Supervision of banking networks using the multivariate Threshold-Minimum Dominating Set (mT-MDS)

Periklis Gogas* **, Theophilos Papadimitriou*, Maria-Artemis Matthaiou*

* Department of Economics, Democritus University of Thrace, Greece

** Department of Economics and Administration, Neapolis University Pafos, Cyprus

Corresponding author: pgkogkas@econ.duth.gr

Abstract

The global financial crisis of 2008, triggered by the collapse of Lehman Brothers, highlighted a banking system that was widely exposed to systemic risk. The minimization of the systemic risk via a close and detailed monitoring of the entire banking network became a priority. This is a complex and demanding task considering the size of the banking systems: in the US and the EU they include more than 10000 institutions.

In this paper, we introduce a methodology which identifies a subset of banks that can: a) efficiently represent the behavior of the whole banking system and b) provide, in the case of a failure, a plausible range of the crisis dispersion. The proposed methodology can be used by the regulators as an auxiliary monitoring tool, to identify groups of banks that are potentially in distress and try to swiftly remedy their problems and minimize the propagation of the crisis by restricting contagion. This methodology is based on Graph Theory and more specifically Complex Networks. We termed this setting a "multivariate Threshold – Minimum Dominating Set" (mT–MDS) and it is an extension of the Threshold – Minimum Dominating Set methodology (Gogas e.a., 2016). The method was tested on a dataset of 570 U.S. banks: 429 solvent and 141 failed ones. The variables used to create the networks are: the total interest expense, the total interest income, the tier 1 (core) risk-based capital and the total assets. The empirical results reveal that the proposed methodology can be successfully employed as an auxiliary tool for the efficient supervision of a large banking network.

 doi:10.20944/preprints202205.0065.v1

Key words: Complex networks, Minimum Dominating Set, Banking supervision, monitoring optimization

1. Introduction

The global financial crisis of 2008, also known as the subprime mortgage crisis, is classified as the world's worst crisis since the Great Depression of 1930. The initial collapse of Lehman Brothers triggered a cascading effect of failures exposing the high systemic risk embedded in a highly interrelated and interdependent banking network. Systemic risk in a banking system is defined as the risk of destabilization of the whole system, caused by the failure of a single or a small set of banking institutions. The collapse of Lehman Brothers threatened the viability of many other large financial institutions. The ones that eventually survived received significant subsidies (through bailout programs) under the Troubled Assets Relief Program (TARP) implemented by the government. The same was not true for a set of smaller institutions that were left to collapse.

Historically, following each banking crisis, a series of Acts were introduced aiming to stabilize the banking system and avoid a future re-occurrence of a crisis. The Banking Act of 1933, commonly known as the Glass-Steagall Act, was signed by President Franklin Roosevelt at an attempt to restore the confidence in the U.S banking system. The bill was designed to reform the banking system by imposing a dichotomy between commercial and investment banking in order to reduce risk. It also intended to allow for a safer and more effective use of banks' assets, to regulate interbank control and prevent banks form conducting speculative operations. This boosted disintermediation, the practice of financing directly from capital markets without the intermediation of banks, reducing banks' share in total financing. As a result, banking institutions pursued the abolishment of the Glass-Steagall Act and several decades later, the U.S Congress passed the Gramm-Leach-Bliley Act of 1999, which was signed into law by President Clinton. The Gramm-Leach-Bliley Act waved the preexisting barriers in financial markets. Now, banking institutions were again free to engage in securities trading and insurance contracts that helped them increase their market share. This newly gained freedom and the inevitable fierce competition with non-banking institutions, led to the rapid development in the markets of new and diverse financial instruments. In effect, the liberalization of the banking sector led to a significant increase in systemic risk as banks are since exposed to investment and other types of risk. Many analysts believe that the 2008 financial crisis is directly linked to the imposition of the Gramm-Leach-

Bliley Act. The lack of separation between commercial and investment banking activities, allows financial institutions to be involved in securities trading not only for their customers but also for themselves, a practice that exposes common depositors to high market risk.

After the 2008 financial crisis, the Obama administration enacted the Dodd-Frank Wall Street Reform and Consumer Protection Act, in 2010, in an attempt to minimize systemic risk, enforce financial sector's transparency and accountability, and implement rules for consumers' protection. However, the provisions of the Dodd-Frank Act did not include the strict separation between commercial and investment banking and thus cannot fully minimize such risk in the banking sector.

It is essential for the regulators to swiftly pinpoint incidents of bank distress. A prompt identification of instances of increased systemic risk can help minimize the policy reaction time. This can help minimize the contagion and defuse the propagation of a potential financial crisis. Thus, an effective and continuous monitoring of financial institutions is necessary for the maintenance of a solvent and stable banking system. Increased supervision and strict regulation of the banking system is also required by the new Basel Accord (Basel III) by the Basel Committee on Banking Supervision, introduced in the U.S. in 2013, and implemented in 2018.

