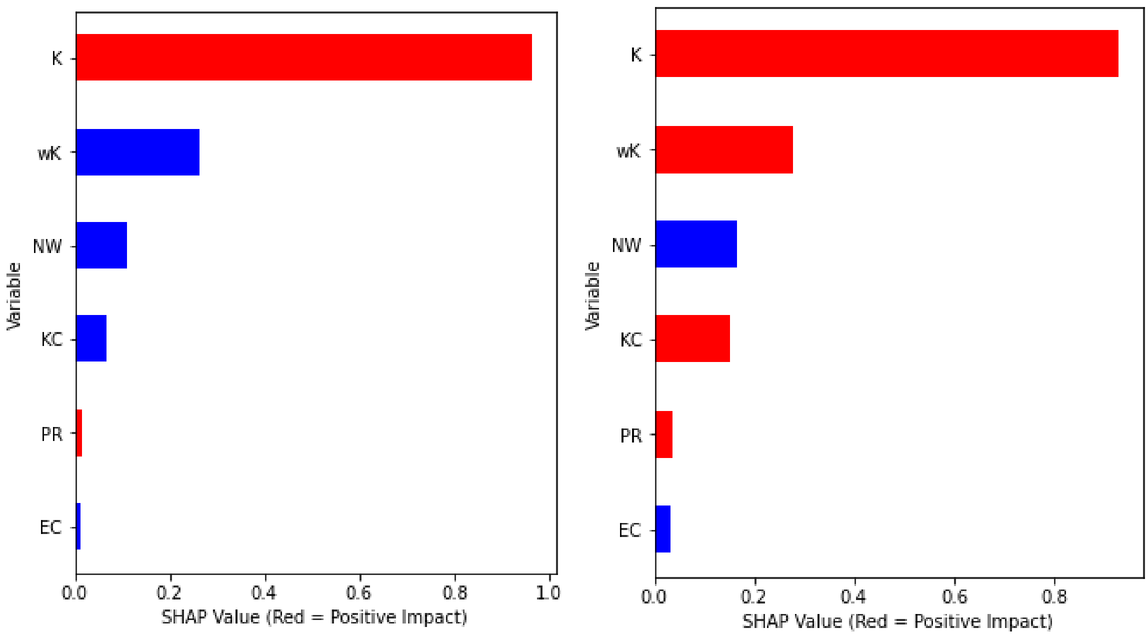

Figure 10: *Features' average absolute SHAP. Red (blue) bars denote a positive (negative) correlation between SHAP values and the feature values. The predicted output is the INC of banks in the bank-firm network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) have been excluded.*

On the other hand, the INC of firms in the bank-firm network is driven by a mix of variables. The main determinant is the core number, whose impact is positive (Figures 11 and 12). The PageRank and the weighted degree are also important drivers of the INC of the firms, being the impact of the former (latter) negative (positive).

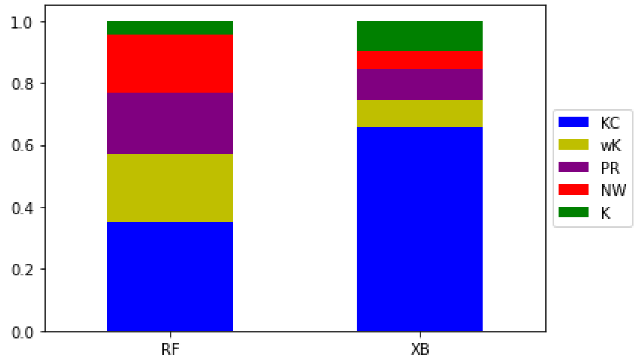

Figure 11: Importance of the features to the prediction of the INC for the firms in the bank-firm network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) have been excluded.

