TALLINN UNIVERSITY OF TECHNOLOGY
Faculty of Information Technology

IDN70LT

Mihhail Verhovtsov 120029IAPM

INVESTIGATING OWNERSHIP STRUCTURE OF BANKS WITH NETWORK THEORY USING BANKSCOPE DATABASE 2003−2013

Master's thesis

Supervisor: Innar Liiv
Doctor of Philosophy
Associate Professor

Tallinn 2016

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

IDN70LT

Mihhail Verhovtsov 120029IAPM

PANKADE OMANDISTRUKTUURI UURIMINE VÕRGUSTIKE TEOORIAGA BANKSCOPE ANDMEBAASIGA 2003−2013.

magistritöö

Juhendaja: Innar Liiv
Filosoofiadoktor
Dotsent

Tallinn 2016

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Mihhail Verhovtsov

09.05.2016

Abstract

Ownership relations between banks are rarely studied as a standalone network in the literature, which focuses more on interbank lending networks and general investment markets. This thesis fills this gap by building bank ownership graph from Bankscope data for the period of 2003–2013. The first goal of this study is to describe the topology and properties of the network, including the ones of country-level view and accounting for the time dimension. Visualization and network metrics are used for this part. The second goal is to find the bridge between balance sheet variables of banks and their structural positions. To this aim, social network analysis and statistical methods are used. Additionally, it is tested whether centrality measures can be used to predict changes in bank's wealth.

The results show that bank ownership network is formed around bank groups and is similar to traditional banking network in terms of low clustering coefficient, short average path and negative assortativity. The key difference is that bank groups are loosely coupled. In the country-level view, nodes are more tightly coupled and fit core-periphery model. Some centrality measures strongly correlate with bank sheet variables, confirming that key players in the network are the wealthiest banks and countries.

This thesis is written in English and is 47 pages long, including 6 chapters, 23 figures and 6 tables.

Annotatsioon

Pankade omandistruktuuri uurimine võrgustike teooriaga Bankscope andmebaasiga 2003−2013.

Pankade omandistruktuur on harva uuritud eraldiseisva võrguna kuna kirjandus fokuseerib pigem pankadevahelise laenuvõrgu ja üldise investeerimisturu peale. Käesolev töö täidab seda puudust, ehitades pankade omandivõrku Bankscope andmebaasi andmete põhjal perioodiks 2003−2013. Uuringu esimene eesmärk on kirjeldada selle võrgu topoloogiat ja omadusi, kaasa arvatud riikide tasemel ning arvestades ajadimensiooniga. Selle jaoks on kasutatud visualiseerimist ja võrgustike teooriat. Teine eesmärk on leida seos pankade bilansiaruande muutujate ja struktuurse positsiooni vahel. Selleks on rakendatud sotsiaalvõrgustike analüüs ja statistilised meetodid. Lisaks on testitud, kas tsentraalsuse mõõdikute abil on võimalik prognoosida panga heaolu muudatusi.

Tulemused näitavad, et pankadevaheline omandivõrk on moodustatud pangagruppide ümber ning on sarnane traditsiooniliste pankade võrkudega madala klasterdamise koeffitsiendi, lühike keskmise tee ja negatiivse assortatiivsuse poolest. Põhiliseks erinevuseks on nõrk pangagruppide omavaheline sidestatus. Samuti, riikide taseme vaade võrgust on rohkem sidestaatud ja sobib tuum-perifeeria mudelile. Mõned tsentraaslsuse mõõdikud korrelleeruvad bilansiaruande muutujatega, kinnitades et võrgu võtmepangad on kõige rikkamad.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 47 leheküljel, 6 peatükki, 23 joonist, 6 tabelit.

Acknowledgements

I would like to express my gratitude to my supervisor Innar Liiv for his advice on the direction of this thesis, useful suggestions on methodology and supportive attitude. Furthermore, I would like to thank Laivi Laidroo for being my guide to the world of economic and financial research. Finally, I would like to thank Leo Võhandu for his valuable ideas and insights into banking world.

List of abbreviations and terms

CSV	Comma-separated values, a plain-text tabular data format.
JSON	JavaScript Object Notation, a lightweight human-readable data-interchange format.
PNG	Portable Network Graphics, a bitmap image file format.
SNA	Social Network Analysis
XLS	Microsoft Excel file format, a spreadsheet file format.
Blockmodel	A reduced representation of the graph, where nodes are collapsed into block depending on the chosen node set partitioning.
Block modeling	The process of creating a blockmodel.
IQR	Interquartile range, the difference between the third and the first quartiles in descriptive statistic.

Table of contents

List of figures

List of tables

1. Introduction

1.1. Bank ownership network

Banks are owned fully or partially by private or public shareholders, some of which are also banks. A common example is when a head bank owns subsidiaries in other countries. Another example is consortium banks, which are subsidiary banks created by numerous other banks for funding a specific project [1]. Berger et al. [2] list several reasons for banks to own other banks, calling it a *multiple banking relationship*. According to them, ownership of another bank is necessary when one bank cannot provide all the needs of a firm e.g. due to size, geographic dispersion and service type diversity. Other reasons concern minifying risks and overcoming economic problems. This ownership relationship between banks forms a *bank ownership network*.

Bank ownership network is rarely studied in the literature on its own. Instead, researchers mainly focus on the more dynamic interbank lending network. In these works, bank ownership type is taken into account only as an attribute. Therefore, it would be interesting to fill the gap in the literature by looking solely on the interbank ownership network and comparing it to other financial networks that banks are involved in.

More general papers [3] consider bank ownership network as a part of market investment network (i.e. treating banks as stocks and including all types of shareholders). This approach does not take into account differences in behavior of banks and other market entities. This thesis focuses on interbank network only, treating ownership as a kind of relationships.

Ownership is a relation that has influence on both the owner and the asset. It creates an information channel between the two entities which can influence their behavior. Headquarter banks pass group-wise strategy to their subsidiaries. Additionally, there can be an exchange of management technologies and know-how [4].

Another notable role that ownership relation plays occurs during economic shocks, specifically within the capital market of a bank group. When subsidiary's financial health deteriorates, it can receive financial help from head bank. Headquarter provides its asset with a loan or equity and in turn has to withdraw money from international financial markets. In order to improve group performance, headquarter can shift its funds from one subsidiary to another [5]. However, empirical evidence shows that head banks tend to act as a lender of last resort due to the desire to maintain their reputation [6].

Bank ownership has a number of variables that can classify it, including the number and the types of shareholders. Studies show that there is a difference in behavior between state-owned and private banks and also between domestic and foreign ownership.

In general, banks are established, bought and sold depending on the global and local economic situations. Therefore, the properties of bank ownership network may reflect upon the financial wealth of its entities.

1.2. Research questions

The aim of this thesis is two-fold. First, the structure of European bank ownership network is investigated using shareholder data for the period of 2003−2013 provided by Bankscope database. In addition to individual bank level, the network is studied on bank group (concern) and country levels by using block modeling technique. The study uses both visualizations and aggregated metrics in order to gain insight into topological structure and properties of the networks.

Studying statistical and structural properties of the bank ownership network over period of decade and comparing it to other networks like interbank lending market gives us information on its formation, functionality and stability. Drawing from these findings, the second part of this thesis tries to link the obtained results with economic situation. This is done by testing whether bank financial indices (balance sheet variables) are related to bank's structural position. By taking the time axis into account, another idea that one can predict changes in bank's wealth by observing changes in structural position is also tested.

The overall motivation for studying bank ownership network is to supplement previous studies, both theoretical and empirical. In this regard, this thesis is similar in its goals with the work of Craig and von Peter [7], who analyzed the structure of interbank lending network and its relation to financial indices of banks. The overview of related work shows that information on bank ownership is required by many Economic studies and this thesis can provide such ground knowledge from network-theoretical point of view.

1.3. Related work

Banking networks

As mentioned previously, much research is focused on bank lending market rather than on interbank ownership relationships. However, such literature helps this thesis by establishing a comparison ground and describing methods of analysis that can be reused in given research. Other examples of studied financial networks that are related to banks include international trade network, investment networks and interbank clearing networks.

Hale [8] views lending as an action that "establishes a relationship and produces information flows between the lender and the borrower, which in turn facilitate further lending". From this perspective, lending is similar to ownership as a bilateral relation, but is much more dynamic.

Minoiu and Reyes [9] analyzed global banking in 1978−2009 using network metrics of interconnectedness such as centrality, connectivity, and clustering. They concluded that lending network is unstable, contains structural breaks and its metrics are volatile. The global crisis of 2008−2009 caused large perturbation preceded by a stable period.

A thorough exploration of interbank payment flow topology was undertaken by Soramäki et al. [10]. They describe the network as having a low average path and low connectivity with scale-free degree distribution. They also note the existence of a large tightly connected core of banks.

Boss et al. [11] analyzed the network structure of Austrian interbank lending market and found that "the banking network has the typical structural features found in numerous other complex real-world networks: a low clustering coefficient and a short average path length". Additionally, "banks first have links with their head institution, whereas the few head institutions have links with each other. A consequence of this structure is that the interbank network is a small world with a very low degree of separation between any two nodes in the system". Their goal for studying network structure was to determine how it affects the financial stability properties of the banking system as a whole.

Craig and von Peter [7] propose a core-periphery model of tiered interbank market structure with an intermediation layer. According to them, smaller banks lend to each other through money bank centers. This hierarchical model was fit onto empirical observations of German banking system in the period from 1999 to 2012. They conclude that the interbank market structure is very stable and unlikely to be formed in a standard random network. As a way to explain formation principles, they show that bank's structural position and balance sheet variables are linked.

Bank ownership

According to the empirical study of three shareholder networks (a superset of bank ownership network), conducted by Garlaschelli et al. [3], the distribution of both degrees and assets display power-law tails. Moreover, the exponents of these distributions are in constant relation with each other across all three markets.

Bank ownership is also explored in itself. Micco et al. [12] correlated bank ownership and bank performance based on the data of 50 000 observations for 119 countries over the 1995−2002 period. They concluded that bank performance depends on whether it is state- or private-owned, foreign or domestic and located in developing or industrial country.

A similar research by Nicolò et al. [13] studies joint effects of bank ownership and market structure on bank's risk profiles. It finds that risk profiles depend on the ownership types and their market shares. Additionally, it provides information on the market shares of banks by ownership, revealing that the majority of banks were state-owned during the period of 1994−2003.

Financial crisis and contagion model

The global financial crisis of 2008 heated up interest in bank network structures in the context of contagion model. Researchers studied the role of the network structure in contagion effect. Caccioli et al. [14] considered a model of contagion that took bank (lending) network structure heterogeneity into account. Namely, they pointed out the heterogeneity of degree and asset distribution, power law distribution of balance sheet size and disassortative nature of the network.

Allen et al [15] used Bankscope data to test several hypothesis about the effects of foreign and government ownership on bank lending behavior during a crisis.

Chinazzi et al. [16] studied International Financial Network at country-level in order to see if its topological properties changed after 2008 financial crisis. They concluded that, although the topology did change, the disassortative, core-periphery structure still remained. Their research used methods similar to the ones used in this paper, like exploring correlations between network statistics, graph visualization and plotting network metrics over a period of time.

1.4. Outline of the thesis

The remainder of this thesis consists of five chapters. The second chapter outlines the methodology used in this research, particularly, complex and social network analysis. The third chapter describes the input data used to build the ownership graph and documents all the trade-offs and transformations made in order to create a clean and analyzable dataset. The fourth chapter looks at the bank ownership network topology and its properties. To this aim, visualizations and graph metrics are used. The fifth chapter studies the relation between ownership network structure and financial indices. Finally, the sixth chapter concludes all the research.

2. Methodology

This chapter describes the analysis methods used in this research alongside with their implementation details. Network theory is extensively used throughout this work. As a discipline, network theory was developed outside of the economic studies and is steadily finding its application in the analysis of financial networks. The use of social network analysis, which is an addition to traditional methods, can be regarded as a novel approach.

2.1. Complex networks analysis

The concept of network is fundamental to representing many physical, social and biological phenomena. In broad terms, networks are graphs representing relations between discrete items. In this thesis, the terms *network* and *graph* are used interchangeably. Newman [17] gives a review of network types and methods used to study them. In general, network analysis involves studying network metrics and also measures of individual nodes. Below is a short description of some network properties.

Node degree is the number of node's edges. In directed graphs, one can distinguish *in-degree* as the number of incoming edges and *out-degree* as the number of outgoing edges. Node degree can also be weighted. *Average degree* is the average of all node degrees in a graph. Corresponding average in- and out-degrees can also be calculated. *Degree distribution* is important for graph classification. Many real-world networks have power-law degree distribution and are sometimes referred to as scale-free networks [17].

Average path length is the average of all shortest paths between all pairs of nodes. *Diameter* of the network is the longest of all shortest paths. Many real-world networks have small diameters, meaning that all nodes are relatively closely connected. Such networks are called small-world networks.

Clustering coefficient measures the likelihood that the two neighbors of the node are connected. It is a way to detect groups that have a high density of ties. *Global clustering coefficient* considers triplets of nodes instead of pairs. *Average clustering coefficient* is the average of local clustering coefficients. Both metrics indicate overall clustering in the network. In this thesis, average clustering coefficient is used.

Assortative mixing or *assortativity* is the preference of similar network nodes to attach to each other. Although similarity can be measured differently, similarity by node degree is commonly used. *Assortativity coefficient* is defined as a Pearson's correlation coefficient of degree between pairs of linked nodes. A positive value shows that the network is *assortative* i.e. similar nodes indeed tend to connect to each other, while the negative value shows that the network is *disassortative*.

Graph metric calculation algorithms were implemented by the author using *JGraphT* library [18] to model the graph. Calculated metrics include average path length, network diameter, network density, average degree (both binary and weighted), assortativity coefficient and average clustering coefficient. Calculation of these metrics treated input graph as undirected. View pattern, supported by JGraphT, was used to provide algorithms with data relevant to the analyzed year.

2.2. Social network analysis

A social network is a set of individuals (persons, groups, organizations or social entities) that interact with each other in some way. Examples of interaction include friendship between people, communication in a group, intermarriages between families, business relationships between companies and transactions between corporations. The core idea of social networks is that individuals are affected by their neighborhood. Social network analysis (SNA) can identify important (influential) individuals, discover communities or identify actors that are similar is some way [19].

SNA assigns individual properties called *centrality measures* to all nodes in the graph. They identify the most important nodes in the network and are calculated based on the node's neighborhood. The meaning of importance depends on both network's semantic and chosen centrality type. There is no "best" centrality measure. Moreover, there are a number of centralities designed to fit different observed phenomena. The most common centralities, known from works of Freeman [20] [21] and Bonacich [22], are:

- *Degree centrality*, which ranks nodes by their degree. The most important ones are those which have the highest degree. In case of directed network, in-degrees and out-degrees can be used to calculate separate measures.

- *Closeness centrality*, which highlights nodes which have the lowest average of all shortest paths to all other nodes, therefore being “close” to all other nodes.
- *Betweenness centrality*, which is a measure that is based on the number of shortest paths that pass through the node. Nodes with high betweenness centrality ranking are crucial for information flow.
- *Eigenvector centrality*, which measures the influence of a node in a network by identifying nodes that are connected to many other well-connected nodes.

Definitions of these centrality measures can be extended to weighted graphs. Also, variations of these definitions exist to deal with edge direction, disconnected nodes and other features of networks.

This research uses only degree and closeness centralities. Other centrality measures like betweenness centrality and eigenvector centrality were also considered in the early stages, but were omitted for different reasons. Betweenness centrality [23] turned out to be too chaotic to provide sensible results. This is, probably, due to the incompleteness of data and sampling effects magnified by filtering. Eigenvector centrality [24] was found to be very similar to degree centrality and was omitted for brevity. Instead, in addition to the graph with binary links, it was decided to bring in two weighted versions of the graph and, correspondingly, two weighted versions of degree and closeness centralities. Section 5.1 describes chosen weights in detail.

Implementations of centrality calculating algorithms were provided by open-source *jgrapht-sna* library [25], which utilizes JGraphT graph model. Degree centrality was implemented as described in [20]. Weighted degree centrality was implemented as described in [26]. Dangalchev closeness centrality [27] was used for unweighted closeness centrality measures, while implementation as described in [26] and based on the code from *jgrapht-sna* was used for weighted one.

In case of weighted closeness centrality, the traditional path shortness as measure of influence did not fit the model well. Instead, a stronger path was the preferable one as influence transfers better through stronger (larger) ownerships. Therefore, weighted closeness centrality algorithm was inverted to find nodes with the strongest (longest) average path compared to other nodes. Following formula was used to calculate weighted closeness centrality $C_w(n)$ for each node n:

$$C_w(n) = \sum_{v \in V, v \neq n} sp_n(v) \cdot \frac{|V_n|}{|V|}$$

Here V is the set of all nodes, V_n is the set of nodes reachable by node n, $sp_n(v)$ is the longest path from node n to node v, equal to 0 for unreachable nodes. This definition also promotes highly connected nodes. The code of implementation is available in Appendix G.

2.3. Block modeling

Block modeling is an intuitive graph reduction technique, where graph nodes are collapsed into blocks according to the partitioning of the node set. Figure 1 shows an example of this process for nodes are partitioned by color. In the reduced graph each partition of the original nodes is represented by a single node. Graph edges are collapsed and consolidated correspondingly. In the process of block modeling node and edge attributes like weight are aggregated.

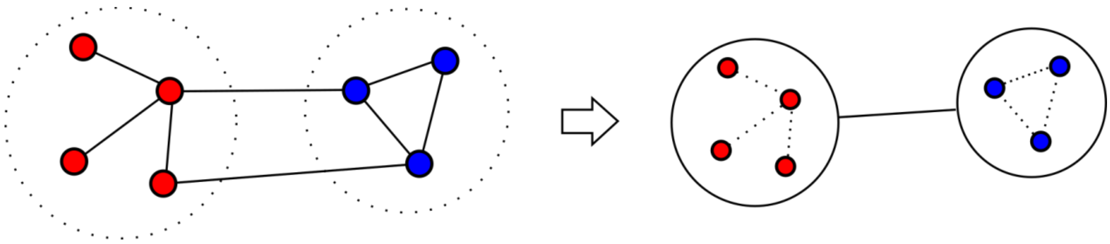

Figure 1. Example of block modeling by color.

The reason for using block modeling is to better understand interactions between different partitions of the network like bank groups or countries. In this sense, block model is a higher-level view of the original graph.

2.4. Statistical methods

Both metrics and measures obtained in the process of network analysis can be treated to common statistical methods of research like looking at distributions of series and calculating Pearson's correlation coefficient [28] between pairs of series. This is particularly of use when it is necessary to aggregate the individual measures of nodes into a single variable.