According to the new Basel Accord framework, the regulatory authorities are responsible to: a) implement an extensive supervision of the banking system, and b) mitigate the effects of possible crises and limit contagion. A significant concern, apart from finding the appropriate monitoring tools, is the appointment of such a regulatory authority. Many researches support that the supervision of the entire banking system should be vested to a single authority. Vives (2000) and Blinder (2010) support that a single authority can: a) establish credible systems, b) achieve economies of scale and c) reach financial stability, by taking advantage of the economies of scale between Lender of Last Resort (LOLR) facility, supervision and monetary policy. Boyer and Ponse (2012) support that a single supervisory authority preserves a more effective supervision of the banking system, than the one achieved by multiple authorities. Following the idea that a single authority is competent to optimize the supervision of the entire banking system by providing a timely and efficient intervention, E.U. leaders decided in October 2012 to assign the supervision of the whole European banking network to a single authority, the ECB.

In this paper, we introduce the multivariate Threshold – Minimum Dominating Set (mT-MDS) methodology and then use it to group the banking network in neighborhoods according to their financial health. The mT-MDS is an extension of the T-MDS methodology, that is especially designed to treat multivariate networks, since the T-MDS can handle only univariate ones. The methodology creates a multilayered network and distills the multivariate information in one binary network using a Boolean operator. The multilayered network is build using the variables that were identified in Gogas, et al. (2018) for banking bankruptcy forecasting. The empirical results reveal that the neighborhoods efficiently classify solvent and failed institutions. We propose the mT-TMDS method, as an auxiliary monitoring mechanism, an extra layer on banking supervision.

The rest of the paper is organized as follows. In Section 2, we present a literature review of studies related to ours. In Section 3, we present the multivariate T-MDS model. In Section 4, we present our dataset, while in Section 5, we report the empirical results. In Section 6, we draw the conclusions.

2. Literature Review

Systemic risk is spread through the multiple interrelations between financial institutions in a banking network. A simple, concise and efficient way to model this system is by using a Complex Networks representation: each bank is represented by a node and the interrelation between two banks is represented by an edge linking them. The theory of Complex Networks provides a set of tools able to examine the structure of economics networks. Mantegna (1999) and Hill (1999) were the first to apply Complex Networks in economics systems. More specifically, Mantegna (1999) uses the MST in his attempt to study the hierarchical structure of the New York Stock Exchange, while Hill (1999) compared price levels across countries using the same methodology. The application of complex networks in economics and more particular in banking has grown expeditiously during the last years. There are several studies that examine the risk of contagion.

The seminal paper of Allen and Galle (2000) investigates the cascading effect of a banking crisis on a network of regions or economic sectors. The authors showed that two cases are resilient to a liquidity shock: a) the case of a complete interbank market (i.e., a market where every bank is connected with to all the other banks of the network) and b) an incomplete interbank market with low degree of interconnectedness. Conversely, in the case of an incomplete interbank market with high degree of interconnectedness the liquidity shock may spread to the whole network. Similar conclusions are drawn from the studies of Leitner (2005) and Gaia and Kapadia (2010), where the results revealed that as the network becomes denser, systemic risk drops and the influence of an institution's default is negligible as the losses of the failed institution will be spread and engrossed from the rest institutions of the network.

In order to examine in depth, the origins of systemic risk, a number of papers analyze the topology of real word networks in an effort to identify their features. Minoi and Reyes (2011), analyse the topology of the global banking network, formed of financiall flows during and after periods of financial stress. The findings show that a number of structural breaks in the network, indicates the waves of capital flows before and after crises. Network's centrality falls at the begining and after a debt crisis. In the study of Tabak et al. (2014), the authors introduce the directed clustering coefficient as a measure of systemic risk in complex networks. The results reveal that the network is not exposed to systemic risk, more specific the clustering coefficient and domestic interest rates reveal negative correlation: as interest rates increase, the banks decrease their exposure to the system. Thurner et al. (2003), examine the impact of a network's structure on the wealth of the economy, concluding that a highly connected network, is more stable since it is not exposed to large wealth changes. Kuzubas et al. (2014) use centrality measures: betweenness centrality, closeness centrality and Bonacich's centrality, to assess the network's connectivity and identify the systemically important institutions. The results reveal that the centrality measures are adequate for the identification and the observance of systemically important financial institutions.

Furthermore, there are studies using the balance-sheet based technique to explore and evaluate the interrelations of banking institutions and their level of influence in the overall banking system. These studies test the influence of credit relations, in different banking systems. Their aim is to explore how mutual claims between banks can affect the propagation of contagion. Upper and Worms (2004) study the German banking system, Cocco et al. (2003) analyze the Portuguese, Wells (2004) explores the U.K. interbank market, Furfine (2003) studies the U.S. banking system, Nguyen (2004) tests the Belgian interbank market and Sheldon and Maurer (1998) study the Swiss banking system. The results of the above studies coincide and reveal that the default of a single institution is not capable to trigger the collapse of the entire system, though it is able to influence a quite small part of the network. All studies reach to the same conclusion: banking systems are robust to the failure of a single institution. Finally, Chan-Lau (2010) also uses a balance-sheet based approach to examine whether the financial crisis of 2008 that was triggered by an institution’s failure and it spread to the largest part of the world, was an aftermath of institutions’ interbank exposure and their externalities with too-connected-to-fail institutions. The results reveal that when shocks in the network jeopardize banks' solvency, they can be characterized as sources of financial contagion worldwide.