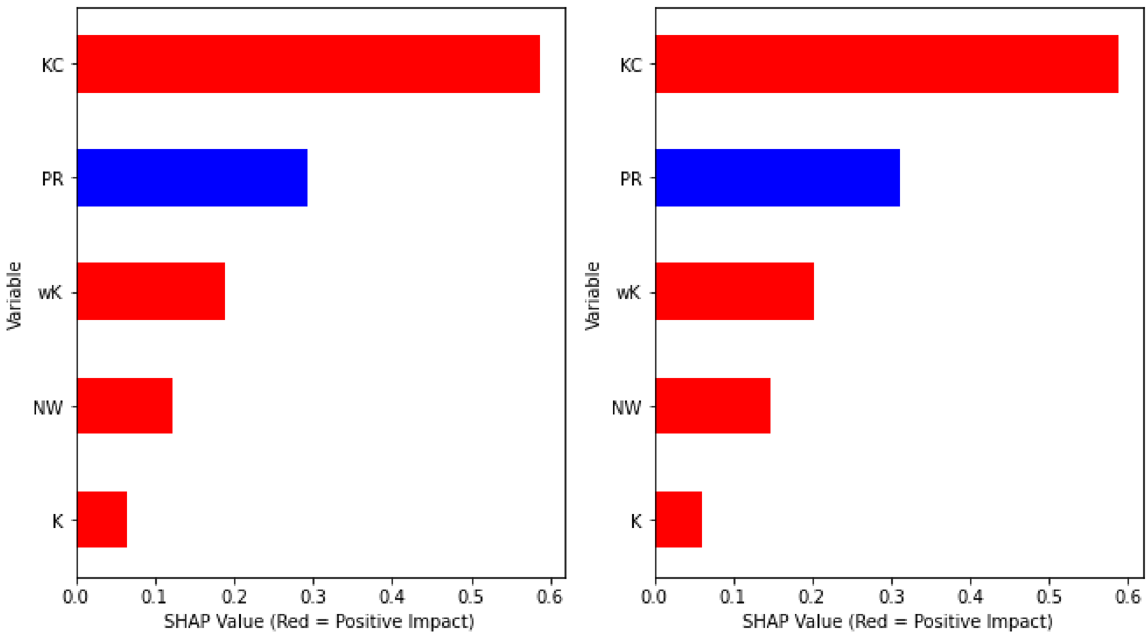

Figure 12: *Features' average absolute SHAP. Red (blue) bars denote a positive (negative) correlation between SHAP values and the feature values. The predicted output is the INC of firms in the bank-firm network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) have been excluded.*

Interbank network The number of data instances is 1,982 for borrowers and 1,944 for lenders (average of 123.9 and 121.5 per period, respectively). Similar to the banks in the bank-firm network, the R^2 is higher for the lenders in the interbank network (Figure 13). Also in this case the INC of the lenders is mainly determined by their degree (Figures 14 and 15). The importance of the features in determining the INC of the borrowers is distributed much more evenly. Closeness centrality, in- and out-degree, and clustering coefficient are among the main determinants (Figures 16 and 17). Therefore, as a general result, we have that degree is the main driver of the INC of lenders in financial networks, while the INC of borrowers has not a clear main determinant.

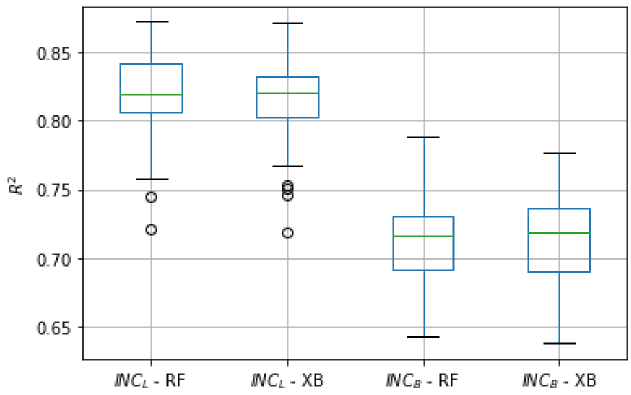

Figure 13: *R^2 of the regressions for the interbank network.*

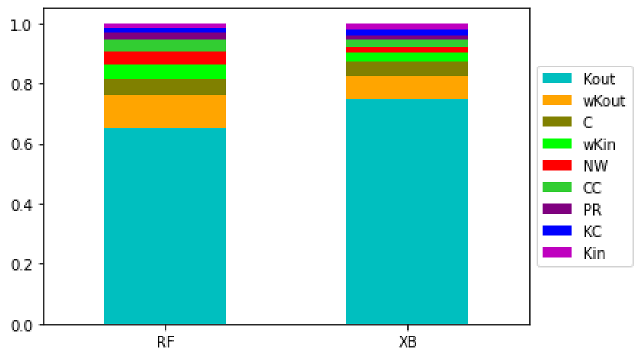

Figure 14: *Importance of the features to the prediction of the INC for lenders in the interbank network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) have been excluded.*

