TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
20/2019

# From Econophysics to Networks: Structure of the Large-Scale Estonian Network of Payments

STEPHANIE RENDÓN DE LA TORRE

TTÜ PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Science
Department of Cybernetics
This dissertation was accepted for the defense of the degree of Doctor in Physical Sciences on 13/02/2019

**Supervisor**: Prof. Dr. Jaan Kalda
School of Science, Department of Cybernetics
Tallinn University of Technology
Tallinn, Estonia

**Opponents** Dr. Marco Patriarca
National Institute of Chemical Physics and Biophysics
Tallinn, Estonia

Dr. Michael Isichenko
SVP at Jefferies

**Defense of the thesis**: 16/05/2019, Tallinn

**Declaration:**
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology, has not been previously submitted for doctoral or equivalent academic degree.

Stephanie Rendón de la Torre

---------------------------------------------------------------------
Signature

European Union
European Regional
Development Fund

Investing
in your future


ISSN 2585-6898 (publication)
ISBN 978-9949-83-418-1 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9949-83-419-8 (PDF)

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
20/2019

# Majandusfüüsikast võrgustikesse: Eesti maksevõrgustike struktuur

STEPHANIE RENDÓN DE LA TORRE

TTÜ KIRJASTUS

# Contents


List of Publications ... 6

Author's Contribution to the Publications ... 7

Introduction ... 8

Motivation and contribution ... 9

Objectives and outline of the thesis ... 11

Approbation of the results ... 13

1. Topologic structure of economic complex networks: Payments network of Estonia ... 14
1.1 Structure and components: Analytics metrics ... 15
1.1.1 Activity patterns ... 15
1.1.2 Statistical properties and common metrics ... 17
1.2 Scale degree distributions in other known networks ... 20
1.3 Components of networks ... 21
1.3.1 Importance of nodes: strength ... 25
1.4 Robustness of the network ... 28

2. Detecting overlapping community structure ... 33
2.1 Structures of communities ... 35
2.2 Distribution functions ... 37

3. Multifractal networks ... 41
3.1 Fractal network analysis ... 42
3.2 Multifractal network analysis ... 44

Conclusions ... 49

List of Figures ... 52

List of Tables ... 54

References ... 55

Acknowledgements ... 59

Abstract ... 60

Kokkuvõte ... 61

Appendix ... 63

Curriculum vitae ... 144

Elulookirjeldus ... 146

## List of Publications

The thesis is based on five academic publications which are referred to in the text as Paper I, Paper II, Paper III, Paper IV and Paper V. Four papers are indexed by the ISI Web of Science. The list of the publications, on the basis of which the thesis has been prepared:

I **Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2016). On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. *Chaos, Solitons & Fractals*, 90, 18−27 DOI:10.1016/j.chaos.2016.01.018.

II **Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2017). Fractal and multifractal analysis of complex networks: Estonian network of payments. *The European Physical Journal B*, 90. DOI: 10.1140/epjb/e2017-80214-5.

III **Rendón de la Torre S**., Kalda J. (2018) Review of structures and dynamics of economic complex networks: Large scale payment network of Estonia. In: Zengqjang C., Dehmer M., Emmert-Streib F., Shi Y. (eds.), *Modern and interdisciplinary problems in network science.* Taylor & Francis CRC Group, USA, 193-226 https://www.crcpress.com/Modern-and-Interdisciplinary-Problems-in-Network-Science-A-Translational/Chen-Dehmer-Emmert-Streib-Shi/p/book/9780815376583.

IV **Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2019) Detecting overlapping community structure: Estonian network of payments. *Proceedings of the Estonian Academy of Sciences*, 68(1) 79-88. DOI:10.3176/proc.2019.1.08

V **Rendón de la Torre S**., Kalda J., Kitt R. (2019) Specific statistical properties of the strength of links and nodes of the Estonian network of payments. *Proceedings of the Estonian Academy of Sciences,* 68(3). Manuscript (in press). DOI:10.3176/proc.2019.3.02

## Author's Contribution to the Publications

Contribution to the papers in this thesis are:

I I performed the calculations, visualization and most of the analysis of the results, drafted the manuscript, and acted as corresponding author of the manuscript.

II I performed the calculations, visualization and most of the analysis of the results, drafted the manuscript, and acted as corresponding author of the manuscript.

III I performed the calculations, visualization and most of the analysis of the results, drafted the manuscript and acted as corresponding author of the manuscript.

IV I performed the calculations, visualization and most of the analysis of the results, drafted the manuscript and acted as corresponding author of the manuscript.

V I performed the calculations, visualization and most of the analysis of the results, drafted the manuscript and acted as corresponding author of the manuscript.

# Introduction

The networks science approach for financial and economic systems has potential to go further on the frontiers of research. In recent years, a part of the main focus of research has tilted towards the discovery and understanding of the underlying financial, social and economic systems' structures through the use of the tools of complex networks science. In this context, the network approach has two sources of origin: one source originates from economics, finance and sociology while the second source originates from computer science, physics, complexity and mathematics. The convergence point of both sources of origin attempts to combine economy and complex systems studies and this approach can be translated into a graph representation of economic systems in order to study how interactions among the components of the graph occur whatsoever the nature of the relations between the components is.

Nowadays, complex networks are a central concept where an intuitive path that fuses economy with complex systems studies emerges. Complex networks can be biological, technological, economic, social, and cultural, among other types. The physical approach has become increasingly important during recent years regarding complex networks structures studies.

The field of complex networks is developing at a fast pace and has already made significant progress towards the design of its own framework with the purpose of unravelling the organizing principles that govern complex networks and their evolution. Particularly, the structures and dynamics of complex networks have attracted considerable attention from the research community in recent years.

## Motivation and contribution

During the last decade, important efforts have been made towards two things: 1) the improvement of our understanding of the topological structures underlying complex networks and 2) the unveiling of large-scale characteristics which belong to such systems. Particularly those efforts have been focused towards theoretic properties of networks and generic models that are able to represent systems statistically. However, with the help of the latest technological advances, increasing data sets of unstudied systems are emerging and there is a generic need to study newer and real complex networks which will nurture new model developments in the future and help in the construction of the theoretical framework of complex networks theory.

The neologism *econophysics* was first coined by H. Eugene Stanley in a Statphys conference in 1995 held in Kolkata, India. Mantegna and Stanley (2000) defined *econophysics* as a multidisciplinary field that denotes the activities of physicists who work on economic problems in order to test a variety of new conceptual approaches derived from physical sciences. Much has been studied and developed in econophysical studies since then. Mainly, those studies originated from models of statistical mechanics. Similarly, problems related with distribution of income, wealth and economic returns in financial markets have been addressed in research papers, and mostly these topics are related with the insufficiency to explain non-Gaussian distributions and scaling properties which are empirically detected by traditional economic theoretic approaches. Some of the most relevant outcomes of the research accomplished in the area of econophysics are related to: 1) The detection and explanation of power-law tails observed in the distributions of different types of financial data, 2) the existence of certain underlying universalities in the behaviour of individual market agents and 3) the detection of similarities between financial time series and natural phenomena.

Network science is an active interdisciplinary field of research that originates from branch mathematics: graph theory, extended into different directions including towards economics, statistical mechanics, computer science, neuroscience, sociology, transportation, ecological systems and biology. With complex networks it is possible to describe the structure of any system, when the system is suitable to be represented as a graph.

“Complexity”, may refer to the quality of a system or to a quantitative characterization of a system. As a quality of a system, it refers to what makes a system complex and it has something to do with the ability to understand a system; it refers to the existence of emergent properties which appear as a consequence of the interactions of the components of the system (Standish, 2008). An example of a property that emerges as a consequence of global organizational structure of a network is the “small world” property, which is characterized by small average path length and a high number of triangles in the network. As a quantitative characterization of a system, “complexity” is used as a quantity when referring to something that is more complicated than another thing; it refers to the quantity of information needed to specify the system. For real-world networks a huge amount of information is needed to describe a system, such as the number of nodes, links, degree correlations, degree distributions, clustering coefficients, diameter, betweenness, centralities, community structure, average or shortest paths, communication patterns and other quantities. In random networks the only information needed to describe their structure is the number of nodes and the probabilities for linking pairs of nodes.
The network representation of real networks is called “complex networks” because of two reasons. Firstly, because there are characteristics that arise as a consequence of the global

topological organization of the system and secondly because these structures cannot be trivially described like in the cases of random or regular graphs (Estrada, 2011).

The theoretical framework behind complex networks is continuously developing, advancing at a fast pace, and has already made significant progress towards unravelling the organizing principles that govern complex networks structures and their dynamics. Studies related with: topological features, dynamical aspects, community detection, network phenomena and particular properties of networks have been the focus of attention of extensive research in the last couple of decades (Reka and Barabási, 2002; Dorogovtsev and Mendes, 2003; Furuya and Yakubo, 2011, Newman 2010; Palla, Barabási and Vicsek, 2007; Watts and Strogatz, 1998).

Networks play an important role in a wide range of economic and social phenomena. The use of techniques and methods from graph theory has permitted economic networks studies to expand the knowledge and give insights into economic and social phenomena in which the embeddedness of individuals or agents in their social or economic interrelations cannot be neglected (König and Battiston, 2009). For example, Souma et al., (2006) studied a shareholder network of Japanese companies where the authors analysed the companies' growth through economic networks. Other examples of interesting applications of complex networks in economics are provided by the regional investment or ownership networks where European company-to-company investment stocks show power-law distributions that allow predicting the investments that will be received or made in specific regions, based on the connectivity and transactional activity of the companies (Battiston et al., 2007; Glattfelder et al., 2009). Nakano and White (2007) have shown that analytic concepts and methods related with complex networks can help to uncover structural factors that may influence the price formation for empirical market-link formations of economic agents. Reyes et al., (2008) used a weighted network analysis focused on using random walk betweenness centrality to study why high-performing Asian economies have higher economic growth than Latin-American economies. Complex network-based approaches are very useful and provide means by which to monitor complex economic systems and may help in providing better control in managing and governing these systems. Another interesting line of research is related to network topology as a basis for investigating money flows of customer driven banking transactions. A few recent papers describe the actual topologies observed in different financial systems (Lublóy, 2006; Inaoka et al., 2004; Soramäki et al., 2007; Boss et al., 2004) Some other works have focused on economic shocks and robustness in economic complex networks (Iory and Jafarey, 2001; Iori et al., 2007).

## Objectives and outline of the thesis

The main focus of my doctoral studies has been to study general and particular properties of complex networks through analysis that consists of different experiments on a unique, interesting and new economic network: the large-scale Estonian network of payments.

Networks can be studied from different points of view, for example: from a local, global or mesoscale perspective. The contribution of my doctoral studies is to explore such approaches by using different methodologies and experiments while studying the economic development of Estonia.

This thesis presents an extensive study that contributes to the field of complex networks (particularly to economic complex networks studies) by adding empirical evidence in favour of economic networks with a new study case. The study is done thanks to the application of known network methods. In this work I investigate the structure of the large-scale network of payments of Estonia:

1. Global and local topology
2. Community structures
3. Fractal and multifractal properties

The data set was obtained from Swedbank's databases. Swedbank is one of the leading banks in the Nordic and Baltic regions of Europe. The bank operates actively in Estonia, Latvia, Lithuania and Sweden. All the information related to the identities of the nodes is very sensitive and thus will remain confidential and unfortunately cannot be disclosed. The data set is unique in its kind and very interesting since ~80% of Estonia's bank transactions are executed through Swedbank's system of payments; hence, this data set reasonably reproduces the structure of the Estonian economy and can be used as a proxy of it. The data set utilized, contains the best possible information available and describes fairly accurately the tendencies of money transactions. This data set comprises domestic payments (company-to company electronic transactions) of the year 2014. The network consists of 16,613 nodes, 2,617,478 payment transactions, and 43,375 links. In this economic network, the nodes represent Estonian companies and the links represent payments done between the companies.

The main objectives of this thesis are as follows:

- To study the structure (topology) of the network of payments of Estonia.
- To study the structure and functionality of this large network and to expand the knowledge of the local organization of its components by detecting community structures in this network.
- To study the fractality and multifractality of the network.

To achieve these objectives, Chapter 1 mainly follows the results of Paper I, III and V. Chapter 1 presents a topological analysis of the structure of the large-scale Estonian network of payments. To achieve this, I study scale-free properties, network components and patterns, statistical properties and robustness of the network.

In Chapter 2 I discuss about the community structure of the Estonian network of payments and scale-free properties at a mesoscale level. This Chapter mainly follows Paper III and Paper IV. Chapter 2 also addresses the analysis of the global structure of the network through the distribution functions of four basic quantities.

Next, I study the fractal and multifractal structure of the network in Chapter 3 by following the results of Paper II. I present a fractal scaling analysis by calculating the fractal dimension of the network and its skeleton. Then I use a sandbox algorithm to calculate the spectrum of the generalized fractal dimensions and mass exponents in order to study the multifractal behaviour.

Paper III mainly consists of a review of the results obtained in Papers I, II and IV which was published as a chapter of a book. This Paper consists of a review on the structures and characteristics of the large-scale Estonian network of payments.

# Approbation of the results

The basic results described in this thesis have been presented by the author at the following international conferences:

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. Estonian network of payments. Poster presentation at the *5th Ph.D. School-Conference on Mathematical Modeling of Complex Systems* (20-30 July, 2015, Patras, Greece)

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. On the topologic analysis of economic complex networks: Swedbank's network of payments in Estonia. Oral presentation at *Data Science Challenges* (14-17 October, 2015, Torino, Italy).

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. Communities' detection and evolution: Estonian economic network of payments. Poster presentation at the *26$^{th}$ STATPHYS IUPAP International Conference on Statistical Physics* (18-22 July, 2016, Lyon, France).

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. Communities detection and dynamics: Estonian economic network of payments. Poster presentation at *International Conference Statistical Physics SIGMAPHI 2017 (and School of Statistical Physics)* (6-14 July, 2017, Corfu, Greece).

**Rendón de la Torre S.,** Kalda J., Kitt R., Engelbrecht, J. Fractal and multifractal analysis of complex networks: Estonian network of payments. Poster presentation at *NetSci-X 2018 International School and Conference on Network Science* (5-8 January, 2018
Hangzhou, China)

**Rendón de la Torre S**., Kalda J., Kitt, R. Specific statistical properties of the strength of links and nodes: Estonian network of payments. Poster presentation at *Statphys 27* (8-12 July, 2019, Buenos Aires, Argentina)

# 1. Topologic structure of economic complex networks: Payments network of Estonia

The focus of this Chapter and the underlying Paper I is to present the first topological analysis of the economic structure of an entire country based on payments data: Estonia's network of payments. I focus on analysing interesting structural properties of such network, with a particular emphasis on topologic components. I study the scaling-free and structural properties of the network. Also, I examine statistical characteristics, components and patterns. I identify the hubs of the network and perform a simulation of resiliency against a random and a targeted attack of the nodes. I found that by identifying and studying the links between nodes it is possible to perform vulnerability analysis of the Estonian economy with respect to economic shocks.

A random network is the most basic model of all network formations and it is based on the assumption that a fully random process is responsible for the structure of the links in a network. The properties of random network models (Erdös and Rényi, 1959) provide rich insight into the characteristics of complex networks. Firstly, I focus on analysing some interesting structural properties of the network, with special interest on topologic components. Graph theory definitions not introduced in this Chapter can be found in (Dorogovtsev and Mendes, 2003; West, 2003).

Random network models are useful benchmarks for comparing empirical networks and have the ability to identify the elements that are a result of randomness and the ones that can be rooted to other factors. Some properties of random networks that are useful for studying generic networks are, for example: the distribution of links across nodes, connectivity in terms of paths, distances within networks, shortest-average paths, diameter, etc.