We presented a large collection of papers focused in measuring the robustness of a banking network and its exposure to systemic risk. Our approach differs as it is oriented on the optimization of network monitoring. More specifically, we propose the introduction of a new monitoring layer to the existing systems for bank supervision. In this layer, the regulator will identify a reduced version (small subset) of the initial network that: a) contains an adequate amount of the total information and b) is easier to monitor and analyze. We merge these two targets in one under the term ***representation goal***. We investigate the interrelations between banking institutions using tools from Complex Networks theory and more specifically Graph theory. In Gogas, e.a. (2016) we introduced a two-step methodology termed T-MDS for the identification of a small subset of nodes able to represent the entire network. In our setup the nodes are the economic entities (banks) and the edges define temporal similarity as described by the correlation. We showed that the T-MDS banks can be used as distress sensors for the whole banking network.

3. Proposed Methodology

In this paper, we introduce the multivariate Threshold-Minimum Dominating Set (mT-MDS), a multivariate extension of the T-MDS. The mT-MDS shares the same goals with the T-MDS, though it works in a multivariate environment. Indeed, the main limitation of the T-MDS methodology lies in its univariate design: the network is built based only on one variable and the representation goal may be unreachable in such information-wise poor environment. Banking interconnections are complex, and they are better described by more than one variable. With this need in mind, here we introduce and employ the multivariate T-MDS. We begin with some basic concepts of the Graph Theory.

Definition 1: A **graph** G consists of a collection V of **nodes** and a collection E of **edges**, for which we write G = (V, E). An edge $e_{ij} \in$ E is said to join the nodes *i* and *j*. The structure can be extended by assigning numerical values (often called weights) to each edge representing quantitative relations v_{ij}.

To fully comprehend the methodology and the explanation of the empirical results, the following key concepts must be defined:

- **Isolated node** is a node that is not connected to any other node in the network.
- **Interconnected node** is any node that is connected to at least another node in the network.

In our experiments the nodes represent economic entities (banks), the edges represent the existence or not of a relation (similarity) between two nodes, and the weights of the edges measure the relation between the two nodes (in our tests we calculated the temporal similarity between nodes using the Pearson's correlation coefficient *r*). To identify the smallest subset of nodes that can represent the whole network we will use the notion of the Dominating Set.

Definition 2: *A* ***dominating set*** *of a graph* G *is a subset of nodes* DS ⊆ V *such that every node* $v \notin$ DS *is directly connected to at least one member of the* DS. *The members of the* DS *are called* ***dominant nodes***.

So, if we identify a dominating set of a graph, any node of the graph is either a) a dominant node or b) a node directly connected to a dominant node. This is a particularly important property to us, since it means that, given that the edges of a graph describe high similarity between the nodes, the whole network can be represented by the dominating set. Indeed, by the definition of DS, every non-dominant node is directly connected with at least a dominant node by an edge that describes high similarity. The nodes directly connected with a dominant node form the *neighborhood* of the dominant node. The dominant node of a neighborhood will serve as the representative node of the group.

A network can have many DS. The ones with minimum cardinality are called Minimum Dominating Sets (MDS). The MDS can be identified through a minimization process:

Define the variables $x_i \in \{0,1\}, i = 1, 2, \dots, n$ representing the nodes' membership to the MDS (1 describes the dominance and 0 describes the non-dominance of a node). These variables in vector form are $\mathbf{x} = [x_1, x_2, \dots, x_n]$. The Minimum Dominating Set is identified by minimizing:

$$\min_{\mathbf{x}} \sum_{i=1}^{n} x_i$$

subject to

$$x_i + \sum_{j \in N(i)} x_j \geq 1, \;\; i = 1, \dots n$$

where $N(i)$ represents the set of adjacent nodes of node i. The assumption is straightforward: the node i can be a) a node of the MDS ($x_i = 1$) or b) adjacent to at least a MDS node ($\exists j \in \mathrm{N}(i): x_j = 1$). In any case the l.h.s of the constraint is equal or greater than 1. The imposed constraint preserves that every node of the network will be represented from the MDS subset.

In real data networks, all the edges may not describe high similarity between its endpoint nodes. Couples of low correlated entities may exist in the network. These low correlation edges are undesired in the identification of the MDS; we wish to monitor every neighborhood through its dominant node, and this is achieved only if the edges describe high correlation. So, we add a thresholding step on the correlation level marked as weight on the edges, eliminating the low correlation ones from the network. After this step, all the edges depict the desired to the representation goal, high similarity.