Though there is no common agreement upon the range and names of different strength levels of Pearson's correlation coefficient, this thesis uses definitions shown in Table 1 to describe statistical results.

Table 1. Interpretation of Pearson's correlation coefficient

Coefficient value	Strength of relationship
−1.0 to −0.5 or 1.0 to 0.5	Strong
−0.5 to −0.3 or 0.3 to 0.5	Moderate
−0.3 to −0.1 or 0.1 to 0.3	Weak
−0.1 to 0.1	None or very weak

Apache Commons Math library [29] was used to calculate correlation coefficients between series of measures as well as to find p-values.

As mentioned previously, several related works [12] [16] use similar methods of analysis, including statistical ones. Correlation of centrality measures to other node measures is not uncommon method. This approach was undertaken by Abbasi et al. [30] and Soheili et al. [31] to study the co-authorship network. They show that there is a positive significant correlation between some centrality measures and performance indices that are measured independently of network structure. Abbasi et al. concluded, that professional social network can be used to predict the future performance of researchers. Correspondingly, this work tests if bank ownership network can be used to predict performance of banks.

2.5. Visualizations

Graph visualizations featured in this paper [Appendix A] are generated by the interactive web panel created for this research [32]. For this purpose, all graph models along with calculated measures were exported into JSON format using *Jackson* library [33]. The interactive panel was created using HTML5, JavaScript and the *D3.js* visualization library [34]. Working with the graph interactively allowed detecting and highlighting particular features of the network. While the panel renders graph for each year of the studied period, for brevity, only snapshots for one year (2006) are featured in this paper.

In addition to network visualizations, time dimension required some resourcefulness in representing other analysis results. For example, graph metrics are shown as line charts rendered with help of *JFreeChart* library [35]. In order to show distributions of various measures during the analyzed period of 2003−2013, 3D-histograms were used, rendered by author’s own charting library. Correlation matrix shows a bar chart of correlation coefficients and a corresponding bar chart of p-values for each year of the observed period for each pair of series. In-degree/out-degree plot, featured in Section 4.2, is also a snapshot of interactive panel created using *D3.js.*

3. Data

Bank ownership network was created from empirical data provided by the Bankscope database. First of all, this chapter introduces the source of the data and describes different datasets that were used. It then goes on to explain how these datasets were combined to create the ownership network. Next, the post-processing steps are documented, including all of the trade-offs and assumptions made to bring the data into analyzable form. Finally, the transformation of the original graph into three derivative ones by the means of block modeling technique is discussed.

2.1. About the Bankscope database

Bankscope database provides information about public and private banks throughout the world. It is a product of Bureau van Dijk, which specializes on gathering information about companies. Bankscope, as per its overview webpage, "combines comprehensive financial statements with a wide range of other banking intelligence including ratings, an analysis model, bank structures, news, AML [i.e. Anti-Money Laundering] documentation and banking research. [...] Bankscope has information on 32,000 banks and is the definitive tool for bank research and analysis" [36].

From a researcher's perspective, Bankscope provides an interface to query its multidimensional database and output search results in a range of formats.

2.2. Bank shareholders and financial indices

This research combines several datasets obtained through Bankscope. The first dataset is the list of all shareholders per bank, including share sizes for each year during the period of 2003–2013. Ownership data prior to 2003 was available, but proved to be too incomplete for inclusion in the analysis. This dataset represents a list of European banks, which contains a sub-list of their shareholders (i.e. owners). Banks from the sub-list (owners) may also occur in the first list (owned). Equal banks from the list and sub-lists were merged into one graph node.

The second dataset consists of bank financial indices, also known as balance sheet variables, for the banks in the first dataset. These values are unconsolidated and there is a separate sheet for each year of the observed period.

The two datasets were parsed and combined into one dataset, represented as a directed graph. Nodes of this graph are the banks and links are ownership relations between them (shareholding relation). A link from one node to another means that the former holds a share of the latter. Additional data (share sizes, financial indices) is kept as attributes of corresponding nodes and links. Due to the dynamic nature of the graph, a link between banks exists if there was an ownership relation at least as short as one year within the observed period. The analysis accounts for this case by using weighted measures and view pattern which treat links with zero weights as non-existent.

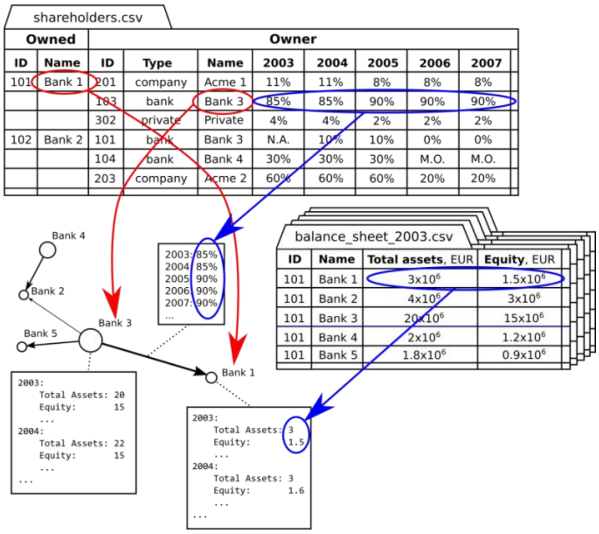

Figure 2. Combing shareholder data and balance sheets into ownership graph

All query results came in form of Excel spreadsheet documents (XLS), containing both original query description and the obtained result. In order to process the data, Bankscope query output was converted into CSV files. Data reading and processing was done by a Java program which scanned input CSV files and built an in-memory model of the ownership network using *JGraphT* library. This code is available in Appendix H. The process of combining data into the graph is graphically explained in Figure 2.

2.3. Attribute description and interpretation

Bank shareholder dataset required some interpretation and unification of share sizes. Besides values expressed as numeric percentage, there were abbreviations for some common values as well as missing values. Table 2 lists these exceptions and describes the way they were interpreted.

Table 2. Share size value interpretation

Input value	**Interpretation**	**Description**
empty cell	0%	No value or missing
n.a.	0%	No value or missing

Input value	Interpretation	Description
-	0%	No value or missing
MO	50%	Majority Owned
WO	100%	Wholly Owned
NG	0%	Negligible
CQP1	50%	50% + 1 share

It should be noted that empty values were interpreted as 0% share according to the previously described approach of links in the dynamic graph. An alternative would be to represent each year with a separate graph. However, this path was not selected because JGraphT allows the use of *view* pattern to create a one-year snapshot representation of the graph.

The list of financial indices provided by the corresponding dataset is listed in Table 3.

Table 3. List of financial indices (balance sheet variables) that occur in the dataset

Attribute	Description
Total assets	The sum of current and long-term assets owned by the bank.
Equity	Total assets minus total liabilities; net worth.
Net loans	Total loans to customers, reduced by possible default losses and unearned interest income.
Gross loans	The total amount of issued credits given to banks during the accounting period.
Total customer deposits	The sum of all customer deposits.
Net income	Net profit; all income minus all expenses.
Pre-impairment operating profit	Operating profits before impairment charges.

Banks from both datasets were matched by Bankscope's *BvD ID* – a unique bank number that does not depend on bank's name.

2.4. Data adjustments

A closer inspection revealed problems with data completeness. First, some banks from the shareholders dataset were missing their counterpart from the balance sheets dataset. This was solved by removing these banks. Next, a large number of shareholder relations had associated share size values partially missing. It was first attempted to fix the gaps by interpolating values. After that, shareholding relationships that still had too little share values (less than 9) were filtered out.

The resulting graph consisted of multiple components with many nodes completely disconnected due to the previous filtering of links with missing share size values. There was one a large central component and a number of small ones. The graph was further filtered to leave out all components besides the largest one.

Table 4 shows the number of nodes and links before and after the adjustment steps. One can see that the initially large dataset was downsized into a much smaller, but analyzable dataset.

Table 4. Adjustment steps and affected/left node/link count.

#	Adjustment	Affected		Left	
		Nodes	**Links**	**Nodes**	**Links**
0	Loading initial graph	-	-	11834	21486
1	Removing links with too many missing values (less than 9 values present)	0	20063	11834	1423
2	Fixing gaps in share size values	0	1423	11834	1423
3	Removing banks with missing size data	1726	199	10108	1224
4	Removing all disconnected banks	8747	0	1361	1224
5	Removing all but the largest component	942	760	419	464

Additionally, total share size was calculated for each owned bank in order to validate the data. This sum was expected to be 100% in case of complete data. In the original graph, only 393 banks had sum of all shares for each year equal to 100% and for 8317 banks this sum was either 100% or zero. In the final graph, out of 419 nodes only 80 had complete data and 170 had either 100% sum of shares or none.

Social network studies often suffer from data sampling and data errors. There are works [37] that researched the stability of centrality measures in error-prone networks. The general conclusion was that these measures change gradually and therefore can be treated as reliable. Other works have considered the effects of sampling on centrality measures and found that centrality measures can also hold for sampled data.

2.6. Block modeling by bank group and by country

In order to stress out certain structural features of the bank ownership network, three higher-level graphs were derived from the original one using block modeling – two bank group-level graphs and one country-level graph. The code that performs block modeling is available in Appendix I.

Partitioning by bank group

Partitioning by bank group was done based on the bank's name. The list of name patterns for each bank group was composed manually. Matching process took possible name variations and abbreviations into account. For example, banks with names "RBS Bank (Polska) SA" and "Royal Bank of Scotland ZAO (The)" were both attributed to "Royal Bank of Scotland" group. While this method is not ideal, it still allows for understanding the role of bank group in the network as shown in Chapter 4.

In the process of node consolidation, financial indices were also consolidated. Share size percentages and their absolute values were aggregated upon edge consolidation as well. Edge directions were respected. Produced blockmodel is also a directed graph. Table 5 lists all bank groups featured in this thesis.

Table 5. List of bank groups and the number of banks they consist of.

Group	Count	Group	Count
Other	237	FGA	5
Credit Agricole	37	UniCredit	4
Caja Rural	36	Societe Generale	4
BNP Paribas	12	Commerzbank	3
ProCredit	9	WGZ-Bank	2
Deutsche Bank	9	Norddeutsche Landesbank	2
Skandinaviska Enskilda Banken	9	LBB	2
KBC	7	Hypothekenbank Frankfurt	2
ING Bank	7	Santander	2
NLB	7	Intesa Sanpaolo	1
Volkswagen	7	Banco Comercial	1
HSBC	6	UBS	1
Royal Bank of Scotland	6	Erste	1

Some of the banks in the dataset are standalone, meaning that they do not belong to any group as detected by the algorithm. It was decided to produce two bank group-level graphs – one where all standalone banks are consolidated into a surrogate group "Other" and one where all standalone banks are left as they are. The reason for both approaches is explained in Chapter 4. Whenever the research does not mention which group-level graph it is referring to, the one with separate standalone banks should be assumed. This version of the graph contains 262 nodes and 303 links.

Partitioning by country

Partitioning by country was based on the country code attribute of the bank taken from the original dataset. Consolidation of nodes was done according to similar rules as in case of bank groups with one exception that the resulting graph is undirected. Resulting graph consists of 35 nodes and 84 links.

Table 6 lists all the countries that participate in the analysis and the corresponding number of banks they contain. It also allows for assessing the sampling level of the data. For example, Estonian Banking Association lists 14 commercial banks as of year 2014 [38], with four of them having at least 10% share of the market. Only one bank is featured in the filtered graph.

Table 6. List of countries and the number of banks they consist of.

Country	**Count**	**Country**	**Count**	**Country**	**Count**
Germany	136	Ireland	6	Slovakia	2
France	74	Austria	5	Turkey	2
Spain	49	Switzerland	5	Albania	1
Portugal	19	Russia	5	Cyprus	1
Netherlands	15	Serbia	4	Denmark	1
Belgium	14	Bulgaria	3	Moldova	1
Italy	11	Hungary	3	Norway	1
Luxembourg	11	Romania	3	Lithuania	1
Czech Republic	11	Sweden	2	Ukraine	1
Slovenia	9	Bosnia & Herzegovina	2	Estonia	1
Poland	8	Greece	2	Latvia	1
United Kingdom	7	Macedonia	2		

4. Network structure

This chapter focuses on the topology of the bank ownership network including derived block models. First of all, visualizations of the networks [Appendix A] are examined to get the general picture and find possible structural patterns. Next, networks are characterized by their graph metrics [Appendix B]. Finally, the conclusion section discusses the results and provides comparison to other networks.

4.1. Visualization of network structure

Bank-level graph

The renderings of bank ownership network [Figure 6–Figure 8] represent a tiered sparse graph with two-three levels deep partial hierarchies. Some nodes appear disconnected because the link that connects them has zero weight in that particular year. Renderings with different link weights demonstrate that the majority of shares are very small. Absolutely weighted links show that the real value of assets is even lower. Link weights are described in Section 5.1.

The most prominent structural patterns are the star-shaped clusters which usually have larger banks in the center. In these clusters, the majority of links is either outgoing or incoming. These patterns appear through the whole period of observation and are relatively stable. Closer examinations reveal that they are formed by either bank groups, where one central bank directly owns its smaller subsidiaries, or by consortium banks. Child banks of head banks, as seen in the renderings, do not tend to own other banks in general, but some still do. Figure 3 features an example of such bank group cluster formed around "BNP Paribas" head bank. Names suggest that this French bank owns a few subsidiaries in other countries. The depth of ownership is at most two.

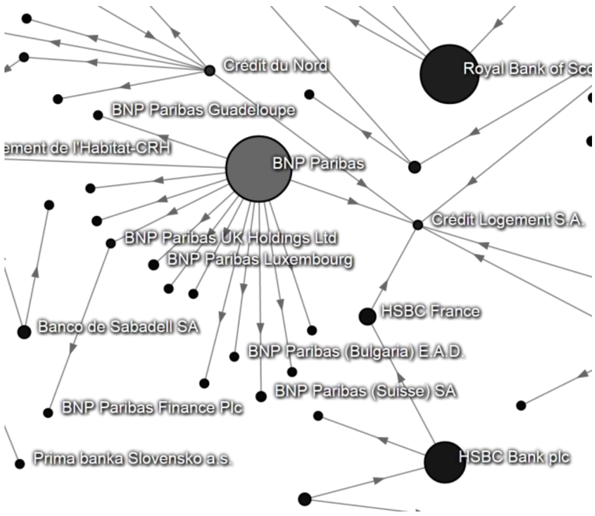

Figure 3. Example of bank group cluster

The clusters of partial hierarchy seemingly have low connectivity with the exception of German banks, which form a more tightly connected cluster (located north to “Liquiditäts-Konsortialbank GmbH” on bank-level rendering).

Group-level graph

The seemingly important role of bank group is explored in the renderings of group-level network with unconsolidated standalone banks [Figure 9−Figure 10]. This graph features less star-shaped patterns, but they still present around bank groups and also around consortium banks. This suggests that groups own some banks that are treated as standalone in this block model. For example, Societe Generale bank in Figure 6 has a large number of child banks and even some grandchildren, whereas in Figure 9 it is combined into a “Societe Generale” bank group with fewer standalone banks.

Group-level graph also shows that bank groups seemingly do not connect to each other directly, but through intermediary standalone banks. In order to validate this claim, the blockmodel with all standalone banks collapsed into a single node is rendered as shown in Figure 4. This visualization makes it clear that bank groups rarely own banks from other groups directly. However, almost all of the bank groups are related to some standalone banks.

Country-level graph

Country-level network visualization [Figure 11−Figure 12] shows a different and seemingly much denser graph than the previous ones. This graph features a tightly connected core and peripheral countries. In the core, Germany and France stand out both by the number of connections and by their total asset size. The peripheral countries are smaller by total asset size and have only one or two connections to the core.

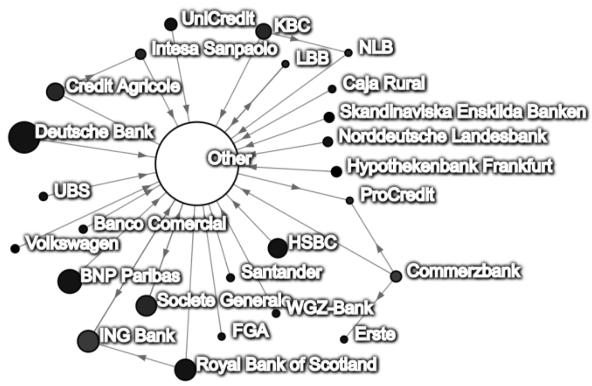

Figure 4. Group-level ownership network as year 2006 with standalone banks consolidated into a single node.

4.2. Network metrics

Treating metric values

As mentioned in Chapter 3, the set of banks under study is sampled from a larger dataset, which, in turn, does not claim to be a full dataset of European banks. Eventually, this affects the absolute values of network metrics. However, this is tolerable because these values change gradually, preserving distribution and other ratios. These values give us some indication of the magnitude of network metrics.

Network metrics are presented as line graphs [Appendix B], showing values for each year of the observed period. The slope of line segments and the local minimums and maximums are subject to graph processing settings like filtering thresholds and therefore should be used carefully in the analysis. This is why this thesis does not rely on this information in conclusions. The goal of using line chart is to show that, aside from some anomalies, metrics stay more or less the same over the time.

Bank-level graph

The diameter of the bank ownership network stays roughly between 12 and 18 [Figure 13c]; the average path length is between 5 and 7 [Figure 13b]. Despite the small average path length, the diameter makes ownership network a non-small-world network [39]. Another metric that hints a non-small-world network is low clustering coefficient.

The density of the network is low [Figure 13d], making the network a sparse graph (also confirmable visually). The average degrees are also low because the majority of banks have a degree of 1 or 2 and only a few banks have larger degrees. The distribution of degrees (in form of degree centrality) is studied in Section 5.2.

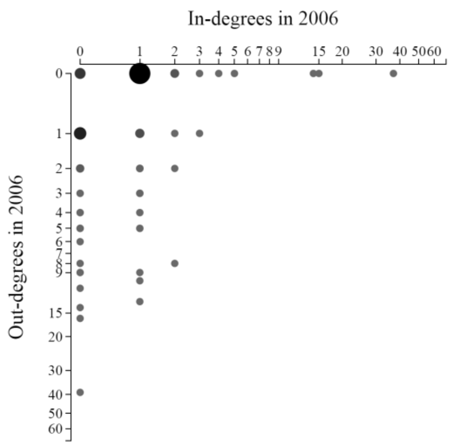

Figure 5. Plot of in- vs out-degree in bank-level network as of 2006. Dot size is proportional to pair frequency.

When talking about degrees in directed graph, it is also important to look at differences between in- and out-degrees. Figure 5 shows all the pairs of in- and out-degree. Most frequently a bank is owned by one bank and owns no other banks.