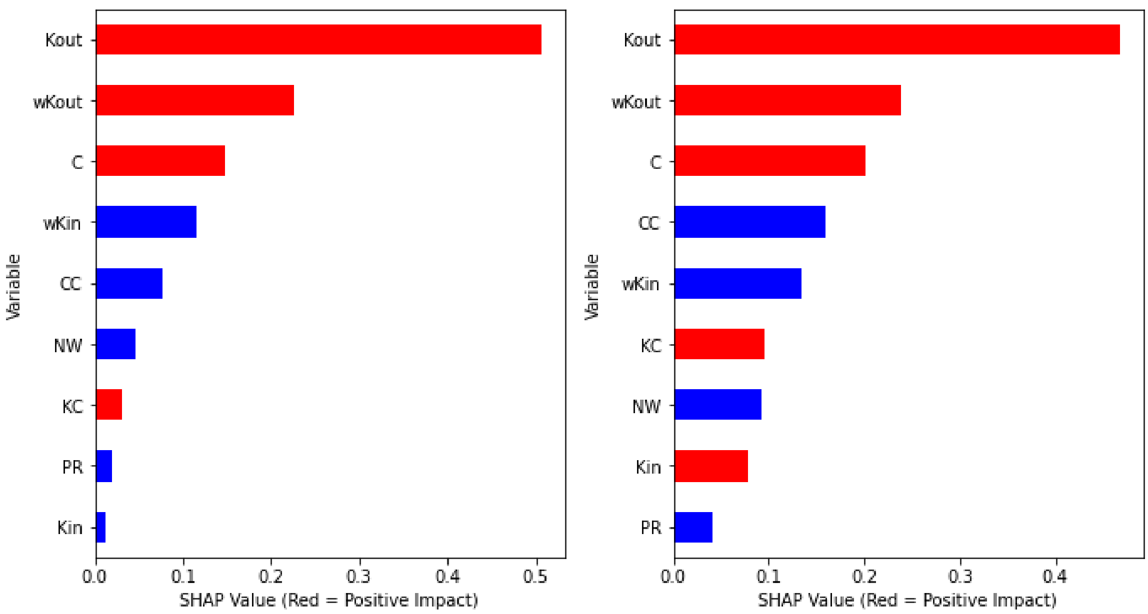

Figure 15: *Features' average absolute SHAP. Red (blue) bars denote a positive (negative) correlation between SHAP values and the feature values. The predicted output is the INC of lenders in the interbank network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) have been excluded.*

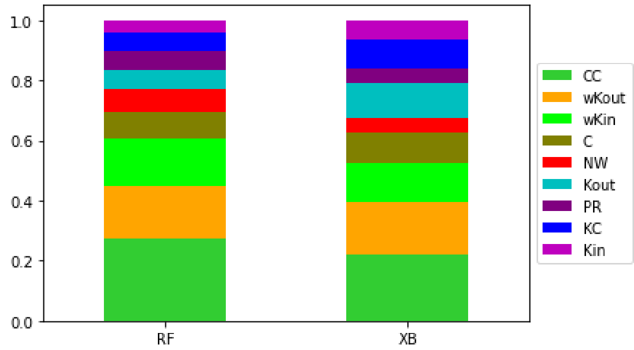

Figure 16: *Importance of the features to the prediction of the INC for borrowers in the interbank network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) have been excluded.*