During the thresholding isolated nodes may appear in the network. Nodes with all weight edges below the selected threshold become isolated. Even, if the focus of this study is the representation of the interconnected banks, the isolated banks are important as well since they are the members of the banking network that exhibit unique idiosyncratic behavior. Any isolated node can only be represented by itself and is a member of the T-MDS by definition.

This methodology was introduced in Gogas, e.a. (2016), termed as Threshold-Minimum Dominating Set (T-MDS) and defined with the following steps:

- **Preprocessing:** Create the network using Pearson's correlation coefficient of the selected variable.
- **Step 1**. Apply a threshold on the weights and eliminate the edges with weights less than the threshold.
- **Step 2**. Identify the MDS nodes on the resulting network.

We claim that every dominant node in the T-MDS can be perceived as a sensor for its directly connected neighborhood nodes, since they are connected with edges that survived the thresholding step (i.e. they are highly similar/correlated). Thus, all the nodes in a neighborhood are expected to exhibit a similar behavior. In our empirical section, we examined the solvency of the banking institutions. If the dominant node is expected to fail (or not), so are the neighboring nodes. Empirical results in Gogas e.a., 2016 backed-up this theory in the univariate T-MDS methodology.

In this paper, we extend the T-MDS concept to multiple variables. We call a ***multivariate Threshold – Minimum Dominating Set*** (mT-MDS) the three-step methodology for the identification of the network's most representative nodes in a multilayered network:

- **Preprocessing:** For each variable, create the correlation-based network. The network is a layer in the multilayered framework.
- **Step 1:** Impose a threshold in every layer to eliminate the low correlation edges, producing thresholded layers. The threshold level may vary across layers.
- **Step 2:** For every couple of nodes combine the edge information from all layers using some operator. In our experiments we used the conjunction Boolean operator: an edge is created when an edge exists between the two nodes in every layer. This step results in the creation of a single unweighted binary network.
- **Step 3:** Identify the MDS.

In Figure 1 we depict graphically the representation of the mT-MDS.

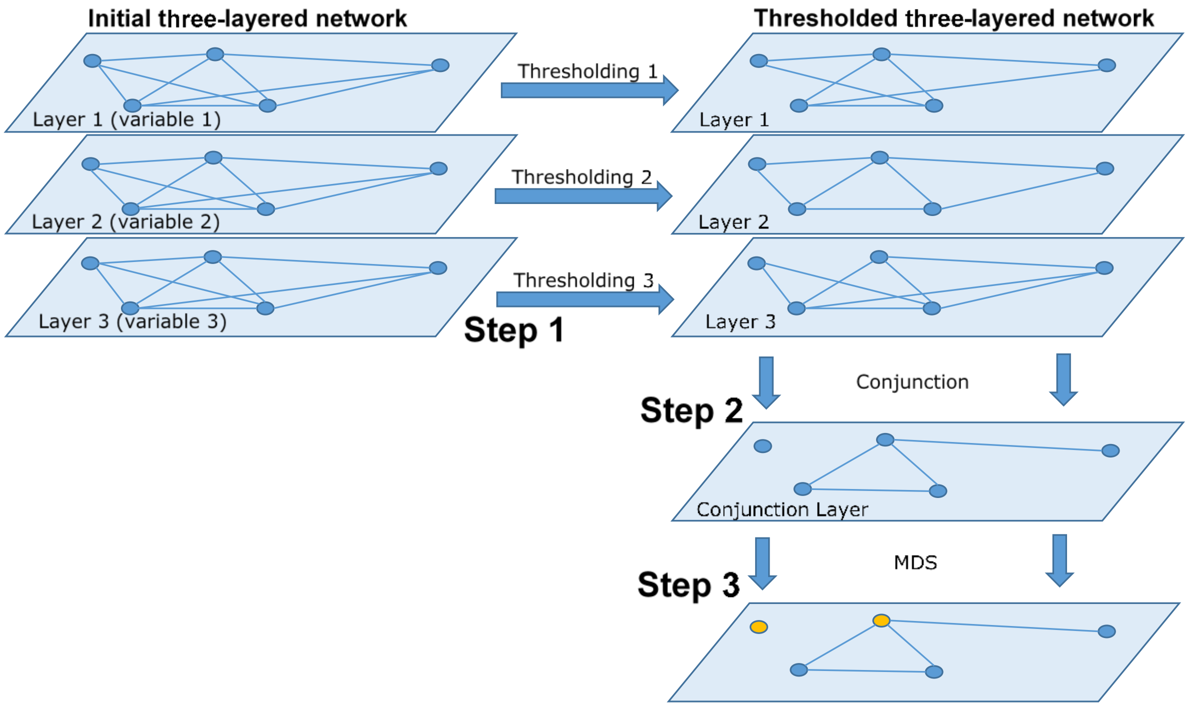

Figure 1: A graphical representation of the mT-MDS

Initially, there is an edge between every node (regardless of the correlation level). In Step 1, the thresholding eliminates the low correlation edges in each layer of the network. After this point, the edges have no weights and describe the high correlation of its endpoints. Its absence means the opposite. The information of all layers is combined in a single binary network in Step 2, through the Boolean conjunction[1]. An edge in the resulting network, describes nodes that are highly correlated in all layers, thus, for every variable. The final step is the identification of the Dominant Nodes in the network and their respective neighborhoods. In section 5, we will show the superiority of this method over the simple T-MDS.