Assortativity coefficient is negative, with value closer to 0 than to −1. This indicates that connections between low-degree banks prevail over the connections between high-degree central banks and low-degree child banks.

Group-level graph

As group-level bank graph is basically a bank-level graph with one or two path levels removed around centers of bank groups, all metrics [Figure 14] have roughly the same values as their bank level counterpart.

Country-level graph

The diameter of country-level graph is around 5 [Figure 15c], the average path is slightly above 2 [Figure 15b]. Clustering coefficient is also high [Figure 15e]. Country-level graph fits the definition of a small world network. Confirming visual inspections, country level graphs are many times denser [Figure 15d], but they are still by definition sparse graphs.

Average degree is between 4 and 5, which is higher than on previous graph. Negative assortativity coefficient [Figure 15f], which is now close to −0.5, is mainly due to small-degree peripheral countries connecting to tightly interconnected central core.

4.3. Conclusion

Graph visualizations reveal that bank ownership graph is a tiered network with smaller banks gathering around larger banks. This result is consistent with findings of Craig and von Peter [7] about interbank lending market tiers and money center banks. However, the ownership network does not fit into the core-periphery model mentioned in the same work as well as in others like [10]. Central banks do not form a tightly connected ownership core, but are rather indirectly related to each other. This is logical, because while central banks can exchange payment flows, they all cannot own each other. This is the key difference between lending and ownership relation. Thus, the ownership network can be said to be multi-core, where bank groups basically play the role of cores. In addition to that, consortium banks also act as cores.

Bank groups with their direct assets are hierarchical structures (trees) that are loosely connected to each other, mainly through standalone banks. This brings out the importance of bank groups in the formation of bank ownership structure. Moreover, the absence of tight channels that transmit ownership shocks between bank groups makes ownership structure less prune to the risk of contagion and more stable. This idea can be supported by the fact that the global financial crisis of 2008 did not bring much change into the ownership structure, as both the interactive panel and metrics show. This is in contrast to large perturbations that happened in other types of banking networks during the global crisis of 2008.

Bank group hierarchical structure explains the small average path. The nature of the ownership limits the freedom of subsidiaries also limiting the number of levels in the network. Negative assortativity coefficient is indicative of financial networks. Loose coupling of the groups may explain large diameter. According to [40], large diameter may suggest that the formation of links between nodes is influenced by spatial proximity. This means that a bank is more likely to own a bank from same or neighboring country. This fits the reasoning behind *multiple banking relationship* mentioned in Chapter 1.

As for the country-level bank ownership network, the picture is different. First of all, it shows properties of small-world network. Secondly, it features a core comprised of tightly connected countries (Germany, France, Netherlands and Belgium) and periphery. Well-established banks have created many subsidiaries in neighboring countries, but only a few reached out to the geographically more remote countries.

5. Relation to financial indices

The results of the previous chapter show that the network is built around money centers (large banks and countries with large banking sector). As a next logical step in the research, this chapter seeks to determine the role financial attributes play in the formation of ownership networks. Bank-level and country-level networks are analyzed separately.

Section 5.1 describes the use of different weighted indicators in the analysis of the graphs. Section 5.2 looks at the distributions of centrality measures. The next section examines distributions of financial indices. Section 5.4 tests how centrality measures and financial indices correlate to each other when viewed year-wise. This answers the question whether larger or more successful banks are actually more central ones. Section 5.5 treats changes in centrality measures and financial indices during the period of 2003–2013 as time series and correlates them. This experiment tests if changes to centrality measures can predict changes to bank's or country's financial wealth.

5.1. The use of weighted graphs

As share size is a measure of link in the range between 0 and 100%, treating it simply as a binary relation may hide important insights. Therefore, further research uses three link weight approaches: one unweighted (i.e. binary) and two weighted ones. This effectively means that three versions of each network (bank- and country-level) were used in the analysis.

The first approach to weighted links is the use share sizes. As described in Chapter 1, an entity can have many owners, each owning some share of the whole. While share sizes are expressed in percentages, it is convenient to normalize it to the range from 0 to 1. Using share-weighted links gives the advantage of detecting changes in share sizes. In case of binary links, only changes in the ownership (whether owns or not) can be detected.

The second approach to weighted links is to use absolute values in euros, calculated as owned bank's total assets size multiplied by share size. For brevity, this approach is referred to as assets-weighted. Using assets-weighted links allows for detecting finer changes in the node's neighborhood. For example, assets-weighted centralities are affected when the share size remains the same, but the financial wealth of the owned bank changes. Also note that bank's own total asset size does not already contain child banks because unconsolidated total asset index was used.

5.2. Distributions of centrality measures

Bank-level network

Unweighted degree centrality distribution [Figure 16a] represents normalized degree distribution. Consistent with network metrics (Chapter 4), low values dominate the histogram. Degree of 1 is the mode and other popular values are 0 and 2. The histogram features several clear outliers – banks with an exceptionally large number of degrees. Therefore, this distribution may be described as a heavy-tailed power law distribution.

Share-weighted degree centrality distribution [Figure 16b] is bi-modal: one larger mode is around 0 and another one, smaller one, is around value of 1 (i.e. 100% share). This means that there is a large amount of very small shares (first mode) and also a significant amount of maximum shares (100%) contributing to the second (smaller) mode. In case of unweighted graph, these two modes are collapsed into one. Additionally, larger degree centrality measures are more evenly distributed than in unweighted version. This indicates that there are banks, which hold a lot of small shares and banks that hold a smaller number of large shares. Share-weighted links smooth out this difference.

Asset-weighted degree centrality [Figure 16c] is affected by uneven distribution of assets in the network, as Section 5.3 later shows. The histogram shows a long tail of outliers with rich neighborhood and a steeper curve for poor neighborhood than in case of share-weighted degree centrality.

The distribution of unweighted closeness centrality measure [Figure 16d] is multimodal, with modes varying over the observed period. This behavior may be related to the dynamic nature of some links. The majority of the nodes have closeness centrality measure at most half-size of the maximum value.

The introduction of weight introduces very different distributions with only one mode and very steep curve [Figure 16e, Figure 16f]. This is true for both share-weighted and assets-weighted closeness centrality, with the latter being steeper. Intuitively, head banks and their children are connected by short paths. Therefore, closeness centrality should be on par for all banks. However, it seems that the “lengths” of the paths in the network are drastically affected by the weights of relations.

Country-level network

On the country level, unweighted degrees [Figure 17a] are spread more gradually among the countries. Still, the distribution is right-skewed as the number of peripheral countries is larger than the number of core ones. Weight affects centralities in a similar way as in case of bank-level network: the distributions follow heavily-tailed power laws [Figure 17b, Figure 17c]. Assets-weighted degree centrality has a higher mode.

Unweighted closeness centralities [Figure 17d] are concentrated around middle values. The distribution tends to be slightly left-skewed. This behavior conforms to low average-path and small diameter properties of the network. Weighted versions of closeness centralities [Figure 17e, Figure 17f] have power law distributions just like their counterparts in the bank-level network.

5.3. Distributions of financial indices

Histograms of financial indices of bank-level [Figure 18] and country-level [Figure 19] networks follow a heavy-tailed power laws. In this regard, they are similar to degree centrality distributions. The distributions in general are quite stable during the whole period.

In case of some bank-level indices like operating profit [Figure 18f] and equity [Figure 18b], the frequency of very small values is decreasing towards less small values, making the curve less steeper. Most noticeable changes happen in the tail section during the global financial crisis of 2008. The maximum values of net income [Figure 18e] and net loans [Figure 18c] decrease fall down after peaking in previous year.

The fluctuations of frequencies in country-level distributions are seemingly more chaotic. However, global financial crisis affects net income in similar way. In case of net loans, the decrease has a lag of two years. Same lagged decrease can be found in the tail of total customer deposit distribution [Figure 19d].

5.4. Year-wise correlations

Previous sections showed that both centrality measures (at least weighted ones) and financial indices follow power laws with some outliers. This suggests that these measures may have a correlation. This and the next sections are testing this hypothesis.

As mentioned in Chapter 2, the meaning of centrality measure depends on the network under study and the nature of relations. Based on the observations of network structure (Chapter 4), it can be assumed that centrality measures should indicate the hierarchical place of the bank in the ownership network. For example, head banks most certainly have high degree centrality measures. In case of country-level graph, the assumption is that centrality measure would classify countries as either central or peripheral. Therefore, significant correlation with financial indices would mean that financial wealth defines the position and the type of the bank.

Experiment design

For the first experiment, each year is analyzed as a separate slice. The goal is to determine **if centrality measures of banks in bank ownership network correlate to bank's financial indices**. A similar goal is set for country-level network: **do centrality measures of countries in country-level bank ownership network correlate to country's consolidated bank financial indices?**

In order to answer these questions, Pearson's correlation matrices were calculated for all pairs of measures of all nodes within one year. For each year and level, a separate matrix was calculated.

Bank-level results

Figure 20 shows the correlation matrix for bank-level network, where each pair's correlation coefficients are represented as bar charts below diagonal, showing values for each year of the observed period. Bar charts above the diagonal show corresponding p-values.

Correlation coefficients show that unweighted degree, share-weighted degree and assets-weighted degree are all differently behaving measures. Degree and share-weighted degree centralities have moderate to strong correlations. Asset-weighted degree does not correlate well with share-weighted degree and is almost uncorrelated to degree centrality.

In general, unweighted degree centrality weakly (sometimes moderately) correlates with financial indices. The highest correlation coefficient is with total assets. Unweighted closeness centrality, which is moderately correlated to degree centrality, has almost no correlation with financial indices.

In contrast, weighted values tend to show stronger correlations with financial indices. Share-weighted degree centrality has a moderate to strong correlation with total assets, equity, net loans, total customer deposits and operating profit. Unexpectedly, assets-weighted degree centrality has weaker correlations in comparison to share-weighted one, with the exception of equity and net income.

Share-weighted closeness centrality has very strong correlation with its assets-weighted counterpart. Therefore, they behave very similarly, both having moderate correlations with equity.

All financial indices strongly correlate to themselves, except for correlations with net income and operating profit, which have varying coefficient values. Table shows particularly strong correlation coefficients between total assets and net loan amount, and between total assets and total customer deposit amount.

The stability of these correlation coefficients, as shown in the matrix, varies from measure to measure. For example, correlations of unweighted and share-weighted degree centralities with other measures (except for net income and operating profit) are quite stable. On the contrary, all correlations with net income and operating profit are unstable and feature perturbations around the year 2008.

Country-level results

The results for country level [Figure 21] are comparable to bank level and feature stronger correlations in general. Unweighted centralities do not correlate with financial indices. Unweighted degree centrality is more similar to share-weighted degree centrality than to asset-weighted one. Share-weighted degree centrality has strong correlation with total assets, moderate to strong correlation with equity, net loans and total customer deposits. The correlations with net income and operational profit are unstable, varying from none to strong correlation. Asset-weighted degree centrality has a stable strong correlation with equity and moderate correlation with other financial indices.

Financial indices on country-level also correlate well with themselves. The strongest correlations are between country's banks' total amount and net loans and between total amount and total customer deposits. This is the same case as with bank-level network.

5.5. Period-wise correlations

The distribution histograms of different measures (Sections 5.2−5.3) show that, while the larger pictures stay fundamentally the same, there is some fluctuation to values. The question is whether these changes in different time series of measures and indices are somehow related. Such relation would directly explain the formation of the network structure and would allow for making predictions about financial wealth based upon network's structural changes. This is of great use to a large group of people, from investers to economists.

Experiment design

This test is about the dynamic nature of the network. Measures of individual nodes in each year of the observed period are treated as time series. Therefore, the concrete question under study is **whether time series of centrality measures of banks in bank ownership network correlate to time series of financial indices.** The same question is applicable to country-level network.

Pearson's correlation coefficients for the same pair of series for all nodes form a set. Its cardinality matches node count. Next, the *distribution* of each set is analyzed to see if the data suggests that a particular pair of measures tends to correlate. The goal of the analysis is to detect the presence or absence of relation rathen than to find its probability.

Bank-level network

Box plots of the distributions of correlation coefficients [Figure 22] show that the ranges of the correlation coefficients are large. Some of the distributions also feature outliers. The interquartile range (IQR) is around 0.5 and sometimes less. Both median and average values stay close to each other around the center of IQR.

Judging by the median and average values of these distributions, the correlations are weak ones (0.1−0.3). The strongest correlation of centrality and financial index is between assets-weighted degree centrality and total assets.

Country-level network

The ranges of distributions of correlation coefficient between time series of measures in country-level network [Figure 23] are smaller than their bank-level counterparts. Taking median and average values as the most representative ones, these distributions show weak, moderate and even strong correlations.

The strongest correlations are exhibited by assets-weighted closeness centrality and total assets, assets-weighted closeness centrality and equity, assets-weighted degree centrality and total assets.

5.6. Conclusion

As expected, unweighted degree centrality has a heavy-tailed power law distribution both in bank- and country-level networks. This is due to the heterogeneous mixture of nodes, which is found and described in previous chapter. The power law also strictly applies to financial indices: the vast majority of the nodes are very “poor” and there are only a few very “rich” ones. This may also be reason why the distribution of weighted centrality measures is universally the same: the weight overrides the centrality part of the measure.

Weighted degree centrality strongly and positively correlates with financial indices, meaning that larger and richer banks are more central ones, i.e. have more connections. However, it doesn’t seem that the number of connections defines the wealth of the bank in case of bank-level network. Bank groups of similar form may be in different wealth categories depending on a country, for example. On country level, same kinds of correlations are even stronger.

Financial indices very strongly and positively correlate with each other except for those which depend on the year (net income and operation profit). This suggests that these financial parameters can be used to describe the system.

The results also show that centrality measures cannot in general be used to predict changes in financial wealth – corresponding time series do not correlate. The laws of bank ownership network formation are more subtle. They are subject to factors other than simply buying or selling bank assets.

6. Summary

This thesis set out to study ownership structure of banks. The dynamic network of banks enriched with data on financial indices was created from Bankscope database and included years from 2003 to 2013. After necessary adjustment and filtering steps, network theory was applied to the distilled graph (largest connected component) of 419 nodes and 464 links. Additionally, analysis was conducted on derivative versions of the network that represented bank group-level and country-level views of the original network.

The results of visual study as well as network metrics show that bank ownership network is a sparse disassortative graph, with degree distributions similar to scale-free network, but a large diameter. The reason for large diameter may be due to the spatial factor in the formation of the network. Properties of the network are stable during the observed period.

The topology of bank ownership network shows properties similar to interbank lending network, but contains some key differences. The similarity is probably due to the fact that a relation between banks, whether it is ownership or payment flow, does not occur randomly, but is bound by similar laws. Child banks lend from their owners – parent banks. To some degree, lending network and ownership network coincide. The difference is due to the nature of these relations. The ownership is by definition more hierarchical than lending: banks of equal size usually do not own one or another, but they can lend from each other.

Bank ownership network, as it turns out, is comprised of bank groups and consortium banks – star-shaped bank clusters with one central bank – and standalone banks. These groups are loosely related to each other through standalone banks. These partial hierarchies explain the network metrics as well as the measures of individual nodes. Distribution of centrality measures and financial indices indicate that smaller in any sense banks form a vast majority.

Bank’s degree centrality is directly related to its financial wealth: the largest banks own more assets than smaller banks. Weighted degree centrality also shows that rich banks have also richer neighborhood. Changes to the neighborhood of a bank do not reflect upon the financial wealth of the bank in a predictable way.

On country level, ownership structure is different: the graph is a much denser scale-free network that fits core-periphery model. Countries with wealthier banking sector comprise the core. Degree centrality measure is also the most significant one in describing this network, correlating well with aggregated financial indices.

To sum up, the main contributions of this thesis are:

- A thorough study of bank ownership network that describes its topology, properties, driving factors, evolution over a decade and relation to economic situation. This fills the gap in related literature and provides ground knowledge for future research.
- An application of novel analysis methods to Bankscope database and financial networks, particularly the use of social network analysis.
- An extensive use of visualization techniques to show the evolution of network's structure alongside its properties, which may be applicable in other fields of study where dynamic graphs are involved.

Regarding the future work on the subject, a deeper study of the behaviors discovered by this research would be valuable. Such study would require more knowledge of economy, finance and banking and should try to propose explanations to the observed results. They should take more bank attributes into account, e.g. ownership type.

Another direction is cross-validation of the results obtained in this research on both larger and smaller levels of networks. The fact that ownership network was based on European banks only should also be taken into account. European Union implements a single banking market which might be very specific to this region. In other parts of the world, the overall picture may vary. Such additional research might bring out differences in both ownership networks structure and its reaction to the crisis. Finally, one may compare bank ownership network to other bank networks like lending network and explore correlations between them.

References

[1] Investopedia. [Online]. http://www.investopedia.com/terms/c/consortium-bank.asp

[2] A.N. Berger, L.F. Klapper, M.S. Martinez Peria, and R. Zaidi, *Bank ownership type and banking relationships*.: J. Finan. Intermediation, 2007.

[3] Garlaschelli D., Battiston S., Castri M., Servedio V.D.P., Caldarelli G., "The scale-free topology of market investments," *Physica A: Statistical Mechanics and its Applications*, vol. 350, no. 2, pp. 491-499, 2005.

[4] Charles P. Kindlebereger, *American Business Abroad*. New Haven: Yale University, 1969.

[5] J.C. Stein, "Internal Capital Markets and the Competition for Corporate Resources," *Journal of Finance*, vol. 52, no. 1, pp. 111-133, 1997.

[6] R. De Haas and I. Naaborg, "Foreign Banks in Transition Countries: To Whom Do They Lend and How Are They Financed?," *Financial Markets, Institutions & Instruments*, vol. 15, no. 4, pp. 159-199, 2006.

[7] Craig B., von Peter G., "Interbank tiering and money center banks," *Journal of Financial Intermediation*, vol. 23, no. 3, pp. 322–347, 2014.

[8] G Hale, "Bank relationships, business cycles, and fi nancial crises," *Journal of International Economics*, vol. 88, pp. 312-325, 2012.

[9] Minoiu C, Reyes J, "A network analysis of global banking: 1978-2009," 2011.

[10] Soramäki K., Bech M.L., Arnold J., Glass R.J., Beyeler W.E., "The topology of interbank payment flows," *Physica A: Statistical Mechanics and its Applications*, vol. 379, no. 1, pp. 317–333, 2007.

[11] Boss M., Elsinger H., Summer M., Thurner S., "Network topology of the interbank market," *Quantitative Finance*, vol. 4, no. 6, pp. 677-684, 2004.

[12] Micco A., Panizza U., Yañez M., "Bank Ownership and Performance," 2004.

