Cluster and network analysis for business models in the Mexican banking system*

Luis O. L. Escobar-Farfan,†1, Serafin Martinez-Jaramillo1,2, Manuel Ramos Francia2, and Santiago Garcia-Verdu1,2

1Banco de México
2Center for Latin American Monetary Studies

Abstract

The recent crisis showed that it is necessary for regulators to understand the roles that individual banks play in the financial system; this can be done in many different ways depending on the information at hand. In this work we undertake two possible forms for identifying the role played by a bank in the financial system: cluster and network analysis. Our first goal is to use balance-sheet data in order to identify clusters which can be associated with business models. We employ two clustering techniques suited for data with high dimensionality to identify the clusters and we observe their time evolution. We find four business models which are stable and we are identify the most relevant variables to build the clusters by using random forests. The second aim of this work is to study the role played by the same banks in the banking network, we use Stochastic Block Models for this. The network used to study interconnectedness in the banking system is an exposures network which has been used for previous studies. We find that the network can be described by three different groups of banks with similar connectivity patterns. The first cluster is core1 of the network in which the members are highly connected among them; in the second cluster, core2, the members have a medium level of connectivity and have many connections with core1; in the third group, the periphery, in the members have the lowest level of connectivity and the group

*The views expressed here are those of the authors and do not reflect the views of the Mexican central bank.
†Corresponding author: lescobarf@banxico.org.mx
is weakly connected to the other two clusters; we also study the evolution of the clusters. This is the first study in which both aspects of bank behavior (business models and interconnectedness) have been studied at the same time for such a long time span.

1 Introduction

The financial system and in particular banking systems have become more complex in recent times. The evolution of banks, the arrival of new players and the emergence/transformation of (new) markets have driven this change. These structural changes might have left possible blind spots, with systemic implications not covered by current regulation.

One of the most commonly mentioned causes of the global spread of the 2008 crisis, which started in the US, is the high interconnectedness of the financial system Martinez-Jaramillo et al. (2019). It was then when there was a surge in research related to financial contagion, financial networks and systemic risk metrics. Moreover, after the crisis the term “Too Interconnected To Fail” was conceived.

Macroprudential regulation has been the answer to deactivate future financial crises, at least the ones which are similar to the ones observed in the past. Nevertheless, we still don’t know where the next financial crisis is going to come from and for that reason the financial system is now under constant monitoring.

The global regulatory efforts to strengthen the resilience of the financial system materialized in the famous Basel III framework and the Dodd-Frank Act. Both regulatory efforts developed a good number of new regulation or performed important changes to the existing rules. An important part of such regulatory efforts involved increasing the level of capital or improving its quality. However, there were also initiatives which aimed to reduce systemic risk by changing the structure of the financial network like the reforms on the OTC derivatives Heath et al. (2016) and the regulation on large exposures Batiz-Zuk et al. (2016).

In this context, the concept Globally Systemically Important Bank (G-SIB) was created. The methodology for the determination of the G-SIB status is given in BIS (2017).

The high level of interconnectedness in the banking system is related to the fact that banks interact in many different ways (lending, buying other banks’ securities, engaging in foreign exchange transactions, etc), markets (repurchase agreements,

\footnote{The 2017 list of GSIBs can be consulted in: https://www.fsb.org/wp-content/uploads/P211117-1.pdf}
derivatives among others). These interactions increase the complexity and make hard to evaluate the possible impact that shocks will have in the system. Fortunately, new systemic risk models and network studies have been extremely helpful assisting the authorities to preserve the stability of the financial system.

Contagion models were developed and used in the last decade. Battiston and Martinez-Jaramillo (2018) provides a good account on the network studies on financial contagion. These models rely on a very important input: the interbank network of exposures. Immediately after the crisis only a few jurisdictions had the necessary data to build such network and for that reason, some methods were developed to construct the matrix of interbank exposures from balance sheet data. Anand et al. (2018) summarizes well some of these methods and performs a comparison study using data from many different jurisdictions and types of interbank networks.

In Mexico, after the 1994 Tequila crisis, several government authorities agreed on a agreement to share the data collected from banks and asked a considerable amount of low granularity data from banks. This is why the central bank possess good quality interbank exposures data for a long period of time on a daily basis. The data used for the clustering study come from detailed balance sheet data that banks report to the Mexican supervisor.

In addition to study the roles played by the banks in the banking network, we are also interested in identifying the existence of business models. For this purpose we resorted to machine learning techniques which have recently gained acceptance among central bankers beyond the statistical and econometric models. Machine learning techniques have obtained such prominence in economics and finance due to some of its characteristics like their pattern recognition capacity, object classification and function approximation. These improved capacities in machine learning and the availability of balance sheet data encouraged us to investigate the application of these techniques for business models identification.

The main goal of the present work is to identify homogeneous groups of banks in terms of business models and in terms of connectivity. We resort to clustering analysis for the former and to network theory for the later. Along this line of thought, two important question arise: i) why is important to identify groups of banks with similar business models? and ii) why is important to find groups of banks with similar connectivity patterns?

In general, more homogeneous banking systems, in business models as well as in connectivity terms, are perceived as more fragile León (2017): as in the presence of a shock, banks operating in a similar manner and with similar connectivity patterns could propagate rapidly such an initial shock and this, in turn, could have systemic implications. A situation in which banks with similar business models also
have similar connectivity patterns is a situation which deserves careful analysis and monitoring; fortunately, we have the data and tools in order to do so.

The main contributions of this work are: i) this is the first work, to the best of our knowledge, in which artificial intelligence and network techniques are used to study the banking system homogeneity in terms of business models and interconnected-ness for a long period of time; ii) identify which variables are the most relevant to determine the clusters by resorting to another machine learning technique (random forest).

Artificial intelligence, machine learning and network analysis are changing our world in ways and at a speed unseen before. Therefore, it is important to understand deeply these tools because the financial system already uses them and, as regulators, it is relevant to assess adequately the possible implications in the financial system of a widespread use of these novel techniques.

The rest of the paper is organized as follows: Section 2 describe what do we mean in this paper by business models and also explains the clustering methods used in this work; whereas Section 4 introduces the Financial Networks field and the Stochastic Block Models (SBMs). Section 5 presents the most relevant results of the application of the clustering methods and the SBMs to the Mexican banking system; finally, Section 6 concludes.

2 Business models and clustering methods

Before the Global Financial Crisis (GFC), regulation followed the microprudential approach; however, after the default of Lehman Brothers, it was acknowledged that a macroprudential approach was more adequate for the modern financial system. Regulation has changed since then and the so called Basel III framework developed at the BIS and the Dodd-Frank Act in the US were the most important reactions to the crisis. Even now, banks are categorized by using the regulatory capital ratio in which the numerator is the Tier 1 capital and the denominator are the risk-weighted assets. However, information on the liability side of banks is not commonly used. In this section we will explain what do we refer by business models and will also introduce the clustering methods used in this work.

2.1 Business models

In [Blundell-Wignall et al. (2014)], the authors mention that the first regulatory effort to control business models is the Glass-Steagal law, among its more important features was the separation between traditional banking and investment banking in
order to avoid a similar situation to the 1929 crisis. Nevertheless, the Glass-Steagall law was abolished at the end of the XX century.

A business model is defined as the activities performed by companies, banks, institutions or other types of organizations with the purpose of earning profits from such activities. A business model should define actions and means in order to achieve a profitable organization: potential clients, growth goals, operational plans and so on and so forth. A good business model is key for the success of the organization as along with the business models, procedures and actions are defined to supervise the progress towards achieving the original objectives.

In the specific case of a financial institution a business model is translated into the activities that such institution performs in order to obtain funding, provide liquidity to the system and maximize its profits among many other. Nevertheless, the business model comprises more than just the performed activities, it also involves the intensity, its geographical reach and the target market of such activities. For example, there are important differences among banks’ business models: banks specialized in consumer credit, investment banks, regional banks, niche banks, etc.

The question on why is important to study and monitor banks business models comes again. The answer is that banks engage in risky activities in order to keep up with competition. This, in turn, could endanger not only their individual stability but the whole system’s soundness.

After the GFC, in addition to the well known term Too-Big-To-Fail (TBTF), the new term Too-Interconnected-To-Fail (TITF) was incorporated into the financial authorities jargon. These terms are both associated with moral hazard risk Rowell and Connelly (2012). How are these concepts related to business models? The answer lies in the fact that if institutions which fall under these categories have also similar business models and connectivity patterns, then the banking system is perceived as being fragile.

For example, in Lucas et al. (2017) and Farnè and Vouldis (2017) the authors identify different business models in the European banking system by using clustering methods. In banks Lucas et al. (2017) they find that banks which are TBTF belong to the same cluster. In León (2017) the authors examine the homogeneity in the Colombian banking system using a similar data set but they resort to the agglomerative clustering algorithm whereas we use a particular version of the K-means algorithm, MacQueen (1967).

This work borrow heavily from Farnè and Vouldis (2017) which is similar to ours in the aspect that our study also falls in the category of fat-data because we have more variables than objects (banks) which are subject to the clustering algorithm. Additionally, Farnè and Vouldis (2017) also use a derived method from
the K-means algorithm and use similar regulatory data. In their work the authors find four different groups from 365 banks: wholesale funding, traditional commercial banks, traditional complex banks, securities focused banks and a group of banks with idiosyncratic business models. The authors find that the different groups show different distributions for the risk/return indicators:

- Banks focused on securities have the highest return, capital reserves higher and their portfolios are relatively risky.
- Banks relying in wholesale funding also possess relatively risky portfolios but with high return relative to capital.
- Traditional commercial banks have the less risky portfolios with higher returns than the commercial complex banks.
- Commercial complex banks have the worst risk/return combination.

The authors in Farné and Vouldis (2017) also indicate that a previous selection of the variables used for the clustering method could bias the results. In other terms, the authors opt for an agnostic view and “let the data speak for themselves”. We proceed in the same form in this paper and we do not perform any selection of the variables used for the classification.

In a related work, Lucas et al. (2017) use a complex clustering method based on distributions mixtures to identify groups of banks in Europe, their approach takes into consideration the temporal dimension for the elaboration of the clustering model. The groups are made with a selection of variables by the authors, the time component is incorporated in order to describe the actions taken by the banks after the crisis where the regulator moved to a regime of extremely low interest rates. The authors find six groups out of 208 banks: big international banks, wholesale banks, investment banks, small retail diversified, small local retail and cooperative/mutualist banks.

- The international banks are the largest in asset size, 60% of its profits come from interest paying assets, these banks are the most leveraged ones and posses large derivatives and securities portfolios with important cross-border activity
- The wholesale banks are the second largest and lend to corporations, they trade derivatives on behalf of their clients and their retail funding is small or null.
- Investment banks are the third in size, their profits come from fees on financial trading; half their assets are loans to corporations.
• Local retail banks lend equally to corporations and retail customers, they are less leveraged and well capitalized; these banks have well diversified funding from corporations and retail deposits. They lend to domestic as well as foreign clients.

• Cooperative/mutualist banks are the smallest in the sample but also the most numerous together with the local retail banks.