4. The Data

The examined period spans from 2006Q1 to 2010Q3, a period that includes a major financial crisis where many banks failed. We collected quarterly data from the databank of the Federal Deposit Insurance Corporation (FDIC) and they include 570 US banks.

[1] Different applications may use other operators to create the final networks: Boolean Disjunction, majority ruling or even more complex voting.

From these banks, 429 were solvent and 141 have failed. The 141 failed institutions represent the total number of banks that failed during the 2010-2011 period and have available data for all the studied quarters. We also collected data for the 429 largest, in terms of total assets, solvent banks, to maintain a 3:1 ratio between solvent and failed banks. Gogas, et al. (2018), created a Machine Learning based forecasting model for bank failures for the period 2007-2013 that produces a remarkably high overall accuracy reaching 99.22%. Their model selected from an initial extensive list only two variables: a) Tier 1 (core) risk-based capital over total assets (T1CRC) and 2) Total interest expense over total interest income (TIE). The same two variables were used in the empirical section of our paper, since our goal is to test the presented methodology on banking solvency.

Tier 1 (core) risk-based capital over total assets includes disclosed reserves and equity capital and provides a measure of a bank's capital adequacy. It is a significant ratio considered as a proxy of a bank's financial strength. It provides a relative measure of the amount of financial losses a bank can absorb without requiring new capital injections. Whenever a recapitalization is needed, the new capital can either be added through a common increase of share capital or in times of crisis through a bail-out or a bail-in program. Total assets measure the absolute size of a bank. In times of economic expansion, banks tend to augment their assets mainly by issuing new loan facilities; in times of declining economic activity, they reduce their lending and as a result total assets and their balance sheet is reduced. A direct way that a crisis propagates through the banking system is via interbank loans. A troubled bank that used interbank lending to augment its balance sheet and provide more loan facilities, is unable to repay its financial obligations to other banks and thus the total assets of the lending banks take a hit. In general, banks with higher T1CRC ratios are associated with lower probability of default (Gogas et al. 2018). The total interest expense over total interest income (TIE) ratio provides a proxy of a bank's operational efficiency from its core business, i.e. pooling funds and lending. The nominator includes all interest paid by the bank for its interest bearing liabilities and the denominator includes all interest earned on any type of lending from the assets' side of the balance sheet. An increasing TIE ratio reflects a declining gross profit margin.

5. Empirical Results

Trivially, a network combining two variables is information-wise richer than a univariate one. However, the combination of the two variables through the logical conjunction may result in a significant loss of information[2], rendering the multivariate network appealing in theory but not in practice. In this section, we will compare the univariate networks to the multivariate one and show that the information loss is negligible, while the multivariate network neighborhoods portray better the banking solvency.

5.1 Comparison between the T-MDS and mT-MDS networks

In the univariate T-MDS we create two networks: one based on the Tier1/Total Asset variable and one based on the Total Interest Expense/Total Interest Income variable. In the multivariate T-MDS we create a layer for each variable (in a multilayered framework) and then combine them in one binary network (without weights). In every case the threshold level will be set at 0.8. The composition of the univariate networks after thresholding are shown in Table 1 and 2, and the composition of the multivariate network after thresholding and logical conjunction is shown in Table 3.

Table 1: Composition of the 570 banks network using the variable Tier1/Total Asset. Threshold level = 0.8

	Solvent	*Failed*	*Total*	*%*
Total Banks	429	141	570	
Interconnected	317	129	446	78.3%
Isolated	112	12	124	21.7%
Dominant	47	16	53	

The Tier1/Total Asset based network after thresholding retains 446 of the nodes interconnected (78.3%), while 124 of the nodes (21.7%) are isolated from the rest of the network. The interconnected part is clustered in 53 neighborhoods (47 with a solvent dominant node and 16 with a failed one).

[2] In the Boolean conjunction, the operation is true only if all the operands are true. In our case, true is the presence of an edge and false describes its absence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2022 doi:10.20944/preprints202205.0065.v1

Table 2: Composition of the 570 banks network using the variable Total Interest Expense/Total Interest Income. Threshold level = 0.8

	Solvent	*Failed*	*Total*	*%*
Total Banks	429	141	570	
Interconnected	428	139	567	99.5%
Isolated	1	2	3	00.5%
Dominant	5	6	11	

The Total Interest Expense/Total Interest Income based network is almost unaffected from the thresholding step: only 3 nodes became isolated. The 567 interconnected nodes are represented by 11 dominant nodes (5 solvent and 6 failed ones).