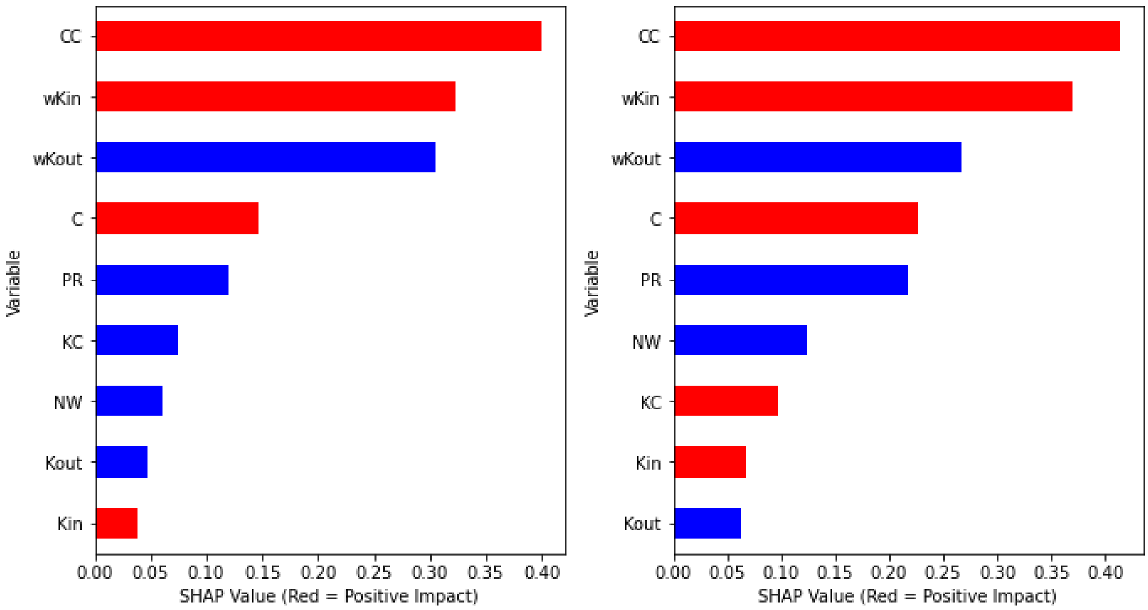

Figure 17: *Features' average absolute SHAP. Red (blue) bars denote a positive (negative) correlation between SHAP values and the feature values. The predicted output is the INC of borrowers in the interbank network obtained through RF (left) and XB (right). Highly correlated variables (correlation higher than 0.9 to another explanatory variable) have been excluded.*

4.3 Nestedness and systemic risk

Saavedra et al. (2011) found nodes with a higher INC are those whose losses are more detrimental to the network's persistence, as well as are the more vulnerable ones. We test this hypothesis for the Brazilian interbank network. Here, the impact of the node on the network resilience and its vulnerability will be measured, respectively, by the concepts of SI and SV, as detailed in Section 3.4.

Both SI and SV are computed for each node and for different levels of the initial shock ζ. Moreover, we compute for each node its total INC $INC_T = INC_B + INC_L$. Finally, we compute the correlation between INC (INC_T, INC_B, and INC_L) and systemic risk (SV and SI).

Considering the total INC, we did not observe the correlation between INC and vulnerability found by Saavedra et al. (2011) (Figure 18, left panel). The correlation between total INC and SV is negative; moreover, it is not significantly different from zero for all levels of the initial shock ζ. However, as in Saavedra et al. (2011), the INC is positively correlated to SI (Figure 18, right panel). Therefore, the nodes that contribute the most to the nestedness of the network are also those that would cause more damage to the network in case of suffering a shock. Also, this correlation is nonlinear concerning ζ.