A graph is a mathematical and symbolic representation of a network and of its connectivity. A simple undirected graph $\mathrm{G}$ is a set of vertices $\mathrm{V}$ connected with edges $\mathrm{E}$, therefore $\mathrm{G} = (\mathrm{V}, \mathrm{E})$. A graph is defined by the structural information contained in its adjacency matrix. A network may have a large arbitrary amount of additional information on top of it: for example, edges can have attributes such as capacity or weight, or it may be a function of other variables. Also, in a network the vertices are called nodes and the edges are called links. Network terminology is generally used when the links transport or send something meaningful between the nodes (like in social, computer, biological, transport, or economic networks).

There are many ways to define the network of payments. In this study I consider more than one definition. In the first definition, an undirected graph is mapped, a symmetric payment adjacency matrix $\mathrm{A}_{\mathrm{NxN}}$ where $\mathrm{N}$ is the total number of nodes in the network, and considering that two companies are connected if they have at least one payment, then $a_{ij}^{u} = a_{ji}^{u}$ and $a_{ij}^{u} = 1$ if there is a transaction between company *i* and *j*, otherwise, $a_{ij}^{u} = 0$ if there is no transaction between companies $i$ and $j$. Diagonal elements are equal to 0 and non-diagonal elements are either 0 or 1.

The links can also represent direction: where the links follow the flow of money. The second definition is a directed graph where the links follow the flow of money, such that a link is incoming to the receiver and is outgoing for the sender of the payment. There are two more matrices, one for the in-degree case and another for the out-degree case.

The weighted connectivity matrix *B* contains the number of transactions between companies *i* and *j*. The elements $w_{ij}$ of the weighted connectivity matrix *C* denote the overall volume exchanged between companies *i* and *j.* The choice of the definition of the matrix representation depends on the focus of the analysis.

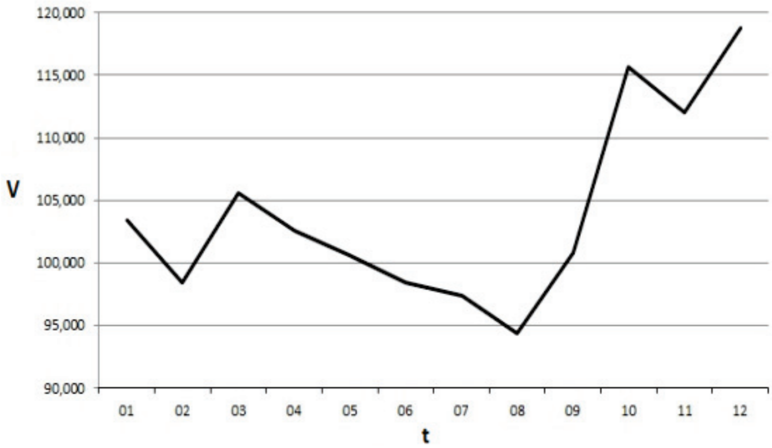
## 1.1 Structure and components: Analytics metrics

### 1.1.1 Activity patterns

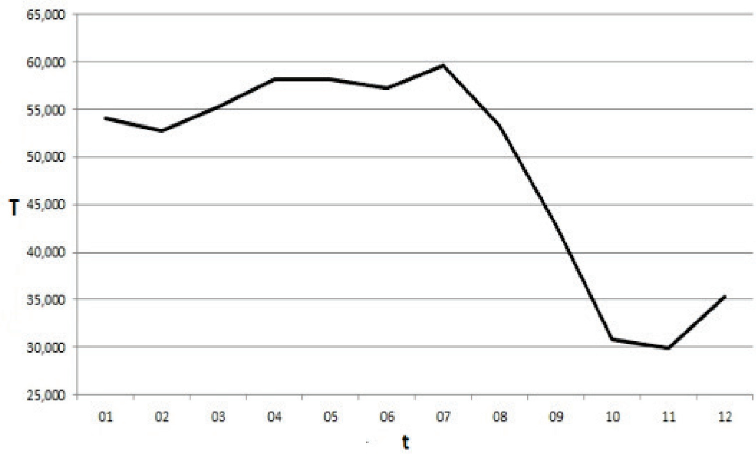
I perform an analysis on the structure of the network and its time evolution to be able to identify any structural changes and compare the emerging patterns. Figure 1(a) shows the monthly volume of transactions during 2014 while Figure 1(b) displays the monthly number of transactions. Figure 1(c) shows the monthly average number of active links as a function of time. Figures 1(b) and 1(c) show that the number of transactions decreases dramatically in the third quarter of the year while the number of active links decreases already in the second quarter.

The plots show that the number of transactions and the active links increase in the last quarter of the year, suggesting that liquidity in the Estonian network of payments increases by the end of the year through increased transaction volumes and payments, and higher than usual number of active counterparties. It is interesting that from August until the end of the year there is a high concentration in the volume of payments while the number of payments diminishes dramatically in the same period of time. These observations indicate that the average number of active companies has decreased 20%, while the volume of transactions has increased 14% and the number of transactions has decreased 66% by the end of the year (compared with the beginning of the year). This indicates that in Estonia companies manage higher volumes of money at the end of the year than at the beginning of the year, while not all the companies remain active by the end of the year. It is observed that there are some seasonal effects characterizing the trends of payments. For example, the fact that there is an impact in summer: transaction-volumes are visibly increased, perhaps linked to the fact that consumer expenses rise (such as traveling or vacation related expenses).

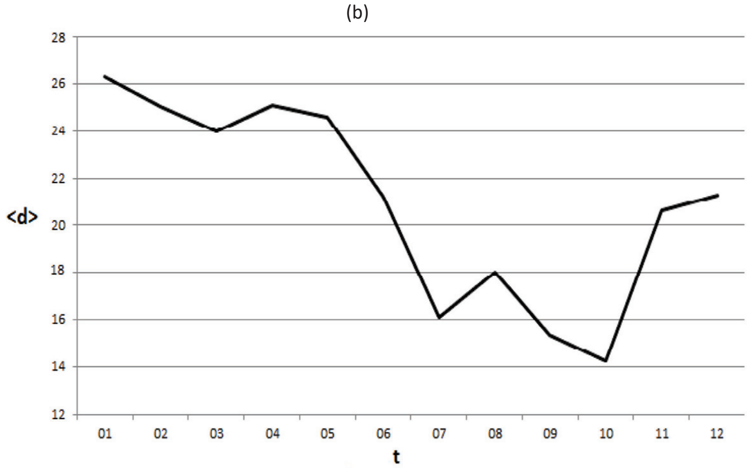
It is not possible to obtain a full explanation of this financial liquidity pattern due to the lack of complete information about the overall financial and commercial activities of the companies in this network. Nonetheless, there are some possible explanations for such patterns. For example, these patterns could be highly affected by business cycles of payments, or by seasonal effects on the liquidity of companies, or by macroeconomic variations such as the changes in monetary policy of the Euro area and Estonia. Another explanation for the increased volume of transactions and increased liquidity at the end of the year is that there might be a generalized release of delayed payments, like when companies try to spend the remaining balances of their annual budgets.




(a)




(b)




(c)

*Figure 1. Time evolution activity of patterns (payments and volumes). (a) Monthly trading volumes of payments V. (b) Monthly number of transactions T. (c) Average degree ⟨d⟩ versus time t. X-axis represents the number of the month of year 2014.*

### 1.1.2 Statistical properties and common metrics

In order to characterize the statistical properties and the underlying structure of the network, I use some common metrics (Standish, 2008; Estrada, 2011; Rendón de la Torre et al., 2016) that combine topological and weighted observables. The most basic properties of a network are the number of nodes N and the overall number of links $k$. The number of nodes defines the size of the network, while the number of links relative to the number of possible links defines the connectivity of a network. Connectivity $(p)$ is the unconditional probability that two nodes are connected by a direct link. For a directed network, connectivity is defined as follows

$$(p) = \frac{k}{n(n-1)} . \qquad (1)$$

The connectivity of the network is 0.13, meaning that the network is sparse and 87% of the potential connections are disabled. Diameter $d$ is the maximum distance between two nodes (measured by the number of links) and this distance is equal to 29. Random networks usually have small diameter values. The differences between the low diameter in similar random networks and real networks like this one could be explained by the preferred money paths that nodes have in the network. Having preferred money paths means that some companies have specific preferences when considering the counterparties they transact with. Intuitively, this makes sense because for a company it is important to choose carefully which counterparties become trading partners, clients, service providers or suppliers and which ones do not. Usually, this decision is based upon determined factors such as geographical location, goals affinity, cost policies, future joint ventures, legal agreements, nature of the business or various other reasons, and it is interesting to notice how this particular feature can be observed through the comparison of the connectivity of the network and a random network.

A path is a sequence of nodes such that each node is linked to the next node along the path by a link. A path consists of $n + 1$ nodes and $n$ links. A path between nodes $i$ and $j$ is an ordered list of $n$ links. The length of this path is $n$. The path length of all node pairs could be represented in the form of a distance matrix. The average path length is the average of the shortest path lengths across all node pairs in the network.

Other simple quantity that can be observed in a network is the number of nodes of a given degree. The degree of a node is the number of neighbours of that node and is defined as

$$k_i = \sum_{j \in \zeta(i)} a_{ij} , \qquad (2)$$

the sum runs over the set $\zeta(i)$ of neighbours of $i$. For example: $\zeta(i) = \{j | a_{ij} = 1\}$.

In a directed network there are two important characteristics of a node, the number of links that end at a node and the number of links that start from the node. These quantities are known as the out-degree $k^o$ and the in-degree $k^d$ of a node, and are defined as

$$k^d = \sum_{j \in \zeta(i)} a_{ij}^d , \quad k^o = \sum_{j \in \zeta(i)} a_{ij}^o . \qquad (3)$$

It is possible to categorize networks by the degree distributions of their tails. In general, real-world networks are very different compared with random networks when referring to their degree distributions. Random networks commonly show Poisson distributions, while real networks might have long tails in the right part of the distribution with values that are far above the mean. Measuring the tail of the distribution of the degree data could be achieved by building a plot of the cumulative distribution function. In real-world networks, it is common to find distributions that follow power laws in their tails:

$$P(k) \sim \sum_{k'=k}^{\infty} k'^{-\gamma} \sim k^{-(\gamma-1)} , \tag{4}$$

where $\gamma$ is the scaling exponent of the distribution and the degree distribution $P(k)$ is the probability that the degree of a node is equal to $k$. This type of distribution is called scale-free and networks with such degree distributions are referred to as scale-free networks. Such distributions have no natural scale and the functional form of the distribution remains unchanged within a multiplicative factor under a rescaling of the random variables. Previous studies (Mandelbrot, 1983; Jeong et al., 2000) have shown that in large scale-free networks, independently of the system and the origin of the components, the probability $P(k)$ that a node of the network interacts with $k$, then other links decays as a power-law, suggesting that there is a tendency for large networks to self-organize into a scale-free state. A degree distribution with power laws is a characteristic commonly seen in complex networks such as in the World Wide Web network, protein-interaction networks, phone calls networks, food webs networks, citation networks, actors-movies networks and it also appears in systems of payments from different banks around the world (Inaoka et al., 2004; Söramaki et al., 2007; Boss et al., 2004).

The average degree of a node in a network is the number of links divided by the number of nodes and is defined as:

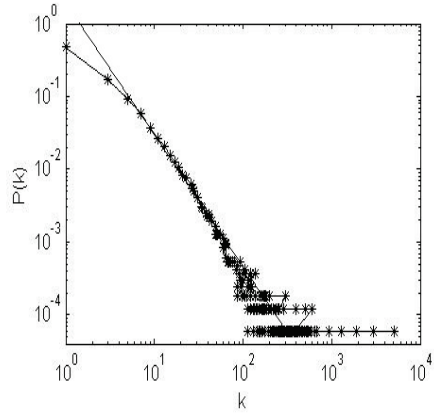
$$\langle k \rangle = \frac{1}{n} \sum k_o = \frac{1}{n} \sum k_i = \frac{m}{n} . \tag{5}$$

The average degree of the Estonian network of payments is 20. Most of the nodes have only 5 or less links, and 45% have only 1 link. Like other real networks, the degree distributions (undirected and directed) of the network of payments follow power laws.

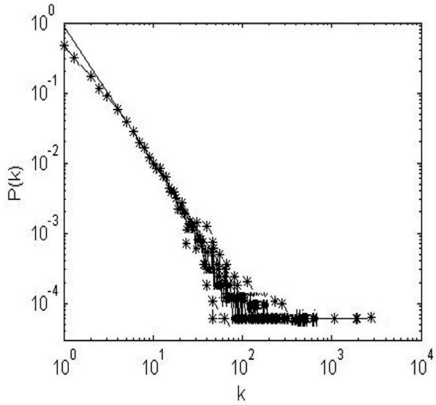
Complex networks can be classified as homogeneous or heterogeneous depending on their degree distributions. Homogeneous networks are identified by degree distributions that follow an exponential decay. In these networks, the distribution peaks at an average $k$ and then decays exponentially for large values of $k$, such as the distributions formed in the random graph model (Erdös and Rényi, 1959) and the small-world model (Watts and Strogatz, 1998) where each node has approximately the same number of links $k$. These have a normal distribution and the majority of the nodes have an average number of connections and only few or none of the nodes have either some or lots of connections. In heterogeneous large networks or scale-free networks, the degree distribution decays as a power law with a characteristic scale. The degree distribution follows a Pareto form distribution where many nodes have few links and few nodes have many links, therefore, highly connected nodes are statistically significant in scale-free networks.

Figure 2(a) shows the cumulative degree distribution of the Estonian network of payments (undirected). A straight line was added as an eye guideline. The distribution in Figure 2(a) follows a power law with the following scaling exponent:

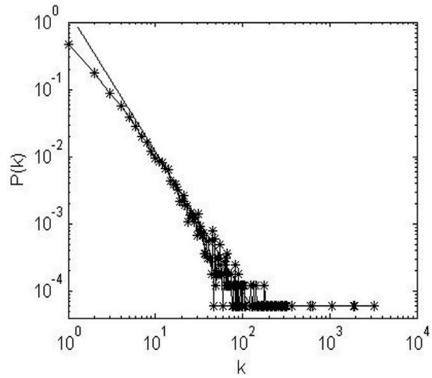
$$P(\geq k) \propto k^{-2.4}. \tag{6}$$




(a)




(b)




(c)

*Figure 2. Degree distributions (a) Empirical degree distribution for the connectivity network of the Estonian network of payments. X axis is the number of $k$ degrees and Y axis is $P(k)$. (b) out-degree distribution of the network, $P(k) \sim k^{-2.39}$. c) Empirical in-degree distribution $P(k) \sim k^{-2.49}$.*

Figure 2(b) shows the out-degree distribution and Figure 2(c) shows the in-degree distribution of this network. In all the distributions, I found regions that can be explained by power laws, and this implies that the network has a scale-free structure.

Weights $w_{ij}$ of the links $i$ and $j$ in a network show the importance of each link. The strength of the nodes is the sum of the weights of all the links. In this network the strength measures the overall transaction volume for any given node. The node-weighted strength is defined as

$$s_i = \sum_{j \in \zeta(i)} w_{ij}, \quad (7)$$

where $w_{ij}$ is the weight of the link between nodes $i$ and $j$ and the sum runs over the set $\zeta(i)$ of neighbours of $i$. The average strength can be calculated as a function of the $k$ number of links of a node to examine the bond between the strength and the degree.