Table 3: Composition of the 570 banks network using both the Tier1/Total Asset and the Total Interest Expense/Total Interest Income. Threshold level = 0.8

	Solvent	*Failed*	*Total*	*%*
Total Banks	429	141	570	
Interconnected	294	119	413	72.5%
Isolated	135	22	157	27.5%
Dominant	56	16	72	

The multilayered network refines the univariate layers to a binary one through the logical conjunction. Thus, for an edge to "survive" the conjunction step, it should appear in every layer of the multilayered network. Consequently, the number of edges in the resulting binary network can be at maximum as many as in the layer with the fewer edges. The same is true for the number of interconnected nodes. Thus, in our case, the upper limit of the interconnected part in the multivariate case is 446 nodes (the interconnected part in the thresholded Tier1/Total Asset network, which is the smallest interconnected part). The combination of the univariate layers in a multivariate one results in 413 interconnected nodes clustered in 72 neighborhoods and 157 isolated ones. The multivariate thresholded network has 33 (5.8%) less interconnected nodes than the Tier1/Total Assets univariate network. Next, we will demonstrate that this did not affect the results quantitatively or qualitatively.

5.2 Comparison of the T-MDS and mT-MDS clustering

In this paper, our main goal is to create a monitoring tool that will be able to swiftly examine the health state of the entire banking network (all neighborhoods) using only information from the representative nodes (dominant nodes). Such a tool, would be fast, easy and low-cost. Thus, it can be used by the supervising authority as an auxiliary monitoring system for the health of the banking sector.

Ideally, we would like to be able to label as solvent or failing every node of a neighborhood based on the state of the dominant node. It must be noted here that a node may belong to more than one neighborhoods. Consequently, in some cases, a node may be labeled as both solvent and failed. We call these nodes “fuzzy”. The fuzzy nodes have the disadvantage that they introduce uncertainty in our monitoring tool. We are unable to forecast the fate of a fuzzy node.

The misclassification of our system (nodes that belong to a single neighborhood and thus have one label, but this is the wrong one) may create two distinct cases: false alarms or missed failures. Obviously the costs of a false alarm to the efficient monitoring of the banking system is far less than a missed failure, i.e. identifying a bank as healthy while it is actually failing. The later, may potentially have cascading effects due to contagion and imply systemic risks to the banking sector. However, we also want to keep the false alarms in low levels. For any insolvent case the regulator orders measures to promptly and efficiently treat it and avoid the undesired outcome; in the false alarm case, these measures are unnecessary and prevent the prosperity of the banking institution. So, we will evaluate each system by counting the number of correctly labeled nodes, the erroneously labeled nodes (both missed failures and false alarms) and the fuzzy nodes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2022 doi:10.20944/preprints202205.0065.v1

Table 4: The efficiency of the Tier1/Total Assets based network.

Tier1/Total Assets				
	Solvent	Failed	Total	%
Correct	165	46	211	47.3%
Fuzzy	120	68	188	42.1%
Wrong	33	14	47	10.6%

The Tier1/Total Assets based network after thresholding has 476 interconnected nodes: 211 (47.3%) of them are labeled correctly, 47 (10.6%) are labeled erroneously, and 188 (42.1%) received both labels thus are marked as fuzzy.

Table 5: The efficiency of the TIE/TII based network.

TIE/TII				
	Solvent	Failed	Total	%
Correct	48	51	99	17.5%
Fuzzy	372	71	443	78.1%
Wrong	9	16	25	4.4%

The TIE/TII based network has 3 isolated nodes and 597 interconnected nodes after the Thresholding step. 99 nodes (17.5%) are correctly labeled, 25 are mislabeled (4.4%), and 443 are fuzzy (78.1%). The performance of the clustering using the TIE/TII variable can be characterized as subpar.

Table 6: The labeling of the multivariate network.

Tier1/Total Assets and TIE/TII				
	Solvent	Failed	Total	%
Correct	241	67	308	74.6%
Inconsistent	40	40	80	19.4%
Wrong	14	11	25	6.0%

The interconnected nodes in the multivariate case are 413, partitioned as follows: 308 are correctly labeled (74.6%), 25 received the wrong label (6%) and 80 received both labels (19.4%) and are thus classified as fuzzy.

The multivariate network produced the most accurate labeling in terms of the absolute number of nodes correctly labeled (308 correctly labeled nodes over 211 on the Tier1/Total Assets network and 99 on the TIE/TII network) and ratios as well (74.6% correctly labeled nodes over 47.3% on the Tier1/Total Assets network and 17.5% on the TIE/TII network). Another noteworthy point is that the multivariate network labeling missed the fewer failures: 11 over 14 and 16 respectively in the univariate networks.