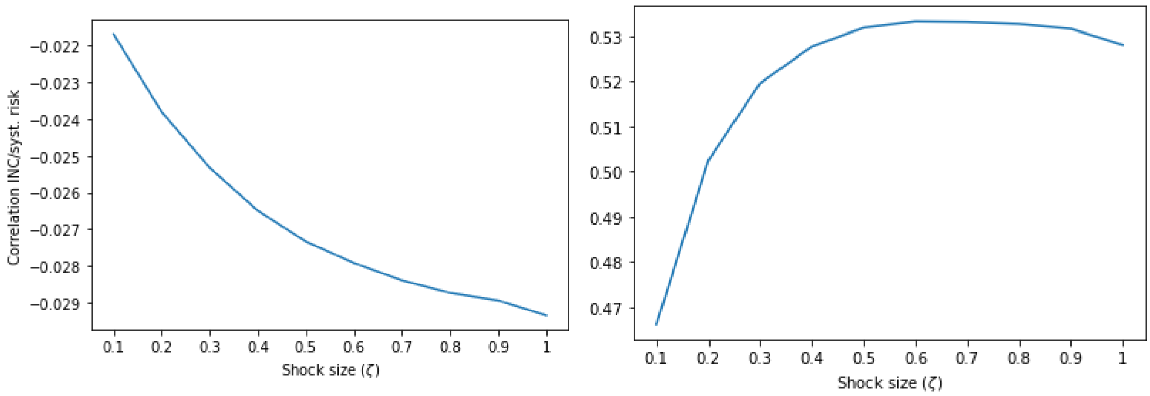

Figure 18: *Correlation between INC_T and SV (left) and SI (right). Except for the left panel, the correlation is statistically different from zero for all levels of ζ (p-value < 10^{-100}).*

Decomposing the INC of the nodes considering both their roles – lender and borrower –, we find that, while INC_L and node vulnerability are positively correlated (Figure 19, left panel), this correlation is negative in the case of INC_B (Figure 20, left panel). In both cases, the absolute value of the correlation increases with the size of the initial shock. Both INC_L and INC_B are positively correlated to SI (Figures 19 and 20, right panel). While in the latter case the correlation increases with ζ, in the former one this relationship is nonlinear, represented by an inverted U-shaped curve.

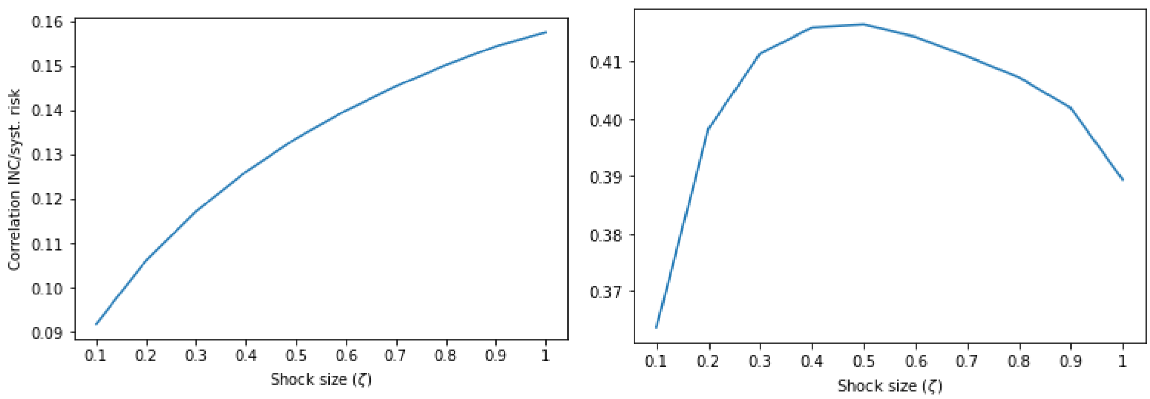

Figure 19: *Correlation between INC_L and SV (left) and SI (right). In both panels, the correlation is statistically different from zero for all levels of ζ (p-value < 10^{-4}).*