I calculated the probability *P(s)* that a company has *k* outgoing or incoming links. As per Figure (3), the distribution of the out-degree volume (strength) follows a power-law decay

$$P(s) \sim s^{-2.32} \quad (8)$$

where the scaling exponent is 2.32. There are some deviations from the power law behaviour but they are sufficiently small. A similar distribution was found in the in-degree volume (strength) distribution (Rendón de la Torre et al., 2016). The power-law tail signals that the probability of finding companies paying out very large quantities of money is small. Moreover, while the companies have an absolute freedom in choosing how much money to pay or the counterparties to whom they interact with, the overall system obeys a scaling law, which is a particular property observed in critical phenomena and in highly interactive self-organized systems. Figure (3) displays the distribution of link weights weighed by the number of payments transacted.

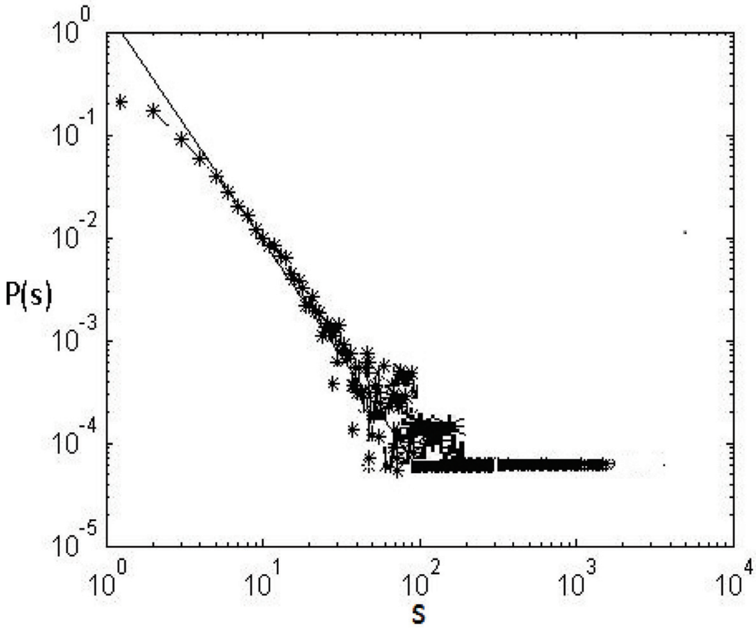



*Figure 3. Node out-degree distribution by strength.*

## 1.2 Scale degree distributions in other known networks

It is common to find scale-free degree distributions in different networks, such as in the World Wide Web, the proteins and interactions network, phone calls, food webs, transportation and in payments systems of different banks. Table 1 shows a list of the power-law exponents obtained from different types of real networks.

*Table 1. Scaling exponents and clustering coefficients for different types of reported networks*

| Type | Network | Exponent | Clustering coefficient * | References |
|---|---|---|---|---|
| Economical | Bank of Japan payments | $\gamma = 2.1$ | - | (Inaoka et al.,, 2004) |
| | US Federal Reserve Bank | $\gamma i = 2.11$<br>$\gamma o = 2.15$ | 0.53 | (Soramäki et al.,, 2007) |
| | Austrian Interbank Market payments | $\gamma i = 1.7$<br>$\gamma o = 3.1$ | 0.12 | (Boss et al.,, 2004) |
| Technological | WWW | $\gamma o = 2.4$<br>$\gamma i = 2.1$ | - | (Albert and Barabási, 1999) |
| | Peer-to-peer network | $\gamma = 2.1$ | 0.012 | (Ripeanu et al.,, 2002) |
| | Digital electronic circuits | $\gamma = 3$ | 0.03 | (Ferrer et al.,, 2003) |
| Social | Film actors | $\gamma = 2.3$ | 0.78 | (Watts and Strogatz, 1998) |
| | Email messages | $\gamma i = 1.5$<br>$\gamma o = 2.0$ | 0.16 | (Ebel et al.,, 2002) |
| | Telephone calls | $\gamma = 2.1$ | - | (Aiello et al.,, 2000) |
| Biological | Protein interactions (yeast) | $\gamma = 2.4$ | 0.022 | (Jeong et al.,, 2001) |
| | Metabolism reactions | $\gamma i = 2.2$<br>$\gamma o = 2.2$ | 0.32 | (Jeong et al.,, 2000) |
| | Energy landscape for a 14-atom cluster | $\gamma = 2.78$ | 0.073 | (Doye, 2002) |

$\gamma i$ = scaling exponent for in-degree distribution. $\gamma o$ = scaling exponent for the out-degree distribution. $\gamma$ = scaling exponent for the connectivity distribution. *Refers to average clustering coefficient.

## 1.3 Components of networks

Nodes can be partitioned into components, according to how the nodes connect with each other. A path is an ordered collection of nodes, each one connected to the next node. A component is a group of nodes such that any two nodes can be connected by a direct or indirect path. A component of an undirected network is a set of nodes such that for any pair of nodes $i$ and $j$ there is a path from $j$ to $i$, meaning that two nodes share the same component if there is a path connecting them.

In a directed network the largest component is known as the Giant Weakly Connected Component (GWCC) in which all nodes connect to each other via undirected paths. The core is the Giant Strongly Connected Component in which the nodes can reach each other through a directed path. The Giant Out-Component (GOUT) comprises the nodes that have a path from the GSCC and the Giant In-Component (GIN) comprises the nodes that have a path to the GSCC. The set of disconnected components (DC) are smaller components. Tendrils are nodes that have no directed path to or from the GSCC, but to the GOUT and or to the GIN (Dorogovtsev and Mendes, 2003). These concepts are shown in Figure 4.

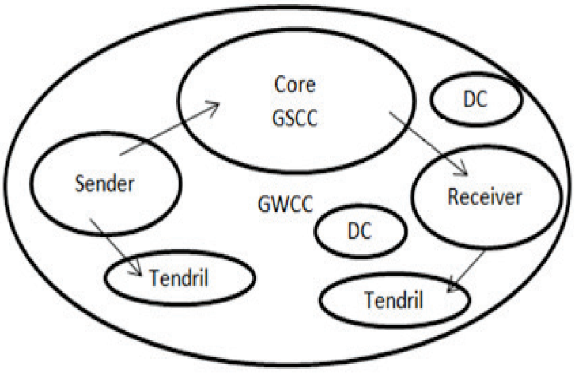



*Figure 4. Components of a directed network.*

All the components in this undirected network show that the GCC is composed of 15,434 nodes which means that 92.8% of the nodes are reachable from one another by following either forward or backward links. This suggests that it is a very well connected network. The remaining 7.2% nodes correspond to 508 DC. If we consider a direct approach, the GSCC contains 24% of the nodes in the system.

Another interesting and fundamental metric is the clustering coefficient of a node. It represents the probability that any two neighbours of a node are connected; it is the density around a node. In this study it indicates whether or not there is a link between two companies that have a common third business party.

$$C(i) = \frac{1}{k_i(k_i - 1)} \sum_{j \neq k} a_{ij} a_{jk} a_{ik}. \tag{9}$$

The average clustering coefficient is the mean of the clustering coefficients $\langle C \rangle$ of each and all the nodes. In this network, the average clustering coefficient is 0.18, and this number suggests there is cliquishness in the network. This means that two companies that are trading partners with a third one, have an average probability of 18% to be trading partners. For visualization purposes, Figure 5 displays the distribution of the clustering coefficient of the network. As seen in the plot, there is high number of unlinked neighbour nodes (45% of the nodes) that might be explained by the large number of nodes with degrees equal to 1 which frequently appear in scale-free networks.

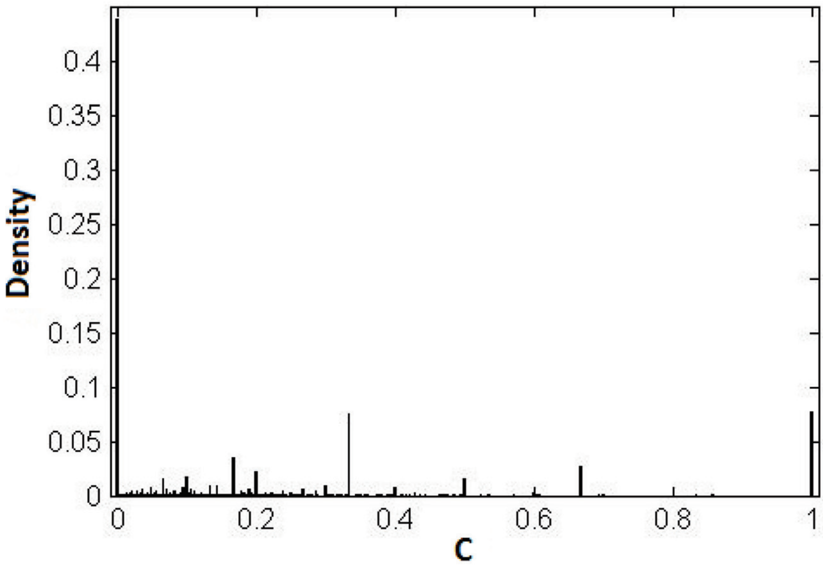



*Figure 5. Distribution of the clustering coefficient of the Estonian network of payments.*

In other words, the clustering coefficient indicates whether there is a link between two companies which have a common trading partner and it gives a way to measure the extent of the reach of intermediary trading. Compared with other real networks this average clustering coefficient is relatively small (see Table 1 for comparison). Compared to other real networks, such as the U.S. Federal Reserve Bank network of payments, the film-actor network, or the metabolism reactions network, the average clustering coefficient is low. A small coefficient is reasonable because it can be interpreted as how companies could see diversification as costly when it means to change or add their trading partners, suppliers or service providers. Business relationships between companies are commonly settled through medium or long term contracts. A company would like to remain doing business with the same parties because it's easier and cheaper than continuously changing them.

As mentioned earlier, the most basic model of networks is the random network model $G(n,p)$ developed by Erdös and Rényi (1959). This model has two parameters: $n$ and $p$ (where $n$ is the number of nodes of the graph and $p$ is the probability to link). The model works under the assumption that there could be a link $i-j$ between two nodes $i$ and $j$ and this assumption holds no matter if the nodes had a common neighbour node before the link was formed. The outcome of the model is the generation of random network graphs with a low clustering coefficient and a low variation in the degrees of the nodes. A random network cannot capture the decreasing nature of the clustering coefficient of the nodes with increase in the node degree because the clustering coefficient of the nodes in this type of network is totally independent of the node degree and is equal to the probability of a link between any two nodes (Barrat et al., 2008).

The betweenness centrality $\sigma(m)$ of a node $m$ is the total number of shortest paths which pass through a given node. It is a measure of the number of paths between other nodes that run through the node $i$; the more paths this node has, the more central is the node $i$ in the network. It indicates whether or not a node is important in the traffic of the network. It was originally introduced as a measure for quantifying the control of a determined human on the communication between other humans in a social network. The nodes with high betweenness control the network.

$$\sigma(m) \equiv \sum_{i \neq j} \frac{B(i,m,j)}{B(i,j)}\,, \tag{10}$$

where $B(i,j)$ is the total number of shortest paths between nodes *i* and *j* and the sum goes over all the pairs of nodes for which at least one path exists, with $B(i,j) > 0$.

The general characteristics and statistics of the Estonian network of payments are listed in Tables 2 and 3. According to Table 2, the average betweenness for the links is 40 and for the nodes is 110, meaning that each company handles on average 110 shortest paths, and the higher is the number of shortest path the more central the company is for the network.

Regarding other statistical measures of the Estonian network of payments, as per Table 2, the average shortest path length $\langle\, l \,\rangle$ is equal to 7.1 (calculated with Dijkstra's algorithm). This network is a "small world" with 7.1 degrees of separation, meaning that on average any company can be reached by another company in just a few links. 93% of the nodes are within 7 links of distance from each other and this suggests that the network of payments is comprised of a core of nodes with whom the other companies interact with. There is a smaller group of 1,081 nodes (6.5% of the total number of nodes in the network) connected by high value links. This group contains weighted links that comprise 75% of the overall value of the funds transferred. A *k*-core in an undirected graph is a connected maximal induced sub-graph which has a minimum degree greater than or equal to *k*. Alternatively, the *k*-core is the (unique) result of iteratively deleting nodes that have degree less than *k,* in any order.

This network showed low connectivity ($\mathrm{C} = 0.13$) but at the same time the network is densely connected (see Table 3). This characteristic is in line with the fact that there are companies that act as hubs and lead to short distances between the other companies. The clustering coefficient is higher than the connectivity, therefore, the network is not random (in a random network the clustering coefficient is equal to the connectivity; a random network is built by adding links randomly to a given set of nodes, thus is an unreal type of network). A random network of a comparable size has a clustering coefficient around 60 times lower than the one in the Estonian network of payments.

*Table 2 Network's characteristics*

| | |
|---|---|
| Total companies analysed (N) | 16,613 |
| Total number of payments analysed | 2,617,478 |
| Total value of transactions | 3,803,462,026 * |
| Average value of transaction per customer | 87,600 * |
| Maximum value of a transaction | 121,533 * |
| Minimum value of a transaction (aggregated) | 1,000 * |
| Average volume of transaction per company | 60 |
| Maximum volume of transaction per company | 24,859 |
| Minimum volume of transaction per company | 20 |

*All money quantities are expressed in monetary units and not in real currencies in order to protect the confidentiality of the data set. The purpose of showing monetary units is to provide a notion of the proportions of quantities and not to show exact amounts of money.

*Table 3 Summary of Statistics*

| Statistic | Value | Components | # of nodes |
|---|---|---|---|
| N | 16,613 | GCC | 15,434 |
| Nbr. of payments | 2,617,478 | DC | 1,179 |
| Undirected Links | 43,375 | GSCC | 3,987 |
| $< k >$ | 20 | GOUT | 6,054 |
| $\gamma o$ | 2.39 | GIN | 6,172 |
| $\gamma i$ | 2.49 | Tendrils | 400 |
| $\gamma$ | 2.45 | Cutpoints | 1,401 |
| $< C >$ | 0.183 | Bi-component | 4,404 |
| $< l >$ | 7.1 | $k$-core | 1,081 |
| $T$ | 0.13 | | |
| $D$ | 29 | | |
| $< \sigma >$ (nodes) | 110 | | |
| $< \sigma >$ (links) | 40 | | |

N = number of nodes. $< k >$ = average degree. $\gamma o$ = scaling exponent of the out-degree empirical distribution. $\gamma i$ = scaling exponent of the in-degree empirical distribution. $\gamma$ = scaling exponent of the connectivity degree distribution. <C> = average clustering coefficient. $< l >$ = average shortest path length. T = connectivity %. D = Diameter. $< \sigma >$ = average betweenness. GCC = Giant Connected Component. DC = Disconnected Component. GSCC = Giant Strongly Connected Component. GOUT = Giant Out-Component. GIN = Giant In-Component.

### 1.3.1 Importance of nodes: strength

Zemp et al., (2014) developed new versions of measures for directed and/or weighted networks which take into account the importance of nodes. In their work these authors showed that by using their measures one can avoid systematic biases created by a higher node density and larger weights of the links. Newman (2004) showed that weighted networks could be analysed by using simple mapping that goes from a weighted network to an unweighted multigraph and that this approach allows using standard techniques for studying unweighted and weighted networks.

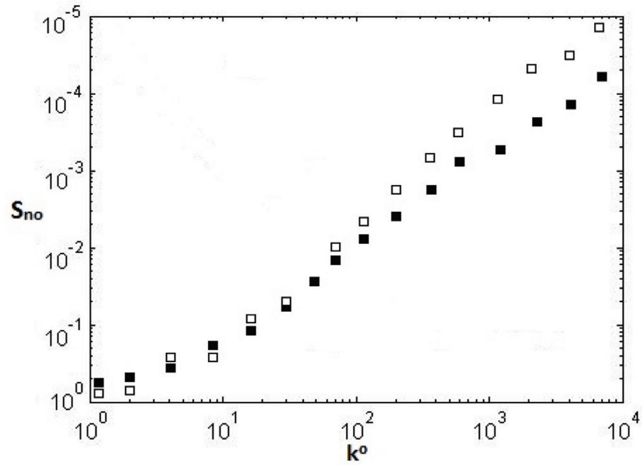
In this subsection of the chapter, I make a characterization of the links by investigating the strength of the interactions of the elements of the network: the link weight of payments and volume of payments. I analyse specific statistical measures of the weighted Estonian network of payments that combine the topology of the relations of the strength of links and nodes and their specific weights with the purpose of investigating beyond the topological architecture of the network and reveal aspects of its complex structure.