5.3 Inside the neighborhoods

Since we established the superiority of the multivariate clustering over the univariate ones, we can analyze the concordance level between the dominant node and its neighbors.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 6 May 2022 doi:10.20944/preprints202205.0065.v1

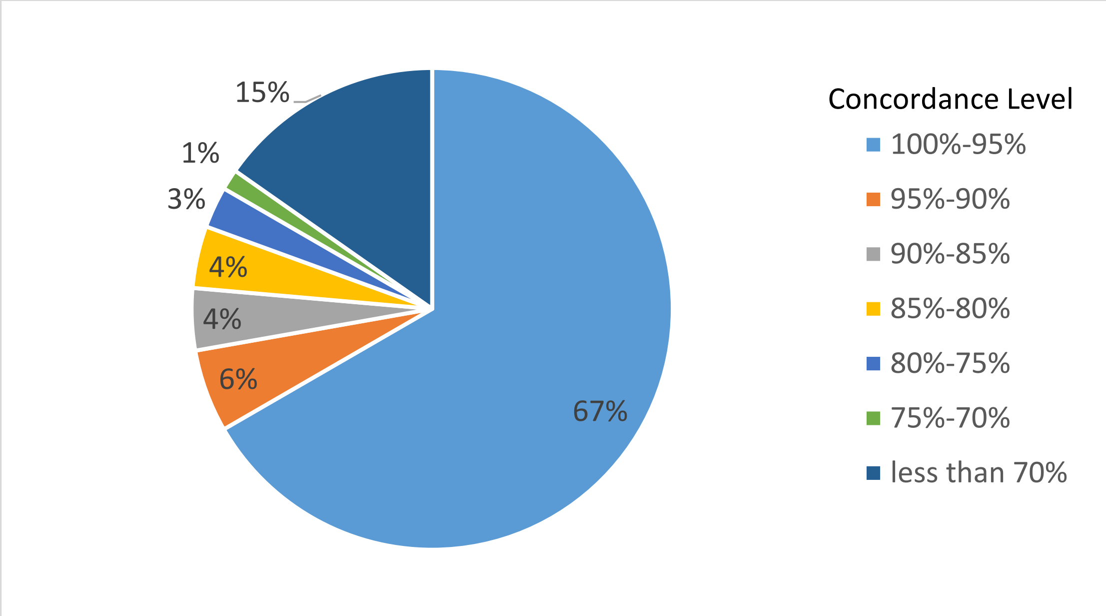

Figure 2: The concordance between the dominant node and its neighbors.

In Figure 2, we show the concordance level inside each one of the resulting mT-MDS neighborhoods. In 48 of the 72 neighborhoods, or 67% of the total, we have a more than 95% concordance between the dominant node and its neighbors.

6. Conclusion

In this paper we presented an extension of an already proposed methodology of Graph Theory to optimize the supervision of complex banking networks. We introduce the multivariate Threshold-Minimum Dominating Set (mT-MDS), to identify a subset of banks with three distinct characteristics: a) small in size b) representative enough for the whole network, and c) able to act as a gauge of its respective neighborhoods. The proposed methodology outperforms the T-MDS method in terms of efficient classification of both solvent and failed banks. We attribute this superiority in the ability to form multivariate networks exploiting information from different variables. Considering the latest consensus which supports that the supervision of the whole banking system should be empowered to a single authority we recommend the mT-MDS method as an auxiliary monitoring tool. The method's efficiency is tested in a network of 570 American banks, 429 solvent and 141 failed. The network is constructed in terms of two variables: the Tier 1 (core) risk based capital over Total Assets and the Total Interest Expense over Total Interest Income. We compare the performance of both methods (mT-MDS and T-MDS) in terms of efficiency and accuracy. The results indicate that the multivariate T-MDS method outperforms the respective univariate T-MDS method in terms of absolute number of labeled nodes and ratios as well.

References

F. Allen and D. Gale (2000), "*Financial contagion*", The Journal of Political Economy, Vol. 108, no 1, pp. 1–33

P. Angelini, G. Maresca, D. Russo (1996), "*Systemic risk in the netting system*", Journal of Banking and Finance, Vo. 20 no. 5, pp. 853-868.

G. Bonanno, G. Caldarelli, F. Lillo, S. Micciche, N. Vandewalle, and R. N. Mantegna (2004), "*Networks of equities in financial markets*", European Physical Journal B, Vol. 38, pp. 363-371.

S.P. Borgatti and M. G. Everett (1999), "Models of core/periphery structures", Soc. Networks 21 375–395.

M. Boss, H. Elsinger, M. Summer, and S. Thurner (2004), "*Network topology of the interbank market*", Quantitative Finance, Vol. 4, no. 6, pp. 677–684.

J. A. Chan-Lau (2010), *"Balance Sheet Network Analysis of Too-Connected-to-Fail Risk in Global and Domestic Banking Systems*", Vol. 10, International Monetary Fund.

M. R. Garey, D. S. Johnson, "*Computers and Intractability: A Guide to the Theory of NP-Completeness*", W.H Freeman and Company, New York, USA, (1979).

P. Gogas, Th. Papadimitriou, M.A. Matthaiou (2016), "*Bank supervision using the Threshold – Minimum Dominating Set*", Physica A, Vol. 451, pp. 23-35.

P. Gogas, Th. Papadimitriou, A. Agrapetidou, Anna (2018), "*Forecasting Bank Failures and Stress Testing: A Machine Learning Approach*" (2018), International Journal of Forecasting, 440-455.