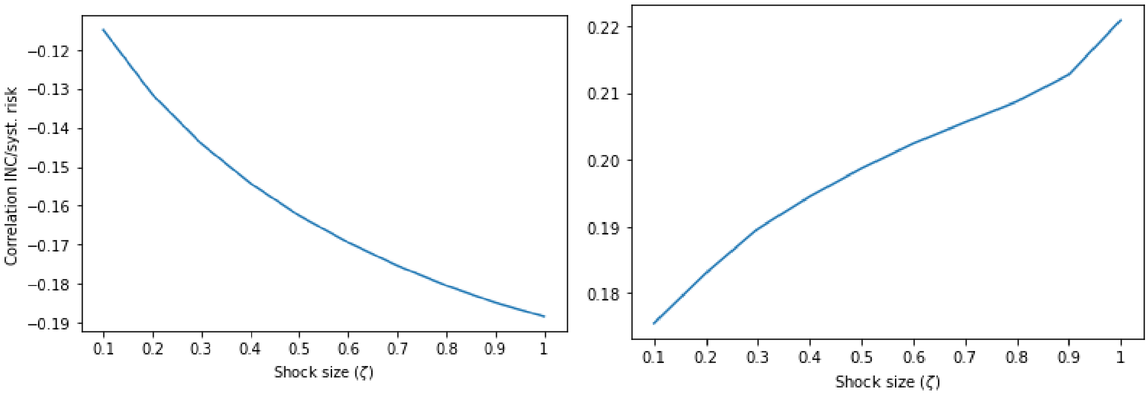

Figure 20: *Correlation between INC_B and SV (left) and SI (right). In both panels, the correlation is statistically different from zero for all levels of ζ (p-value $< 10^{-6}$).*

5 Final considerations

In this study, we assessed the nestedness of the Brazilian financial system. To accomplish this task, we relied on data from two Brazilian financial networks: the bank-firm network and the interbank network. Besides computing the nestedness – measured here by the NODF – of the financial networks for different dates, we also calculated the Individual Nestedness Contribution (INC), which is a measure of the node contribution to the network nestedness. Then, we performed two exercises. In the first one, we assessed the determinants of the INC. This task was performed separately for banks and firms in the bank-firm network, and for lenders and borrowers in the interbank network. In this analysis, we applied machine learning techniques: random forest, XGBoost, and Shapley values.

We concluded, in both networks, the INC of lenders – banks in the bank-firm network and lending banks in the interbank network – is mainly determined by their degree. On the other hand, there is not a single main driver of the INC of borrowers. Core number, weighted degree, and PageRank are the main determinants of the INC of firms in the bank-firm network. The INC of borrowers in the interbank network is mainly driven by closeness centrality, weighted degree (in and out), and clustering coefficient.

Finally, in the second exercise, we assessed the relationship between INC and systemic risk. We computed the correlation between the INC and the systemic impact (SI) – the loss caused in the network by a shock on the node – and systemic vulnerability (SV) – the loss suffered by the node due to a shock in the network – in the interbank network. The INC is positively correlated to the SI. Thus, nodes that contribute the most to the nestedness of the network are those that would cause more damage to the network if they were hit by a shock. The correlation between the total INC and SV is not significantly different from zero. However, while the lending INC is positively correlated to SV, the correlation between the borrowing INC and SV is negative. It means that nodes with a higher lending (borrowing) INC are more (less) vulnerable to shocks on the network. Furthermore, the absolute value of this correlation increases with the size of the initial shock.

This study contributes to the literature on identifying systemically relevant banks through the analysis of the topological features of the financial network. We show the INC is correlated to the systemic importance of banks. Shocks on banks with higher INC would cause a higher loss in the whole system. Moreover, shocks on the system would cause more damage to banks with a greater lending INC, while banks with a higher borrowing INC would be less impacted by such shocks. A natural follow-up study of this paper would investigate the INC as a driver of the systemic importance of the banks, in a model including other explanatory variables.

References

Abramson, G., Trejo Soto, C. A., and Ona, L. (2011). The role of asymmetric interactions on the effect of habitat destruction in mutualistic networks. *PLoS One*, 6(6):e21028.

Alexandre, M., Silva, T. C., Connaughton, C., and Rodrigues, F. A. (2021). The drivers of systemic risk in financial networks: a data-driven machine learning analysis. *Chaos, Solitons & Fractals*, 153:111588.

Almeida-Neto, M., Guimaraes, P., Guimaraes Jr, P. R., Loyola, R. D., and Ulrich, W. (2008). A consistent metric for nestedness analysis in ecological systems: reconciling concept and measurement. *Oikos*, 117(8):1227–1239.

Bardoscia, M., Battiston, S., Caccioli, F., and Caldarelli, G. (2015). Debtrank: A microscopic foundation for shock propagation. *PloS One*, 10(6):e0130406.