[13] Loukoianova E. Nicolò G., "Bank Ownership, Market Structure and Risk," 2007.

[14] Caccioli F., Catanach T.A., Farmer J.D., "Heterogeneity, correlations and financial contagion," *Advances in Complex Systems*, vol. 2, no. 15, 2012.

[15] Allen F., Jackowicz K., Kowalewski O., "The effects of foreign and government ownership on bank lending behavior during a crisis in Central and Eastern Europe," *Wharton Financial Institutions Center Working Paper No. 13-25*, 2013. [Online]. http://mpra.ub.uni-muenchen.de/48059/

[16] Chinazzi M., Fagiolo G., Reyes J.A., Schiavo S., "Post-mortem examination of the international," *Journal of Economic Dynamics and Control*, vol. 37, pp. 1692-1713, 2013.

[17] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, no. 2, pp. 167-256, 2003.

[18] JGraphT, a Java library for working with graphs. [Online]. http://jgrapht.org

[19] Social Network Analysis. [Online]. https://reference.wolfram.com/language/guide/SocialNetworks.html

[20] L. C. Freeman, "A set of measures of centrality based upon betweenness," *Sociometry*, vol. 40, pp. 35–41, 1977.

[21] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social Networks*, vol. 1, no. 3, pp. 215-239, 1979.

[22] P. Bonacich, "Power and Centrality: A Family of Measures," *American Journal of Sociology*, vol. 92, no. 5, pp. 1170–1182, 1987.

[23] Ulrik Brandes, "A faster algorithm for betweenness centrality," *Journal of Mathematical Society*, vol. 25, pp. 163-177, 2001.

[24] P Bonacich, "Some unique properties of eigenvector centrality," *Social Networks*, vol. 29(4), pp. 555-564, 2007.

[25] Social Network Analysis addon for JGraphT. [Online]. https://bitbucket.org/sorend/jgrapht-sna

[26] Opsahl, Tore and Agneessens, Filip and Skvoretz, John, "Node centrality in weighted networks: Generalizing degree and shortest paths.," *Social Networks*, vol. 33(3), pp. 245-251, 2010.

[27] C Dangalchev, "Residual closeness in networks," *Physica A*, vol. 365, pp. 556-564, 2006.

[28] Pearson product-moment correlation coefficient. [Online]. http://en.wikipedia.org/wiki/Pearson_product-moment_correlation_coefficient

[29] Commons Math: The Apache Commons Mathematics Library. [Online]. http://commons.apache.org/proper/commons-math/

[30] Abbasi A., Altmann J., Hossain L., "Identifying the Effects of Co-Authorship Networks on the Performance of Scholars: A Correlation and Regression Analysis of Performance Measures and Social Network Analysis Measures," 2011.

[31] Khademi R., Mansoori A. Soheili F., "Correlation between Impact Factor and productivity with centrality measures in journals of Information science: A social network analysis," *International Journal of Information Science and Management*, vol. 13, no. 1, pp. 21-38, 2015.

[32] Mihhail Verhovtsov. GitHub.org. [Online]. https://mihhail-verhovtsov.github.io/bank-ownership-network-analysis/graph.html

[33] Jackson, a Java library for reading/writing data in JSON format. [Online]. http://jackson.codehaus.org/

[34] D3.js, a JavaScript library for data visualization. [Online]. http://d3js.org/

[35] JFreeChart, a Java library for rendering charts. [Online]. http://www.jfree.org/jfreechart/

[36] Bankscope | Global database for bank financial data inc. statements, ratings & intelligence. [Online]. http://www.bvdinfo.com/en-gb/our-products/company-information/international-products/bankscope

[37] Costenbader E., Valente T. W., "The stability of centrality measures when networks are sampled," *Social Networks*, vol. 25, pp. 283–307, 2003.

[38] Estonian Banking Association homepage. [Online]. http://www.pangaliit.ee/en/banking-information/market-shares-of-estonian-banks

[39] Ed Jr. Pegg. "Small World Network." From MathWorld. [Online]. http://mathworld.wolfram.com/SmallWorldNetwork.html

[40] Small-world network. [Online]. http://en.wikipedia.org/wiki/Small-world_network

Appendix A. Network visualizations

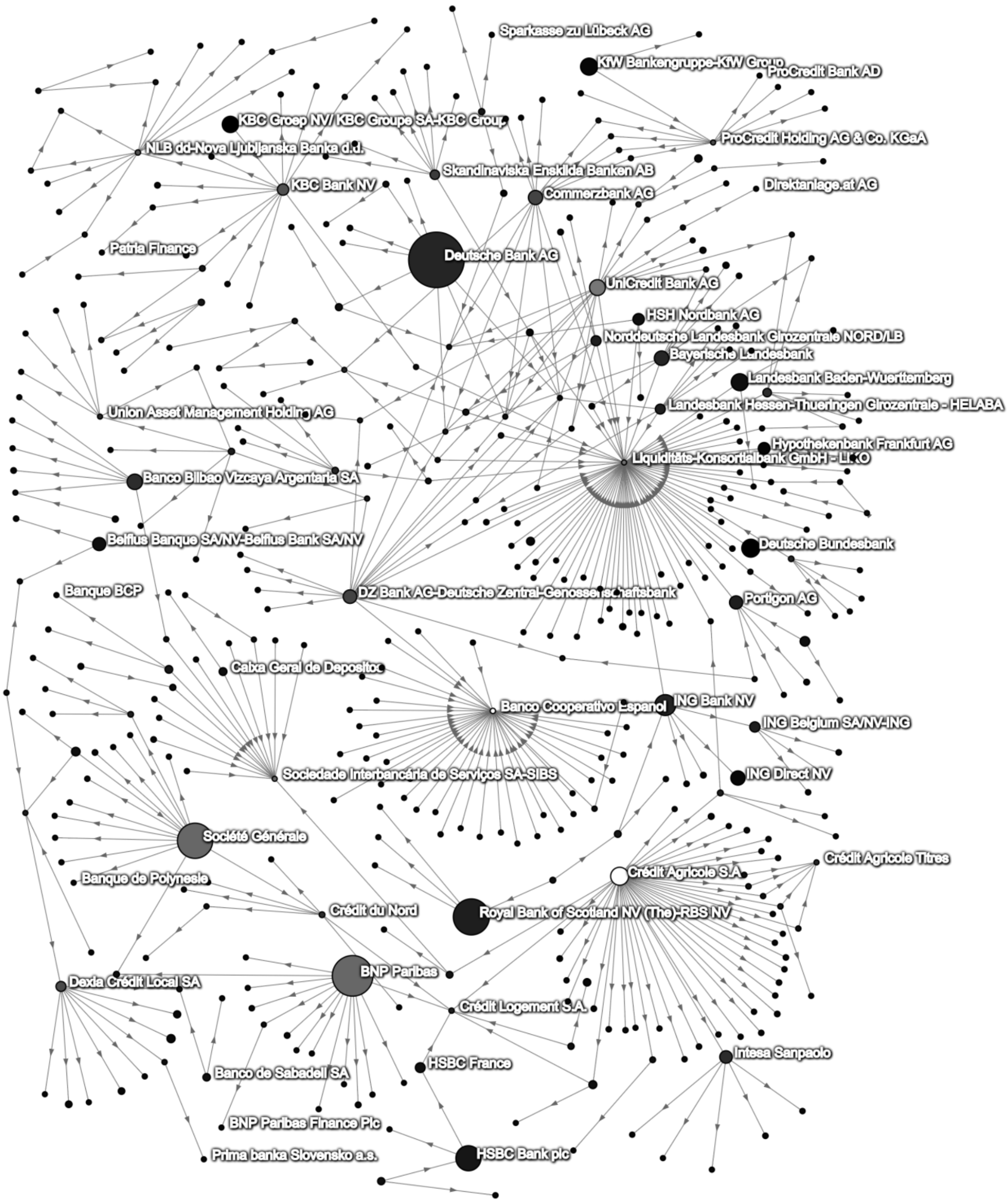

Figure 6. Visualization of bank-level network as of year 2006. Nodes represent individual banks. Node size is proportional to bank's total assets. Node color is proportional to bank's unweighted degree centrality (black-white scale, white is the highest value). Links represent ownership relations with direction from owner to owned.

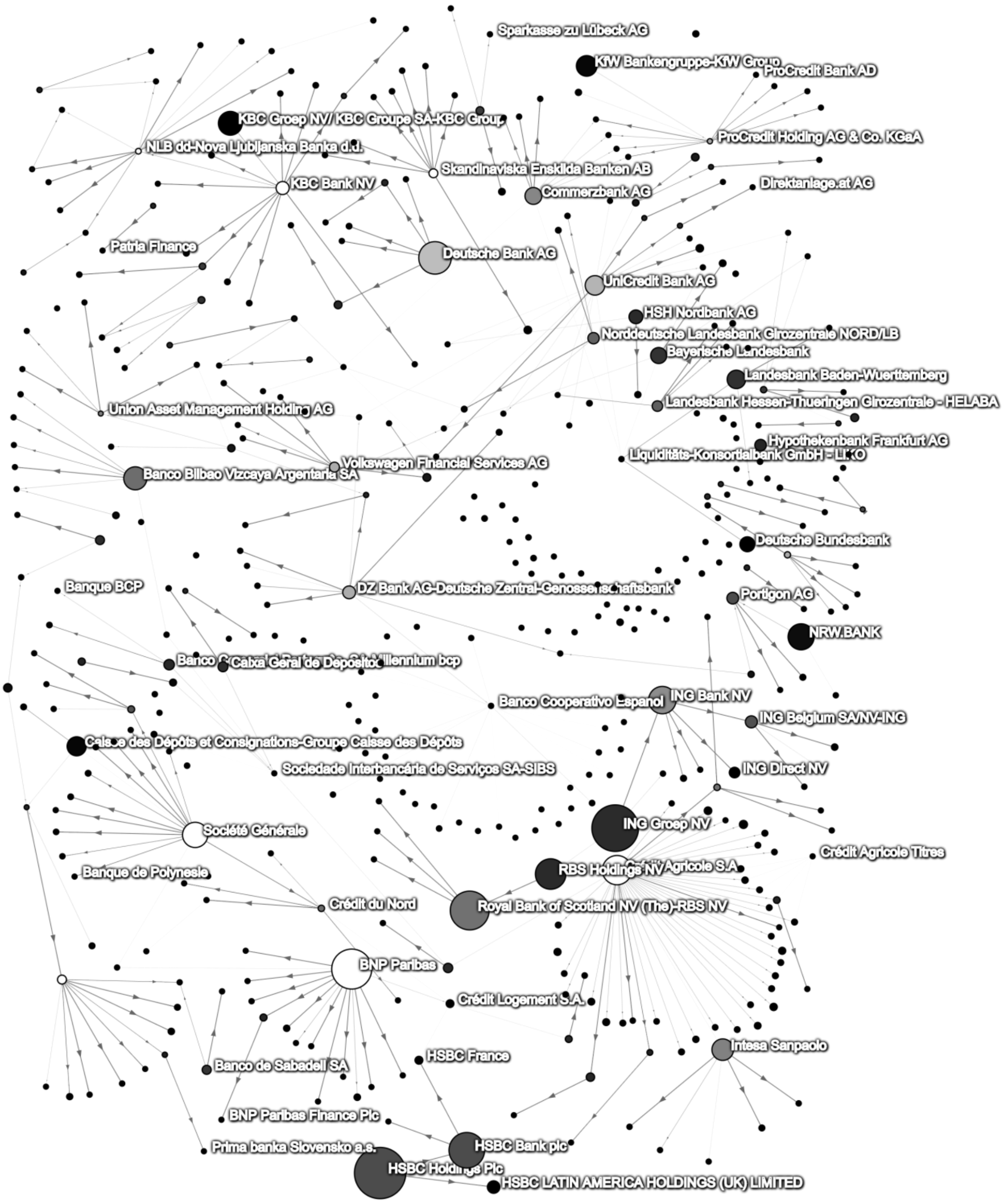

Figure 7. Visualization of bank-level network as of year 2006. Nodes represent individual banks. Node size is proportional to bank's equity. Node color is proportional to bank's share-weighted degree centrality (black-white scale, white is the highest value). Links represent ownership relations with direction from owner to owned. Link thickness is proportional to share percentage.

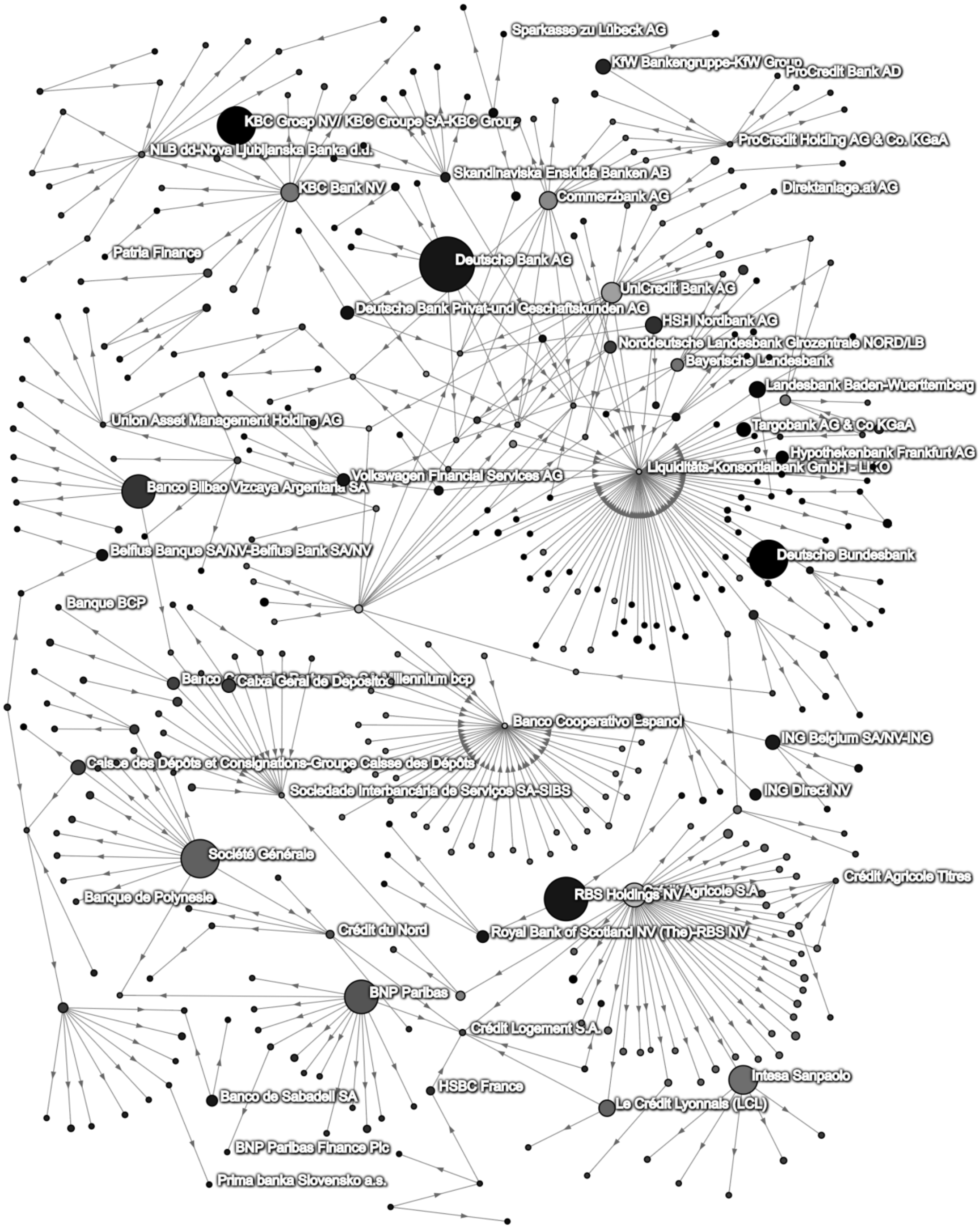

Figure 8. Visualization of bank-level network as of year 2006. Nodes represent individual banks. Node size is proportional to bank's operating profit. Node color is proportional to bank's unweighted closeness centrality (black-white scale, white is the highest value). Links represent ownership relations with direction from owner to owned.

Figure 9. Visualization of group-level network as of year 2006. Nodes represent either individual banks of bank groups. Node size is proportional to bank's or group's total assets. Node color is proportional to unweighted degree centrality (black-white scale, white is the highest value). Links represent ownership relations with direction from owner to owned.

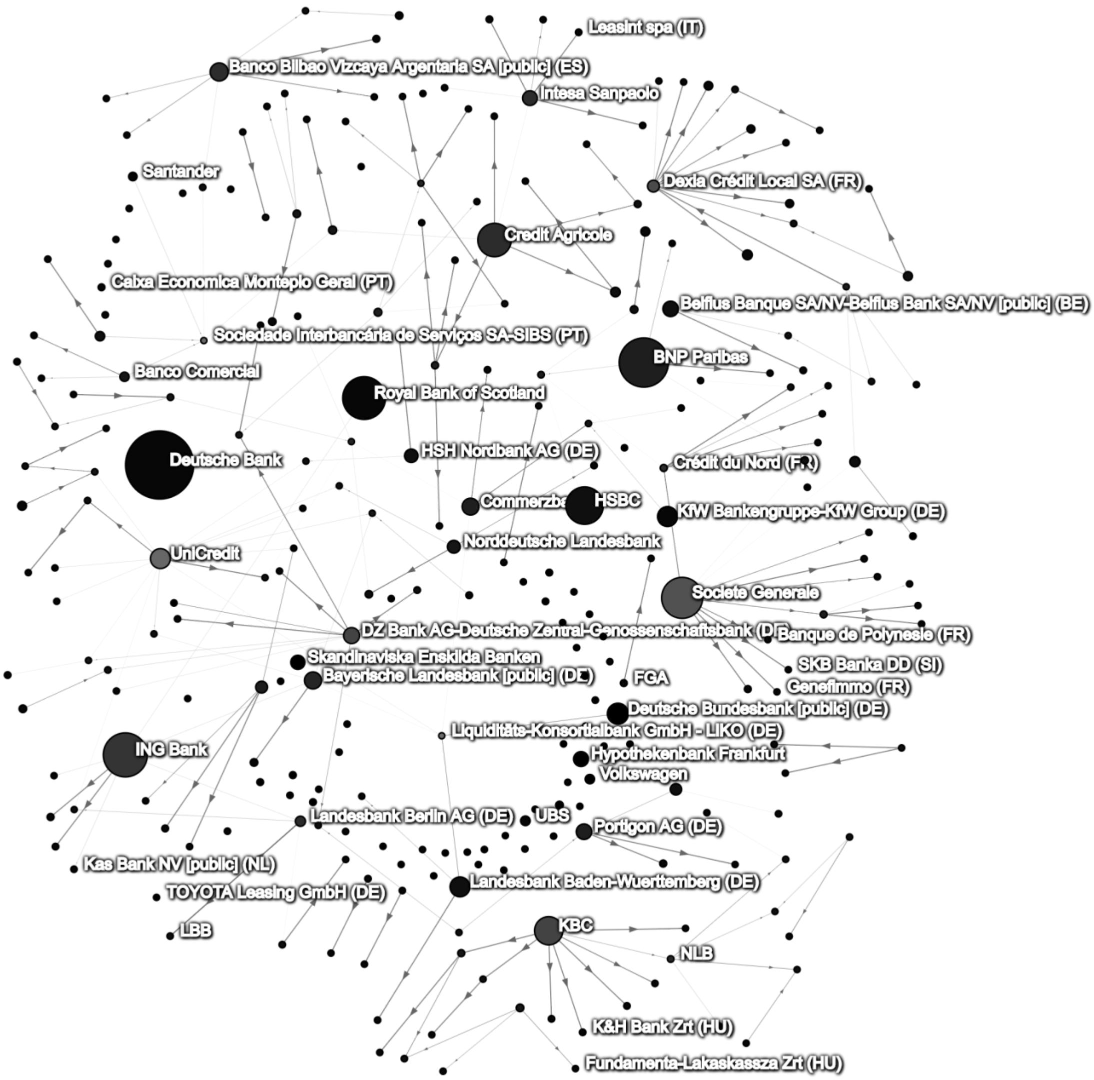

Figure 10. Visualization of group-level network as of year 2006. Nodes represent either individual banks or bank groups. Node size is proportional to bank's or group's equity. Node color is proportional to share-weighted degree centrality (black-white scale, white is the highest value). Links represent ownership relations with direction from owner to owned. Link thickness is proportional to share percentage.