I analysed the bond between strength and degree of a node. Figures 6(a) and 6(b) depict the volume and value (in and out strengths) as functions of degree for both outgoing and incoming links (in-degree and out-degree). The strength $s$ is normalized by dividing it over the average link weight $\langle w_{ij} \rangle$. There is a power-law relationship between the strength and the degree, as follows:

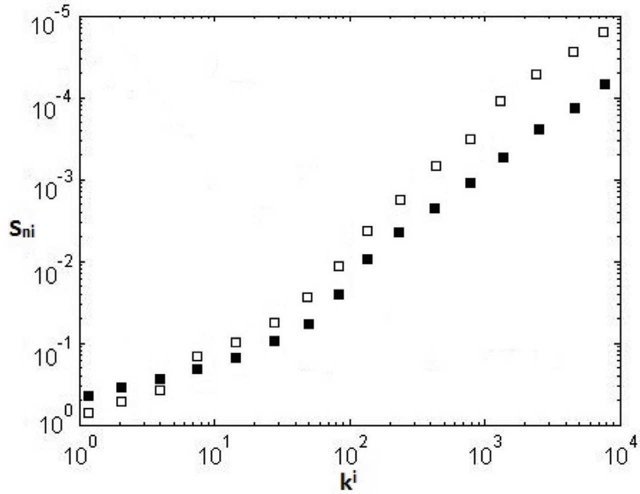
$$s(k) \sim k^{\alpha} \qquad (11)$$

where $\alpha$ is the coefficient of the scaling distribution. The power-law fit of Figure 6(a) has an exponent $\alpha_{vol} = 1.5$, when volume is used as the weight, and $\alpha_{val} = 2.4$ when the value is used instead. These values imply that the out-strength of nodes $S_{no}$ and in-strength of nodes $S_{ni}$ grow faster than the degree $k$ of a node, as seen in Figure 6(a), meaning that the most connected companies execute a higher number of payments with higher values of money than suggested only by their degree. This indicates that if a company has twice as many payments (out links) as another company, it could be expected that such company sends three times the number of payments, and almost five times the total value of payments. Figure 6(b) indicates that the relationships between the in-degree and the in-strength show similar trends like the out-degree and out-strength cases seen in Figure 6(a).

Also, the strength of a node scales with the degree $k$ indicating that highly connected companies have payments of high weights. The strength of a company grows generally faster than its degree. In other words, highly connected companies not only have many payments, but their payments also have a higher than average weight. This observation agrees with the fact that big companies are better equipped for handling large quantities of payments with large amounts of money. Comparable results were found in the cargo ship movements network (Kaluza et al., 2010) and in the airport network (Barrat et al., 2004), and such results may hint or point to the existence of a generic pattern in large-scale networks.




(a)




(b)

*Figure 6. Distributions of strength. (a) Node out-strength as a function of degree. (b) Node in-strength as a function of degree. Empty squares represent values of payments and black squares represent the number of payments.*

For a given node $i$ with connectivity $k_i$ and strength $s_i$, the weights of its links might be of the same order of magnitude $s_i/k_i$, or they can be distributed heterogeneously with some links predominating over others. Then, the participation ratio is defined as follows:

$$H_2^w(i) = \sum_{j \in \zeta(i)} \left[\frac{w_{i,j}}{s_i^w}\right]^2, \quad (12)$$

or equivalently

$$H_2^c(i) = \sum_{j \in \zeta(i)} \left[\frac{c_{i,j}}{s_i^c}\right]^2. \quad (13)$$

Now I define the participation rates to separate outgoing and incoming links. Then, the average participation ratio is calculated as

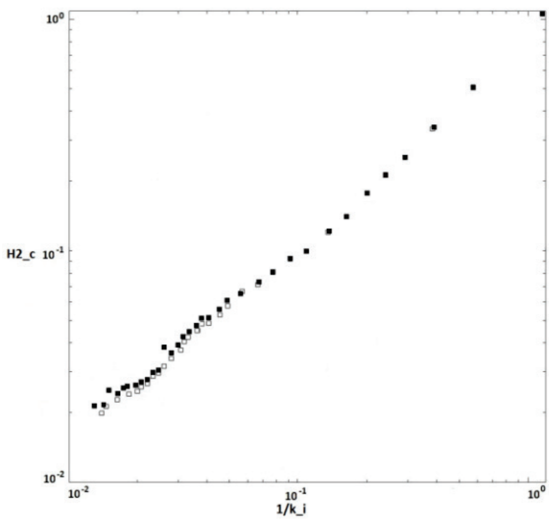
$$H_2^c = \frac{1}{N}\sum_i H_2^c(i), \quad (14)$$

and

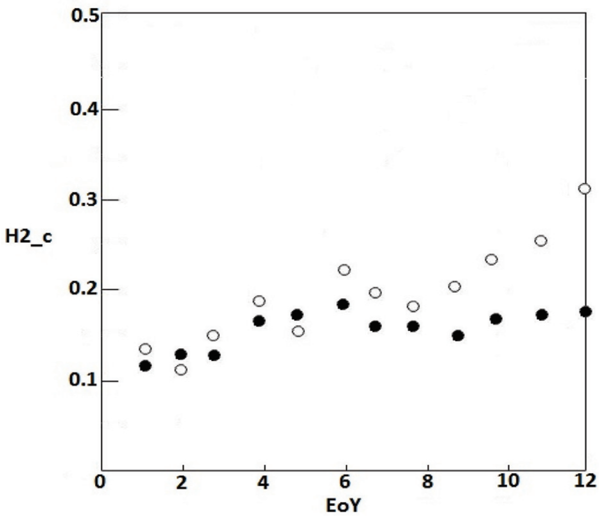
$$H_2^w = \frac{1}{N}\sum_i H_2^w(i). \quad (15)$$

I calculate the participation ratio as a function of a company's inverse degree, where the objective is to identify the links that are used more often than the others. If a low number of weights are dominant then $H_2$ is close to 1 but if all the weights are of the same order of magnitude then $H_2 \sim 1/k_i$. When $H_2^w$ is close to 1, it indicates the existence of preferential interactions between the nodes, meaning that companies prefer to transact with certain companies.

Figure 7(b) shows a plot of the participation ratio $\mathrm{H}_2^c$ as a function of the inverse degree of the nodes. The plot shows the links that are used more often than others. For example, for a degree up to 10 $\mathrm{H}_2^c(\mathrm{i}) \sim 1/\mathrm{k_i}$ and for higher degrees the participation ratio is higher than the inverse degree suggesting there is a disposition in the direction of preferential trading with specific counterparties. Figure 7(b) shows the average participation ratio during the whole year for out-going payments and in-coming payments. By the end of the year the participation ratio for all the payments decreases. Particularly, the participation ratio of the outgoing payments decreases dramatically. This reveals that the preferential linking is limited. By the end of the year, the preference for trading with only certain counterparties became less important. This could be caused by an increased payments/liquidity tendency that could potentially be driven by generalized unspent company annual budgets or delayed payments that were done before the year ended.




(a)




(b)

*Figure 7. Participation ratio plots. (a) Participation ratio as a function of the inverse degree for out-going payments and in-coming payments. Black squares represent out-going payments; white squares represent in-coming payments. (b) Participation ratio as a function of the distance of the end of the year expressed in months for incoming (black circles) and outgoing (white circles) payments.*

## 1.4 Robustness of the network

Previous studies of the structure of the World Wide Web network components (Albert et al., 1999) have focused on analysing the robustness of the GCC against attacks, and it has been found that it is very difficult to destroy the World Wide Web network by using random elimination of links. (Table 3 displays the component sizes of the network of payments, among other statistics).

In complex networks some nodes are essential while others are not, and identifying these essential nodes is a critical task in determined situations. The most essential nodes are those which if removed from the network would cause the whole system to collapse. In order to have a deeper understanding on how the network is likely to behave as a whole in the presence of perturbations, I will address the next question: if a portion of nodes were removed, would the structure of the network become divided into disconnected clusters? How will the network respond to an actual removal of nodes? There are many approaches on how to tackle this problem and locate the “key nodes” in the network, or on how to calculate the optimal percolation threshold of nodes that would break the network into disconnected clusters. Some logical approaches are: high degree node, $k$-core, closeness or eigenvector centralities, however, a common characteristic in these approaches is that they do not necessarily optimize a measure that reflects the collective influence arising from considering the entire influential nodes at once. Under a collective approach, nodes' inherent strength and weakness arise collectively from the configuration of interactions they have with the other components.

Currently, there are many heuristic methods for calculating the optimal percolation threshold of nodes at which the network breaks into disconnected clusters, such as the high degree node, $k$-core, closeness and eigenvector centralities. However, a common characteristic in these approaches is that they do not necessarily optimize a measure that reflects the collective influence arising from considering the entire influential nodes at once. Under a collective influence approach, the inherent strength and weaknesses of the nodes arise collectively from the configuration of interactions that they have with the other components.

Morone and Makse (2015) designed an approach that has proven to perform better than other heuristic methods (such as the high degree node, $k$ −core, closeness and eigenvector centralities). Morone and Makse's algorithm optimizes a measure that can reflect the collective influence effect that arises when taking into account the entire influential set of nodes at once. This algorithm predicts a smaller set of optimal influencer nodes (the group of nodes that destroy the network if they are removed).

The collective influence of a node $\mathrm{CI}$ is defined as the product of the node's reduced degree (the number of its nearest connections $k_i - 1$), and the total reduced degree of all nodes $k_j$ at a distance $\ell$ from it, and is represented as follows:

$$CI_{\ell}(i) = (k_i - 1) \sum_{j \in \partial Ball(i,\ell)} (k_j - 1), \qquad (16)$$

where $\ell$ is defined as the shortest path. $\mathrm{Ball}(\, i, \ell)$ is the set of nodes inside a ball of radius $\ell$ around node i. $\partial \mathrm{Ball}(\mathrm{i}, \ell)$ is the frontier of the ball and comprises the nodes $j$ that are at a distance $\ell$ from $i$. By computing $\mathrm{CI}$ for each node it is possible to locate the nodes with the highest collective influence. The collective influence algorithm addresses the problem of optimal influence in the computation of the minimum structural total number of nodes that reduces the largest eigenvalue of the non-backtracking matrix of the network.

I performed a simulation using the CI, where I calculate the collective influence of a group of nodes as the fall in the size of the Giant Connected Component (GCC) which would occur if the nodes of the GCC were eliminated. The GCC contains 15,434 nodes and this quantity represents 92.8% of the nodes of the whole network.

These results are displayed in Figure 8. The plot shows the GCC when a fraction of the nodes has been removed. The optimal percolation threshold occurs when 6.0% of the nodes are removed and that is the point where $\mathrm{GCC}(Pc) = 0$. This means that there are many companies that execute a large number of payments which in fact have a weak influence in the economic network as a whole. The most influential companies in the network are not necessarily the most connected ones, neither are those which have more intense economic activity. A weak but smart node attack in the Estonian network of payments where only 6.0% of the nodes are removed destroys the whole network of payments, meaning that a few nodes maintain the unity of the whole network.

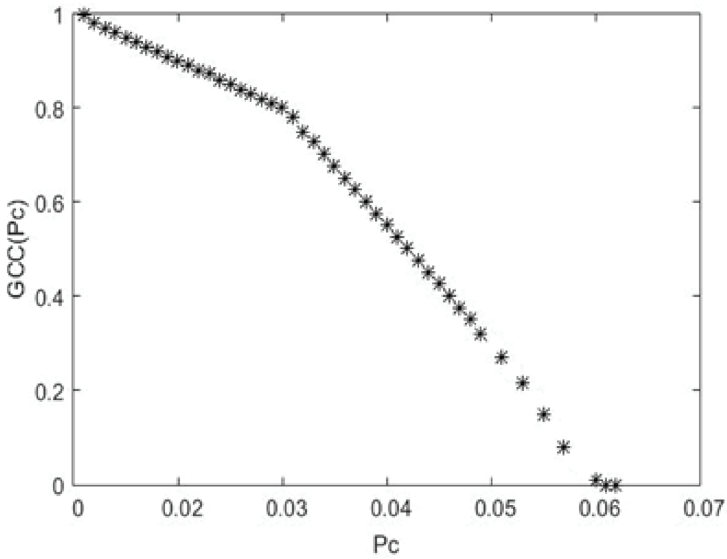
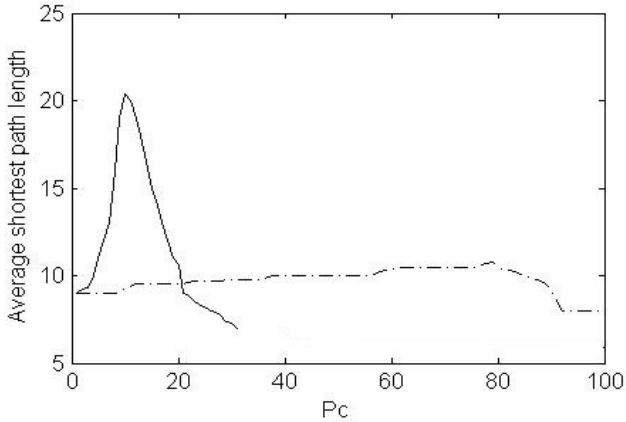


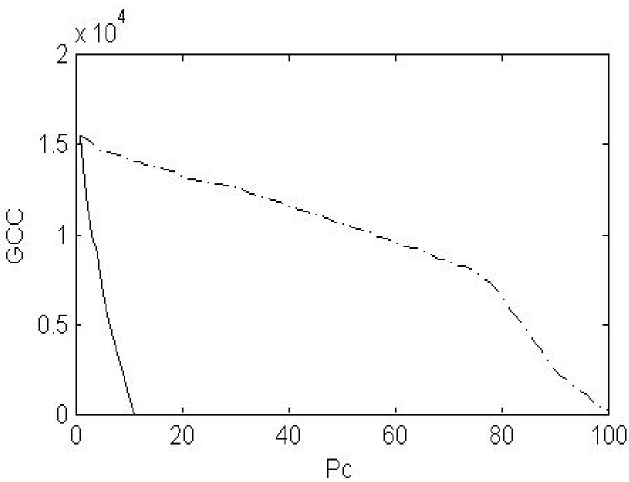

*Figure 8. GCC of the network of payments as a function of the percolation threshold Pc.*

I ran a simulation that shows a random removal of a fraction of the nodes and another simulation considering strategically chosen nodes. I found 1,401 cut-point nodes (Hanneman and Riddle, 2005). In this simulation I have established that 8% of the nodes are necessary to maintain the structure of the network connected. If these nodes are removed from the network, the quantity of the components and the average path lengths between the nodes would increase, leaving the network vulnerable.

I measure the average shortest path length $< l >$ and the relative size of the GCC as functions of the percentage $d$ of deleted nodes (Dorogovtsev and Mendes, 2003; Albert et al., 1999; Cohen et al., 2000). The results are displayed in Figures 9(a) and 9(b). The effect of the targeted removal of nodes causes a quick growth in the average shortest path length until the GCC disappears, $\text{GCC}(Pc) = 0$ at a very low level of targeted damage (less than 10%). I will call this level the percolation threshold $Pc$. It is noticeable that a weak but smart attack destroys the network. In the random removal of nodes simulation the damage is less than in the targeted damage. In the previous chapters I have established that my network of payments has shown scale-free properties. Scale-free networks are resilient to random damage, so it is almost impossible to destroy such network of payments by a random removal of nodes, but if  the exact portion of particularly selected nodes, the network breaks completely. This effect has been seen in financial systems in economic crisis before: companies or banks may declare themselves in bankruptcy and the whole system stays healthy, but if certain organizations declare themselves in bankruptcy then the whole system collapses.