The Mexican banking system is much smaller in terms of assets and also regarding the number of institutions; therefore, we do not expect many groups as it was found in Lucas et al. (2017). Moreover, our work is much closer to Farnè and Vouldis (2017) in terms of the selection of variables and on the clustering method employed.

2.2 Clustering methods

Once we have made the case for the relevance of business models identification, we can now proceed to explain the methodology that we will use. In this paper we resort to machine learning techniques first to identify the groups associated with business models, and second to disentangle which are the most important variables driving the identification of these groups. For the identification of groups we use two clustering methods paired with a dimensional reduction procedure De Soete and Carroll (1994), Vichi and Kiers (2001). For the identification of the most important variables we use a population based version of decision trees known as random forest Breiman (2001).

In machine learning, supervised and unsupervised learning differ on the input used to develop the model. In the case of supervised learning, the input data used to train the model must include the output variable. Whereas in the case of unsupervised learning, this is not the case. The output variable is the variable whose behaviour is modeled and which will be forecasted with the information provided by all the other variables. The identification of business models using machine learning techniques is relatively new. In this line of research we find the works, which we have already mentioned: Lucas et al. (2017), Farnè and Vouldis (2017) and León (2017).

As it was mentioned above, the clustering method used to identify the business models in this work is the well-known k-means algorithm De Soete and Carroll (1994). However, taking into account some relevant issues of the k-means algorithm, we used the factorial k-means method Vichi and Kiers (2001), which was used in Farnè and Vouldis (2017) for the European banking system.
2.2.1 Data

The data used for this study was low granularity balance sheet information from a regulatory report known as the “Catálogo Mínimo.”\footnote{This regulatory report can be loosely translated as “Minimum catalog” which contains balance sheet private information from more than 1,000 accounts collected by the Mexican supervisory agency. The public version of this report can be found at: https://datos.gob.mx/busca/dataset/serie-r01-catalogo-minimo} The time span of the data for this study goes from 2007 to 2017. The generated models are static but we explore the time evolution of the clusters resorting to alluvial diagrams in order to explore the stability and changes in the composition of the groups.

The algorithm used is implemented in the R computer language and comes with the “clustrd” library, \cite{Markos:2018}. The program accepts as input a $n \times p$ matrix, with $n$ being the number of banks and $p$ the number of variables. After eliminating variables which have value 0 for all the institutions, we ended up with around 400 variables. The number of banks also changed from 32 in March 2007 to 48 in June 2017. In a similar way to \cite{Farné:2017} an important pre-processing procedure was performed which consisted in eliminating the variables that are highly correlated. The threshold pairwise correlation value used for eliminating variables was 0.75.

The number of accounts (variables) which are eliminated due to high correlation varies considerably across time, unlike the accounts eliminated due to present a 0 value for all banks. The sporadic cases in which accounts which frequently present 0 value, have a non-zero amount, cause the creation of additional groups in which the members are those banks which have a positive value in such accounts.

In order to explore the optimal level of granularity, we executed the reduced k-means and the factorial k-means on three data sets with different levels of granularity. The key to identifying the level of granularity in the accounts table is in the numeric unique identifier for each account. Each account in the accounts master table has assigned a 12 digits account number. The first digit identifies broad concepts (1=assets, 2=liabilities and 4=capital); the next two digits define a new aggregation level, the next two another level and so on and so forth.

We decided not to include other types of variables\footnote{Like performance related variables} as such variables reflect the effectiveness of the decisions, expressed in the balance sheet accounts, on the business models followed by banks. Additionally, there is a huge discrepancy in the values taken by these metrics among banks. It is worth mentioning again that we followed an agnostic approach for the identification of groups, allowing the data to speak for themselves.
2.2.2 Dimensionality reduction

The clustering algorithm used is designed to work with “fat data”, this means that the number of variables is much larger than the number of objects for which the variables are measured. The input data for the clustering method is a matrix $\mathbf{A} \in \mathcal{M}_{n \times p}$, where $\mathcal{M}_{n \times p}$ is the set of matrices of size $n \times p$, where $n$ represents the number of observations (in our case the number of banks) and $p$ represents the number of variables (balance sheet accounts number) for each bank. Our data have the characteristics for being “fat data” as $p >> n$.

When using a clustering algorithm to high dimensional data a common problem is the visualization of results. A visualization in a space of more than three dimensions is beyond representation; nevertheless, this visualization is useful in order to decide on the best specification of the model, without being an absolute criteria to select a partition of the data. Given this problem a pre-processing of our data is necessary. There are different methods that help to reduce the complexity of a data set of a dimensional space beyond representations that humans can visualize. Among these techniques we can find the principal components analysis which is embedded in the two algorithms that we use to obtain the banks’ groups.

2.2.3 Clustering algorithms

The clustering methods are quite useful as other exploratory methods like scatter plots. These methods belong to the unsupervised learning group within the machine learning literature. Unlike supervised learning, there is no response variable in the input data set and the objective is to get one from the application of these techniques. At the end, the idea is to find groups of observations with similar characteristics. However, the creation of these groups depend critically on the definition of similarity that is used. In this work the similarity among observations is measured with distance metrics. As it is not possible to have graphical representations of the data, it is through principal components analysis and clustering methods that we will be able to generate useful visualizations.

There are several distance measures which are commonly used in machine learning, see Johnson and Wichern (2007) for an introduction. Some of the commonly used measures in machine learning are the euclidean distance, the statistical distance and the Minkowski distance. In this work the similarity among observations will be measured by the euclidean distance, the closer in terms of euclidean distance the banks are, the more similar they will be.

In principle, it could be possible to verify all the possible combinations of groups. However, this is equivalent to a general class of combinatorial optimization problems,
which are NP-hard. Fortunately, there are many heuristics that can be used to find acceptable solutions in polynomial time. Furthermore, some meta-heuristics like genetic algorithms and other population based techniques have been used successfully for these problems.

Several clustering algorithms have been developed in the past like the well-known K-Means algorithm developed by MacQueen (1967). These classic exploratory methods are by no means new but recent computational progress have made them more efficient and this in turn have opened the possibility of using them in larger and more complex data sets.

The clustering methods can be categorized into hierarchical and non-hierarchical grouping methods:

- Hierarchical grouping methods are based on a finite succession of mergers/splits into different groups; there are two ways to proceed within the hierarchical methods, agglomerative and divisive methods. In the agglomerative methods, each observation belongs to its own group and then pairs of clusters are merged based on similarity; whereas in the divisive methods, all the observations belong to one single group and successive splits are performed also based on distance.

- The non-hierarchical methods generate a classification given an initial grouping with no hierarchical relationship among them. There are three main different non-hierarchical methods, namely, single pass methods, re-allocation methods and nearest neighbour methods. The single pass methods are the simplest and start with set of observations as the clusters seeds, then each observation is assigned to the cluster that maximizes its similarity; the most important problem with these methods is that the clusters depend on the order in which the data set is processed. The re-allocation methods start with an initial definition of k points which will act as centroids of each cluster, then some observations are reallocated to clusters if an “improvement” is made. In the nearest neighbour methods the merging of groups is made according to the distance of their nearest neighbours.

In general, non-hierarchical clustering methods are less expensive in computational terms but cannot guarantee to reach the global optimum. Moreover, given

---

4NP-hard is a computational complexity class that contains problems which are as hard to solve as NP problems. The NP computational complexity class contains problems which are solved in non-deterministic polynomial time, this basically means that there is no polynomial time algorithm known to solve these problems.
that non-hierarchical clustering is less expensive, it is being currently used for new large data sets and some of its weaknesses are being tackled with recent research.

The methods that we use in this work are two non-hierarchical, reallocation clustering methods: Factorial K-Means and Reduced K-Means. In the next session we will show the results obtained by applying these two methods to balance sheet data for Mexican banks.

2.2.4 Factorial and Reduced K-Means

The K-means clustering method, MacQueen (1967), aims to separate a set into K different groups maximizing the separation among groups while minimizing the distance within the cluster in relation to the cluster’s centroid. An important disadvantage of the K-Means method in comparison to the hierarchical methods is that the number of groups must be specified beforehand.

The optimality criteria for the K-Means algorithm is the (within-cluster variation), defined as $W(C_k)$ for cluster $C_k$ and represents how different between themselves are the objects within a cluster; the “similarity” metric in our context is the euclidean distance; as a result, $W(C_k)$ represents how distant are the objects that belong to the same group among them. Therefore, the K-Means algorithm consists in solving:

$$\minimize_{C_1, \ldots, C_k} \sum_{k=1}^{K} W(C_k),$$

that is, to find the partition $C_1, \ldots, C_k$ that has the smallest intra-cluster variation for all the groups.

As it was mentioned before, the similarity metric frequently used is the squared euclidean distance, in this way $W(C_k)$ can be defined as:

$$W(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2,$$

where $|C_k|$ is the number of observations in cluster $k$, then, intra-cluster variation is the sum of the distance between each pair of points in the cluster, normalized with the total number of observations in this group. Therefore, the K-Means problem can be reduced to solve:

$$\minimize_{C_1, \ldots, C_k} \sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2.$$
The K-Means algorithm cost function is defined as:

\[ J = \sum_{k=1}^{K} \sum_{i=1}^{N} \| x^k_i - c_k \|^2, \]  

(4)

where \( x^k_i \) is a point \( i \) assigned to group \( k \) and \( c_k \) is the centroid for group \( k \). \( J \) represents the distances among the objects of each of the \( K \) groups.

Nevertheless, solving the above equation for all the possible partitions of size \( K \) is computationally expensive. However, there are algorithms which can find local minimums:

1. Assign the \( n \) points randomly to any of the \( K \) groups.

2. Iteration:
   
   (a) For each cluster the centroid is computed, this is a \( p \)-dimensional vector where each entry is the average of the points in this group.
   
   (b) Each point is assigned to the cluster with the closest centroid.

   This is repeated until the assignments do not change the groups.

It is worth mentioning that given that the algorithm only guarantees convergence to a local minimum, results can change depending on the initial grouping, for this reason is recommendable to execute the algorithm several times with different initial groupings.

The two methods that we use for this study are designed with the K-Means algorithm as a basis; nevertheless, in addition to obtain the groups these methods also reduce the dimensionality of the input data. This is extremely useful for problems like ours in which \( p \gg n \). Both methods belong to the clustering approach in sub-spaces of the variables and assume that the groups centroids are located in a subspace of the \( p \) variables.

The reduced K-Means was first introduced in [De Soete and Carroll] (1994); whereas, the factorial K-Means is presented in [Vichi and Kiers] (2001). In [Timmerman et al.] (2010), the authors compare the two methods in order to understand under which circumstances one method is better than the other. We refer the reader to [Timmerman et al.] (2010) to investigate more on the differences between both methods.
3 Decision Trees and Random Forests

As we mentioned in the previous sections, clustering algorithms belong to the unsupervised learning class of machine learning techniques. Unlike unsupervised learning, in supervised learning there is a variable which will be predicted or approximated from a set of independent variables. A common practice in machine learning is to split the data into a training set and a testing set. Typically, the training set is larger than the testing set. The testing set is used to validate how good the model obtained from the training data.