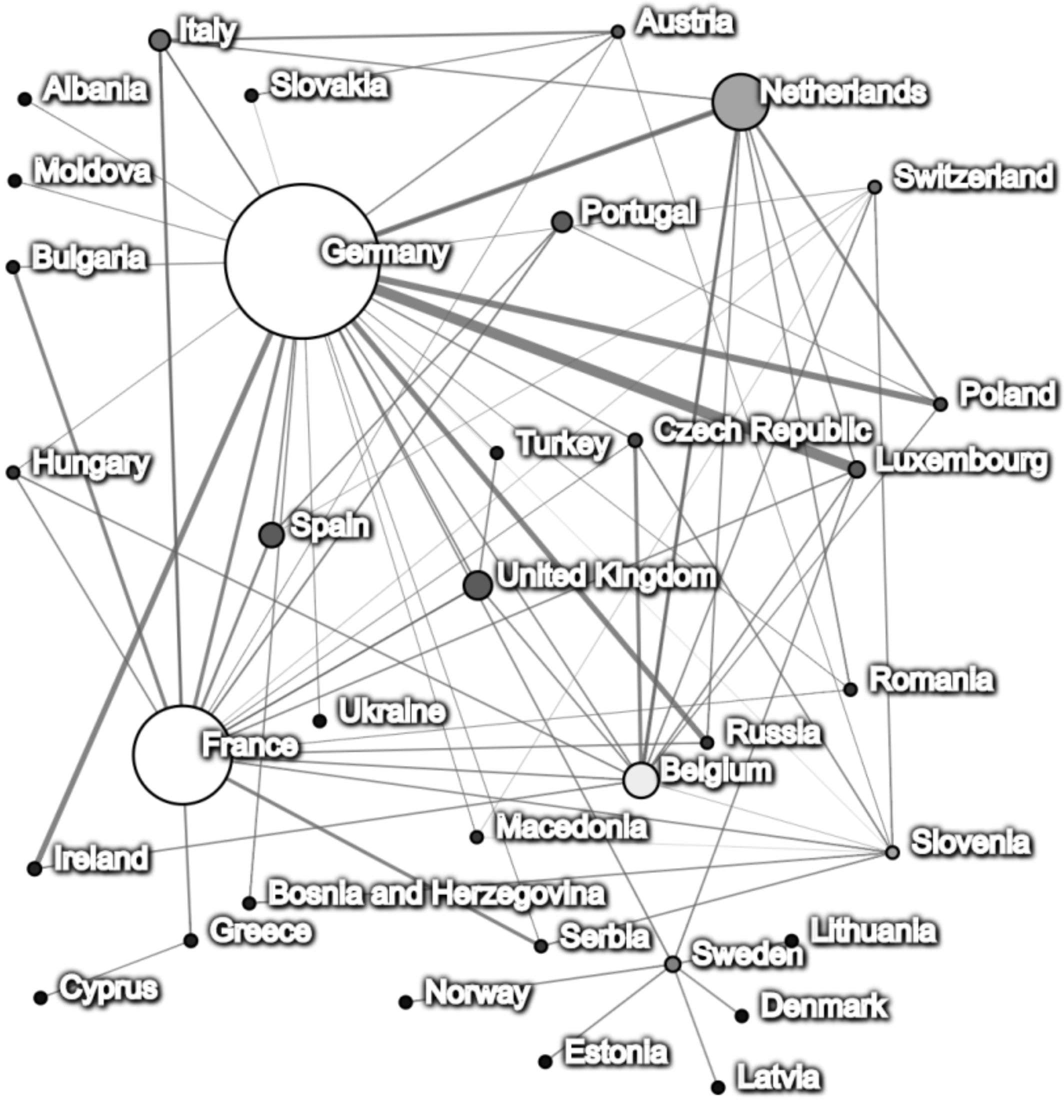

Figure 11. Visualization of country-level network as of year 2006. Nodes represent countries. Node size is proportional to country's aggregated total assets. Node color is proportional to unweighted degree centrality (black-white scale, white is the highest value). Links represent ownership relations. Link thickness is proportional to aggreagted share-percentage.

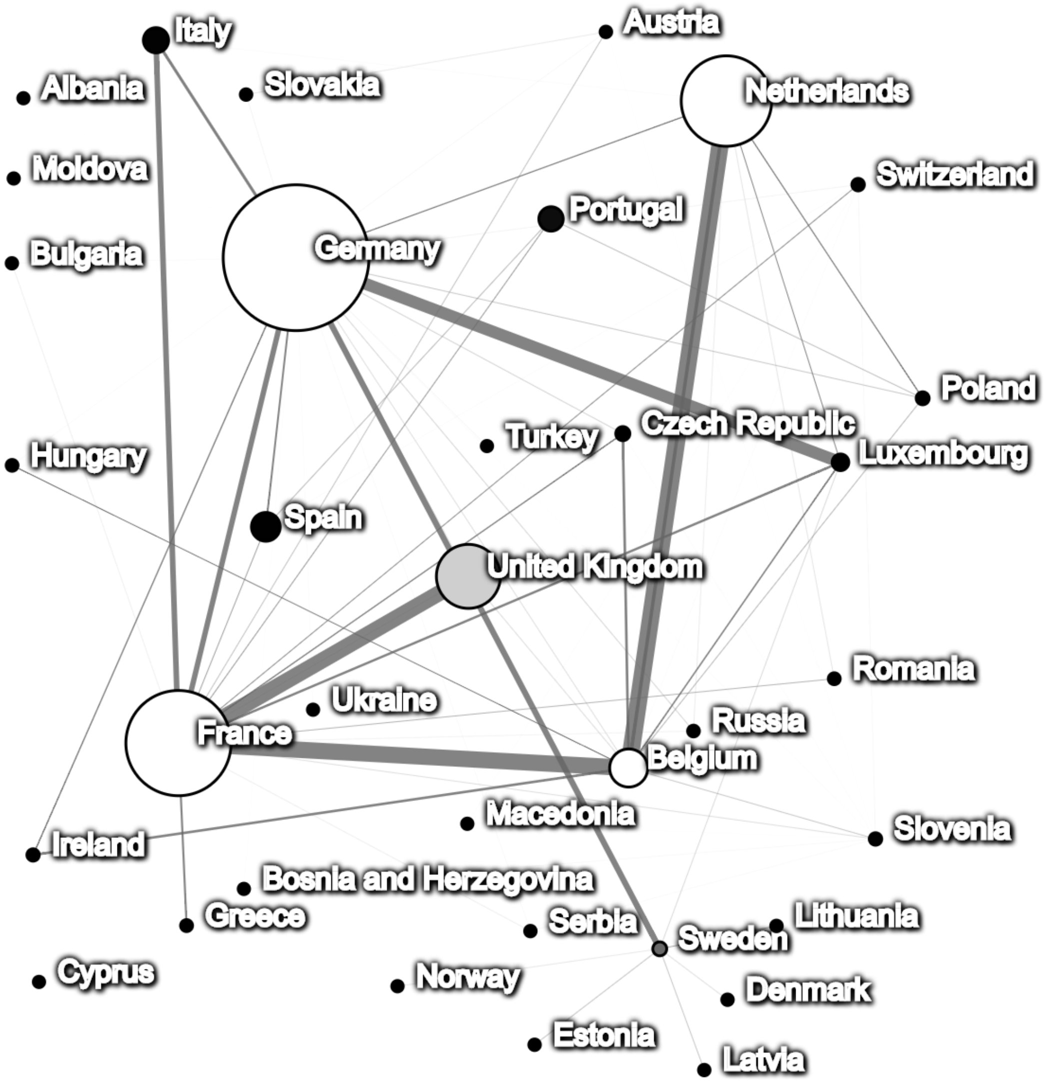

Figure 12. Visualization of country-level network as of year 2006. Nodes represent countries. Node size is proportional to country's aggregated equity. Node color is proportional to bank's assets-weighted degree centrality (black-white scale, white is the highest value). Links represent ownership relations. Link thickness is proportional to absolute share size.

Appendix B. Network metrics

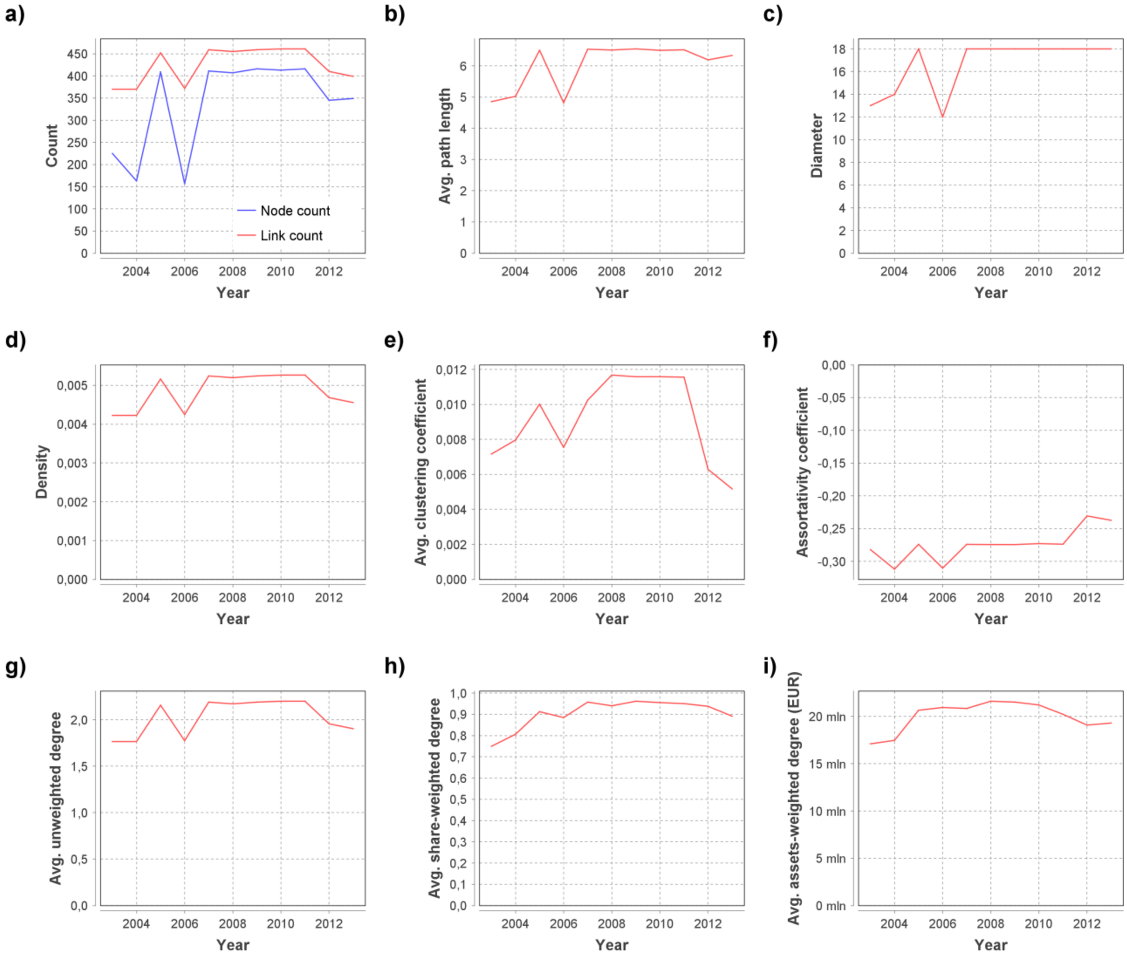

Figure 13. Metrics of bank-level network in 2003−2013. Included metrics are: a) node and link count in largest connected component, b) average path length, c) diameter, d) density, e) average clustering coefficient, f) assortativity coefficient, g) average unweighted degree, h) average share-weighted degree and i) average assets-weighted degree.

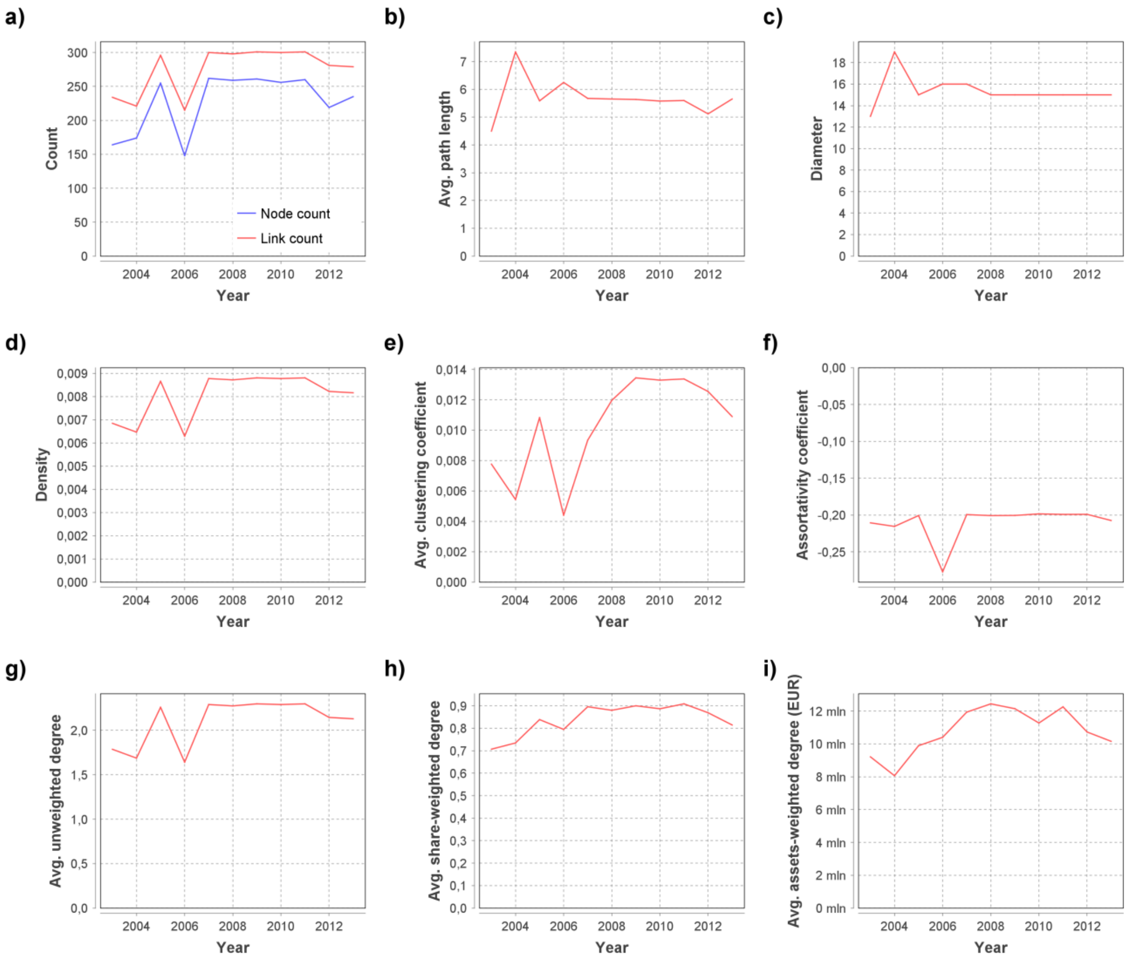

Figure 14. Metrics of group-level network in 2003−2013. Included metrics are: a) node and link count in largest connected component, b) average path length, c) diameter, d) density, e) average clustering coefficient, f) assortativity coefficient, g) average unweighted degree, h) average share-weighted degree and i) average assets-weighted degree.