(a)




(b)

*Figure 9. Plots of the effect of the targeted and random removal of nodes from the network of payments. (a) The average shortest-path length $< l >$ in the GCC plotted against the percentage of removed nodes. (b) The GCC plotted against the percentage of removed nodes. Continuous lines display the effect of the targeted removal of nodes and the dashed lines display the effect of the random removal of nodes. Pc are the percolation thresholds, for each case.*

It is not rare that the GCC in heavy-tailed networks is resilient against random removal of nodes. If the degree distribution of the network is fat-tailed, then this fact determines the topology of the system. However, it might be possible that when removing nodes in a random way, the tail of the degree distribution changes and then the GCC structure would be damaged.

Scale-free networks are commonly observed in a wide array of different contexts such as nature and society. Scale-free networks are resilient to random removal of nodes, but are vulnerable to smart attacks. The Estonian network of payments is a scale-free network (with power laws in the degree distribution) and its own scale-free nature makes it almost impossible to destroy the network by a random removal of nodes, but if the exact portion of particularly selected nodes are removed then the network collapses completely. This "collapsing" effect has been already observed in financial systems when severe economic crisis occur and specific companies or banks crash leading the whole system to break down. An example of this, is the global financial crisis of 2008 that started with the collapse of the famous investment bank Lehman Brothers, followed by Bear Sterns, UBS and other financial entities that dragged the whole global financial system into severe liquidity problems.

# 2. Detecting overlapping community structure

In this Chapter I comment on the overlapping community structure of the Estonian network of payments and the scale-free properties at a mesoscale level. This Chapter also contains the analysis of the global structure of the network through the distribution functions of four basic quantities. This Chapter mainly follows Paper III and IV.

Networks play an important role in a wide range of economic and social phenomena and the use of techniques and methods from graph theory has permitted economic network theory to expand the knowledge and understanding of economic phenomena in which the embeddedness of individuals or agents in their social or economic interrelations cannot be ignored.

Studying the community structure exhibited by real networks is a fundamental step towards a comprehensive understanding of complex systems beyond the local organization of their components. Community detection analysis is essential for understanding the structure and functionality of large networks and it also helps to expand the knowledge on complex networks.

Community detection is a graph partitioning process that provides valuable insight into the organizational principles of networks and is essential for exploring, and among other things, for predicting connections that are not yet observed. Thus far, recent advances of the underlying mechanisms that rule dynamics of communities in networks are limited, and this is why the achievement of an extensive and wider understanding of communities is important. Locating the underlying community structure in a network makes it easier to study the network, and could provide insights into the function of the system represented by the network, as communities often correspond to functional units of systems. The study of communities and their properties also helps in revealing relevant groups of nodes, creating meaningful classifications, discovering similarities or revealing unknown linkages between nodes.

The usefulness of identifying the communities within networks lies in how this information could be used in a practical scenario. In the context of the bank industry the output of the community analysis (based on payments between companies who are customers of a bank) could be used for targeted marketing activities. For example, it could be used at the moment of integrating criteria for creating target groups of customers to whom certain products or lines of products would be offered. Customers in the same community would be included in the same target group and later on after one offer is made to them it would be possible and interesting to assess the contagion effects of the product acquisition among customers of the same communities who received the same offer.

Another useful application is for helping to create customer-level segmentations or marketing profiles. To know the community (or communities) a customer belongs to, could be one of the drivers for creating customer profiles or clustering levels. An alternative usage of the output of community analysis is in predictive analytics, for example when building churn models. Churn models usually define a measure of the potential risk of a customer cancelling a product or service and provide awareness and metrics to execute retention efforts against churning. The communities to which the companies/customers belong could be used as variables or features when using logistic regression, random forests or neural network models. Additionally, community detection analysis could be used as input for product affinity and recommender systems. Affinity analysis is a data mining technique that helps to group customers based on historical data of purchased products and is used for cross-selling product recommendations. Another useful and immediate application is in product acquisition propensity models. These models calculate customers' likelihood to acquire a product after an offer is made based on a myriad of variables and with this evidence the sales process can become more efficient.

The majority of previous studies on communities have essentially been devoted to the description of structures inside the communities and their applications: communities representing real social groupings (Traud et al., 2000; González et al., 2007; Palla et al., 2007) communities in a co-authorship network representing related publications of particular topics (Pollner et al., 2006), protein-protein interaction networks (Lewis et al., 2010), communities in a metabolic network representing cycles and functional units in biology (Guimera and Amaral 2005; Ravasz et al., 2002) and communities in the World Wide Web representing web pages with related contents (Dourisboure et al., 2007).

Regarding community studies on economic networks and their applications, Vitali and Battiston (2014) studied the community structure of a global corporate network and found that geography is the major driver of organization within that network. Fenn et al., (2009) studied the evolution of communities of a foreign exchange market network in which each node represents an exchange rate and each link represents a time-dependent correlation between the rates. By using community detection, they were able to uncover major trading changes that occurred in the market during the credit crunch of 2008. Other related economic studies have focused on the overlapping feature of communities, such as in (Piccardi et al., 2010; Bóta and Kresz, 2013).

Most of the algorithms for community detection can be classified as divisive, agglomerative or optimization-based methods, and each method has specific strengths and weaknesses. Previous studies on communities based on divisive and agglomerative methods consider that structures of communities can be expressed in terms of separated groups of clusters (Newman, 2004; Yang et al., 2016; Hopcroft et al., 2004; Scott, 2000) but most of the real networks are characterized by well-defined statistics of overlapping communities. An important limitation of the popular node partitioning methods is that a node must be in one single community whereas it is often more appropriate to attribute a node to several different communities, particularly in real-world networks.

An example where community overlapping is commonly observed is in social networks where individuals typically belong to many communities such as: work teams, religious groups, friendship groups, hobby clubs, families or other similar social communities. Moreover, members of social communities have their own sub-communities resulting in a very complex web of communities (Derényi et al., 2005). The phenomenon of community overlapping has been already noticed by sociologists but has barely been studied systematically for large-scale networks (Gavin et al., 2002; Devi and Poovammal, 2016; Xie et al., 2013; Ding et al., 2016).

Networks have sections in which the nodes are more densely connected to each other than to the rest of the nodes in the network, and such sub-sections are called communities. Communities might exist in different networked systems, such as economics, sociology, biology, engineering, politics and computer science. There is no unique definition of community in the existing literature. Definitions change depending on the author and the type of study, and precisely one of the core issues in community detection is the lack of a unified definition of what is a community.

I use the Clique Percolation Method (CPM) definition because such algorithm allows overlapping nodes among communities, a condition that arises when a node is a member of more than one community. In economic systems, the nodes could frequently belong to multiple communities; therefore, forcing each node to belong to a single community could result in a misleading characterization of the underlying community structure.

An overlapping community graph is a network that has links between communities. Moreover, it is a representation of a network that denotes links between communities, where the nodes represent the communities and the links are represented by the shared nodes between communities. In my study the nodes represent communities and the links represent shared nodes between communities. CPM is based on the density of links and the definition of community for this algorithm is local and it is not too restrictive. Overlapping communities arise when a node is a member of more than one community. CPM is based on the assumption that a community comprises overlapping sets of fully connected sub-graphs and detects communities by searching for adjacent cliques. A clique is a complete (fully connected) subgraph. A $k$-clique is a complete sub-graph of size $k$ (the number of nodes in the sub-graph). Two nodes are connected if the $k$ -cliques that represent them share $k-1$ members.

The method begins by identifying all cliques of size $k$ in a network. When all the cliques are identified, then a $N_c\ x\ N_c$ clique-clique overlapping symmetric matrix $\boldsymbol{O}$ can be built, where $N_c$ is the number of cliques and $\boldsymbol{O}_{ij}$ is the number of nodes shared by cliques $i$ and $j$ (Everett and Borgatti, 1998). This overlapping matrix $\boldsymbol{O}$ encodes all the important information needed to extract the $k$-clique communities for any value of $k$. In the overlapping matrix $\boldsymbol{O}$ rows and columns represent cliques and the elements are the number of shared nodes between the corresponding two cliques. Diagonal elements represent the size of the clique and when two cliques intersect they form a community. For certain $k$ values, the k-clique communities form such connected clique components in which the nearby cliques are linked to each other by at least $k-1$ adjacent nodes. In order to find these components in the overlapping matrix $\boldsymbol{O}$, one should keep the entries of the overlapping matrix which are larger than or equal to $k-1$, set the others to zero and finally locate the connected components of the overlapping matrix $\boldsymbol{O}$. The formed communities are the identified separated components (more details on the Clique Percolation Method can be found in Palla et al., 2005).

## 2.1 Structures of communities

An overlapping community graph is a representation of a network that denotes links between communities, where the nodes represent the communities and the links are represented by the shared nodes between communities. For visualization purposes and in order to draw a readable map of the network, Figure 10 shows a graphic view of a representative section of the overlapping network of communities where big and small communities can easily be distinguished. This image depicts 25 overlapping communities and each circle represents a node which in turn represents an overlapping community. The links represent the shared nodes between the communities. The size of the nodes characterizes the size of each community. For example, the big node in the middle represents a community with 61 companies.

*Figure 10. Visual representation of a section of the overlapping network of communities (Estonian network of payments). The circles (nodes) represent communities and the black lines between them represent shared nodes between communities.*

A parameter $k$ needs to be chosen and the optimal choice is a problem in any cluster analysis. The parameter $k$ affects the constituents of the overlapping regions between communities. The larger the parameter $k$ is, the less the number of nodes which can arise in the overlapping regions. When $k \rightarrow \infty$, the maximal clique network is identical to the original network and no overlap is identified. The choice of $k$ will depend on the network. It is observed from many real-world networks, that the typical value of $k$ is often between 3 and 6 (Shen, 2013). Figure 11 shows a plot of the number of communities and the average size of the communities at different $k$ values. When $k$ increases the number of communities decreases while the size of the communities increases rapidly. When $k$ decreases the number of communities increases rapidly while the size of the communities remains low.

I tested different values of $k$ ranging from 3 to 10 and a posteriori chose $k$=5 because when $k$<5 a high number of communities arises and the partitions become very low and giant communities appear (with sizes of more than 3200); at the level $k$=5 a rich partition with the most widely distributed cluster sizes set for which no giant communities appear is obtained.

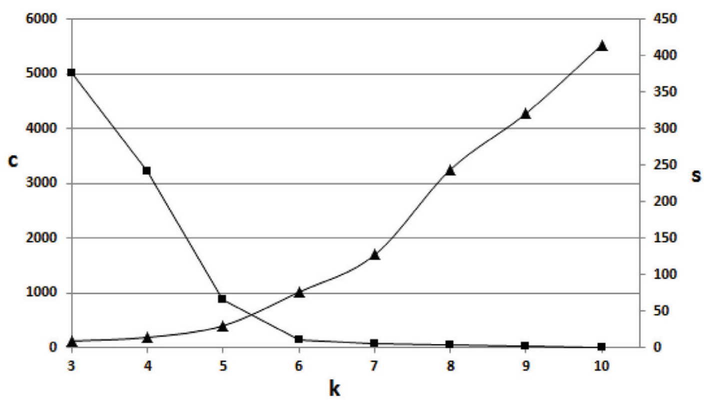



*Figure. 11 Plot of the average size of community (s) and number of communities (c) as k increases. Squares represent the number of communities and triangles represent the size of the communities.*

In order to study and characterize the global community structure of my network, I investigated the distribution functions of the following four elementary quantities: community size $P(s)$, overlap size $P(s_o)$, community degree $P(d)$ and membership number $P(m)$. In general, the nodes in a network can be characterized by a membership number which is the number of communities a node belongs to. This means that for example, any two communities may share some of their nodes which correspond to the overlap size between those communities. There is also a network of communities where the overlaps are represented by the links and the communities are represented by the nodes, and the number of such links is called: community degree. The size of any of those communities is defined by the number of nodes it has.

## 2.2 Distribution functions

The community size distribution is an important statistic that describes partially the system of communities. Figure 12 displays the cumulative distribution function of the community size $P(s)$ and it shows the probability of a community to have a size higher or equal to $s$ calculated over different points in time, where $t$ is the time in months. The overall distribution of community size resembles a power-law $P(s) \propto s^{\alpha}$, where $\alpha$ is the scaling exponent, and a power-law is valid nearly over all times $t$. The scaling exponent (calculated by maximum likelihood estimators) when *t*=3 is -2.8 (included for eye guideline) and Equation 17 is:

$$P(s) \propto s^{-2.8} \qquad (17)$$

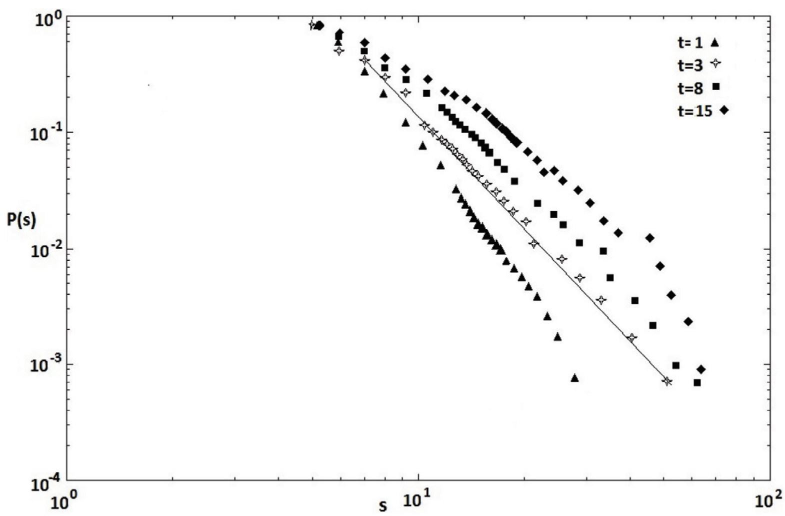



*Figure 12. Cumulative community size distribution at different times t.*

The sizes of the communities at *t=1* are smaller than in the rest of the months; as time increases the size increases, particularly the size of the largest communities. The shapes of the power-laws observed in the community size distributions of Figure 12 suggest there is no characteristic community size in the network. The distribution at different moments in time follows similar decaying patterns, but in general, the scaling tail is higher as *t* increases. A fat tail distribution implies that there are numerous small communities coexisting with some large communities (Newman, 2004; Clauset et al., 2004). Figure 13 shows statistics of the community sizes across time and according to the plot, both the standard deviation of community sizes and the average size of communities increased with time.

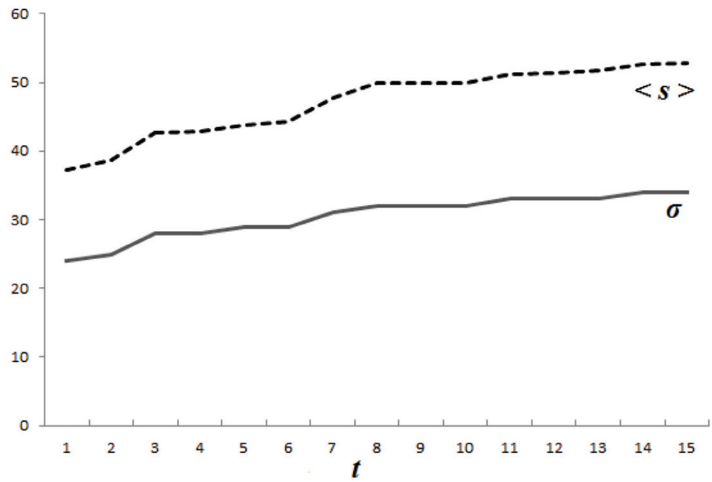



*Figure.13 Statistics of community size. <s> is the average community size. σ is the standard deviation of the size of communities at different times t.*

In a network of overlapping communities, the overlaps are represented by the links and the number of those links is represented by the community degree $d$. Then, the degree $d$ is the number of communities another community overlaps with. Figure 14 shows the cumulative distribution of the community degrees in the network. There are some outstanding community degrees in the end of the tail and these include communities that cluster the majority of the biggest customers from the network. The central part of the distribution decays faster than the rest of it. There is an observable curvature in the log-log plot, however no approximation method fitted the distribution. Figure 14 shows that the maximum number of degrees $d$ is 63 and corresponds to a relatively small quantity of nodes.