G. Hoggarth, R. Reis, V. Saporta (2002), "*Costs of banking system instability: some empirical evidence*", Journal of Banking & Finance, Vol. 26, no. 5, pp. 825-855.

L. Huang, Y. C. Lai, and G. Chen (2008), "*Understanding and preventing cascading breakdown in complex clustered networks*". Physical Review E, Vol. 78, no. 3, 036116

H. Inaoka, Hideki Takayasu, T. Shimizu, T. Ninomiya, K. Taniguchi (2004), "*Self-similarity of banking network*", Physica A, Vol. 339, no. 3, pp. 621-634

G. Iori, S. Jafarey, and F. G. Padilha (2006), "*Systemic risk on the interbank market*", Journal of Economic Behavior and Organization, Vol. 61, no. 4, pp. 525–542

G. Iori, G. D Masi, O. V Precup, G. Gabbi, and G. Caldarelli (2008), "*A network analysis of the Italian overnight money market*", Journal of Economic Dynamics and Control, Vol. 32, no. 1, pp 259 – 278

S. Kumar, and N. Deo (2012), "*Correlation and network analysis of global financial indicies*", Physical Review E, Vol. 86, pp. 026101(1)-026101(8)

J. Lorenz, S. Battiston, and F. Schweitzer (2007), "*Systemic risk in a unifying framework for cascading processes on networks*", European Physical Journal B, Vol. 71, no. 4, pp. 441–460

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

Y. Leitner (2005), "*Financial networks: contagion, commitment and private sector bailouts*", Journal of Finance, Vol. 60, no. 6, pp. 2925-2953

S. Lyocsa, T. Vyrost, and E. Baumohl (2012), "*Stock market networks: The dynamic conditional correlation approach*", Physica A, Vol. 391, pp. 4147-4159.

C. Furfine (2003), *"Interbank Exposures: Quantifying the Risk of Contagion,"* Journal of Money, Credit and Banking, Vol. 35, no. 1, pp. 111–128.

R. N. Mantegna (1999), "*Hierarchical Structure in Financial Markets*", European Physical Journal B, Vol.11, pp. 193-197.

C. Minoiu, J. A. Reyes (2011), "*A networks analysis of global banking: 1978-2009*", No. 11/74, International Monetary Fund.

J.C Nacher, T. Akutsu (2012), *"Dominating scale-free networks with variable scaling exponent: Heterogeneous networks are not difficult to control"*, New Journal of Physics, Vol. 14, no.7, pp. 1-24

A. Nagurney, and S. Siokos (1997), "*Financial Networks – Statics and Dynamics"*, Springer, Heidelberg, Germany.

J.-P. Onnela, K. Kaski, and J.Kertesz (2004), "*Clustering and information in correlation based financial networks*", European Physical Journal B, Vol. 38, pp. 353-362

Th. Papadimitriou, P. Gogas, B.M. Tabak (2013), *"Complex Networks and Banking Systems Supervision", Physica A: Statistical Mechanics and its Applications,* **Vol. 392, no.** 19, pp. 4429 – 4434

L. Sandoval (2012), "*Pruning a minimum spanning tree*", Physica A, Vol. 391, pp. 2678-2711

J. Schleich, H. Thi, P. Bouvry (2011), "*Solving the minimum M-dominating set problem by a continuous optimization approach based on DC programming and DCA*", Journal of Combinatorial Optimization, Vol. 24, no. 4, pp. 397-412

G. Sheldon, and M. Maurer, (1998), "*Interbank Lending and Systemic Risk: An Empirical Analysis for Switzerland,*" Revue Suisse d'Economie Politique et de Statistique, Vol. 134, pp. 685-704.

M. Steen (2010), "*Graph Theory and Complex Networks*", Maarten van Steen, Amsterdam, Netherlands,

B. M. Tabak, T. R. Serra, and D. O. Cajueiro (2010), "*Topological properties of stock market networks: The case of Brazil*", Physica A, Vol.389, pp. 3240-3249

B.M Tabak, M Takami, J.M.C. Rocha, D.O. Cajuero (2014), "*Directed clustering coefficient as a measure of systemic risk in complex networks*", *Physica A: Statistical Mechanics and its Applications*, Vol. 394, pp. 211-216.

S. Thurner, R. Hanel, S. Pichler (2003), "*Risk trading, network topology and banking regulation*", Quantitative Finance, Vol. 3, no. 4, pp. 306-319.

M. Tumminello, T. Di Matteo, T. Aste, and R.N. Mantegna (2007), "*Correlation based networks of equity returns sampled at different time horizons*", European Physical Journal B, Vol.55, pp. 209-217

C. Upper, and A. Worms, (2004), "*Estimating Bilateral Exposures in the German Interbank Market: Is there a Danger of Contagion?*" European Economic Review, Vol. 48, no. 3, pp. 827-849

S. Wells, (2004), "*Financial Interlinkages in the U.K. Interbank Market and the Risk of Contagion*," Working Paper No. 230 (Bank of England)