Bascompte, J. and Jordano, P. (2013). *Mutualistic networks*. Princeton University Press.

Bascompte, J., Jordano, P., Melián, C. J., and Olesen, J. M. (2003). The nested assembly of plant–animal mutualistic networks. *Proceedings of the National Academy of Sciences*, 100(16):9383–9387.

Bonacich, P. (1972). Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Bustos, S., Gomez, C., Hausmann, R., and Hidalgo, C. A. (2012). The dynamics of nestedness predicts the evolution of industrial ecosystems. *PloS one*, 7(11):e49393.

Cristelli, M. C. A., Tacchella, A., Cader, M. Z., Roster, K. I., and Pietronero, L. (2017). On the predictability of growth. *World Bank Policy Research Working Paper*, (8117).

De Benedictis, L. and Tajoli, L. (2011). The world trade network. *The World Economy*, 34(8):1417–1454.

Friedman, J., Hastie, T., Tibshirani, R., et al. (2000). Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *Annals of statistics*, 28(2):337–407.

Ghanbari, R., Jalili, M., and Yu, X. (2018). Correlation of cascade failures and centrality measures in complex networks. *Future Generation Computer Systems*, 83:390–400.

Gleich, D. F. (2015). PageRank beyond the web. *Siam Review*, 57(3):321–363.

Jonhson, S., Domínguez-García, V., and Muñoz, M. A. (2013). Factors determining nestedness in complex networks. *PloS One*, 8(9):e74025.

Kalair, K. and Connaughton, C. (2021). Dynamic and interpretable hazard-based models of traffic incident durations. *arXiv preprint arXiv:2102.08729*.

König, M. D., Tessone, C. J., and Zenou, Y. (2014). Nestedness in networks: A theoretical model and some applications. *Theoretical Economics*, 9(3):695–752.

Kuzubas, T. U., Omercikoglu, I., and Saltoglu, B. (2014). Network centrality measures and systemic risk: An application to the Turkish financial crisis. *Physica A*, 405:203–215.

Lee, S. H. et al. (2016). Network nestedness as generalized core-periphery structures. *Physical Review E*, 93(2):022306.

Lundberg, S. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

Mariani, M. S., Ren, Z.-M., Bascompte, J., and Tessone, C. J. (2019). Nestedness in complex networks: observation, emergence, and implications. *Physics Reports*, 813:1–90.

Martinez-Jaramillo, S., Alexandrova-Kabadjova, B., Bravo-Benitez, B., and Solórzano-Margain, J. P. (2014). An empirical study of the Mexican banking system's network and its implications for systemic risk. *Journal of Economic Dynamics & Control*, 40:242–265.

Patterson, B. D. and Atmar, W. (1986). Nested subsets and the structure of insular mammalian faunas and archipelagos. *Biological journal of the Linnean society*, 28(1-2):65–82.

Payrató-Borras, C., Hernández, L., and Moreno, Y. (2019). Breaking the spell of nestedness: The entropic origin of nestedness in mutualistic systems. *Physical Review X*, 9(3):031024.

Payrató-Borràs, C., Hernández, L., and Moreno, Y. (2020). Measuring nestedness: A comparative study of the performance of different metrics. *Ecology and evolution*, 10(21):11906–11921.

Saavedra, S., Reed-Tsochas, F., and Uzzi, B. (2009). A simple model of bipartite cooperation for ecological and organizational networks. *Nature*, 457(7228):463–466.

Saavedra, S., Stouffer, D. B., Uzzi, B., and Bascompte, J. (2011). Strong contributors to network persistence are the most vulnerable to extinction. *Nature*, 478(7368):233–235.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.

Shoham, Y. and Leyton-Brown, K. (2008). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.

Staniczenko, P. P., Kopp, J. C., and Allesina, S. (2013). The ghost of nestedness in ecological networks. *Nature communications*, 4(1):1–6.

Tacchella, A., Cristelli, M., Caldarelli, G., Gabrielli, A., and Pietronero, L. (2012). A new metrics for countries' fitness and products' complexity. *Scientific reports*, 2(1):1–7.