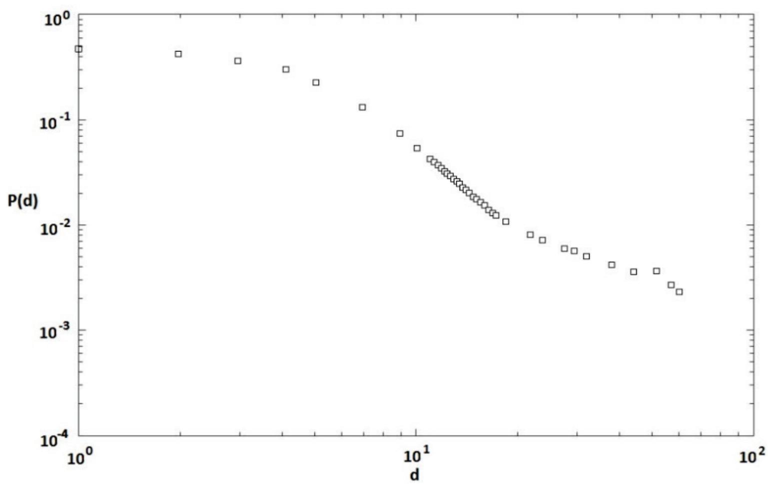



*Figure 14. Cumulative distribution of community degrees d.*

A node $i$ of a network can be characterized by a membership number $m_i$, which is the number of communities to where the node $i$ belongs to. Figure 15 shows the cumulative distribution of the membership number $m_i$. The distribution follows a power-law where no characteristic scale exists. The largest membership number found in the network was 10, meaning that a company can belong to maximum of 10 different communities simultaneously. Figure 15 shows that the fraction of nodes that belong to many different communities is quite small, while the fraction of nodes belonging to at least one community is high. For example, when $m = 1$ the percentage of nodes that belong to at least one community is 50%, while the percentage of nodes that belong simultaneously to 10 communities ($m$=10) is extremely small. The rest of the communities belong to two or more communities. The companies that overlap with 10 communities belong to the energy and water services industries. The majority of the nodes that have $m \neq 1$ have a degree that is less than $k - 1$, meaning they are weakly connected.

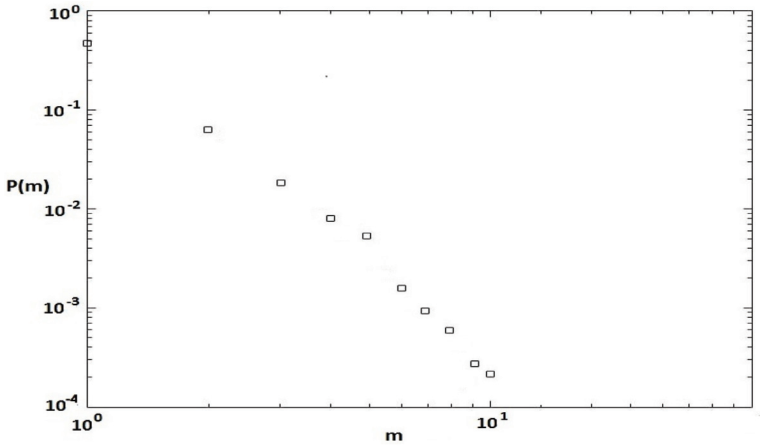



*Figure 15. Cumulative distribution function of the membership number $m_i$.*

The range in which the communities overlap with each other is also an important property of the Estonian network of payments. The overlap size is defined as the number of nodes that two communities share. $P(s_o)$ is the proportion of overlaps larger than $s_o$.

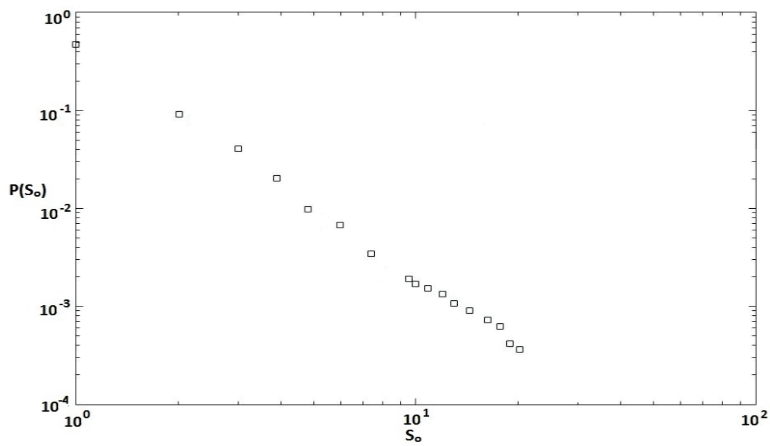



*Figure 16. Cumulative distribution function of the overlap size $s_o$.*

Figure 16 shows the cumulative distribution function of the overlap size. In general, although the extent of overlap sizes is limited, the data is close to a power-law dependence, meaning that there is no characteristic overlap size. The largest overlap size is 22, however at $s_o \geq 9$ the number of overlapping nodes becomes small.

In Chapter 1 I discussed the scale-free properties of the degree distributions of the Estonian network of payments. It is interesting to observe that the scale-free property is also preserved at a higher level of organization where overlapping communities are present. In this Chapter I have analysed the community structure of my network by using the Clique Percolation Method. I found that there are scale-free properties in the statistical distributions of the community structures. Size, overlap and membership distributions follow shapes that are compatible with power-laws. Power-law distributions have already appeared in this network at a global scale at the level of nodes (Rendón de la Torre et al., 2014), and in this community structure study I have shown that power-laws are present at the level of overlapping communities as well. This study adds to the existing literature on complex networks by presenting the first overlapping community analysis of a country's network of payments.

An immediate application and utility of the community detection results of this study is that they could be used in targeted marketing activities. The output is a list of nodes and the community classification where the nodes belong to. This could be used as input for predictive analytical models such as product acquisition propensities, churn propensities, product affinity analyses, for creating marketing profiles or customer segmentations and for creating customer target lists for product offering (in an effort to propagate consumer buzz effects). Further applications for community detection in similar economic networks could involve strengthening relationships between companies of the same community for improving performance of the whole network, or for identification of patterns between companies and tracking suspicious business activities.

A question that remains open for future research is to investigate if the similarities in communities' features amongst different complex networks arise randomly or if there are any unknown properties shared by all of them. Another line of research that remains open for the future is to study the plausibility of predicting changes in a payment network through communities' detection analysis.

## 3. Multifractal networks

In the late 60's Benoit Mandelbrot was the first to coin the term "fractal" and he also was the first one in describing the fractal geometry of nature (Mandelbrot, 1983). Since then the fractal approach has been widely spread and used in extensive research studies related with the underlying scaling of different complex structures, including networks.

Whether a single fractal scaling spans, or not, all the constituents or areas of a system, is a fundamental issue that helps in distinguishing when a system is multifractal or just fractal. One scaling exponent is enough to characterize completely a monofractal process. Monofractals are considered homogeneous objects because they have the same scaling properties branded by one singularity exponent. Instead, a multifractal object requires several exponents to characterize its scaling properties. Multifractals are inherently more complex and inhomogeneous than monofractals and portray systems with high variations or fluctuations that originate from specific characteristics.

Fractal and multifractal analysis helps to reveal the structure of all kinds of systems in order to have a better understanding of them. In particular, both the fractal and the multifractal approaches have many different interesting applications in economy. An interesting line of research is related with the relevance and applicability of fractal and multifractal analysis in social and economic topics. Regarding social studies, Lu et al., 2004 showed the importance of road patterns for urban transportation capacity based on fractal analysis of such network. In this study, the authors were able to link the fractal measurement with city mass measurements. A few recent studies have focused on the analysis of the changes of multifractal spectra across time to assess changes in economy during crisis periods (Fotios and Siokis, 2014). Some other studies have focused on gathering empirical evidence of the common multifractal signature in economic, biological and physical systems (Pont et al., 2009).

Fractal analysis helps to distinguish global features of complex networks, such as the fractal dimension. However, the fractal formalism is insufficient to characterize the complexity of many real networks which cannot be described by a single fractal dimension. Furuya and Yakubo (2011) demonstrated analytically and numerically that fractal scale-free networks may have multifractal structures in which the fractal dimension is not sufficient to describe the multiple fractal patterns of such networks, therefore, multifractal analysis rises as a natural step after fractal analysis.

Multifractal structures are abundant in social systems and in a variety of physical phenomena. Inhomogeneous systems which do not follow a self-similar scaling law with a sole exponent could be multifractal if they are characterized by many interweaved fractal sets with a spectrum of various fractal dimensions. Multifractal analysis is a systematic approach and a generalization of fractal analysis that is useful when describing spatial heterogeneity of fractal patterns (Song et al., 2015). Multifractal network analysis requires taking into account a physical measure, like the number of nodes within a box of specific size in order to analyse how the distribution of such number of nodes scales in a network as the size of the box grows or reduces. In the last years, numerous algorithms for calculating the fractal dimension and studying self-similar properties of complex networks have been developed and tested extensively (Palla et al., 2005; Zhou et al., 2007; Gallos et al., 2007; Schneider et al., 2012; Eguiluz et al., 2003). Song et al., (2007) developed a method for calculating the fractal dimension of a complex network by using a box-covering algorithm and identified self-similarity as a property of complex networks (Song et al., 2005). Additionally, several algorithms and studies on multifractal analysis of networks have been proposed and developed recently (Li et al., 2014; Liu et al., 2015; Wei et al., 2013; Wang et al., 2012).

In this sub-section of the Chapter I analyse fractal and multifractal properties of the large scale economic network of payments of Estonia. I perform a fractal scaling analysis by estimating the fractal dimension of the network of Estonian payments and its skeleton. Then, I study the multifractal behaviour of the network by using a sandbox algorithm for complex networks to calculate the spectrum of the generalized fractal dimensions $D(q)$ *an*d mass exponents $\tau(q)$.

## 3.1 Fractal network analysis

According to Song et al., (2005), the box-counting algorithm is an appropriate method for studying global properties of complex networks. The fundamental relation of fractal scaling is based on the box-covering method which counts the total number of boxes that are needed to cover a network with boxes of a certain size. The box-covering method is equivalent to the box-counting method widely used in fractal geometry and is a basic tool for measuring the fractal dimension of fractal objects embedded in Euclidean space (Feder, 1998). However, an Euclidean metric is not well defined for networks, thus I use the networks' adaptation (Wang et al.,, 2012) of the random sequential box-covering algorithm (Kim et al., 2007) in order to calculate the fractal dimension of the network and its skeleton. This method involves a random process for selecting the position of the centre of each box. $N_B(r_B)$ is the minimum number of boxes needed to tile the whole network, where the lateral size of the boxes is the measure of radius $r_B$ as follows:

$$N_B(r_B) \sim r_B^{-d_B}, \quad (18)$$

where $d_B$ is the fractal dimension. If I measure the number of $N_B$ for different box sizes, then it is possible to obtain the fractal dimension $d_B$ by obtaining the power law fitting of the distribution. The algorithm selects a random node at each step, and this node is the seed that will be the centre of a box. Then I search the network by distance $r_B$ from the seed node and cover all the nodes that are located within that distance, but only if they have not been covered yet. Later, I assign the newly covered nodes to the new box; if there are no more newly covered nodes then the box is removed. This process is repeated until all the nodes of the network belong to boxes. Before using the algorithm I calculate the skeleton of the Estonian network of payments.

One of the main challenges of complex network studies is the identification of critical structural features that are underneath the network's complexity. This is related with the basic concept of: the distinctive character of a whole is inside just a few of its parts, for example in specific colours and shapes of a painting, particular notes or tunes in a song or certain keywords in a text or speech. This basic concept is also true for complex networks, where only a few parts of the whole network reflect the most important properties of it. For example, in large-scale networks only a small number of links are critical for the network to exist as a whole. A skeleton network is generally smaller than the original and it reproduces all the fundamental properties of the whole because it contains the essence of the network. Grady et al., (2012) analysed the network of international flight connections and discovered that the skeleton network consists of just 6.76% of the original network. The skeleton network concept can be used to detect epidemic propagations of disease when indicating which individuals are key participants in a social network or it can be useful when describing ecosystems to identify the species that should not be damaged at all to avoid jeopardizing the whole network.

The concept of skeleton was first introduced by Kim et al., (2004). The skeleton is a particular type of spanning tree based on the link betweenness centrality (a simplified quantity to measure the traffic of networks) that is entrenched beneath the original network. The skeleton provides a shell for the fractality of the network and is formed by links with the highest betweenness centralities. Only the links that do not form loops are included. The remaining links from the original network which are not included in the skeleton are local shortcuts that contribute to loop formation, meaning that the distance between any two nodes in the original network may increase in the skeleton. A fractal network has a fractal skeleton beneath which is distressed by these local shortcuts but it preserves fractality. For a scale-free network the skeleton also follows a power-law degree distribution where the degree exponent might differ slightly from that of the original network. When studying the origin of fractality in networks, actually the skeleton is more useful than the original network itself due to its unsophisticated and simplistic tree structure (Goh et al.,, 2006). In general, the skeleton preferentially collects the sections of the network where betweenness is high and this preserves the structure and simplifies its complexity. Therefore, by looking at the properties of the skeleton it is easier to appreciate the topological organization of the original network.

In order to calculate the skeleton structure of a complex network, the link betweenness of all the links in the network has to be calculated. The betweenness centrality of a network (of a link or a node), is defined as follows:

$$b_i = \sum_{j,k \in \mathrm{N}, j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \quad (19)$$

where $\mathrm{N}$ is the total number of nodes, $n_{jk}$ is the total number of shortest-paths between nodes $j$ and $k$. $n_{jk}(i)$ is the total number of shortest-paths linking nodes $j$ and $k$ that passes through the node $i$. In order to perform the fractal scaling analysis, I used Dijkstra's algorithm (Gibbons, 1985); then I used the box-covering algorithm to calculate the fractal dimension of the network and the skeleton to compare both values.

I present a fractal scaling analysis by using the box-counting algorithm expressed in Equation 20 and I calculated the fractal dimension of the network and its skeleton. Figure 17 shows a visualization of the graph representation of the skeleton of the network. The box-covering method yields a fractal dimension $d_{Bs}$= 2.32 ± 0.07 for the skeleton network and for the original network the fractal dimension is $d_{Bo}$= 2.39 ± 0.05.

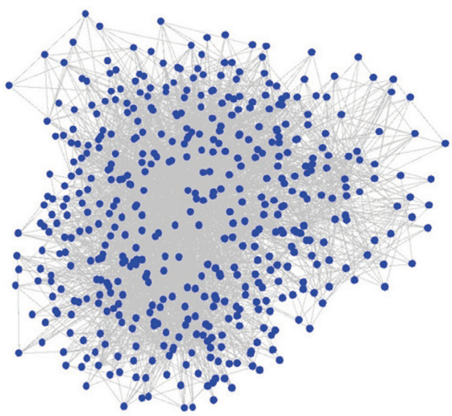


*Figure 17. Graph representation of the skeleton of the Estonian network of payments.*

The comparison of the fractal scaling in the network and its skeleton structure revealed its own patterns according to the fractality of the network. Figure 18 shows a fractal scaling representation of the network and its skeleton, where the fractal dimension is the absolute value of the slope of the linear fit.