Supervised learning methods can be very flexible and sophisticated but this could work against the model interpretability. Moreover, flexible models can deliver very good performance with the training data but at the same time perform poorly with the testing data; this phenomenon is the well-known problem of overfitting.

Having as depending variable the group that each bank belongs to and the balance sheet accounts as independent variables, we use Random Forest [Breiman (2001)], to identify the independent variables, and their importance, that participate in the classification of the institutions.

Decision trees [Breiman et al.] (1984) are the basic block for random forest [Breiman (2001)]. Decision trees can be used for prediction as well as for classification, the former are known as regression trees and the later as classification trees. In the rest of the session we will borrow the techniques’ description and notation from Hastie et al. (2009) and James et al. (2014).

3.1 Regression Trees

Let’s start with a matrix $A_{n \times p}$ and a vector $y$, where each entry of vector $y$ is the dependent variable of each vector $X_{1 \times p}$. The goal of the tree is to split the space generated by the predictor variables $X_1, X_2, \ldots, X_p$ in $J$ exclusive regions $R_1, R_2, \ldots, R_J$. For all the observations that fall into one of the $J$ regions, the same prediction value is assigned: the average of the response variable $y$ of the objects in each region.

The best partition of the predictor variables space is the one that minimizes the residual sum of squares (RSS) given by:

$$\sum_{j=1}^{J} \sum_{i \in R_j} (y_i - \hat{y}_{R_j})^2,$$

where $\hat{y}_{R_j}$ is the average of the response variable in region $j$.

Nevertheless, as the number of independent variables ($p$) grows, the feasibility of computing all the possible combinations of regions decreases; additionally, our data
has high dimensionality (around 400 variables).

As a result it is necessary to resort to a different alternative: the recursive binary splitting. This algorithm follows a top down approach because at the top of the tree, all the observations belong to a single region and in the following steps splits the predictor space in two branches further down. It is also greedy in the sense that tries to find the optimum at each step, in our case finding the partition that produces the largest drop in RSS.

On each step the variable $X_j, j \in \{1, \ldots, p\}$ and a cutpoint $s$ that reduces the most the RSS. In this form, the predictors’ space is split into the regions:

$$R_1(j, s) = \{X|X_j < s\} \ y \ R_2(j, s) = \{X|X \geq s\}$$

(6)

the values for $j$ and $s$ are selected such that minimize:

$$\sum_{i:x_i \in R_1(j, s)} (y_i - \hat{y}_{R_1})^2 + \sum_{i:x_i \in R_2(j, s)} (y_i - \hat{y}_{R_2})^2,$$

(7)

where $\hat{y}_{R_i}$ is the average of the observations that belong to region $R_i$.

This procedure is performed until there are no more predictor variables or until a stop criteria is met. On each step only the resulting branches are divided and not the entire space. Finally, the average of the $y$ variable of the training observations in the same region is assigned to each new observation.

### 3.2 Classification Trees

Under the same assumptions on the data structure it is possible to use decision trees with the goal of classifying objects according to its characteristics (variables $X_1 \ldots X_p$); the main difference is that the response variable is discrete ans represents a group of observations.

For the elaboration of a classification tree the same recursive binary splitting is used. Then, when the predictions are made for the new variables with the adjusted model, the most common class in the region is assigned to each new object. Nevertheless, given that the variable that is being predicted is not continuous, it is not possible to use the RSS as the criteria for the optimization. Now, there are two options to build the tree: the GINI index (Equation 8) and the entropy (Equation 9).

Taking $\hat{p}_{mk}$ as the proportion of observations in region $m$ (branches of the tree) that belong to class $k$ ($k$ in known before as it is the number of classes in the data), the GINI index is defined as:
\[ G = \sum_{k=1}^{K} \hat{p}_{mk}(1 - \hat{p}_{mk}). \] (8)

The GINI index is also known as the node purity, a small value of the index means that the majority of observations belong to the same class.

Entropy is defined as:

\[ D = -\sum_{k=1}^{K} \hat{p}_{mk} \log(\hat{p}_{mk}). \] (9)

As in the case of the GINI index, if entropy is small, the majority of observations are from the same class.

The tree building process is the same that for the regression trees with the only difference that on each step the selected partition is the one that presents the minimum purity value.

### 3.3 Random Forests

Decision trees present some problems associated with their own nature, linked to the Bias-Variance Trade-Off (Geman et al. (1992)). Given that the partitions of the dependent variable space are non-linear under decision trees, and that it is possible to consider as many as are wished; the resulting models could be very complex and present overfitting or bias.

To overcome these problems, the random forests were developed from the resampling methods like the Bootstrap aggregation or Bagging. It is well-known that the variance of the average of \( n \) observations \( Z_1, \ldots, Z_n \), each one with variance \( \sigma^2 \) is given by \( \frac{\sigma^2}{n} \), that is, averaging the variance of the \( n \) observations reduces the variance.

In the same way, it is possible to build many training sets as needed, and build a model with each set and average the results from all of them. This is what random forests do. If there are \( B \) different training sets, each one built using bootstrap, it is possible to get \( \hat{f}^1(x), \hat{f}^2(x), \ldots, \hat{f}^B(x) \) fittings, and with them the following model is obtained:

\[ \hat{f}_{bag}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{sb}(x), \] (10)

having a lower variance than the individual variance of the \( B \) fitted models.
Bootstrap aggregation consists of building $B$ training sets using bootstrap, from which $B$ decision trees are built that will be averaged to obtain the model $\hat{f}_{\text{bag}}$. It is important to mention that each of the $B$ decision trees are built taking all the predictor variables and are not pruned.[5]

In the case of a classification problem an observation is classified in the category that has majority in the $B$ trees. One convenient characteristic from decision trees is that to increase parameter $B$ does not lead to overfitting, as the error converges as $B$ grows.

Once we have the models it is necessary to estimate how good the model is, for this purpose there is a result which shows that each tree uses two thirds of the $n$ observations, the remaining third is known as the observations Out Of Bag (OOB) and are used to compute the average error. Given that there are $B$ trees and each one uses around two thirds of the observations, there are $\frac{B}{3}$ predictions for each observation, which can be averaged (for regressions) or count and get the category with the majority (for classifications), keeping only one prediction for each observation; these are called OOB predictions.

When a bootstrap aggregation is used to build an average model, the available $p$ predictive variables are used; these variables can have difference importance for the classification of the observations and all the trees have a bias towards the most important. As a consequence, all the trees will be very similar among each other and this, in turn, can lead to highly correlated predictions. For this reason, the Random Forests were created.

In random forests the decision trees do not have the same set of prediction variables as for a training set obtained by using bootstrap, $m$ independent variables are selected at random, where $m < p$. This is done to guarantee that the trees will be different and the predictions will have lower correlation. It is common to find in the literature $m \approx \sqrt{p}$.

An important characteristic of random trees is that it is possible to obtain the relevance of the independent variables for the prediction of the dependent variable. On each of the $B$ trees of the forest there exists the possibility of obtaining the improved performance in relation with the reduction of the RSS or the GINI index. This value is averaged for the $B$ trees and the importance is obtained. A larger value for means greater importance for the dependent variable.

---

[5]Decision trees are pruned in order to reduce complexity, in order to do so some branches are removed while preserving its performance. [James et al. (2014)].
4 Financial Networks and Stochastic Block Models

Banks and financial institutions interact in many different ways and in many different markets simultaneously. Given the current surge in social networks and the well-acknowledged fact of the large degree of interconnectedness in the financial system [Martinez-Jaramillo et al. (2019), it seems now strange that it was not like that before the recent global crisis. Moreover it is now well accepted that microprudential regulation is not enough in order to guarantee financial stability.

It is now common to find many studies from central banks, financial authorities or academics that study one or several aspects of the structural properties of their respective financial system and markets [Battiston and Martinez-Jaramillo (2018), Anand et al. (2018) and Fricke and Lux (2015)].

Network models in finance are now well established and it seems that it will be still the same in the near future. In particular, these models are used to measure systemic risk, solvency and funding contagion and also in stress testing. More recently, network models have been also used to understand and model interactions in many different markets and activities [Poledna et al. (2015), Molina-Borboa et al. (2015), de la Concha et al. (2017) and Carmona and Martinez-Jaramillo (2019)].

Figure 1 shows two visualizations of the Mexican banking network. The color and size of nodes is related to its number of connections and the thickness of the links is related to the amount of the exposure. From a simple view of the plots, it is clear that the banking exposures network have had important structural changes. In the context of systemic risk, it is important to know the most relevant institutions taking into account additional aspects to size alone. One important feature of financial institutions is interconnectedness and network models are the most adequate approach to measure it.

4.1 Data

The data for the network study in this paper comes from a database at Banco de México which is used for contagion and systemic risk studies. The database comprises information from several regulatory reports. This dataset contains the daily current exposures among banks from 2005 on wards and from 2011 contains the daily current exposures among financial intermediaries like brokerage houses, investment funds, pension funds and foreign financial intermediaries. This database has been used extensively for monitoring purposes and also in many published studies including [Canedo and Jaramillo (2009), Martinez-Jaramillo et al. (2010), Martinez-]
Figure 1: Interbank market network, each circle represents a Mexican bank, whereas links represent exposures.

The current exposures in such a dataset used in this paper come from the following bank activities:

- Interbank call money loans
- Cross holding of securities
- Derivatives trading
- Foreign exchange transactions

The reader is referred to Martinez-Jaramillo et al. (2014) and Poledna et al. (2015) for a detailed exposition of the exposures database.

4.2 Financial Networks

Nowadays there are many good texts on network theory, we refer the reader to Barabási and Pósfai (2016) and Newman (2010) for excellent introductions to the area. We also borrow some of the descriptions of the main concepts and notation from them and from Martinez-Jaramillo et al. (2014).
Undirected graph

We define an undirected graph as the pair:

\[ G = [V, E], \]  \hspace{1cm} (11)

where \( V \) is a non-empty finite set, and \( E \) is the set of unordered pairs \( E = \{(i, j) | i, j \in V\} \).

The elements of \( V \) are known as vertices, and the elements of \( E \) are known as edges that represent the relationships in the graph. That is, a network \( G \) is the set of vertex \( V \) connected by the edges in \( E \).

Directed graph

The previous definition can be extended to define a directed graph by adding an order to the pairs \((i, j)\), that is, a directed graph is the set

\[ D = [N, A], \]  \hspace{1cm} (12)

where \( N \neq \emptyset \) is finite, and \( A \subseteq N \times N \).

The elements of \( N \) are known as nodes, and the elements in \( A \) arcs; the relationship \((i, j)\), with \( i, j \in N \), implies the existence of an arc with direction from \( i \) to \( j \), that does not imply that the relationship \( j \) to \( i \) exists.

Undirected network

The above definitions are extended to define our principal object of study: financial networks. The terms graph and networks are almost identical conceptually; however, in this work we will use the term network from now on to refer to directed and undirected graphs with an additional important characteristic: arcs have an assigned value called weight. In mathematical terms weights are represented by a function \( w: E \to \mathbb{R} \), that is, the value \( w(i, j) = w_{ij} \) is the weight of the connection between vertex \( i \) and \( j \).