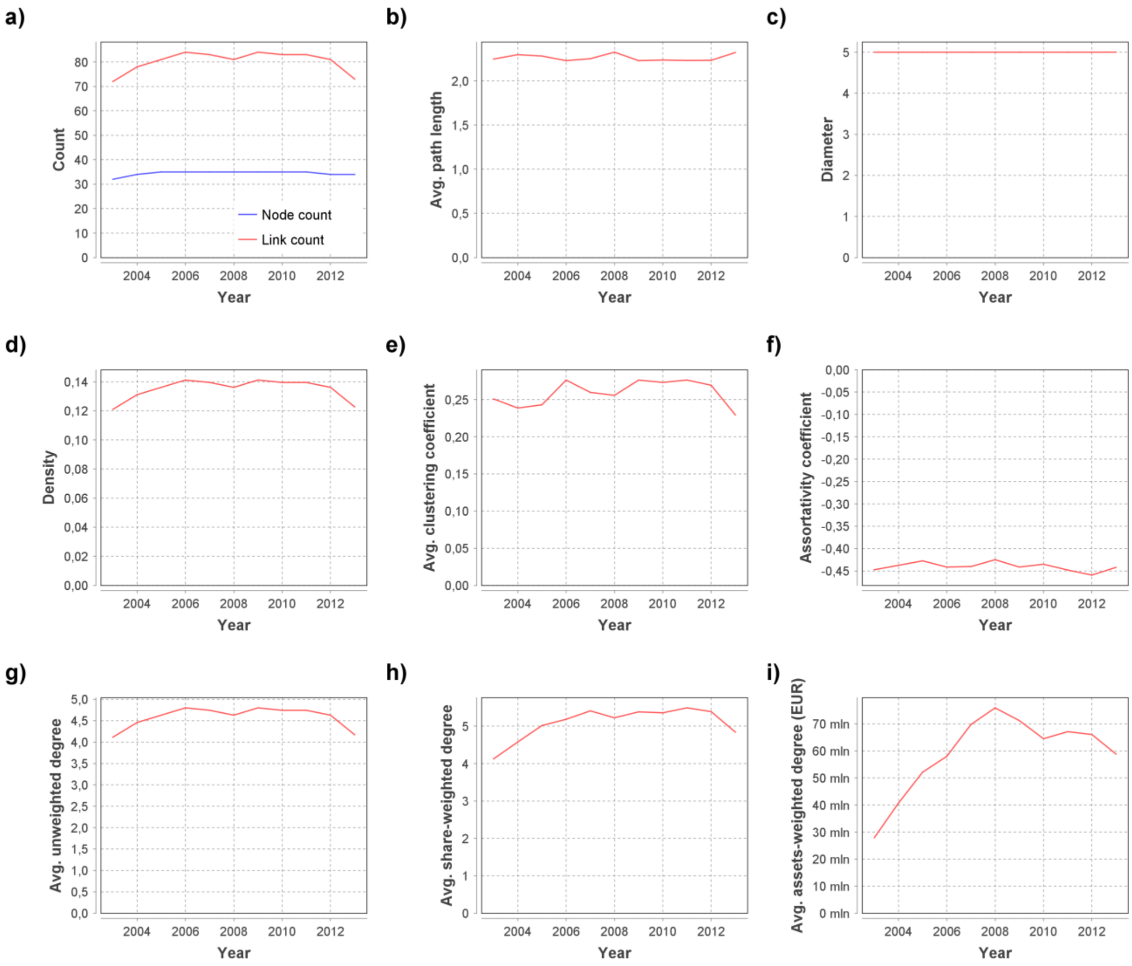

Figure 15. Metrics of country-level network in 2003−2013. Included metrics are: a) node and link count in largest connected component, b) average path length, c) diameter, d) density, e) average clustering coefficient, f) assortativity coefficient, g) average unweighted degree, h) average share-weighted degree and i) average assets-weighted degree.

Appendix C. Distribution of centrality measures

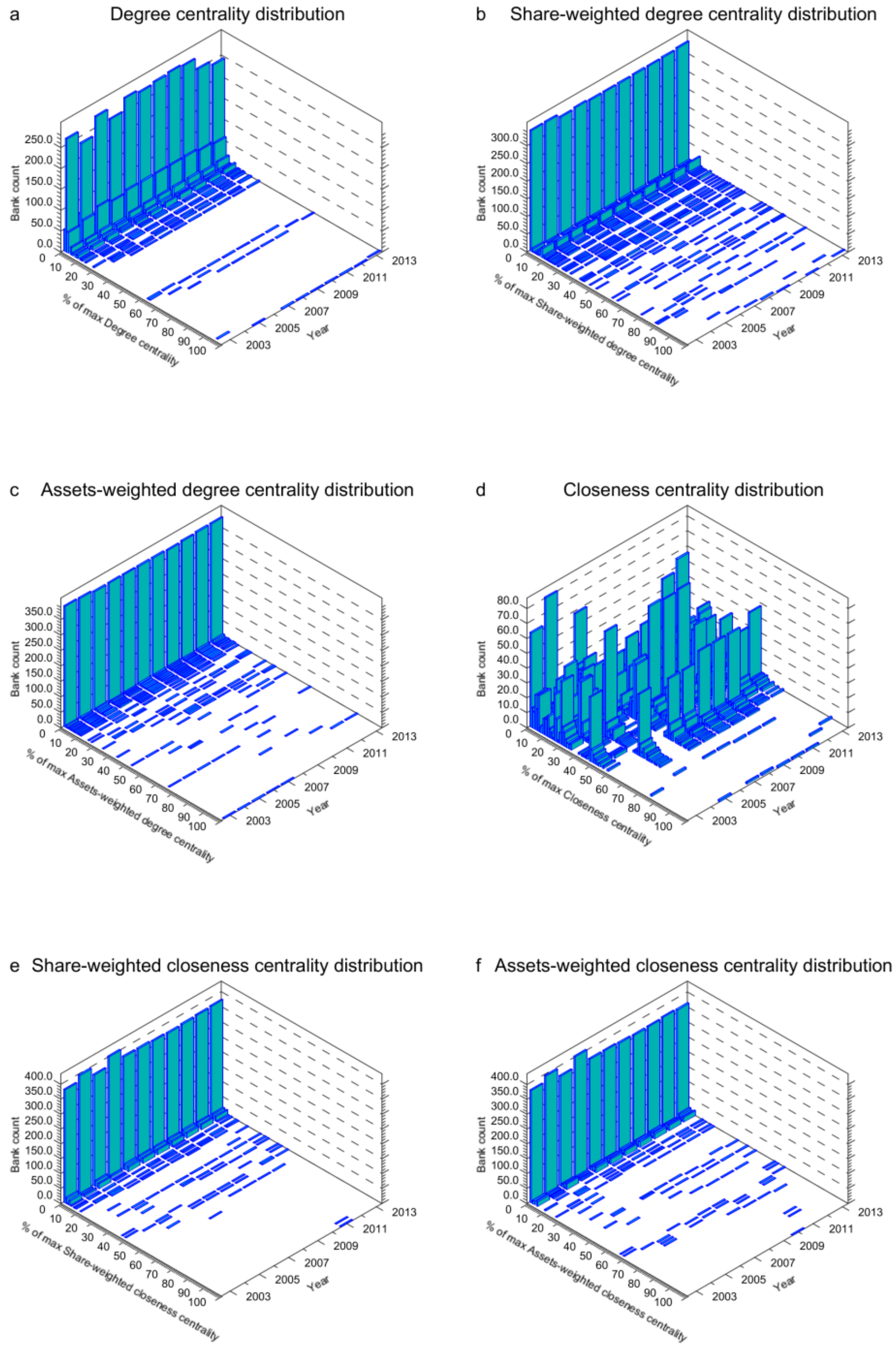

Figure 16. Distribution of centrality measures in bank-level network in 2003−2013.

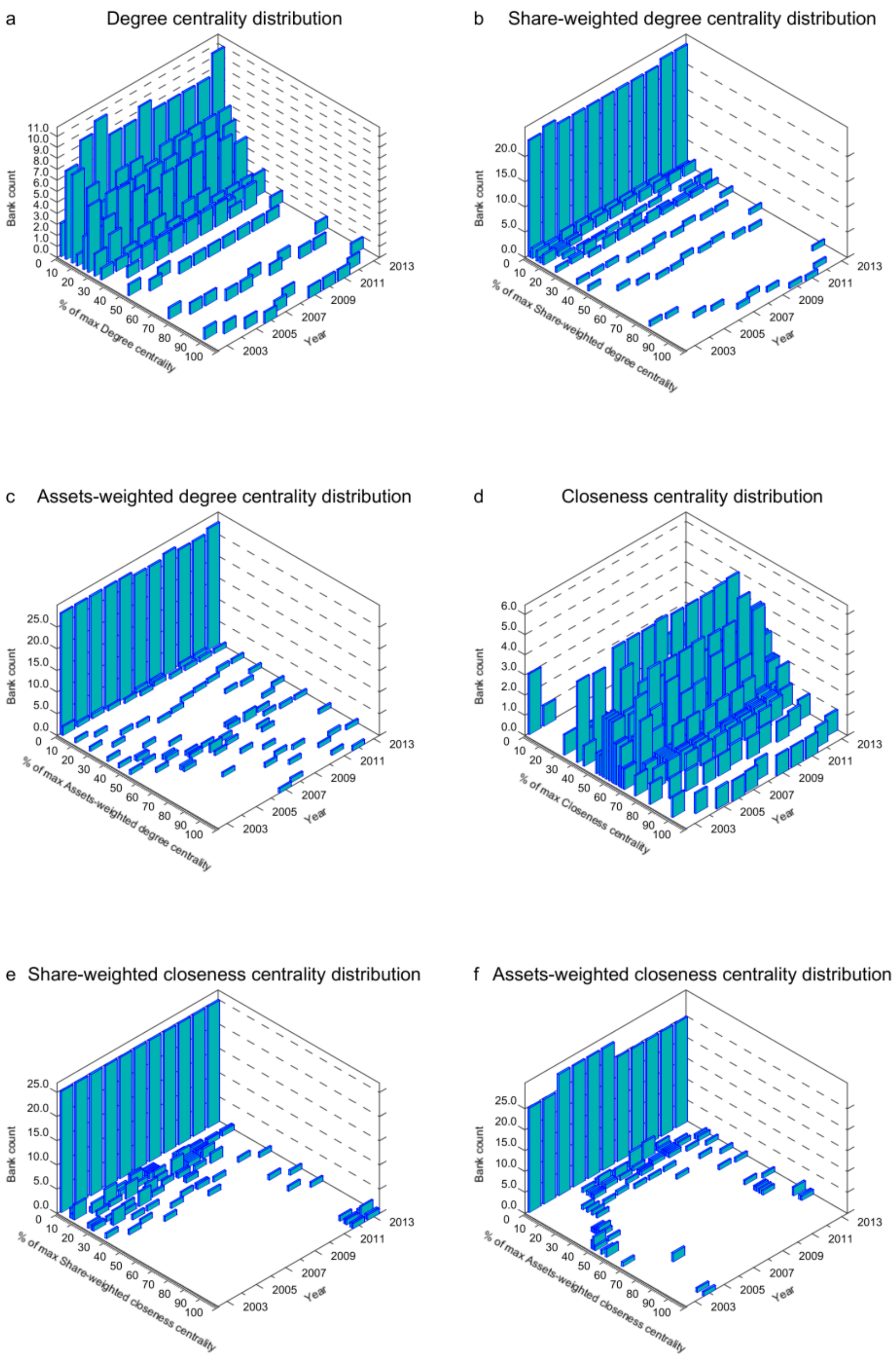

Figure 17. Distribution of centrality measures in country-level network in 2003−2013.

Appendix D. Distribution of bank financial indices

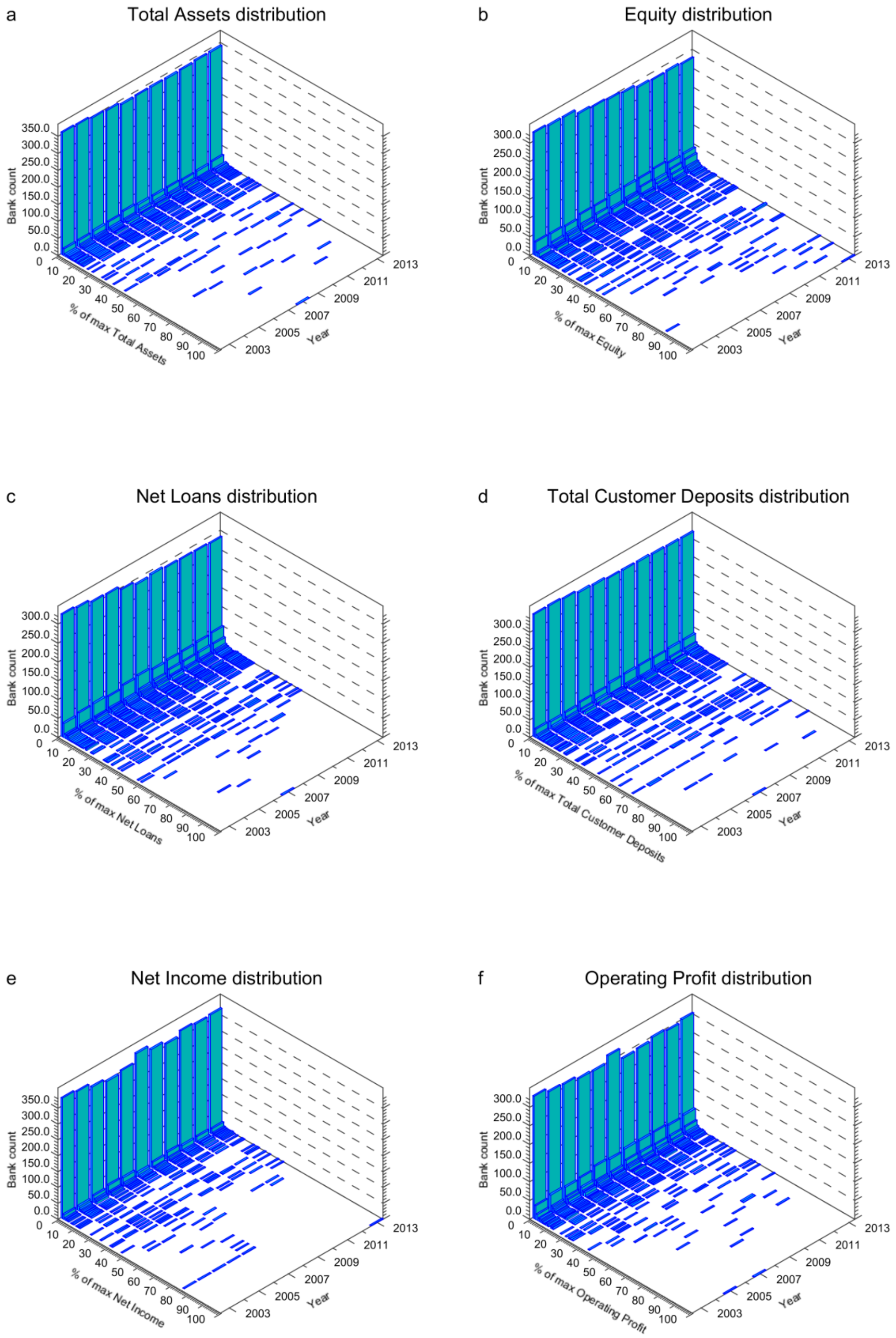

Figure 18. Distribution of financial indices in bank-level network in 2003−2013.

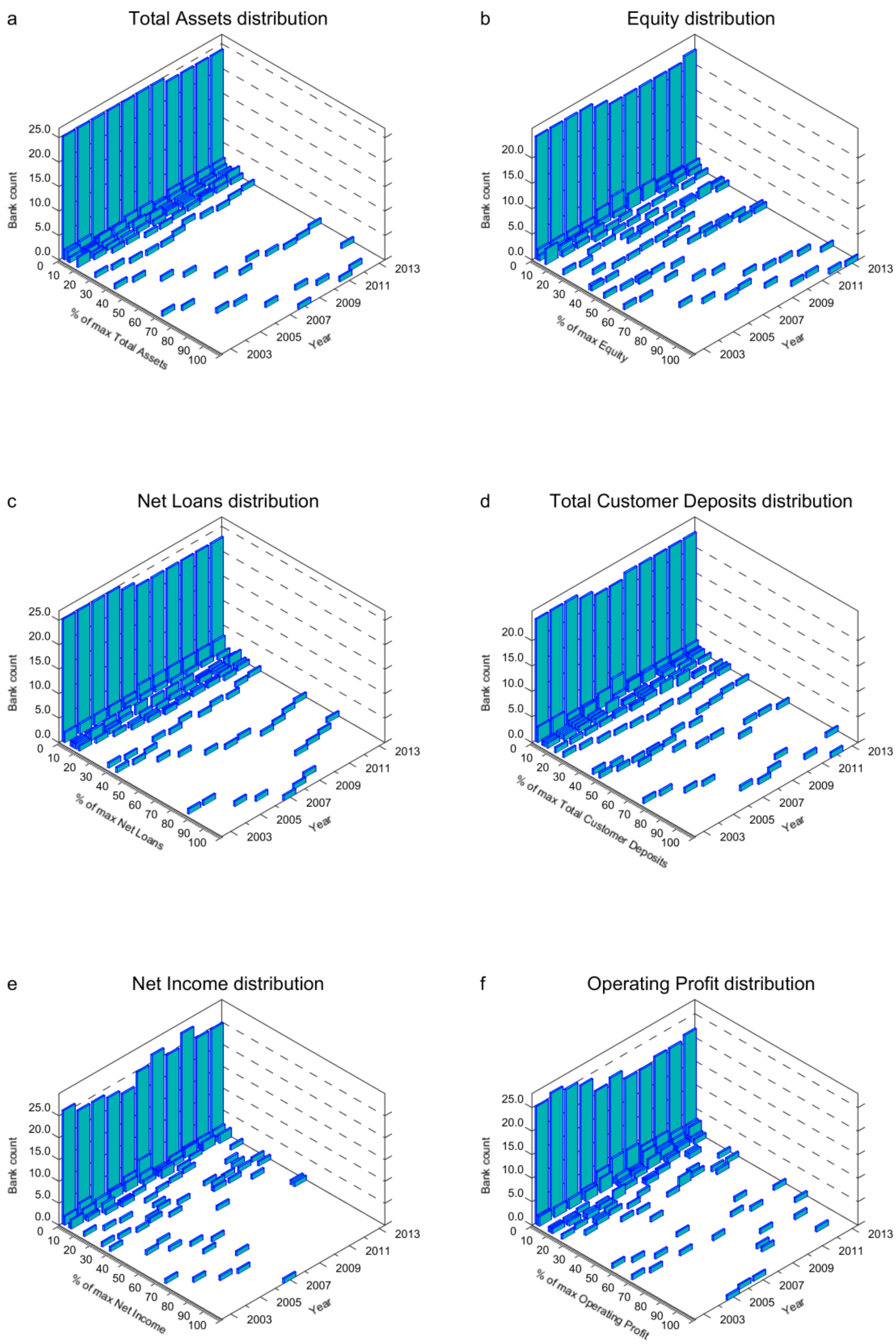

Figure 19. Distribution of financial indices in country-level network in 2003−2013.