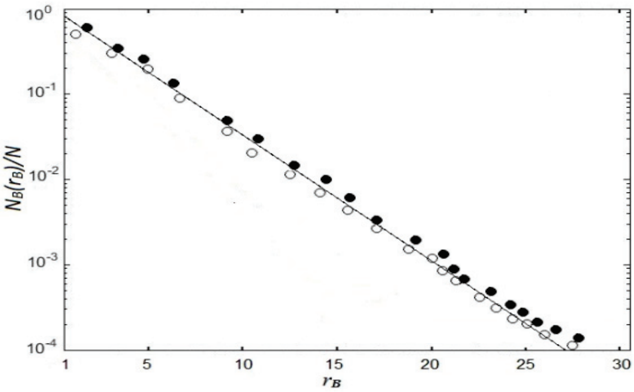



*Figure 18. Fractal scaling representation of the network. The original network (o) and the skeleton network (●). The straight line is included for guidance and has a slope of 2.3. The analysis includes only the giant connected cluster of the network.*

As seen in the plot of Figure 18, the respective number of boxes needed to cover both networks is very similar but not identical, actually more boxes were needed for covering the skeleton. The largest distance between any two nodes in the network of payments is 29, while the largest distance between any two nodes in the skeleton network is 34.

## 3.2 Multifractal network analysis

Scale-free networks are commonly observed in a wide array of different contexts of nature and society. In the first sub-section of this Chapter I have shown that the Estonian network of payments has scale-free properties characterized by power-law degree distributions

In general, multifractality is expected to appear in scale-free networks due to the fluctuations that occur in the density of local nodes. Tél et al., (1989) introduced a sandbox algorithm based on the fixed-size box-counting algorithm (Halsey et al.,, 1986) which was used and adapted for multifractal analysis of complex networks by Liu et al., (2015). In order to determine the multifractal dimensions of the Estonian network of payments, I chose this adapted sandbox algorithm because it is precise, efficient and practical. Moreover, a study by Song et al., (2015) has shown that this algorithm gives better results when it is used in unweighted networks.

The fixed-size box-counting algorithm is one of the most known and efficient algorithms for multifractal analysis. For a given probability measure $0 \leq \mu \leq 1$ in a metric space Ω with a support set E, I consider the following partition sum:

$$Z_{\varepsilon}(q) = \sum_{\mu(B)\neq 0} [\mu(B)]^q \,, \qquad (20)$$

where the parameter $q \in$ R, and describes the moment of the measure. The sum runs over all the different non-overlapping (or non-empty) boxes B of a given size ε that covers the support set E. From this definition, it is easy to obtain $Z_{\varepsilon}(q) \geq 0$ and $Z_{\varepsilon}(0) = 1$. The function of the mass exponents $\tau(q)$ of the measure μ is defined by:

$$\tau(q) = \lim_{\varepsilon \to 0} \left( \frac{\ln Z_{\varepsilon}(q)}{\ln \varepsilon} \right). \tag{21}$$

Then, the generalized fractal dimensions $D(q)$ of the measure $\mu$ are defined as follows:

$$D(q) = \frac{\tau(q)}{q-1}, q \neq 1, \tag{22}$$

and

$$D(1) = \lim_{\varepsilon \to 0} \frac{Z_{(1,\varepsilon)}}{\ln \varepsilon}, q = 1, \tag{23}$$

where

$$Z_{1,\varepsilon} = \sum_{\mu(B) \neq 0} \mu(B) \ln \mu(B). \tag{24}$$

The generalized fractal dimensions $D(q)$ can be estimated with linear regression of $[\ln Z_{\varepsilon}(q)]/[q-1]$ against $\ln \varepsilon$ for $q \neq 1$, and similarly a linear regression of $Z_{1,\varepsilon}$ against $\ln \varepsilon$ for $q = 1$. $D(0)$ is the fractal dimension or the box-counting dimension of the support set E of the measure $\mu$. $D(1)$ is the information dimension and $D(2)$ is the correlation dimension.

For a complex network, a box of size B can be defined in terms of the distance $l_B$, which corresponds to the number of links in the shortest-path between two nodes. This means that every node is less than $l_B$ links away from another node in the same box. The measure $\mu$ of each box is defined as the ratio of the number of nodes that are covered by the box and the total number of nodes in the whole network.

Multifractality of a complex network can be determined by the shape of $\tau(q)$ or $D(q)$ curves. If $\tau(q)$ is a straight line or $D(q)$ is a constant, then the network is monofractal; similarly if $D(q)$ or $\tau(q)$ have convex shapes, then the network is multifractal. A multifractal structure can be identified by the following signs (Grassberger and Procaccia, 1983): multiple slopes of $\tau(q)$ vs $q$, non-constant $D(q)$ vs $q$ values and $f(\alpha)$ vs $\alpha$ value covers a broad range (not accumulated at nearby non-integer values of $\alpha$).

Firstly, I calculate the shortest-path distance between any two nodes in the network and map the shortest-path adjacency matrix $B_{NxN}$ using the payments adjacency matrix $A_{NxN}$. Then I use the shortest-path adjacency matrix $B_{NxN}$ as input for multifractal analysis. The central idea of the sandbox algorithm is simply to select a node of the network in a random fashion as the centre of a sandbox and then count the number of nodes that are inside the sandbox. Initially, none of the nodes has been chosen as a centre of a box or as a seed. I set the radius $r$ of the sandbox which will be used to cover the nodes in the range $r \in [1, d]$, where $d$ (diameter) is the longest distance between nodes in the network and radii $r$ are integer numbers. I ensure that the nodes are chosen randomly as centre nodes by reordering the nodes randomly in the whole network. Depending on the size N of the network, I choose $T$ nodes in random order as centres of $T$ sandboxes; then I find all the neighbouring nodes within radius $r$ from the centre of each box. I count the number of nodes contained in each sandbox of radius $r$, and denote that quantity by $S(r)$. I calculate the statistical averages $\langle [S(r)^{q-1}] \rangle$ of $[S(r)^{q-1}]$ over all the sandboxes $T$ of radius $r$.

The previous steps are repeated for each of the different values of radius $r$ in order to obtain the statistical average $\langle [S(r)^{q-1}] \rangle$ and use it for calculating linear regression. The generalized fractal dimensions $D(q)$ of the measure $\mu$ are defined by

$$D(q) = \lim_{r \to 0} \frac{ln\langle [S(r)/S(0)]^{q-1} \rangle}{ln(r/d)} \frac{1}{q-1}, q \in R, \quad (25)$$

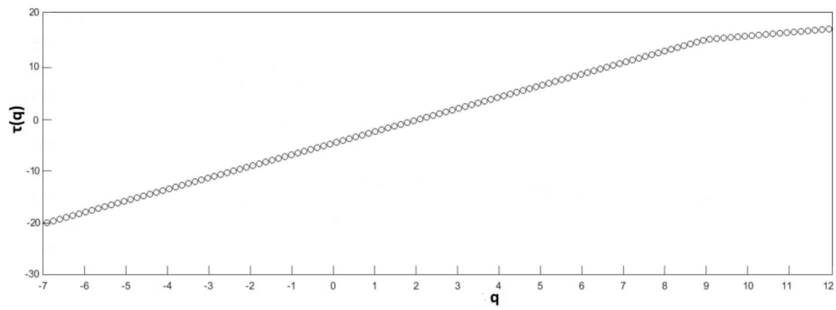
or rewritten as

$$ln(\langle [S(r)]^{q-1} \rangle) \propto D(q)(q-1)\, ln(r/d) + (q-1) ln(S_0), \quad (26)$$

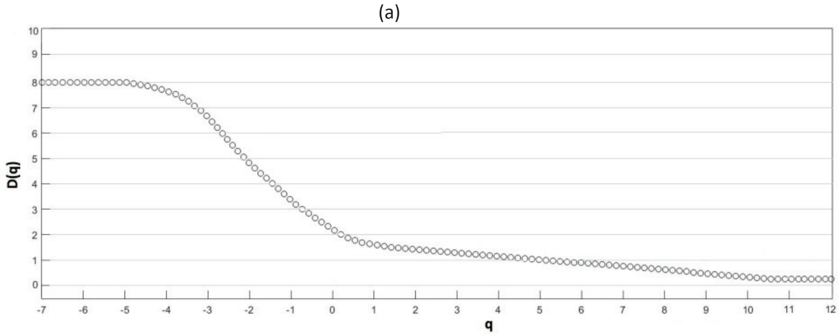
where $S(0)$ is the size of the network and the brackets mean taking statistical average over the random selection of the sandbox centres. I run the linear regression of $\ln(\langle [S(r)]^{q-1} \rangle)$ against $(q-1)\ln(r/d)$ to obtain the generalized fractal dimensions and similarly, calculate the linear regression of $\ln(\langle [S(r)]^{q-1} \rangle)$ against $\ln(r/d)$ to obtain the mass exponents $\tau(q)$. From the shapes of the generalized fractal dimension curves, I can conclude if multifractality exists or not in this network.

Linear regression is an important step to obtain the correct range of radius $r \in [r_{\min}, r_{\max}]$ that is needed to calculate the generalized fractal dimensions (defined by Equations 25 and 26) and the mass exponents (defined by Equation 21). I found an appropriate range of radii $r$ within the range of the interval located between 2 and 29 of the linear regression, thus selected this linear fit scaling range to perform multifractal analysis (I set the range of $q$ values from -7 to 12).

I calculated $\tau(q)$ and the $D(q)$ curves using the sandbox algorithm by Liu et al., (2015) and based upon the shapes obtained from the spectrum in Figures 19(a) and 19(b), it can be seen that the curves are non-linear, suggesting that the network is multifractal.




(a)




(b)

*Figure 19. (a) Plot of mass exponents $\tau(q)$ as function of q. (b) Plot of generalized fractal dimensions $D(q)$ as function of q. Curves indicated by circles represent numerical estimations of the mass exponents and generalized fractal dimensions, respectively.*

In Figure 19(b), the $D(q)$ function decreases sharply after the peak reaches its end when $q = -4$. This could be interpreted as the high densities around the hubs in the network. The hubs have a high number of links connected to them; therefore the density of links around the sections near the hubs is higher than in other parts of the network. These hub nodes or important companies have a noticeable larger amount of business partners (for example: customers, suppliers or any other business parties that interact financially) than the rest of the companies in the network have, and it is interesting to observe that this characteristic can be explored and identified by looking at the values of $D(q)$ spectra. The multifractality seen in this network reveals that the system cannot be described by a single fractal dimension suggesting that the multifractal approach provides a better characterization; hence, this means that the Estonian economy is multifractal.

*Table 4 Comparison of the maximum values of D(q) in different networks*

| Network | Number of nodes | Highest *D(q)* | Reference |
|---|---|---|---|
| Pure fractal network | 6222 | 2.8 | (Li et al., 2014) |
| Small world network | 6222 | 6.6 | (Li et al., 2014) |
| Semi fractal network | 6222 | 3.1 | (Li et al., 2014) |
| Sierpinski weighted fractal network | 9841 | 2.0 | (Song et al., 2015) |
| Cantor dust weighted fractal network | 9841 | 3.2 | (Song et al., 2015) |
| High-energy theory collaboration weighted network | 8361 | 6.0 | (Song et al., 2015) |
| Astrophysics collaboration weighted network | 16706 | 6.2 | (Song et al., 2015) |
| Computational geometry collaboration weighted network | 7343 | 5.1 | (Song et al., 2015) |
| Barabási & Albert model scale-free network | 10000 | 3.6 | (Liu et al., 2015) |
| Newman and Watts model small-world network | 10000 | 4.8 | (Liu et al., 2015) |
| Erdös-Rényi random graph model | 10000 | 3.9 | (Liu et al., 2015) |
| Barabási & Albert model scale-free network | 7000 | 3.4 | (Wang et al., 2012) |
| Random network | 5620 | 3.5 | (Wang et al., 2012) |
| Random network | 449 | 2.4 | (Wang et al., 2012) |
| Protein-Protein interaction network: Human | 8934 | 4.9 | (Wang et al., 2012) |
| Protein-Protein interaction network: Arabidopsis thaliana | 1298 | 2.5 | (Wang et al., 2012) |
| Protein-Protein interaction network: C. elegans | 3343 | 4.5 | (Wang et al., 2012) |
| Protein-Protein interaction network: E. coli | 2516 | 4.1 | (Wang et al., 2012) |
| Small world network | 5000 | 3.0 | (Wang et al., 2012) |
| Estonian network of payments | 16613 | 7.8 | (Rendón de la Torre et al., 2016) |

The quantity $\Delta D(q)$ describes the changes in link density in this network. I use $\Delta D(q) = D(q)_{max} - lim\, D(q)\, to$ observe how the values of $D(q)$ change across the spectrum. From Figure 19(b) it was found that $\lim D(q) = 0.37$ and $D(q)_{max} = 7.8$ and this means that $\Delta D(q) = 7.43$. A large $D(q)$ value means that the link distribution is very irregular, suggesting there are areas near the hubs where the links are densely grouped contrasting with areas where the nodes are connected with only a few links. In this network this means that just a few companies have the role of hubs, while the rest are just small participants of the payments network. Table 4 shows a comparison of the maximum values of $D(q)$ in different networks.

In this Chapter I presented the first multifractal analysis of a complex network of payments. I studied specific fractal and multifractal properties of a novel and unique network: the Estonian network of payments. In this study, I presented a fractal scaling analysis where I identified the underlying skeleton structure of the network. I calculated its fractal dimension and compared it with the fractal dimension of the original network. I found that the skeleton network had a slightly smaller fractal dimension than the original network. This comparison, between the fractal scaling in the original network and the corresponding skeleton network reveals that there are only slightly distinct patterns according to the fractality in the network. This means that the skeleton network preserves the structure very well while simplifying the complexity of the network. Then, the skeleton network captures the general structure of the network and by observing the properties of the skeleton, an easier visualization of the topological organization of the network can be achieved.

Fractal analysis helps to calculate and understand the fractal dimension of complex networks. However, it is necessary to describe and characterize the multiple fractal patterns which cannot be described by a single fractal dimension, thus I also performed a multifractal analysis on the Estonian network of payments. Multifractal analysis allows the calculation of a set of fractal dimensions, particularly the generalized fractal dimensions. I examined the general multifractal structure and explored some statistical features of this network. In order to study the multifractal structure, I calculated the spectrum of the mass exponents $\tau(q)$ and the generalized fractal dimensions $D(q)$ curves, using a sandbox algorithm for multifractal analysis of complex networks adapted by Liu et al., (2015). This algorithm is based on the fixed-size box-counting algorithm developed by Tél et al., (1989). The sandbox algorithm utilized in this study could also be used to explore and characterize other economic networks.

My results indicated that multifractality exists in the Estonian network of payments, and this suggests that the Estonian economy is multifractal (from the point of view of networks). I found large values of $D(q)$ spectra and this means that the distribution of links is quite irregular in the network, suggesting there are specific nodes which hold densely connected links, meanwhile other nodes hold just a few links. This type of structure could be relevant when specific critical events occur in the economy that could threaten the whole network.

It is important to continue observing, describing and analysing the structures and characteristics of economic complex networks in order to be able to understand their underlying processes or to be able to detect patterns that could be useful for predicting or forecasting events and trends. The addition of evidence through empirical studies of economic networks represents a step forward towards the knowledge on the universality and the unravelling of the complexity of economic systems.