We define an undirected network as:

\[ R = [V, E, w], \]  \hspace{1cm} (13)

where \([V, E]\) is an undirected graph as in (11)
Directed network

In a similar way, we define an undirected network as:

\[ R = [N, A, w], \]  

where \([N, A]\) is a directed graph as in [12]

Adjacency matrix

There are two main representations for a network. The first one is the adjacency list, which is an ordered list of the form \((i, j)\), for example, the list of pairs \((1, 2), (2, 3), (3, 4), (4, 1)\) define a network. However, the most common form to represent a network is the adjacency matrix \(A\). The matrix representation is quite convenient as many of the metrics for describing these structures are described and implemented by using matrix algebra. For an undirected graph, the adjacency matrix is defined as:

\[
A_{ij} = \begin{cases} 
1 & \text{if } (i, j) \in E \text{ or } (j, i) \in E, \\
0 & \text{otherwise}
\end{cases}
\]  

Given that for a directed graph, arcs have a direction between nodes, the definition of the adjacency matrix can be split in two parts, as there are entering arcs and outgoing arcs.

We define the entries of the outer adjacency matrix as:

\[
a^+_{ij} = \begin{cases} 
1 & \text{if } (i, j) \in E, \\
0 & \text{otherwise}
\end{cases}
\]  

In a similar fashion, the entries of the inner adjacency matrix are defined as:

\[
a^-_{ij} = \begin{cases} 
1 & \text{if } (j, i) \in E, \\
0 & \text{otherwise}
\end{cases}
\]  

In general, we have that \(A^+ = (A^-)^T\). From now on we will use the terms vertex and node interchangeably as well as the terms link, arc and edge.

Neighbors

From the adjacency matrix, we define the neighbors of node \(i\) in an undirected graph \(G = [V, E]\) as the nodes for which an edge connecting them exists in \(E\). Therefore, the neighbors of node \(i\) are defined as:
In a directed network, the inner neighbors of \( i \) are those that have an arc directed to \( i \), that is, the set
\[
N^{-}(i) = \{ j \in V : a_{ij} = 1 \}.
\] (19)

Similarly, the outer neighbors of \( i \) are those that have an arc departing from \( i \):
\[
N^{+}(i) = \{ j \in V : a_{ij}^{+} = 1 \}.
\] (20)

**Weighted matrix**

Now, we will define the weighted matrices for directed networks. First, we define the outer weighted matrix \( W^{+} \) for a directed network \( R = [N, A, w] \) as:
\[
W^{+}_{ij} = \begin{cases} 
  w^{+}_{ij} & \text{if } (i, j) \in A, \\
  0 & \text{otherwise}
\end{cases}
\] (21)

Furthermore, the inner weighted matrix \( W^{-} \) is defined as:
\[
W^{-}_{ij} = \begin{cases} 
  w^{-}_{ij} & \text{if } (j, i) \in A, \\
  0 & \text{otherwise}
\end{cases}
\] (22)

In general in the present work, the direction of an arc is associated with the flow of money. In the case of the exposures network, the entry \( i,j \) of the matrix \( W^{+} \) represents the amount of money that institution \( i \) is exposed to institution \( j \), that is, the amount of money that \( j \) owes to \( i \). In the case of payment systems networks, the entry \( i,j \) represents the money sent from bank \( i \) to bank \( j \).

Similarly to the adjacency matrices, we have that \( W^{+} = (W^{-})^{T} \). Additionally, we can define matrix \( W = W^{+} + W^{-} \), in which its entries represent the total amount of the activity between a pair of institutions.

### 4.2.1 Network structural metrics

Using the matrices defined above it is possible to compute the main structural metrics that represent a network, such metrics will help to describe the most important characteristics of the financial system, as well as the participation of the individual institutions in such a structure.

According to Silva et al. (2016), it is possible to split the structural metrics into three categories:
• **Strictly local metrics**: Metrics that describe only to the node they represent; these metrics do not take into account information from the neighborhood or the whole network. These metrics are useful to identify important nodes in a network.

• **Cuasi-local metrics**: These are also individual node metrics; however, these metrics do take into account the characteristics of the neighborhood of each node. Cuasi-local metrics are useful to determine the importance of a node, taking into account the importance of its neighbors.

• **Global metrics**: These metrics use information from the whole network and are useful to reveal systemic characteristics of the network.

Detailed descriptions and formal of the structural metrics used in this paper can be found in Martinez-Jaramillo et al. (2013).

### 4.2.2 Network components

Given that one of the goals in the network theory field is to obtain relevant information about complex structures, it is convenient to characterize the network in terms of components with different connectivity patterns and characteristics that allow for a separate study of each component.

The notion of component is associated with the possibility of having some parts of the networks disconnected from the rest. In order to define connectivity there are two important concepts that we must explain before:

• **Walk.** A walk of length \( n \) is a sequence of edges \((e_1, e_2, ..., e_{n-1})\) for which the sequence of \( n \) vertices \((v_1, v_2, ..., v_n)\) obeys the rule \( e_i = (v_i, v_{i+1}) \) for \( i = 1, 2, ..., n-1 \). In other definitions, the vertices (for a path) and edges (for a trail) must be all different; and the sequences can be finite or not.

• **Directed walk.** For a directed graph, a directed walk is a sequence of arcs \((a_1, a_2, ..., a_{n-1})\) for which the sequence of \( n \) nodes \((v_1, v_2, ..., v_n)\) obeys the rule \( a_i = (v_i, v_{i+1}) \) for \( i = 1, 2, ..., n-1 \). An important difference with the previous definition is that the direction of the arc must be taken into account. Again if all the edges are different it is a directed trail and if all the nodes are different, then we have a directed path. The sequences can be finite or not.

A graph is **connected** if there is a path for all possible pairs of vertices in the graph. A directed graph is **strongly connected** if there is a directed path for all
possible pair of vertices in the directed graph. A graph \( G' = [V', E'] \) is a subgraph of graph \( G = [V, E] \) iff \( V' \subseteq V, E' \subseteq E \) and \( (v_i, v_{i+1}) \in E' \implies v_i, v_{i+1} \in V' \). A component of a graph is a subgraph which is connected.

The component concept can be generalized that of a \( k \)-component. A \( k \)-component is a subset of vertices such that each pair of vertices is connected by at least \( k \) vertex-independent paths\(^6\).

### 4.2.3 Centrality

Financial networks adopted the centrality concept from social networks. We resort to some of the metrics already defined and we will also use some other to define centrality. Centrality is an indicator of the importance of a node in the network; in financial networks, these metrics can be useful to detect the institutions which are more important for the stability of the system.

In Martinez-Jaramillo et al. (2014), the authors, citing Henggeler-Müller (2006) consider that a bank is central if:

- Has many counterparts (degree centrality).
- The amount of its participation in the network is large (strength centrality).
- Its failure could transmit contagion in few steps (closeness and DebtRank centrality).
- Its counterparts are also central (PageRank, Eigenvector and DebtRank centrality).
- There are many paths through it (betweenes centrality).

We refer the reader to Martinez-Jaramillo et al. (2014) to see the definition of some centrality metrics and for a study on centrality for two banking networks in México.

### 4.2.4 Network generative models

In the network theory literature there are some structural models which have been identified in diverse systems: biological, transport, social and financial. These structures present different characteristics depending on the form in which its components interact among them and depending also on the processes that take place on them.

---

\(^6\)Two paths are vertex-independent if they only share the initial and final vertices.
The above mentioned structural and centrality metrics are useful to understand some of the behaviour at the local and global level. However, there are some properties which cannot be fully understood and described by these metrics. That is why the study of mesoscales in networks have become more intensive; in particular the study of the community structures in networks have been in the rise. In general, the structure of a network is determined by an unknown generative process; for this reason, an important part of the research on networks, still focuses in developing models which can reproduce network characteristics in the best possible way.

**Random networks**

The random network model serves as a null model to investigate if a network under study presents the properties of a random system or not. Many empirical works have shown that real networks do not follow a random network model. For example, the Internet motivated the development of the Barabási-Albert model, Barabási and Albert (1999).

There are two general models for random networks depending on the generative model:

- **Model $G(N, L)$**: It is assumed that there are $N$ nodes and a fixed number of links $L$. To simulate a random network with the model $G(N, L)$, on each step a pair of nodes is selected randomly and is connected with one of the $L$ available links. Such a procedure is repeated until the $L$ links are used. It was developed and used by Erdős and Rényi in several works, for example: Erdős and Rényi (1959), Erdős and Rényi (1960).

- **Model $G(N, p)$**: In this model, each pair of nodes is connected with a probability $p$. Is is widely used due to the facility to derive analytically many of its properties thanks to the inclusion of the notion of probability. It was introduced in Gilbert (1959).

On the basis of the model $G(N, p)$, random networks have the following general characteristics:

- The degree follows a binomial distribution that under certain circumstances can be approximated by a Poisson distribution; that is, if $p(k)$ is defined as the probability that a randomly selected node has degree $k$, then the degree distribution is:
\[ p(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}, \quad (23) \]

and its Poisson form:

\[ p(k) = \exp\left( -\hat{k} \right) \frac{\hat{k}^k}{k!}. \quad (24) \]

- Given the average degree, many connectivity regimes might arise: if the average degree is \( \hat{k} > 1 \), the network will have a giant connected component; that is, the majority of nodes will be connected among them with a few disconnected nodes. The point \( \hat{k} = 1 \) is known as the critical connectivity point.

- The average distance in the network is given by \( \hat{d} \propto \frac{\ln(N)}{\ln(\hat{k})} \).

- The average clustering metric is computed as \( \hat{C} = \frac{\hat{k}}{N} \).

For the real networks it has been observed that, in general, the degree does not follow a Binomial-Poisson distribution, the connectivity regimes are not necessarily associated with the average degree and the majority has an average degree greater than 1 and the clustering coefficient is much larger than the one of the random networks. Average distance in random networks is a good approximation to average distance in real networks.

**Scale free networks**

Many of the networks in real phenomena do not behave like random networks, one of the most prominent examples is that of the Internet. In Albert et al. (1999), the authors show that the degree distribution of the Internet is far from being a Binomial-Poisson. While examining the degree distribution of the Internet, the authors observed that while in random networks the existence of highly connected nodes is almost impossible, the Internet showed nodes with extremely high in degree and out degree. These highly connected nodes are known as hubs and are present in the majority of real networks. In Albert et al. (1999) the authors showed that the degree distribution of the Internet can be better approximated by:

\[ p_k \sim k^{-\gamma}. \quad (25) \]

\(^7\)In this work, the authors use for their study the network formed by the web pages instead of the physical Internet network.
The above function is known as the Power-Law distribution, also known as the Pareto distribution. From this distribution it is possible to obtain an useful feature to have an initial inspection and determine how similar are a real network and a random network. By applying the logarithm to the previous expression we have:

\[ \ln(p_k) \sim \gamma \ln(k). \tag{26} \]

This simple manipulation makes evident a linear relationship between the degree and the probability with \( \gamma \) representing the slope. As a result, one of the first visual inspections to determine if a network is random or not is to plot its degree distribution in a log-log scale.