Appendix E. Year-wise centrality measure and financial index correlations

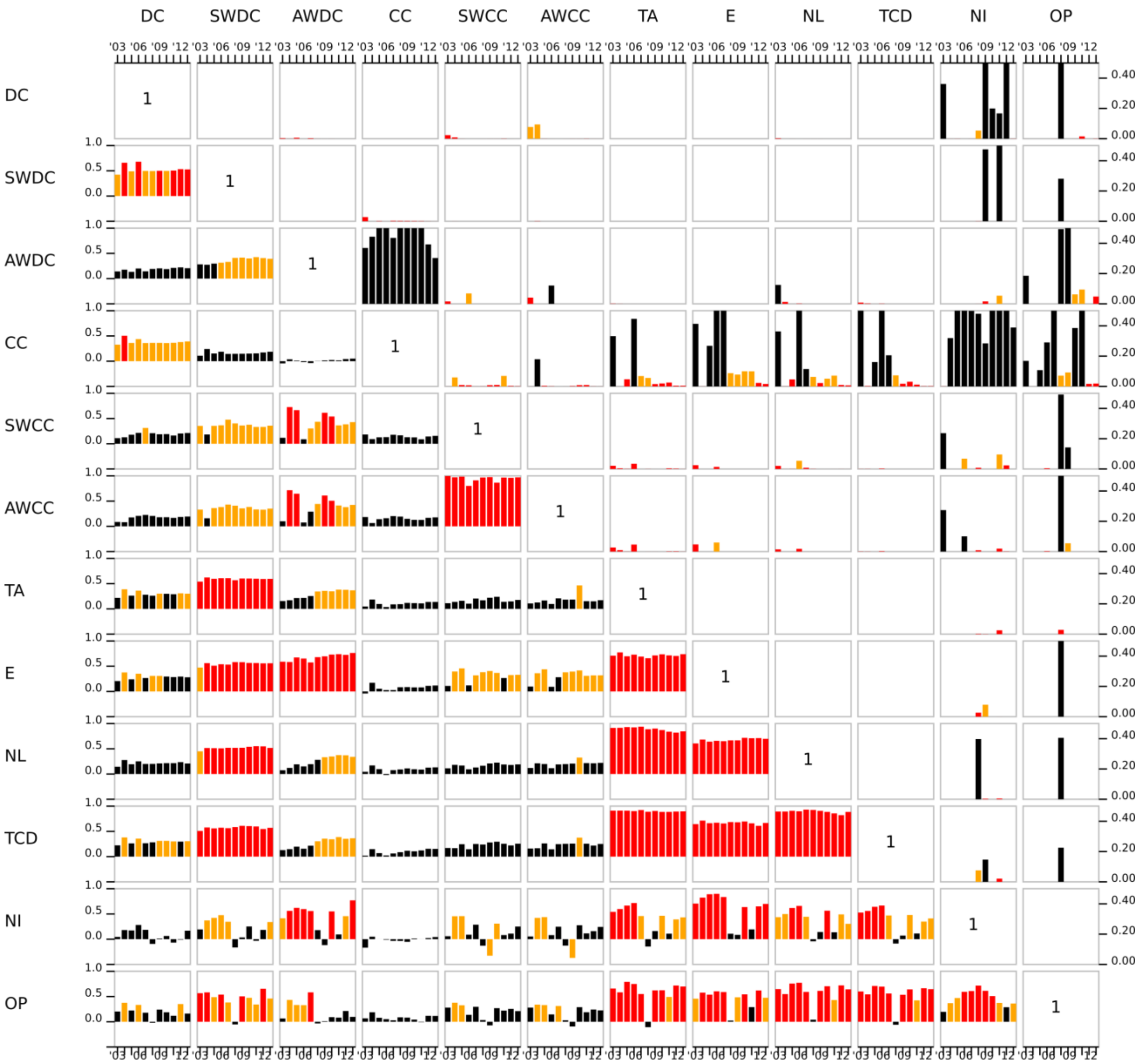

Figure 20. Centrality measure and financial index correlations in bank-level network in 2003−2013. Abbreviations used: DC – Degree centrality, SWDC – Share-weighted degree centrality, AWDC – Assets-weighted degree centrality, CC – Closeness centrality, SWCC – Share-weighted closeness centrality, AWCC - Assets-weighted closeness centrality, TA - Total Assets, E - Equity, NL - Net Loans, TCD - Total Customer Deposits, NI - Net Income, OP - Operating Profit. Below diagonal are correlation coefficients, above diagonal are p-values. Red color denotes correlation coefficients ≥ 0.5 and p-values ≤ 0.05, orange color denotes correlation coefficients ≥ 0.3 and p-values ≤ 0.1.

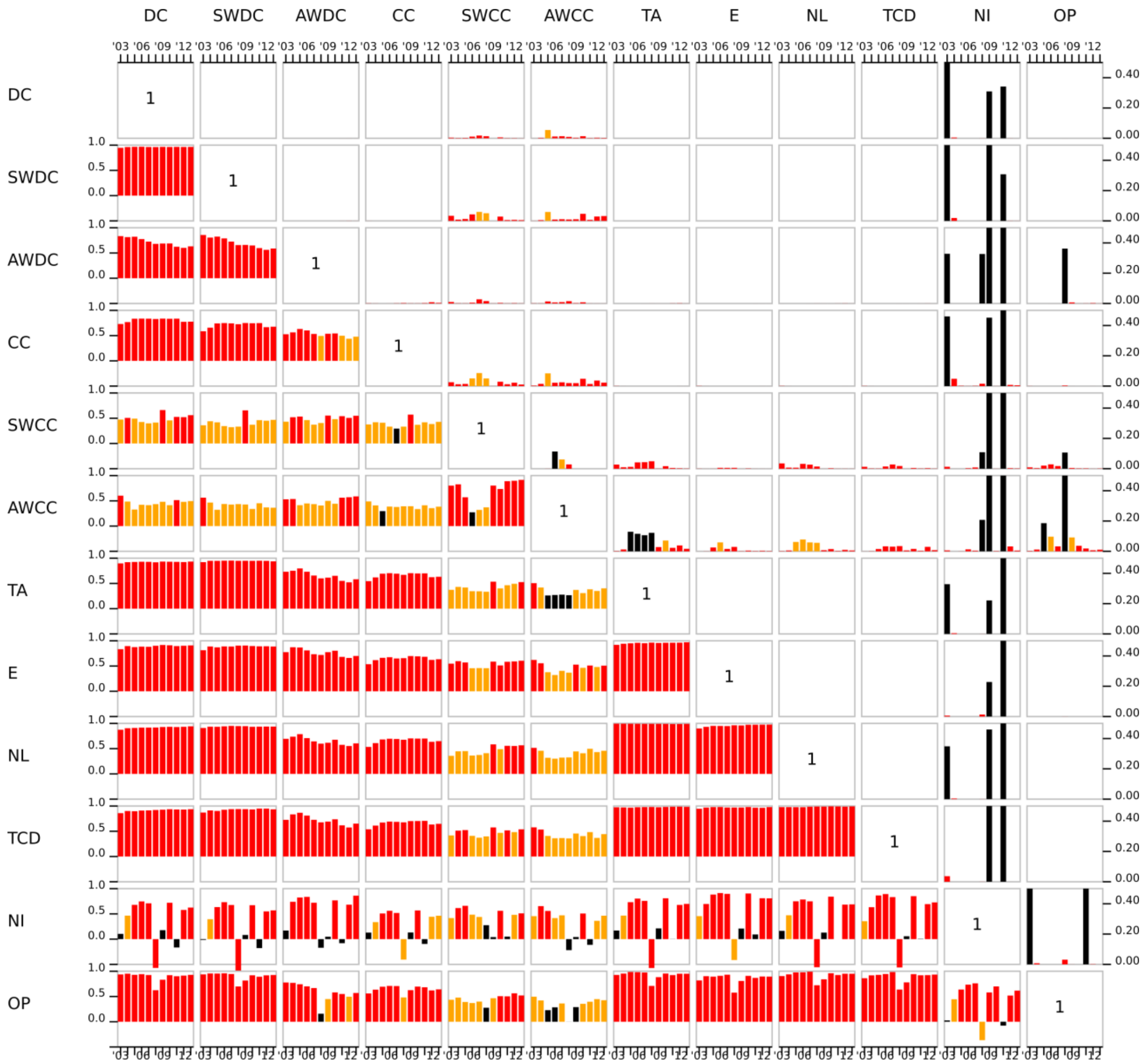

Figure 21. Centrality measure and financial index correlations in country-level network in 2003−2013. Abbreviations used: DC – Degree centrality, SWDC – Share-weighted degree centrality, AWDC – Assets-weighted degree centrality, CC – Closeness centrality, SWCC – Share-weighted closeness centrality, AWCC - Assets-weighted closeness centrality, TA - Total Assets, E - Equity, NL - Net Loans, TCD - Total Customer Deposits, NI - Net Income, OP - Operating Profit. Below diagonal are correlation coefficients, above diagonal are p-values. Red color denotes correlation coefficients ≥ 0.5 and p-values ≤ 0.05, orange color denotes correlation coefficients ≥ 0.3 and p-values ≤ 0.1.

Appendix F. Distribution of centrality measure and financial index period-wise correlations

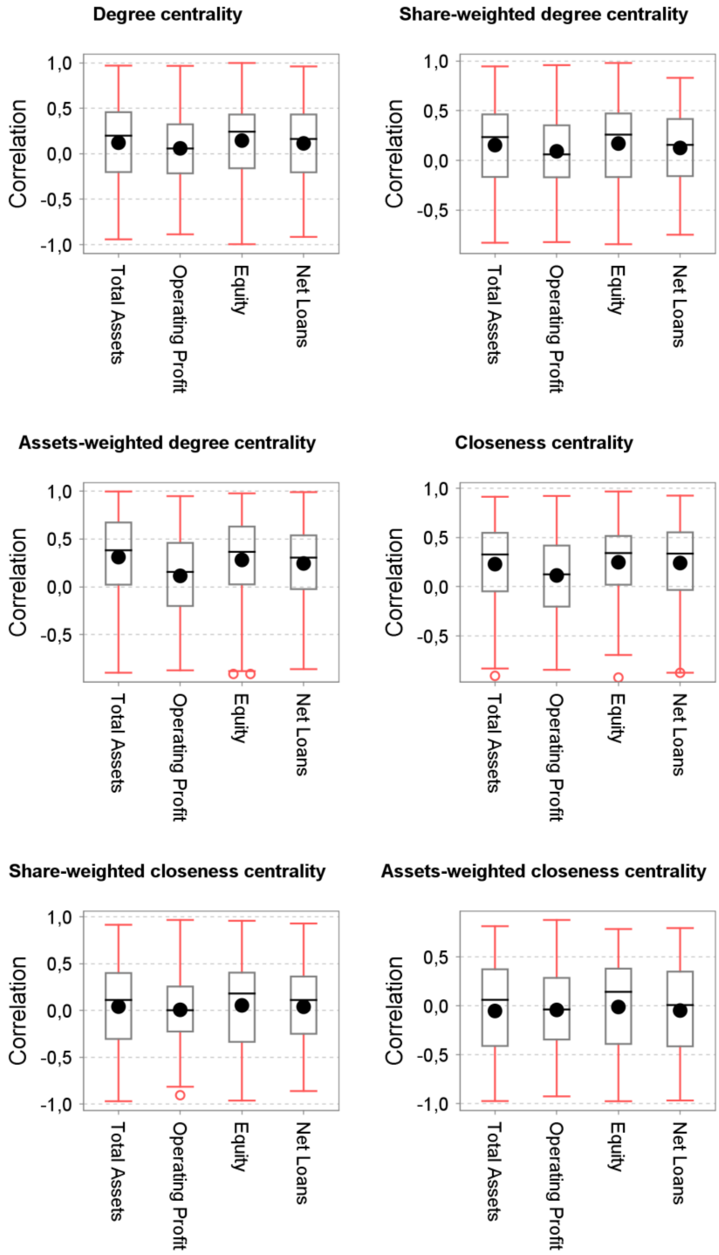

Figure 22. Distribution of correlation coefficients for pairs of series of financial index and centrality measure correlations over the period of 2003−2013 in bank-level ownership network.

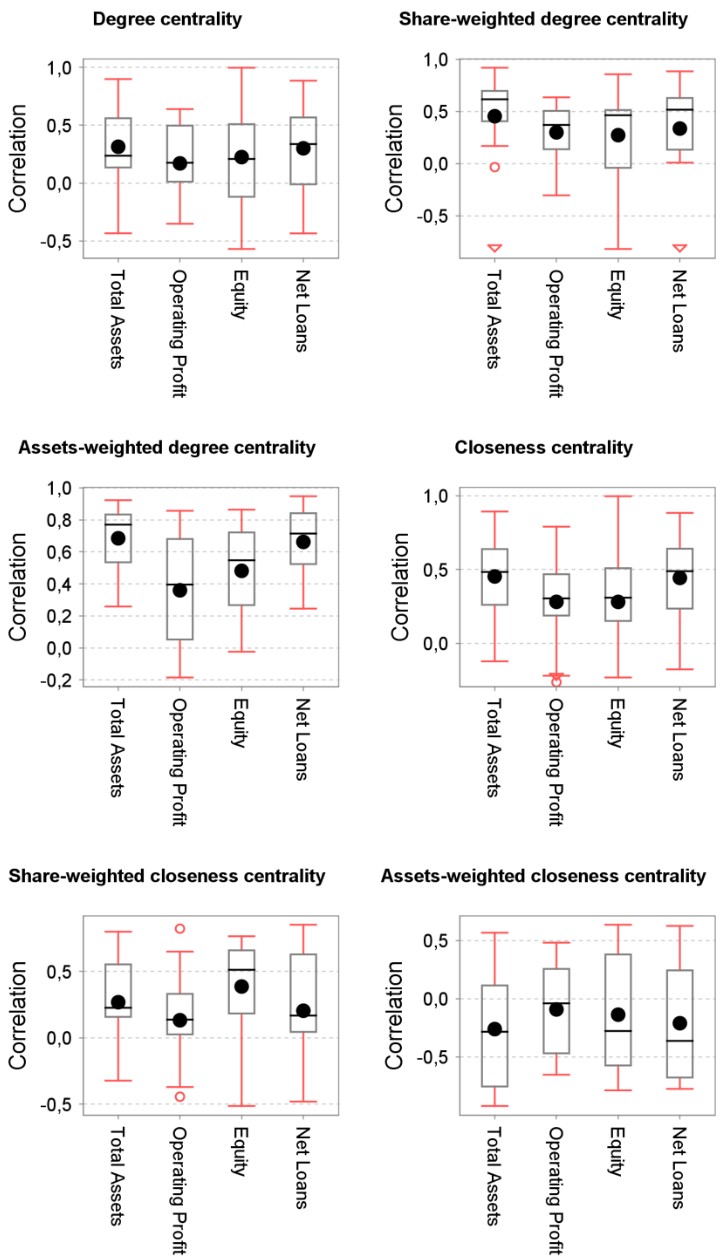

Figure 23. Distribution of correlation coefficients of pairs of series of financial index and centrality measure correlations over the period of 2003−2013 in country-level ownership network.

Appendix G. Implementation of modified weighted closeness centrality algorithm

```
package ee.ttu.datamining.bankscope.algorithms;

import dk.aaue.sna.alg.DijkstraForClosures;
import dk.aaue.sna.alg.PathCostCalculator;
import dk.aaue.sna.alg.centrality.CentralityMeasure;
import dk.aaue.sna.alg.centrality.CentralityResult;
import org.jgrapht.Graph;
import org.jgrapht.GraphPath;
import org.jgrapht.WeightedGraph;

import java.util.HashMap;
import java.util.Map;
import java.util.Set;

/**
 * This implements a closeness centrality for weighted networks, as proposed
in [1]. This algorithm assumes that weights
 * are positive. a weight 0 means absence of edge, and a weight of 10 is
twice as strong as a weight of 5.
 * This particular version treats weight as strength and the stronger the
link the "closer" the node is.
 * <p>
 * See {@link WeightedClosenessCentrality#setAlpha(double)} to control the
alpha parameter (default 1.0).
 * For {@code 0 < alpha < 1}, the number of edges are penalized and for
{@code alpha > 1} the number of edges are
 * favorized.
 * </p>
 * [1] Opsahl, Tore and Agneessens, Filip and Skvoretz, John. Node centrality
in weighted networks: Generalizing
 * degree and shortest paths. In Social Networks 33(3): pp. 245-251,
doi:10.1016/j.socnet.2010.03.006, 2010.
 *
 * @param <V> Node type
 * @param <E> Edge type
 * @author Soren A. Davidsen <soren@tanesha.net>
 * @author Mihhail Verhovtsov
 */
public class WeightedClosenessCentrality<V, E> implements
CentralityMeasure<V> {

    private WeightedGraph<V, E> graph;
    private double alpha = 1.0;

    public WeightedClosenessCentrality(WeightedGraph<V, E> graph) {
        this.graph = graph;
    }

    /**
     * Set the alpha parameter. Controls how much weights counts. For 0 = no
value to weight, 1 = use weight's value,
     * > 1 weight has more value.
     *
```

```java
     * @param alpha see description
     */
    public void setAlpha(double alpha) {
        this.alpha = alpha;
    }

    private class WeightedPathCost implements PathCostCalculator<V, E> {
        @Override
        public double getCost(Graph<V, E> veGraph, GraphPath<V, E>
veGraphPath, E newEdge) {
            return veGraphPath.getWeight() + Math.pow(1.0 /
graph.getEdgeWeight(newEdge), alpha);
        }
    }

    public CentralityResult<V> calculate() {
        Map<V, Double> cc = new HashMap<V, Double>();
        Set<V> V = graph.vertexSet();

        for (V n : V) {

            DijkstraForClosures<V, E> sp = new DijkstraForClosures<V, E>(
                    graph,
                    DijkstraForClosures.STRONGEST_PATH,
                    new WeightedPathCost(),
                    0.0,
                    n
            );

            double sum = 0.0;
            double s = 0.0;
            for (V p : V) {
                // skip reflexiveness
                if (n == p) continue;

                // get length of the path
                Double length = sp.get(p);

                // infinite -> there is no path.
                if (length == null || Double.isInfinite(length)) {
                    continue;
                }
                sum += length;
                s++;
            }

            if (sum == 0.0)
                cc.put(n, 0.0);
            else
                cc.put(n, sum * s / V.size());
        }

        return new CentralityResult<V>(cc, true);
    }

}
```