The networks that are used to obtain communities by an Stochastic Block Model show a degree distribution more similar to random networks than to scale free networks. Scale free networks are those for which the degree distribution follows a Power Law.

There are two forms to define the Power Law distribution for networks: the discrete and the continuous forms.

**Discrete form**

It is known that a node’s degree is an integer \( k \in \{0, 1, 2, \ldots \} \). From Equation \( \ref{eq:25} \) it is possible to see that the probability that a randomly selected node has degree \( k \) is:

\[ p_k = C k^{-\gamma}, \tag{27} \]

where, given that \( p_k \) is a probability distribution it must satisfy

\[ \sum_{k=1}^{\infty} p_k = 1. \tag{28} \]

Then, we obtain:

\[ \sum_{k=1}^{\infty} p_k = \sum_{k=1}^{\infty} C k^{-\gamma}, \tag{29} \]

\[ 1 = C \sum_{k=1}^{\infty} k^{-\gamma}, \tag{30} \]

finally

\[ C = \frac{1}{\sum_{k=1}^{\infty} k^{-\gamma}}, \tag{31} \]
where $\sum_{k=1}^{\infty} k^{-\gamma} = \zeta(\gamma)$ is Riemman’s zeta function. Therefore, the discrete form of the degree distribution for a scale free network is:

$$p_k = \frac{k^{-\gamma}}{\zeta(\gamma)}.$$  \hspace{1cm} (32)

Given that this function is discontinuous in $k = 0$, if needed, it is possible to specify $p_0$ as the percentage of disconnected nodes from the network.

**Continuous form**

If it is assumed that the degree can be any real positive number, the continuous distribution can be derived by following the same steps of the discrete distribution from the equation:

$$p(k) = Ck^{-\gamma},$$  \hspace{1cm} (33)

taking as condition

$$\int_{k_{\min}}^{\infty} p(k)dk = 1.$$  \hspace{1cm} (34)

Then

$$C = \frac{1}{\int_{k_{\min}}^{\infty} k^{-\gamma}dk}.$$  \hspace{1cm} (35)

In Barabási and Pósfai (2016) there is discussion about the impact on the degree distribution of the networks varying $\gamma$. Nevertheless, in most of the networks studied $2 < \gamma < 3$, then, $-\gamma + 1 < 0$, coming back to Equation 35 we have

$$\int_{k_{\min}}^{\infty} k^{-\gamma}dk = \left[ \frac{-\gamma+1}{-\gamma+1} \right]_{k_{\min}}^{\infty} = 0 - \frac{k_{\min}^{-\gamma+1}}{-\gamma+1} = \frac{k_{\min}^{-\gamma+1}}{\gamma-1}.$$  \hspace{1cm} (36)

Then

$$C = \frac{1}{\frac{k_{\min}^{-\gamma+1}}{\gamma-1}} = \frac{\gamma-1}{k_{\min}^{-\gamma+1}} = (\gamma - 1)k_{\min}^{\gamma-1}.$$  \hspace{1cm} (37)

Therefore, the continuous form of the degree distribution for a scale free network is:

$$p(k) = (\gamma - 1)k_{\min}^{\gamma-1}k^{-\gamma}.$$  \hspace{1cm} (38)
From the previous equation it is noted that it is indeterminate for $k = 0$ on $k^{-\gamma} = \frac{1}{k^\gamma}$, in this case $k_{\text{min}}$ is used instead to determine the normalization constant on $C$.

The continuous has a different interpretation to the discrete case as $p(k)$ is not the probability that a randomly selected node has degree $k$; moreover, computing

$$\int_{k_1}^{k_2} p(k)dk,$$ \hspace{1cm} (39)

we have the probability that a randomly selected node has degree $k$ such that $k_1 \leq k \leq k_2$.

For directed networks, the degree distribution for the inner and out degree can follow a Power Law each.

The main difference between random and scale free networks is the existence of hubs. Figure 2\hspace{1cm}8 shows in the left panel the Poisson distribution vs a Power Law, the right panel shows how the power law allows for high degree nodes whereas for the Poisson there is a pronounced fall after relatively small values for the degree.

This feature has been found in a great variety of real networks suggesting the universality of the scale free property. Another characteristic of the scale free networks is that given the existence of highly connected nodes, these structures are vulnerable to targeted attacks but robust to random failures.

### 4.2.5 Core-Periphery structures

In this section we will discuss a particular hierarchical model applied in complex networks in general but also in financial networks: the core-periphery model. In order to define it, we need the important concept of assortativity Newman (2002). Assortativity can be measured in relation to many node-level variables. One commonly used metric is the degree. It is said that nodes in a network are mixed in an assortative form if nodes with similar values are connected among themselves.

A core-periphery structure is a special case derived from an assortative node degree case in a network. According to Newman (2010) when a network presents degree assortativity, highly connected nodes form a compact substructure known as the “core”. The weakly connected nodes form a less dense component considered as the “periphery” in relation to the core. This model is present also in many different networks like social, international relations and economic networks.

8The plots were produced taking the same parameters to simulate the data needed for the plot from Barabasi and Posfai (2016), page 120.
Figure 2: Comparison between the tails of the degree distribution obtained from simulations of random networks (Poisson) and scale free (Power-Law)

In Craig and von Peter (2014) the authors provide an economic interpretation of the core-periphery structure which has been found very useful for interbank networks. In their work the authors propose a model to separate the nodes in two components: a core and a periphery. They apply their model to an interbank exposures network among German banks. The main findings are that the exposures network behaves very differently to a random network and that there is a group (the core) of banks that perform an intermediation role in such market. In their proposal, the authors propose to compute the deviation from an ideal core-periphery structure by computing the sum of missing links in core and the links among periphery nodes.

Figure 3 shows the error for the Mexican interbank exposures network. This figure shows that there are important deviations from the ideal structure, in particular around the financial crisis. This is interesting as it means that the intermediation role changed for the nodes in the core and that some periphery banks interacted more actively among them. Moreover, the error has an upward trend which shows that the Mexican interbank networks is departing from such structural model.

In van Lidth de Jeude et al. (2018) a different model to determine the core-
The periphery structure with the advantage that as result of its construction provides an statistical significance for the partition found represented by a p-value. We will show the comparison between the two methods, in the meanwhile we present in Table 4.2.5 we present the statistical description of the p-value obtained for the Mexican interbank exposures network for the period of study, the table shows that the significance is always below 0.05 which indicates that the partition is adequate.

The value 100 indicates one case for which the algorithm couldn’t obtain the partition.

Figure 4 shows the comparison for a sample of banks between the two methods, it is possible to see that there are some important differences.

It is important to notice that for some institutions both methods differ considerably; however, for some others both methods identify some banks as part of the core consistently. For example, banks 16, 18 and 19 are considered to be part of the core by both methods for the whole period of study.
### 4.3 Stochastic Block Models

Once covered the identification of groups in relation to its business models, another important aspect is to uncover groups of banks which behave similarly in relation to its interbank relationships. To uncover this structure we resort to one of the many techniques for community detection in networks: Stochastic Block Models (SBMs) Nowicki and Snijders (2001).

One of the main tools used in SBMs is Bayesian Statistics (see Gelman et al. (2014) for an introduction). The second tool that is also used for SBMs is Gibbs sampling Geman and Geman (1984). For space reasons we will not discuss further these tools but the reader can find plenty of introductory works in both topics.

There are many forms to detect communities in networks, the majority are based in a notion of distance. However, as it happens for the clustering methods, there could be important failures to capture the connectivity patterns in complex networks. Many of the community detection methods are subject to overfitting and could also be affected by random variations in the connectivity of a system.

As a result, new methods for community detection were developed to tackle the problem of random variations by using probability theory proposing models that translate the concept of structural equivalence into a probabilistic framework.

All the methods which use the probability to identify communities in a network use as assumption the existence of set of \( n \) nodes, for which is possible to observe the variable \( y_{ij} \) called the relationship from \( i \) to \( j \).

The most successful approach to model the existent relationships in a networks has been to assume that such relationships can be modeled conditioning its probability of existence to a vector \( x \) of node attributes, in this case, the \( x_i \) entry \( i \in \{1, 2, ..., n\} \) represents the community to which node \( i \) belongs to. In this, it is possible to model the relationship

\[
y = (y_{ij})_{1 \leq i \neq j \leq n},
\]  

#### Table: Quantile Value

<table>
<thead>
<tr>
<th>Quantile</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Min.</td>
<td>(2.300000000000000e - 153)</td>
</tr>
<tr>
<td>1st Qu.</td>
<td>(4.715000000000000e - 133)</td>
</tr>
<tr>
<td>Median</td>
<td>(1.245000000000000e - 127)</td>
</tr>
<tr>
<td>Mean</td>
<td>(2.38095238095238e + 00)</td>
</tr>
<tr>
<td>3rd Qu.</td>
<td>(8.243750000000000e - 117)</td>
</tr>
<tr>
<td>Max.</td>
<td>(1.000000000000000e + 02)</td>
</tr>
</tbody>
</table>
Comparison of results for a sample of banks

1 = core, 0 = periphery

Figure 4: Comparison between the Craig and von Peter vs the Surprise method to find the core-periphery structure for a sample of banks in Mexico.

conditional to the vector \( \mathbf{x} = (x_1, \ldots, x_n) \), where \( i \) belongs to the set \( C = \{1, 2, \ldots, c\} \), that contains all the possible groups to which the node can belong to. This is equivalent to think that the probability that two nodes \( i \) and \( j \) form a link in the network depends on the community that they belong to. In the case that the classes in \( C \) are known and it is also known to which class each node belongs to, this approach is known as stochastic block models a priori.

In [Holland et al. (1983)], the authors generalize the theory for directed networks, assuming that \( y_{ij} \) and \( y_{ji} \) are independent and not necessarily equal, conditioning to the classification vector \( \mathbf{x} \). This model is known as the pair dependant block-modeling. In this model it is said that two nodes are stochastic equivalent if their connectivity probability is the same for nodes in their same group and also for nodes in other groups and this is satisfied if they belong to the same group.

In [Snijders and Nowicki (1997)] the authors propose a posterior model for undirected networks that only allows for two classes \( C = 1, 0 \). However, in [Nowicki and Snijders (2001)] the authors generalize the model for directed networks for an arbitrary number of groups. In this work, it is not possible to resort to a prior approach
as we do not know the groups’ structure; therefore, it is needed a posterior approach to find the groups’ structure by using the relationships \( y \), known from the adjacency matrices described in the data section. We will borrow from [Nowicki and Snijders (2001)] the notation to describe the SBM used in this work.

The model is described as follows:

- \( N \), denote the set of all node pairs \( (i, j) \), \( i, j \in \{1, \ldots, n\} \).
- Each node \( i \) belongs to a class \( k \), \( k \in \{1, \ldots, c\} \).
- We only observe the relationship matrix \( Y \), and we want to estimate the groups vector \( X \).

First, the joint probability of the group of each node is defined, these are the result of the random variables \( X_i \). In this way we have:

\[
P(X_i = x_i, X_2 = x_2, \ldots, X_n = x_n) = \theta_{m}^{c}, \quad (41)
\]

were \( m_i \) represents the number of nodes which belong to class \( i \) and \( \theta_i \) represents the probability that a node belongs to class \( i \).

As it has already being mentioned, one of the strong assumptions is that the existence of a connection between two nodes depends on the class they belong to, in this form, given the vector \( X = x \) we have that:

\[
P(Y_{ij} = 1 | X = x) = \eta(x_i, x_j), \quad (42)
\]

where \( \eta(k, h) \) is the probability of connection that depends on the grouping of \( i \) and \( j \) and also:

\[
\sum_{k, h \in \mathcal{C}} \eta(k, h) = 1. \quad (43)
\]

Moreover, the probability \( \eta(k, h) \) can be split in \( \eta(k, k) \) and \( \eta(k, h) \), that represents the probability of connection between two nodes from the same group and between nodes of different groups respectively.

Then, taking into account the previous partition it is possible to write:

\[
P(y | x, \theta, \eta) = \left( \prod_{1 \leq k < h \leq c} (\eta(k, h))^{e(k, h)} \right) \times \left( \prod_{k=1}^{c} (\eta(k, k))^{e(k, k)} \right), \quad (44)
\]

where \( e(k, h) \) is the number of existing relations between blocks \( k \) and \( h \) that is computed as:
\[ e(k, h) = \sum_{(i,j) \in N} 1\{y_{ij} = 1\} 1\{x_i = k\} 1\{x_j = h\}. \] \hspace{1cm} (45)

The model that is needed is given by the joint distribution of the relationships and the groups structure, i.e. the joint distribution of \( (Y, X) \), which is given by:

\[
P(y, x \mid \theta, \eta) = \theta_{m_1}^{m_1} \cdots \theta_{m_c}^{m_c},
\]

\[
\times \left( \prod_{1 \leq k < h \leq c} (\eta(k, h))^{e(k, h)} \right),
\]

\[
\times \left( \prod_{k=1}^{c} (\eta(k, k))^{e(k, k)} \right).
\hspace{1cm} (46)
\]

The three components of the previous equation represent, respectively: the distribution of the nodes groups, the connections between nodes of different classes and the connections between nodes of the same group. The parameters are estimated using Bayesian methods and for each node of the components it is assumed that the prior distribution is the product of independent Dirichlet distributions.

Given that only the structure of \( Y \) is known, the previous probability can be re-written as:

\[
P(y \mid \theta, \eta) = \sum_{x \in \mathbb{C}^n} P(y, x \mid \theta, \eta),
\hspace{1cm} (47)
\]

this is, we sum over all possible combination of classes for the \( n \) nodes in the network. In order to apply the previous equation it is necessary to estimate the parameters \( \theta \) and \( \eta \) using the posterior distribution

\[
f(\theta, \eta \mid y) = \sum_{x} f(\theta, \eta, x \mid y),
\hspace{1cm} (48)
\]

with these parameters it is possible to estimate the classes vector \( x \) using the posterior predictive density

\[
P(x \mid y) = \int f(\theta, \eta, x \mid y) d\eta d\theta.
\hspace{1cm} (49)
\]

Gibbs sampling is used to obtain the conditional probability

\[
f(\theta, \eta, x \mid y).
\hspace{1cm} (50)
\]

34
from which, it is possible to make inference for the two previous equations.

In this work the network analysis package Graph-Tool (Peixoto, 2014) is used. In Peixoto (2017) the author details the methodology used for the Python package.

5 Results and discussion

In this section we present the main results of the application of the clustering methods and also of the stochastic block models. We also provide a discussion on the most important results comparing both techniques at the same time.

5.1 Clustering methods

The groups were determined by using the library “clustrd” Markos et al. (2018) of the statistical language R. The two methods were used, their groupings and their goodness of fit. Additionally, when a dimensionality reduction is being done, it is common to standardize the variables. Nevertheless, instead of opting for the common standardization, we followed Farne and Vouldis (2017) by making each balance sheet account relative to the principal category, that is, an asset account was divided by the total assets of each bank; a liability account was divided by the total liabilities and so on and so forth. This procedure was important as big banks tend to have much larger amounts than small banks and this could bias the determination of the clusters.

In the presentation of the results we will present the method’s fitting using different information sets. These sets will be different regarding the level of aggregation, which in turn implies a different number of variables. For example, at the highest possible level of aggregation we will have the level N1, that comprises the immediate level below the general concepts of assets, liabilities and capital. The following level is N2 which comprises the following level of accounts below the level N1, this level is much larger than the level N1. Additionally, the cases in which the highly correlated variables are eliminated will include the NoCol text.

In Figure 5 it is possible to see the fit to the data with the highest level of aggregation (lowest granularity), N1. The figure presents two panels, panel a) presents the groups including all the variables at level N1; panel b) presents the groups without including the highly correlated variables. From this figure it is possible to see that the fitting is complex as the subspace is tri-dimensional, some of the groups are not well differentiated and the

In order to investigate the impact of increasing the information set used as input for the algorithm, Figure 6 shows the results for the next information level (less
aggregation and lower granularity). From this figure it is possible to see that there are more groups but better separated and if the highly correlated variables are eliminated, the complexity is reduced to two dimensions.

Figure 7 shows the results for a more detailed information set. The groups are well separated and the elimination of correlated variables reduces complexity.

In addition to the presented static view of the results we also explore the time dimension as we have access to the information set for a long period of time. The algorithms used do not incorporate the time dimension, in order to do it, a more sophisticated modelling approach would be required like distribution mixtures.

In order to illustrate the time evolution of the composition of the groups we resort to alluvial diagrams or parallel set plots Rosvall and Bergstrom (2010). Given that the models in which the correlated variables are eliminated produce less groups and are less complex in terms of dimensions we will only show the results for such cases.

Figure 8 shows the change in the number of banks and on the composition of the groups for the information set N1 eliminating the highly correlated variables. The classification provided by the Factorial K-Means algorithm does not seem to make sense as it basically produces 1 large group and two very small groups most of the time.

The incorporation of more information does not produce interesting results as it produces more groups with many changes across time, Figure 9 illustrates this. It is important to say that a large number of groups is not bad in itself, but for such a small banking system in terms of the number of banks doesn’t seem to be right. We
stated clearly at the beginning of this paper that we wanted the data to speak for itself; nevertheless, it could be the case that our specific problem cannot be solved with this clustering methods and for that reason we resorted to the Reduced k-Means algorithm.

Now, the results for the Reduced K-Means case show that this method is a lot more stable in terms of dimensions and number of groups, regardless on the level of aggregation of the input data. Have we stayed with the FKM method wwe wouldn’t have seen this results. This doesn’t mean that we are imposing conditions on the models but rather exploring alternatives as it could be the case that FKM is not a suitable model to identify business models in this specific context.

Figure 10 shows the results for the Reduced K-Means algorithm for the highest information aggregation level. The results show that two dimensions are enough for the identification of the clusters, the number of groups is smaller and the dispersion of the institutions is more realistic.

The second information level shows similar results to the N1 information set, Figure 11. The only difference is the existence of a new group in the case including all the variables and in the case where the highly correlated variables are eliminated.

For the last level of aggregation, Figure 12, the RKM methods reduces the number of groups with respect to the information set N2.

In order to see the evolution of the groups identified by the algorithm we will resort again to the alluvial diagram. Figure 13 shows the alluvial diagram for the N3 information set. In this figure it is shown that there are important changes in the
structure and composition of the groups in December 2007 and March 2008. The number of groups and the banks that belong to such groups is quite stable.

In other diagrams not included here it is possible to see the same diagrams for different information sets\(^9\). Some of these diagrams showed important changes in the structure of the groups for particular dates. The end of 2007 and beginning of 2008 present important changes in the number and composition of groups. At the end of the period of study it is possible to see basically three groups and some banks which changed their business model and were integrated into a different group. Some diagrams also showed important changes in September 2009; it is worth mentioning that in August 2009 it was reported a drop in the Mexican GDP of 8%. All this shows that banks do change their behavior during some important economic events and some of those changes remain stable in time.

5.1.1 Random Forests

Once we know to which group each institution belongs, then it is necessary to understand what is causing institutions to belong to a group. For this we resorted to random forest in order to determine the importance of the balance sheet accounts for the groups.

The random forest were implemented using the library VSURF (Genuer et al. (2018), which also uses the library randomForest (Liaw and Wiener (2002)). Such

\[^9\]Plots are available upon request.
libraries allow for multiple adjustments and to obtain the importance of the balance sheet accounts through a large number of simulations, improving the estimations; moreover, the library VSURF is designed for the selection of variables making use of the importance that the random forests assign to the independent variables of a data set.

Using the random forests technique, we obtained the evolution of the importance for the balance sheet accounts of the three models obtained with the RKM clustering method eliminating highly correlated variables. The library that we are using makes the selection automatically.

In Figures 14, 15 and 19 we present the metric used to determine the importance of a variable: mean decrease in precision. This metric measures the negative impact in the precision of a classification tree if the variable is removed; then, the larger the value, the more important the variable is for the classification.

For visualization purposes we use two criteria to select the variables presented in Figures 14, 15 and 19: the frequency of the variable in the period of study and the average ranking in the classification provided by the random forests. Then, only the variables which appear in at least 80% of the dates with the lowest average ranking will be shown.

In the following plots we will be able to visualize the evolution of the most important variables according to the aforementioned criteria. In order to avoid unnecessary
text, the plots present a numeric ID for each variable, which can be matched with the name of the variable in the accompanying tables.

For the highest aggregation level, N1, Figure 14 shows that the most important variables for the classification are the contributed capital and capital gained; after them, current credit portfolio, traditional deposits and securities investments have big jumps. Taking into account dates with important jumps, the jump for the current credit portfolio and securities investment during the second half of 2009 is worth emphasizing. This change in the business model might be related to the Mexican GDP fall around those days.

The important balance accounts could have information on the business models, in particular the credit portfolio, traditional deposits and securities investment. Nevertheless, given that these accounts are very general, the information provided by them is still small. For example, regarding the current credit portfolio, in México there are banks that specialize in commercial credit, others in consumer credit or mortgages; therefore, the credit credit portfolio account does not provide sufficient information to identify business models.

As a consequence of the above result we decided to perform the same exercise for different levels of data granularity. Figure 15 presents the plot of variables relevance for the account level N2.

It is worth notice that the majority of selected series are sub-levels of the accounts
Figure 10: RKM for the information level N1 last available date

shown in the previous (higher) level, namely:

1. Contributed capital:
   - Increment by updating social capital
   - Social capital

2. Capital gained:
   - Increment by updating capital reserves
   - Increment by updating gains from previous fiscal years
   - Capital reserves

3. Traditional deposits:
   - Term deposits

4. Securities Investments:
   - Trading book

5. Permanent investments:
   - Subsidiaries
6. Interbank loans:

- Long term interbank loans

One important aspect of the above result is that there is consistency between the results of the determination of the variables importance at the two different levels. If this was not the case the interpretations as well as the justification would be much more difficult to achieve. Once again, the capital accounts are present and are also the most important ones. However, there are not sub-accounts of the credit portfolio.