Appendix H. Graph building from CSV data

```
package ee.ttu.datamining.bankscope;

import ee.ttu.datamining.bankscope.data.BankAttributes;
import ee.ttu.datamining.bankscope.data.ShareSizes;
import ee.ttu.datamining.bankscope.data.ShareholderAttributes;
import ee.ttu.datamining.bankscope.graph.Share;
import ee.ttu.datamining.bankscope.graph.OwnershipGraph;
import ee.ttu.datamining.bankscope.graph.Shareholder;
import org.apache.commons.csv.CSVFormat;
import org.apache.commons.csv.CSVRecord;

import java.io.FileReader;
import java.io.IOException;
import java.io.Reader;
import java.util.HashMap;
import java.util.Map;

public class CsvGraphImporter {

    // Creates in-memory graph model from CSV
    public OwnershipGraph importGraphFromCsv(String csvFilename) throws
IOException {
        OwnershipGraph graph = new OwnershipGraph("banks");
        // Index of nodes (banks) by bvdIdNr
        final Map<String, Shareholder> nodes = new HashMap<>();

        int line = 0;

        try (Reader in = new FileReader(csvFilename)) {
            Iterable<CSVRecord> records =
CSVFormat.EXCEL.withHeader().parse(in);
            // Hold current owned bank node
            Shareholder bankNode = null;

            for (CSVRecord record : records) {
                line++;
                // Check if this row has new current owned bank
                if (!record.get("ID").isEmpty()) {
                    // Read bank data
                    BankAttributes bank = new
BankAttributes().fromCsvRecord(record);

                    // Get or create this bank (we could have encountered it
among shareholders)
                    bankNode = nodes.computeIfAbsent(record.get("BvD ID
Number"), s -> new Shareholder(s));
                    bankNode.bankAttributes = bank;

                    graph.addVertex(bankNode);
                }

                final String shareholderId = record.get("Shareholder - BvD ID
number");
                // Skip if there is no shareholder information on this line
                if (shareholderId == null || shareholderId.isEmpty())
continue;
```

```
                // Read shareholder (owner) data
                ShareholderAttributes shareholder = new
ShareholderAttributes().fromCsvRecord(record);

                // Read share sizes for each year
                ShareSizes shareSizes = new
ShareSizes().fromCsvRecord(record);

                // Check is this shareholder is state
                if (isPublicAuthority(shareholder)) {
                    // Check if current bank is (mostly) a public bank
                    bankNode.bankAttributes.publicBank =
shareSizes.isMajor();
                }

                // Aggregate shares for Bankscope data validation
                bankNode.totalShareSizes.aggregate(shareSizes);

                // Add shareholder only if it is a bank
                if (isBank(shareholder)) {
                    Shareholder shareholderNode =
nodes.computeIfAbsent(shareholderId, s -> new Shareholder(s));
                    shareholderNode.shareholderAttributes = shareholder;

                    graph.addVertex(shareholderNode);

                    // Avoid self-loops
                    if (shareholderNode.id.equals(bankNode.id)) {
                        //log("Self-loop detected: %s and %s",
shareholderNode, bankNode);
                        continue;
                    }
                    // Create an edge between bank and its shareholder
                    Share edge = graph.addEdge(shareholderNode, bankNode);
                    if (edge != null) {
                        edge.shareSizes = shareSizes;
                    }
                }
            }
        }
        return graph;
    }

    // Checks if shareholder is bank
    protected boolean isBank(ShareholderAttributes shareholder) {
        return shareholder.type.equals("Bank");
    }

    // Checks if shareholder is a public authority (government)
    protected boolean isPublicAuthority(ShareholderAttributes shareholder) {
        return shareholder.type.equals("Public authority, State,
Government");
    }

}
```

Appendix I. Input data processing

```
package ee.ttu.datamining.bankscope;

import ee.ttu.datamining.bankscope.data.BankSizeMap;
import ee.ttu.datamining.bankscope.data.ShareSizes;
import ee.ttu.datamining.bankscope.graph.OwnershipGraph;
import ee.ttu.datamining.bankscope.graph.Share;
import ee.ttu.datamining.bankscope.graph.Shareholder;
import org.jgrapht.alg.ConnectivityInspector;

import java.io.FileNotFoundException;
import java.io.IOException;
import java.io.PrintWriter;
import java.util.HashMap;
import java.util.Map;
import java.util.Objects;

import static ee.ttu.datamining.bankscope.data.BankGroups.detectBankGroup;
import static ee.ttu.datamining.bankscope.data.Countries.countries;
import static ee.ttu.datamining.bankscope.graph.Shareholder.Level.BANK_GROUP;
import static ee.ttu.datamining.bankscope.graph.Shareholder.Level.COUNTRY;
import static ee.ttu.datamining.bankscope.utils.GraphUtils.consolidateEdges;
import static ee.ttu.datamining.bankscope.utils.GraphUtils.consolidateNodes;
import static ee.ttu.datamining.bankscope.utils.LoggingUtils.debug;
import static ee.ttu.datamining.bankscope.utils.LoggingUtils.log;
import static java.util.Comparator.comparing;

public class GraphProcessor {

    public enum Level {BANKS, BANK_GROUPS, COUNTRIES}

    protected OwnershipGraph graph;
    protected Level level;

    public GraphProcessor(OwnershipGraph graph, Level level) {
        this.graph = graph;
        this.level = level;
    }

    public static GraphProcessor importGraphFromCsv(String csvFilename)
throws IOException {
        OwnershipGraph graph = new
CsvGraphImporter().importGraphFromCsv(csvFilename);
        return new GraphProcessor(graph, Level.BANKS);
    }

    public GraphProcessor withSizes(final BankSizeMap bankSizeMap) {
        graph.vertexSet().forEach(n -> {
            if (!bankSizeMap.containsKey(n.id)) {
                if (bankSizeMap.getByLabel(n.getLabel()) != null) {
                    debug("Size data was not found by BvD ID %s, but found by
name %s", n.id, n.getLabel());
                } else {
                    debug("Size data was not found for %s - %s", n.id,
n.getLabel());
                }
            }
```

```java
            n.bankSizes = bankSizeMap.get(n.id);
        });
        return this;
    }

    public GraphProcessor prepare() {
        graph.printStatistics("after importing");

        //graph.filterShareSizes(10.0);
        graph.removeSharesWithTooManyBlanks(9);
        graph.printStatistics("after removing shares with too many blanks");
        graph.fixShareGaps();
        graph.removeBanksWithoutSizes();
        graph.printStatistics("after removing banks without balance sheet
data");
        graph.removeDisconnectedBanks();
        graph.updateAbsoluteShareSizes();

        graph.printStatistics("after processing");

        ConnectivityInspector<Shareholder, Share> connectivityInspector = new
ConnectivityInspector<>(graph);
        log("Component count: %d",
connectivityInspector.connectedSets().size());
        //graph.printComponentSizes(connectivityInspector);
        graph.filterComponents(connectivityInspector, 1);
        graph.printStatistics("after removing mutual ownerships");

        graph.removeMutualOwnerships();

        graph.printStatistics("after filtering");

        return this;
    }

    public GraphProcessor consolidateByGroup(boolean consolidateUnknown) {
        log("\nConsolidating by bank group...");
        final OwnershipGraph consolidatedGraph = new OwnershipGraph("groups"
+ (consolidateUnknown ? "-other" : "-standalone"));
        final Map<String, Shareholder> consolidatedNodes = new HashMap<>();
        graph.vertexSet().forEach(n -> {
            final String group = detectBankGroup(n, consolidateUnknown);
            if (consolidateUnknown && group.equals("Other")) {
                System.out.println(n.getLabel() + " - " + n.getCountry());
            }
            Shareholder consolidatedNode =
consolidatedNodes.computeIfAbsent(group, id -> new Shareholder(id, group,
BANK_GROUP));
            consolidateNodes(consolidatedNode, n);
            consolidatedGraph.addVertex(consolidatedNode);
        });
        graph.edgeSet().forEach(e -> {
            String sourceGroup = detectBankGroup(graph.getEdgeSource(e),
consolidateUnknown);
            String targetGroup = detectBankGroup(graph.getEdgeTarget(e),
consolidateUnknown);

            if (!Objects.equals(sourceGroup, targetGroup)) {
```

```
                Shareholder source = consolidatedNodes.get(sourceGroup);
                Shareholder target = consolidatedNodes.get(targetGroup);
                Share groupEdge = consolidatedGraph.getEdge(source, target);

                if (groupEdge == null) {
                    groupEdge = consolidatedGraph.addEdge(source, target);
                    groupEdge.shareSizes = new ShareSizes(true);
                }
                consolidateEdges(groupEdge, e);
            }
        });

        printBlockmodelStatistics(consolidatedGraph);

        return new GraphProcessor(consolidatedGraph, Level.BANK_GROUPS);
    }

    public GraphProcessor consolidateByCountry() {
        log("\nConsolidating by country...");
        OwnershipGraph consolidatedGraph = new OwnershipGraph("countries");
        Map<String, Shareholder> countryNodes = new HashMap<>();
        graph.vertexSet().forEach(n -> {
            Shareholder countryNode =
countryNodes.computeIfAbsent(n.getCountry(), country -> new
Shareholder(country, countries.get(country), COUNTRY));
            consolidateNodes(countryNode, n);
            consolidatedGraph.addVertex(countryNode);
        });
        graph.edgeSet().forEach(e -> {
            String sourceCountry = graph.getEdgeSource(e).getCountry();
            String targetCountry = graph.getEdgeTarget(e).getCountry();
            if (!sourceCountry.equals(targetCountry)) {
                Shareholder source = countryNodes.get(sourceCountry);
                Shareholder target = countryNodes.get(targetCountry);
                Share countryEdge = consolidatedGraph.getEdge(source,
target);
                if (countryEdge == null) {
                    countryEdge = consolidatedGraph.addEdge(source, target);
                    countryEdge.shareSizes = new ShareSizes(true);
                }
                consolidateEdges(countryEdge, e);
            }
        });

        printBlockmodelStatistics(consolidatedGraph);

        return new GraphProcessor(consolidatedGraph, Level.COUNTRIES);
    }

    public GraphProcessor analyze() throws IOException {
        log("Analysing graph with %d nodes and %d links",
graph.vertexSet().size(), graph.edgeSet().size());
        GraphMetricAnalyzer metricAnalyzer = new GraphMetricAnalyzer(graph);
        GraphCentralityAnalyzer centralityAnalyzer = new
GraphCentralityAnalyzer(graph);

        metricAnalyzer.calculateMetrics();
        centralityAnalyzer.calculateCentralities();
```

```java
        return this;
    }

    public GraphProcessor exportToJson(String fileName) {
        new JsonGraphExporter().exportToJson(graph, fileName);
        return this;
    }

    public static void printBlockmodelStatistics(OwnershipGraph blockmodel) {
        try (PrintWriter out = new PrintWriter("output/" +
blockmodel.getName() + "-stats.txt")) {
            blockmodel.shareholders()
                    .sorted(comparing((Shareholder sh) ->
sh.getConsolidated().size()).reversed())
                    .forEach(sh -> out.format("%s\t%d\n", sh.getLabel(),
sh.getConsolidated().size()));
        } catch (FileNotFoundException e) {
            e.printStackTrace();
        }
    }
}
```

```java
package ee.ttu.datamining.bankscope.graph;

import org.jgrapht.alg.ConnectivityInspector;
import org.jgrapht.graph.SimpleDirectedWeightedGraph;

import java.util.Set;
import java.util.function.Predicate;
import java.util.stream.Stream;

import static ee.ttu.datamining.bankscope.utils.LoggingUtils.log;
import static java.util.stream.Collectors.toSet;

public class OwnershipGraph extends SimpleDirectedWeightedGraph<Shareholder,
Share> {

    protected String name;

    protected Predicate<Shareholder> beingDisconnectedBank = n ->
this.inDegreeOf(n) == 0 && this.outDegreeOf(n) == 0;

    public OwnershipGraph(String name) {
        super(Share.class);
        this.name = name;
    }

    public String getName() {
        return name;
    }

    public Shareholder findShareholder(String id) {
        return shareholders().filter(shareholder ->
id.equals(shareholder.id)).findFirst().orElse(null);
    }
```

```
    public void filterShareSizes(double threshold) {
        int count = removeShares(e -> !e.shareSizes.stream().anyMatch((share)
-> share.hasValue() && share.direct >= threshold));
        log("Removed %d shared below %.2f%%", count, threshold);
    }

    public void removeBanksWithoutSizes() {
        int count = removeShareholders(v -> v.bankSizes == null);
        log("Removed %d banks that had no size data", count);
    }

    public void fixShareGaps() {
        long fixed = this.edgeSet().stream().filter(e ->
e.shareSizes.fixGaps()).count();
        log("Fixed gaps in %d shares", fixed);
    }

    public void removeSharesWithTooManyBlanks(final int threshold) {
        int count = removeShares(e -> e.shareSizes.getCount() < threshold);
        log("Removed %d shares that had too many blanks (threshold = %d)",
count, threshold);
    }

    public void removeDisconnectedBanks() {
        int count = removeShareholders(beingDisconnectedBank);
        log("Removed %d banks that are disconnected", count);
    }

    public void filterComponents(ConnectivityInspector<Shareholder, Share>
connectivityInspector, int topN) {
        // Filter all nodes outside of the largest component
        connectivityInspector.connectedSets().stream()
                .sorted((n1, n2) -> n2.size() - n1.size())  // descending
sort by component size
                .skip(topN)                                 // skip the
largest
                .forEach(s -> this.removeAllVertices(s));   // remove all
nodes from each of the rest
    }

    public void printStatistics(String context) {
        long totalNodes = shareholders().count();
        long totalEdges = shares().count();
        long totalBanks = shareholders().filter(n -> n.bankAttributes !=
null).count();
        long totalShareholders = shareholders().filter(n ->
n.shareholderAttributes != null).count();
        long totalOverlaps = totalBanks + totalShareholders - totalNodes;
        long totalValidShares = shareholders().filter(n ->
n.validTotalShares()).count();
        long totalValidSharesOrNone = shareholders().filter(n ->
n.validTotalSharesOrNone()).count();

        log("Total %d nodes (%d owned banks and %d shareholders, %d
overlapping) and %d edges %s", totalNodes, totalBanks, totalShareholders,
totalOverlaps, totalEdges, context);
        log("%d owned banks have total sum of shares ~100%% and %d have
```

```java
~100%% or none", totalValidShares, totalValidSharesOrNone);
    }

    public void removeMutualOwnerships() {
        removeShares(e -> {
            Share f = this.getEdge(this.getEdgeTarget(e),
this.getEdgeSource(e));
            if (f == null) return false;
            log("Mutual ownership from %s to %s, [%s] vs [%s]",
this.getEdgeSource(e).id, this.getEdgeTarget(e).id, e.shareSizes,
f.shareSizes);
            // Remove either less filled one or both
            return f.shareSizes.getCount() >= e.shareSizes.getCount();
        });
        log("Mutual ownerships have been removed");
    }

    public int removeShareholders(Predicate<Shareholder> predicate) {
        Set<Shareholder> shareholdersToBeRemoved =
shareholders().filter(predicate).collect(toSet());
        this.removeAllVertices(shareholdersToBeRemoved);
        return shareholdersToBeRemoved.size();
    }

    public int removeShares(Predicate<Share> predicate) {
        Set<Share> sharesToBeRemoved =
shares().filter(predicate).collect(toSet());
        this.removeAllEdges(sharesToBeRemoved);
        return sharesToBeRemoved.size();
    }

    public void updateAbsoluteShareSizes() {
        shareholders().forEach(shareholder -> {
            this.incomingEdgesOf(shareholder).forEach(share ->
share.shareSizes.updateAbsoluteShareSizes(shareholder));
        });
    }

    public Stream<Shareholder> shareholders() {
        return vertexSet().stream();
    }

    public Stream<Share> shares() {
        return edgeSet().stream();
    }

    @Override
    public String toString() {
        return getName();
    }
}


package ee.ttu.datamining.bankscope.data;

import ee.ttu.datamining.bankscope.graph.Shareholder;
import org.apache.commons.csv.CSVRecord;
```

```java
import java.util.stream.Stream;

import static ee.ttu.datamining.bankscope.data.Period.yearIndex;
import static ee.ttu.datamining.bankscope.data.Period.years;
import static ee.ttu.datamining.bankscope.utils.ConversionUtils.parseDouble;
import static java.lang.String.join;
import static java.util.stream.Collectors.toList;
import static java.util.stream.Stream.of;

/**
 * Data structure that holds share sizes for each year of the observed
period.
 */
public class ShareSizes {

    protected ShareSize[] sizes = new ShareSize[Period.length];

    public ShareSizes() {
    }

    public ShareSizes(boolean zeroed) {
        Double value = zeroed ? 0.0 : null;
        for (int year : years()) {
            sizes[yearIndex(year)] = new ShareSize(value, value);
        }
    }

    public ShareSizes fromCsvRecord(CSVRecord record) {
        for (int year : years()) {
            Double directShare = parseShare(record.get("Shareholder - Direct
% - 12/" + year));
            Double totalShare = parseShare(record.get("Shareholder - Total %
- 12/" + year));
            sizes[yearIndex(year)] = new ShareSize(directShare, totalShare);
        }
        return this;
    }

    public ShareSize get(int year) {
        return sizes[yearIndex(year)];
    }

    public boolean has(int year) {
        return get(year) != null;
    }

    public void set(int year, ShareSize shareSize) {
        sizes[yearIndex(year)] = shareSize;
    }

    public Stream<ShareSize> stream() {
        return Stream.of(sizes);
    }

    public long getCount() {
        return of(sizes).filter(ShareSize::hasValue).count();
    }
```

```java
    private Double parseShare(String value) {
        if (isBlank(value) || value.equals("-") || value.equals("n.a.")) {
            return null;
        }
        if (value.equals("WO")) { // wholly owned
            return 100.00;
        }
        if (value.equals("MO")) { // Majority owned
            return 50.00;
        }
        if (value.equals("NG")) { // Negligible
            return 0.00;
        }
        if (value.equals("CQP1")) { // CQP1 = 50% + 1 share
            return 50.00;
        }
        // Ignore everything else that is not a number
        return parseDouble(value);
    }

    public boolean fixGaps() {
        Double previousWeight = null;
        boolean gap = false;
        boolean fixed = true;
        for (int i = 0; i < sizes.length; i++) {
            ShareSize shareSize = sizes[i];
            if (shareSize.weight == null && previousWeight != null) {
                gap = true;
            } else if (shareSize.weight != null) {
                if (gap) {
                    int j = i - 1;
                    // Fill the gaps...
                    while (sizes[j].weight == null) {
                        // .. with average value
                        sizes[j--].weight = (shareSize.weight -
previousWeight) / 2;
                    }
                    gap = false;
                    fixed = true;
                }
                previousWeight = shareSize.weight;
            }

        }
        return fixed;
    }

    public void aggregate(ShareSizes additionalShareSizes) {
        for (int year : years()) {
            get(year).aggregate(additionalShareSizes.get(year));
        }
    }

    public void updateAbsoluteShareSizes(Shareholder shareholder) {
        for (int year : years()) {

get(year).calculateAbsoluteWeight(shareholder.bankSizes.get(FinancialIndex.TO
```

```
TAL_ASSETS, year));
        }
    }

    public boolean isMajor() {
        return of(sizes).allMatch(share -> share.direct == null ||
share.direct >= 50);
    }

    @Override
    public String toString() {
        return join(", ", of(sizes).map(sz -> String.format("%.2f",
sz.weight)).collect(toList()));
    }

    private static boolean isBlank(String in) {
        return in == null || in.trim().isEmpty();
    }

}
```