The third level of granularity (Figure 11) shows the that the sub accounts of the important accounts at the previous level are also important. Moreover, a sub sub account of the current credit portfolio appears to be important: unrestricted...
commercial credit. Also relevant is the presence of securities investment but on the liability side of the balance sheet.

The interbank liabilities accounts are important at the three different levels of aggregation.

The change of the balance sheet side for some accounts show that the results are sensible to the incorporation of more information, this is an important issue which deserves further research.

5.1.2 Distribution of the most important accounts by group

The alluvial diagrams showed that the banks’ business models are not static and change depending on the economic environment. For this reason it is hard to perform dynamic clustering analysis. This is also true for the determination of the important variables, as a result we will only show the results of the variables importance for the last available date (June 2017) of the study period.

Figure 17 shows the distribution of the most important accounts as determined by the random forests for the highest level of aggregation N1, for this plot, and the following two, the tables with the variable names used for the evolution of the importance also applies. By looking at the distribution of contributed capital and earned surplus is clear why these variables are selected as the most important, as the distinction among the three groups is very clear.
The third (interbank loans) and fourth (current credit portfolio) accounts have a direct and intuitive interpretation for the business models and it is possible to see well differentiated the three groups (despite some overlapping).

Securities investment is the seventh most important account and although the banks have similar distributions on each group there are institutions that invest more heavily on securities and we can see it on the tail of the distribution.

The accounts of the second information level N2 shown in Figure 18 make evident that the social capital account (a sub account of contributed capital) is relevant for the formation of groups.

Regarding the variables with a straightforward interpretation from the business models perspective, it is possible to see that traditional deposits is the third most important account, this is interesting as this account provides information about the banks funding profile. Specifically, term deposits are the deposits which are considered as important.

The fourth account, derivatives securities shows an important concentration in one group with respect to the other groups. Lastly, the seventh account, current commercial credits portfolio, complements traditional deposits and derivatives in group three which is the best balanced group regarding its balance sheet, whereas groups one and two show very different distributions in relevant accounts from the
In one hand, the distributions of traditional deposits, credit portfolio and interbank loans show that banks in group two have a model in which they give greater importance to deposit taking and credit provision. On the other hand, group one shows a preference for derivatives and securities investment.

Finally, Figure 19 shows a sub sub account from capital in the first place of importance: earned capital - last year results - results to be applied. Nevertheless, additional sub accounts provide new information to define business models.

Second and third places (term deposits from the general public y derivatives in the trading book) confirm the hypothesis made with the accounts from level N2: whereas group one has a more homogeneous balance sheet, groups two and three
Figure 15: Most important variables according to random forest for the RKM N2 SinCol clusters.

differentiate in the type of deposits and on derivatives.

It is clear that moving towards more granular information levels the results are more intuitive in business models terms. Nevertheless, when all the period of study is considered, the most important accounts do not change when one moves from the highest level of aggregation (N1) to the following one (N2) and this does not happen if one goes from level N2 to level N3.

5.2 Structural network metrics

Here we report some of the results for some standard network metrics which provide a good idea about some of the structural properties of the interbank exposures network in Mexico. In Figures 20 and 21 we observe the evolution of some average metrics already presented before.

It is important to note an upward trend for some of the series: affinity, closeness, betweenness, lending and borrowing strength. This upward trend is associated with the banking system’s growth (strength) and an increase in connectivity (affinity, closeness and betweenness).

Other two metrics which showed an upward trend are the HHI indexes as borrower and as lender; this means more concentrated borrowing and lending in terms
of the amount. The average Debrank has shown a strong downward trend, this is related to regulatory changes and the high level of capitalisation in the Mexican banking system.

Dis-aggregating the results, it is possible to examine the evolution of the structural metrics for individual institutions in order to identify important changes for individual institution and structural changes in the behaviour of institutions. Figures 22 to 25 show the evolution of the centrality metrics for the Top 10 banks as well as for the Bottom 10 banks. For example, the series corresponding to closeness and DebtRank (Figure 22) it is clear that the closest institutions behave in a similar way with an upward trend; on the other hand, the bottom part of the ranking is more heterogeneous. regarding the DebtRank, there is a group of institutions which can be considered the most systemic well differentiated and that one of them consistently behaves as the most systemic. The bottom part shows that entering banks are perceived more systemic and that with longer interaction their systemicness is reduced.

In the case of Strength and Degree (Figure 23) institution “B 16” is the one with more connections and the one with the largest strength. The set of institutions which are in the Top 10 by the DebtRank centrality is almost the same than the most central institutions by degree and strength. The less connected and with less
Table 3: Name of the most important variables from the RKM_N3_SinCol adjustment

<table>
<thead>
<tr>
<th>ID</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>420302000000</td>
<td>Capital Gain - Previous Years Results - Results Due To Accounting Changes (4)</td>
</tr>
<tr>
<td>420301000000</td>
<td>Capital Gain - Previous Years Results - Results Pending To Apply (4)</td>
</tr>
<tr>
<td>211101100000</td>
<td>Traditional Deposits - Term Deposits - From General Public (2)</td>
</tr>
<tr>
<td>130107000000</td>
<td>Current Loan Portfolio - Commercial Credits - Without Restrictions (1)</td>
</tr>
<tr>
<td>293101000000</td>
<td>Deferred Credits - Commissions Charged On Initial Loan Grant (2)</td>
</tr>
<tr>
<td>221406030100</td>
<td>Securities Trading - Derivatives - For Negotiation Purposes - Options - Valuation (2)</td>
</tr>
<tr>
<td>221406030000</td>
<td>Securities Trading - Derivatives - For Negotiation Purposes - Options (2)</td>
</tr>
<tr>
<td>221406000000</td>
<td>Securities Trading - Derivatives - For Negotiation Purposes (2)</td>
</tr>
<tr>
<td>190302000000</td>
<td>Other Assets - Advanced Payments (1)</td>
</tr>
<tr>
<td>230203000000</td>
<td>Interbank Loans - Short Term - Loans From Development Banks (2)</td>
</tr>
</tbody>
</table>

participation the set of institutions is more heterogeneous.

Concentration is an important concept in risk management and the Herfindahl-Hirschman index is commonly used to measure it. Figure 24 shows the evolution of the index for the Top and Bottom 10 banks. It is important to note that bank 16 (the one with the highest DebtRank) is among the less concentrated institutions on the lending side. This is also true for the rest of banks in the Top 10 for the DebtRank centrality metric.

Finally, Figure 25 shows the evolution for the Betweeness and PageRank centrality metrics. It is possible to see again in the Top 10, banks which are also important in terms of other centrality metrics. The Bottom 10 banks by Betweeness are those banks which basically do not play an intermediation role in the interbank exposures network.

Network metrics for the core and the periphery

In van Lidth de Jeude et al. (2018) the authors present an statistical procedure to detect core-periphery structures and we used it for the Mexican interbank exposures networks. The results are statistically significant for these networks and we can see (gráficas 26 and 27) clear differences between the averages of the main network metrics separating the core and the periphery. In general institutions in the core are more central than those in the periphery.

Interesting to note is that the DebtRank shows a clear difference after 2014 or the core and the periphery. In the case of the HHI it is possible to see that in

---

10The index takes values in the interval [0, 1]; however, in Figure 24 seems to show values greater than 1, this is only an effect of the smoothing applied to the series.
Table 4: Name of the most important variables in the last date for the first level of granularity

<table>
<thead>
<tr>
<th>ID</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>410000000000</td>
<td>CONTRIBUTED CAPITAL (4)</td>
</tr>
<tr>
<td>420000000000</td>
<td>CAPITAL GAIN (4)</td>
</tr>
<tr>
<td>230000000000</td>
<td>INTERBANK LOANS AND FROM OTHER INSTITUTIONS (2)</td>
</tr>
<tr>
<td>130000000000</td>
<td>CURRENT LOAN PORTFOLIO (1)</td>
</tr>
<tr>
<td>150000000000</td>
<td>FORECLOSED ASSETS (1)</td>
</tr>
<tr>
<td>135000000000</td>
<td>PAST-DUE LOAN PORTFOLIO (1)</td>
</tr>
<tr>
<td>120000000000</td>
<td>INVESTMENT IN SECURITIES (1)</td>
</tr>
<tr>
<td>280000000000</td>
<td>DEFERRED INCOME TAXES AND PROFIT SHARING (2)</td>
</tr>
<tr>
<td>115000000000</td>
<td>MARGIN ACCOUNTS (DERIVATIVES) (1)</td>
</tr>
<tr>
<td>179000000000</td>
<td>OTHER PERMANENT INVESTMENTS (1)</td>
</tr>
<tr>
<td>190000000000</td>
<td>OTHER ASSETS (1)</td>
</tr>
<tr>
<td>140000000000</td>
<td>OTHER RECEIVABLE ACCOUNTS (1)</td>
</tr>
<tr>
<td>290000000000</td>
<td>DEFERRED CREDITS AND ANTICIPATED PAYMENTS (1)</td>
</tr>
</tbody>
</table>

Table 5: Name of the most important variables in the last date for the second level of granularity

<table>
<thead>
<tr>
<th>ID</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>410100000000</td>
<td>CONTRIBUTED CAPITAL - CAPITAL STOCK (4)</td>
</tr>
<tr>
<td>420100000000</td>
<td>CAPITAL GAIN - CAPITAL RESERVOIRS (4)</td>
</tr>
<tr>
<td>211100000000</td>
<td>TRADITIONAL DEPOSITS - TERM DEPOSITS (2)</td>
</tr>
<tr>
<td>221400000000</td>
<td>SECURITIES TRADING - DERIVATIVES (2)</td>
</tr>
<tr>
<td>230200000000</td>
<td>INTERBANK LOANS - SHORT TERM (2)</td>
</tr>
<tr>
<td>241300000000</td>
<td>ACCOUNTS PAYABLE - CREDITORS DUE TO COLLATERALS RECEIVED IN CASH (2)</td>
</tr>
<tr>
<td>130100000000</td>
<td>CURRENT LOAN PORTFOLIO - COMMERCIAL CREDITS (1)</td>
</tr>
<tr>
<td>120100000000</td>
<td>INVESTMENT IN SECURITIES - TITLES TO NEGOTIATE (1)</td>
</tr>
<tr>
<td>240900000000</td>
<td>ACCOUNTS PAYABLE - CREDITORS DUE TO OPERATION SETTLEMENT (2)</td>
</tr>
<tr>
<td>140900000000</td>
<td>ACCOUNTS RECEIVABLE - CREDITORS DUE TO OPERATION SETTLEMENT (1)</td>
</tr>
</tbody>
</table>

2008 and the beginning of 2009 the index was higher for the institutions in the core as borrowers; this means that the institutions in the core had their funding more concentrated during the global financial crisis and shortly on its aftermath.

In general institutions in the periphery have a higher HHI than institution in the core; this can be explained precisely by the intermediation role that institutions in the core perform, allowing them to have many more connections and to distribute exposures in a more diversified form. This is also reflected by the difference in betweeness centrality.

The most important thing to note in the network results is that there is a case to study this structure with a similar spirit to the clustering methods: is there a case for identifying groups of banks?
Table 6: Name of the most important variables in the last date for the third level of granularity

<table>
<thead>
<tr>
<th>ID</th>
<th>Variable</th>
</tr>
</thead>
<tbody>
<tr>
<td>42003010000000</td>
<td>CAPITAL GAIN - PREVIOUS YEARS RESULTS - RESULTS PENDING TO APPLY (4)</td>
</tr>
<tr>
<td>21110100000000</td>
<td>TRADITIONAL DEPOSITS - TERM DEPOSITS - FROM GENERAL PUBLIC (2)</td>
</tr>
<tr>
<td>22140600000000</td>
<td>SECURITIES TRADING - DERIVATIVES - FOR NEGOTIATION PURPOSES (2)</td>
</tr>
<tr>
<td>13040700000000</td>
<td>CURRENT LOAN PORTFOLIO - COMMERCIAL CREDITS - WITHOUT RESTRICTIONS (1)</td>
</tr>
<tr>
<td>19003900000000</td>
<td>OTHER ASSETS - DEFERRED CHARGES, ANTICIPATED PAYMENTS - INTANGIBLE ASSETS (1)</td>
</tr>
<tr>
<td>14069500000000</td>
<td>ACCOUNTS RECEIVABLE - CREDITORS DUE TO OPERATION SETTLEMENT - DERIVATIVES (1)</td>
</tr>
<tr>
<td>16001050000000</td>
<td>PROPERTY, PLANT AND EQUIPMENT - PROPERTY, PLANT AND EQUIPMENT - COMPUTING EQUIPMENT (1)</td>
</tr>
<tr>
<td>16001060000000</td>
<td>PROPERTY, PLANT AND EQUIPMENT - PROPERTY, PLANT AND EQUIPMENT - PROPERTY (1)</td>
</tr>
<tr>
<td>24090200000000</td>
<td>ACCOUNTS PAYABLE - CREDITORS DUE TO OPERATION SETTLEMENT - INVESTMENT IN SECURITIES (2)</td>
</tr>
<tr>
<td>13010701000000</td>
<td>CURRENT LOAN PORTFOLIO - COMMERCIAL CREDITS - WITHOUT RESTRICTIONS - OTHER (1)</td>
</tr>
</tbody>
</table>

In addition to the core-periphery study just presented we will resort to a more general technique, Stochastic Block Models, in which we will not restrict the hierarchical structure of the network to consists of only two groups and we hope to understand the roles played by banks in terms of connectivity to complement our view on business models.

5.3 Stochastic Block Models

The degree distribution is one of the most important features of a network. Given that we had to decide whether to use the SBMs with or without degree correction we plotted the empirical degree distribution and contrast it against the Poisson and Power Law distributions. This was done by using the observed degree distribution to obtain two parameters: \( \lambda \) to simulate the Poisson distribution and \( \alpha \) to simulate the Power Law distribution. As it can be seen in Figure 28, the empirical distribution reveals nodes with larger degree than some of the ones following a Poisson distribution but with considerably lower degree than some of the ones following a Power Law distribution.

With the evidence provided by the previous exercise, we decided to use SBMs without degree correction as the empirical distribution shows that the network does not present enough heterogeneity for the degree correction.

In order to visualize the stability of the groups obtained by the SBM we use the same alluvial diagram as the ones used for the illustration of the clusters obtained using balance sheet information. Figure 29 shows that changes in the groups are more frequent than for the clustering methods. However, the number of groups is
more stable during the period of study.

In Figure 29 it is possible to see that after to the financial crisis the number of groups identified is reduced from five groups only six months before to only two at the end of 2008, contrary to what was seen with the number of business models that increased for the same period. In Martinez-Jaramillo et al. (2014) the authors find that some of the connectivity structural metrics did change after the crisis, we get the same message here. Another important aspect to mention is that we rarely see a two groups network, pointing out to a more general tiered structure than the core-periphery structure suggested in Craig and von Peter (2014).

In order to explore the differences in the groups, we resort to the distributions of the centrality metrics per group. We do not resort to clustering methods to find the groups based on the centrality metrics, instead we used random forests to detect the most relevant metrics and from there we study the distributions. In Figure 30 it is possible to see the relevance of the centrality metrics in the formation of groups. It is not surprising to find that the degree metrics are the most relevant followed by closeness.

Now, in Figure 31 it is possible to observe why the degree metrics are the most relevant for the classification as the distributions are well differentiated for each group being cluster 1 the one with the most connected nodes and cluster 3 the one with the less connected nodes. Moreover, cluster 1 is the one with the highest closeness to the rest of the network.

Affinity also provides some insights as group 3 shows quite an heterogeneous distribution, indicating that banks in this group connect with other banks regardless of their degree. It is also interesting that group 2 has the highest affinity, this means that group 3 forms relationships with a diversity of nodes. This evidence contradicts the core-periphery model proposed in Craig and von Peter (2014). All the facts just showed here completely justify, from our point of view, the need to resort to SBMs as a more general and adequate approach to characterize the meso structure of the interbank exposures network.

From the metrics shown in Figure 32, DebtRank distinguishes itself from the rest given that despite that the mean is not very different among groups, group 1 presents considerable more extreme values than the other groups. This implies that the most systemic banks belong to group 1. Therefore in addition to the existence of similar business models, these banks also share similar connectivity patterns.

In the next subsection we will try to put together all the evidence that we have gathered with the two approaches used in this work in order to extract relevant conclusions in terms of financial stability and policy making.
5.4 Discussion

In this subsection we will summarize and discuss all the important findings presented in this section. First we present some important conclusions from the application of the clustering methods:

1. Despite the relative stability of the business models, there are important changes related to the macro financial environment in which banks are immerse.

2. The identification of business models are sensitive to the data set used. Moreover, the available data set could determine the method used to perform the clustering analysis.

3. The clustering and network approaches implicitly rely on an stochastic component, therefore, it is necessary to perform an adequate number of simulations.

The results for the clusters identification and the application of the random forest to determine the relevance of the variables for such identification were already presented in this section.

One of the main objectives proposed for this work was to find evidence on whether the connectivity in the networks changes the way in which banks are grouped, to accomplish that, two different methods were used on two different datasets. Resulting groups from both methods are presented in the three following plots. Despite that we only selected the groups obtained with the first level of granularity we present all three fits to illustrate that no matter how disaggregated the data, the clusters always change after the network structure is included.

At the beginning we remarked how important it is to consider a differentiated regulation and monitoring according to the business model of the banks, and that in fact bank’s behaviour as part of a complex system is influenced by its interaction with the rest of the participants of such system (plots 33, 34 and 35).

As an example, the second group given by the SBM is constantly fed by institutions in all three groups obtained with RKM (excepting the groups corresponding to the second level of granularity), which may indicate that the influence of the interaction between banks is important for the grouping structure.

Furthermore, for banks which could be “naturally” classified in the same cluster (e.g. investment banks or D-SIB’s) there is a discrepancy between the two methods as can be seen in the same three figures, since there are institutions that constantly move from group 1 with RKM to the second group from SBM.

For the second level of aggregation we obtain one more group with the RKM method; however, as it can be seen in Figure 34 there are two institutions which are
found in a different group for information levels two and three. It is possible to see these institutions on the relevant accounts for the RKM N2, Figure 36; it is evident that the Social Capital account extreme values are responsible for the formation of this fourth group.

From Figures 34, 35, and 36, it is clear that the grouping is different by considering the banks from a systemic perspective. This has important implications in favor of studying the banking system with these two different methods. For example, it could be the case that a bank business model is focused on a very specific activity and at the same time possess a central role in the network. If things go wrong for this bank due to its business model, this could have implications for the whole banking system.

### 6 Conclusions

Taking into account all the previously shown results, the general conclusions of this work are:

- The banks’ business models change accordingly with the economic environment.
- The RKM clustering method is sensitive to the information set and to extreme values from the input variables.
- For the Mexican case, capital accounts seem to be relevant for the groups created by the method for the whole period of study.
- In addition, Figures 14, and 16 reveal that the credit portfolio and other financial assets (securities and derivatives) are very important in some points in time; for example, during the GFC.
- The use of data with higher degree of granularity can provide better insight on the results; for example, Figures 14, and 16 show that relevant accounts in lower levels of granularity are contained in the relevant accounts from lower levels of granularity.
- The groups change using the network structure as banks which would not be considered as important belong to the same group of systemically important banks.
- Comparing the results obtained by the two different methods it is possible to see that banks in Mexico form groups in a non-trivial way. Therefore, it is
necessary to study the banking and the financial system in more than only one dimension.

6.1 Future work

The results obtained in this work are encouraging but there are clear opportunities to improve our results:

- Include explicitly the dynamic aspects of the problem. In Lucas et al. (2017), the authors use a method which takes into account the dynamic aspects of the problem. Nevertheless, they do it for a smaller data set which could result in a computationally unfeasible method in our setting.

- Include weights in the SBMs

- Apply the same methods for a larger banking system or even for an international set of banks. This would reveal if there is a geographical component on the banks’ business models. On the SBMs, the method could be applied for a network that not only contemplates banks.

- The financial system can be better described as a multilayer network, applying such paradigm with the same purpose could result in a more robust grouping structure.

References


Figure 17: Distribution of the percentage that each variable represents in relation to the highest account level (total assets, total liabilities and capital), for the most important variables according to the random forests method for level N1.
Figure 18: Distribution of the percentage that each variable represents in relation to the highest account level (total assets, total liabilities and capital), for the most important variables according to the random forests method for level N2.
Figure 19: Distribution of the percentage that each variable represents in relation to the highest account level (total assets, total liabilities and capital), for the most important variables according to the random forests method for level N3.
Figure 20: Average values for some metrics for the interbank exposures network (1).
Figure 21: Average values for some metrics for the interbank exposures network (2).
Figure 22: Closeness and DebtRank centrality for the Top and Bottom 10 banks.
Figure 23: Strength and degree centrality for the top and bottom 10 institutions.
Figure 24: Herfindahl-Hirschman index as lender and borrower for the Top and Bottom 10 institutions.
Figure 25: Betweenness and PageRank centrality for the Top and Bottom 10 banks.
Relevant average measures differenciating by core and periphery (1)

Figure 26: Average values for the main centrality metrics for the core and the periphery of the Mexican interbank exposures network obtained with the method by Surprise (1)
Relevant average measures differentiating by core and periphery

Figure 27: Average values for the main centrality metrics for the core and the periphery of the Mexican interbank exposures network obtained with the method by Surprise (2)

Figure 28: Degree distribution for the Mexican interbank exposures network.
Figure 29: Alluvial diagram with the groups determined by the SBM

Figure 30: Centrality metrics relevance for the SBMs groups classification.
Figure 31: Strength and degree distribution

Figure 32: Centrality metrics distributions II
Figure 33: Comparison of the groups assigned by the SBM and the RKM for banks using the first level of balance sheet accounts.
Figure 34: Comparison of the groups assigned by the SBM and the RKM for banks using the second level of balance sheet accounts.
Figure 35: Comparison of the groups assigned by the SBM and the RKM for banks using the third level of balance sheet accounts.
Figure 36: Value for the most relevant accounts for the RKM N2 differentiating the two institutions which are grouped in a non existing group for the previous date for the same case.