Further applications and studies could extend this topic by examining the potential factors that drive the strength of the multifractal spectrum. Some applications could involve studying the origin of such factors. Another interesting line of research would be to study the patterns and the changes of the multifractal spectrum across different periods of time. Particularly, it would be interesting to analyse such patterns during determined financial crisis periods for risk pattern recognition purposes. Also, it would be interesting to take into account different probability measures for such kind of multifractal analysis. Another direction of the studies could be to focus on building network models that attempt to forecast country money flows or potential industry growth trends based on transactions data.

# Conclusions

## Summary of the results

The presented studies address global properties and statistics related to the topological structure of the large-scale payments network of an entire country (Estonia) by using payments data. Additionally, I have reviewed some topics related with its community structure and moreover, I have analysed some aspects related to multifractal and fractal properties of complex networks.

Complex networks can be considered as the skeleton of complex systems and they are present in many kinds of social, economic, biological, chemical, physical and technological systems. In the network of Estonian payments I found scale-free degree distributions, small world property, low clustering coefficient, disassortative degrees and heterogeneity properties. Its scale-free structure indicates that a low number of companies in Estonia trade with a high number of companies, while the majority of the companies trade with only few. The clustering coefficient distribution suggests the existence of a hierarchic structure in the network. This network is a small world with just 7 degrees of separation. The connectivity is smaller than the overall clustering coefficient, therefore the Estonian network of payments is not random. The diameter value suggests there is a preference among companies for particular paths of money.

I explored the relations between weighted quantities and their network underlying structures. I investigated the strength of interactions (number of payments and the volumes of payments) and the interconnectivities among these interactions. To achieve this, I did particular experiments, calculated specific metrics, and thus revealed interesting micro-structural features.

I tested the robustness of the network with an approach that focuses on the collective influencer nodes. First, I located the key nodes that prevent the network from breaking into disconnected components. The simulation assumed a targeted removal of key nodes which caused a quick growth in the average shortest path length until the network was destroyed at an optimal percolation threshold of 6%, while in a random removal of nodes the damage was extremely small. This revealed the robustness of this economic network against random attacks but also revealed its vulnerability to smart attacks. The low percentage of the optimal percolation threshold reveals that the most influential companies in the network are not necessarily the most connected ones or those having more economic activity and that a small quantity of companies maintains the unity of the whole network.

Later, I analysed the community structure of the network by using the Clique Percolation Method. I found that there are also scale-free properties in the statistical distributions of the community structure. Size, overlap and membership distributions follow shapes that are compatible with power-laws. Power-law distributions have already appeared in this network at a global scale in the level of nodes, and in this community study I have shown that power-laws are also present at the level of overlapping communities.

An immediate application for the community detection output is that it can be used in targeted marketing activities, as input for predictive analytical models such as in product acquisition propensities, churn, product affinity analyses, for creating marketing profiles or customer segmentations and for creating customer target lists for product offering (in an effort to propagate consumer buzz effects). Further applications for community detection in similar economic networks could involve the identification of patterns between companies, tracking suspicious business activities and strengthening relationships between companies of the same community for improving performance of the whole network.

In the last Chapter, I presented a fractal and multifractal analysis of the network. I identified the underlying structure of the network (its skeleton) and measured the fractal dimension of the skeleton to compare it with the fractal dimension of the original network. Both fractal dimensions were similar but the fractal dimension of the skeleton was slightly smaller. I also analysed the general multifractal structure by calculating the spectrum of the mass exponents ($q$) and the generalized fractal dimension $D(q)$ curves, through a sandbox algorithm for multifractal analysis of complex networks. My results indicated that multifractality exists in the Estonian network of payments, and this suggests that the Estonian economy is multifractal (from the point of view of networks). I found large values of $D(q)$ spectra, which means that the distribution of links is quite irregular in the network, suggesting there are specific nodes which hold densely connected links while other nodes hold just a few links. This type of structure could be relevant when critical events occur in the economy that could threaten the whole network.

It is important to continue studying the structures and characteristics of economic complex networks in order to be able to understand their underlying processes and to be able to detect patterns that could be useful for predicting or forecasting events and trends. The addition of evidence through empirical studies of fractality, multifractality, communities' detection and structural properties of economic networks represents a step forward towards unravelling of the complexity of economic systems.

## Main conclusions proposed to defend

1. I studied the structure of the economic network of an entire country, after extracting the network's topology, characteristics and statistics I conclude that this economic network has scale-free properties (in its degree distributions and statistical distributions of the community structure such as: size, overlap and membership distributions). The network also shows small world characteristics and low clustering coefficient.

2. The network is disassortative in terms of degree. The system shows topological heterogeneity due to its scale-free structure in the degree distributions (few companies in Estonia trade with many parties while the majority trade with only a few).

3. I performed robustness tests on the network: One based on centralities and another test based on collective influencer nodes. In the first analysis the percolation threshold is 8% and in the second is 6%. I found the nodes that prevent the network from breaking into disconnected components. The analysis revealed the robustness of the network against random attacks but it also revealed its vulnerability to targeted attacks. This analysis concludes that the most influential companies in the network are not necessarily the most connected ones or those which have more economic activity. Only a small number of companies maintain the unity of the network.

4. I presented the first multifractal analysis of a complex network of payments where I studied specific fractal and multifractal properties.

5. I identified the skeleton structure of the network (as part of a fractal scaling analysis) where I calculated the fractal dimension. The analysis showed that the both the

fractal dimensions of the skeleton network and the original network are very similar, which means that the skeleton network preserves the structure very well while it simplifies the complexity of the network. This means that one can study a simplified version of the network (skeleton networks) and still capture the general structure of the original network itself.

6. I calculated the spectrum of the mass exponents and the generalized fractal dimensions curves. The results indicated that there is multifractality in the network, suggesting that the Estonian economic system is multifractal.

7. I found large values of $D(q)$ spectra and this means that the distribution of links is quite irregular in the network, suggesting there are specific nodes which hold densely connected links, meanwhile other nodes hold just few links. This type of structure could be relevant when specific critical events occur in the economy that could threaten the whole network.

8. I studied the community structures of the Estonian network of payments by using the Clique Percolation Method. The output of the community detection analysis could be used by the bank for targeted marketing activities or as features for predictive analytical models (propensity models for acquisition of products, or for creating marketing profiles or segmentation).

## Recommendations for further work

It is important to continue observing, describing and studying the structures and characteristics of economic complex networks in order to be able to understand their underlying processes and to detect patterns that could be useful for predicting or forecasting events and trends.

Regarding community structure in economic networks, a question that remains open for future research is to investigate if the similarities in communities' features amongst different complex networks arise randomly or if there are any unknown properties shared by all of them. Another interesting open line of research is to study the plausibility of predicting changes in a payment network through community detection analysis. Further applications in economic networks could involve strengthening relationships between companies of the same community to improve the performance of the whole network, targeted marketing, identification of patterns between companies and tracking of suspicious business activities.

Further applications of multifractal studies in economic networks might involve examining the potential factors that drive the strength of the multifractal spectrum. Some applications could involve studying the origin of such factors. Another interesting line of research would be to study the patterns and the changes of the multifractal spectrum across different periods of time. Particularly, it would be interesting to analyse such patterns during financial crisis periods for risk pattern recognition purposes. Also, it would be interesting to take into account different probability measures for such kind of multifractal analysis. A further direction of the studies could focus on building network models that attempt to forecast country money flows or potential industry growth trends based on data of transactions.

## List of Figures

**Figure 1**. Time evolution activity of patterns (payments and volumes). (a) Monthly trading volumes of payments V. (b) Monthly number of transactions T. (c) Average degree $< d >$> versus time $t$. X-axis represents the number of the month of year 2014.

**Figure 2**. (a) Empirical degree distribution for the connectivity network of the Estonian network of payments. X axis is the number of $k$ degrees and Y axis is $P(k)$. (b) out-degree distribution of the network, $P\ (\ k\ ) \sim\ k^{-2\ .39}$. (c) Empirical in-degree distribution $P\ (\ k\ ) \sim k^{-2\ .49}$.

**Figure 3**. Node out-degree distribution by strength.

**Figure 4**. Components of a directed network.

**Figure 5**. Distribution of the clustering coefficient of the Estonian network of payments.

**Figure 6**. Distributions of strength. (a) Node out-strength as a function of degree. (b) Node in-strength as a function of degree. Empty squares represent value of payments and full squares represent the number of payments.

**Figure 7**. Participation ratio plots. (a) Participation ratio as a function of the inverse degree for out-going payments and in-coming payments. Black squares represent out-going payments, white squares in-coming. (b) Participation ratio as a function of the distance from the end of the year expressed in months for incoming (black circles) and outgoing (white circles) payments.

**Figure 8**. GCC of the network of payments as a function of the percolation threshold $Pc$.

**Figure 9.** Plots on the effect of the targeted and random damage on the network of payments. (a) The average shortest-path length $< l >$ in the GCC plotted against the percentage of removed nodes. (b) The GCC plotted against the percentage of removed nodes. Continuous lines display the effect of the targeted removal and the dashed lines display the effect of the random removal of nodes. $Pc$ is the percolation threshold, for each case.

**Figure 10**. Visual representation of a section of the overlapping network of communities (Estonian network of payments). The circles (nodes) represent communities and the black lines between them represent shared nodes between communities.

**Figure 11**. Plot of the average size of community *(s)* and number of communities *(c)* as *k* increases. Squares represent the number of communities and triangles represent the size of the communities.

**Figure 12**. Cumulative community size distribution at different times *t.*

**Figure 13**. Statistics of community size. $< s >$ is the average community size. $\sigma$ is the standard deviation of the size of communities at different times *t.*

**Figure 14**. Cumulative distribution of community degrees *d.*

**Figure 15.** Cumulative distribution function of the membership number $m_i$.

**Figure 16**. Cumulative distribution function of the overlap size $s_o$.

**Figure 17**. Graph representation of the skeleton of the Estonian network of payments.

**Figure 18.** Fractal scaling representation of the Estonian network of payments. The original network (o) and the skeleton network (●). The straight line is included for guidance and has a slope of 2.3. The analysis includes only the giant connected cluster of the network.

**Figure 19**. (a) Plot of mass exponents $\tau(q)$ as function of $q$. (b) Plot of generalized fractal dimensions $D(q)$ as as function of $q$. Curves indicated by circles represent numerical estimations of the mass exponents and generalized fractal dimensions, respectively.

# List of Tables

**Table 1.** Scaling exponents and clustering coefficients for different types of reported networks. $\gamma i$ = scaling exponent for in-degree distribution. $\gamma o$ = scaling exponent for the out-degree distribution. $\gamma$ = scaling exponent for the connectivity distribution. *Refers to average clustering coefficient.

**Table 2**. Network's characteristics.* All money quantities are expressed in monetary units and not in real currencies in order to protect the confidentiality of the data set. The purpose of showing monetary units is to provide an idea of the proportions of quantities and not to show exact amounts of money.

**Table 3**. Summary of Statistics N = number of nodes. < *k* > = average degree. $\gamma o$ = scaling exponent of the out-degree empirical distribution. $\gamma i$ = scaling exponent of the in-degree empirical distribution. $\gamma$ = scaling exponent of the connectivity degree distribution. <*C*> = average clustering coefficient. < *l* > = average shortest path length. *T* = connectivity %. *D* = Diameter. < $\sigma$ > = average betweenness. GCC = Giant Connected Component. DC = Disconnected Component. GSCC = Giant Strongly Connected Component. GOUT = Giant Out-Component.

**Table 4**. Comparison of the maximum values of *D(q)* in different networks.

## References

Aiello W., Chung F., Lu L. (2000). A random graph model for massive graphs. In: *Proceedings of the thirty-second annual ACM symposium on theory of computing,* 171–180.

Albert R., Jeong H., Barabási A.L. (1999). Error and attack tolerance of complex networks. *Nature*, 406, 378–382.

Albert R., Jeong H., Barabási A.L. (1999). Diameter of the World Wide Web. *Nature*, 401, 130–131.

Barrat A., Barthelemy M., Pastor-Satorras R. Vespignani A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences of USA*, 101, 3747–3752.

Barrat A., Barthélemy M., Vespigniani A. (2008). *Dynamical processes on complex networks*. Cambridge University Press, Cambridge.

Battiston S., Rodrigues J.F., Zeytinoglu H. (2007). The network of inter-regional direct investment stocks across. *Advances in Complex Systems*, 10, 29–51.

Boss M., Helsinger H., Summer M., Thurner S. 2004. The network topology of the interbank market. *Quantitative Finance*, 4, 677–684.

Bóta A., Kresz M. (2013). A high resolution clique-based overlapping community detection algorithm for small-world networks. *Informatica*, 37, 177–187.

Clauset A., Newman M.E.J., Moore C. (2004). Finding community structure in very large networks. *Physical Review E*, 70.

Cohen K., Erez D., Avraham D., Havlin S. (2000). Resilience of the internet to random breakdowns. *Physics Review Letters,* 85, 4625–4630.

Derényi I., Palla G., Vicsek T. (2005). Clique percolation in random networks. *Physical Review Letters,* 94.

Devi J.C., Poovammal E. (2016). An analysis of overlapping community detection algorithms in social networks. *Procedia Computer Science*, 89, 349–358.

Ding Z., Zhang X., Sun D., Luo B. (2016). Overlapping community detection based on network decomposition. *Scientific Reports*, 6.

Dorogovtsev S.N., Mendes J.F.F. (2003). *Evolution of networks*. Oxford University Press, New York.

Dourisboure Y., Geraci F., Pellegrini M (2007) Extraction and classification of dense communities in the web. *Proceeding of the 16th International Conference on the World Wide Web*, 1, 461-470.

Doye J.P.K. (2002). The network topology of a potential energy landscape: A static scale-free network. *Physics Review Letters*, 88.

Ebel H., Mielsch L.I., Bornholdt S. (2002). Scale-free topology of e-mail networks. *Physics Review E*, 66.

Eguiluz V.M., Hernández-García E., Piro O., Klemm K. (2003). Effective dimensions and percolation in hierarchically structured scale-free networks. *Physical Review E* , 68.

Erdös P., Rényi A. (1959). On random graphs. *Publicationes Mathematicae Debrecen*, 6, 290–297.

Estrada E. (2011). *The structure of complex networks*. Oxford University Press, New York.

Everett M.G., Borgatti S.P. (1998) Analyzing clique overlap. *Connections*, 21, 49–61.

Feder J. (1988). *Fractals.* Plenum, New York.

Fenn D., Porter M., McDonald M., Williams S. Johnson N., et al., (2009). Dynamic communities in multichannel data: an application to the foreign exchange market during the 2007–2008 credit crisis. *Chaos*, 19.

Ferrer R., Cancho R.V., Solé R.V. (2003). Optimization in complex networks. Statistical mechanics of complex networks. *Lecture Notes in Physics*, 625, 114–125.

Fotios M., Siokis M. (2014). European economies in crisis: A multifractal analysis of disruptive economic events and the effects of financial assistance. *Physica A*, 395, 283–292.

Furuya, S., Yakubo, K. (2011). Multifractality of complex networks. *Physical Review E*, 84.

Gallos L.K., Song C., Havlin S., Makse H.A. (2007). A review of fractality and self-similarity in complex networks*.* *Physica A*, 386, 686–691.

Gavin A.C. et al., (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 6868, 141–147.

Gibbons A. (1985) *Algorithmic graph theory*. Oxford University Press, Cambridge.

Glattfelder, J.B., Battiston, S. (2009). Backbone of complex networks of corporations: The flow of control. *Physical Review E*, 80.

Goh K.I., Salvi G., Kahng B., Kim D. (2006). Skeleton and fractal scaling in complex networks. *Physics Review Letters*, 96.

González M.C., Herrmann H.J., Kertész J., Vicsek T. (2007). Community structure and ethnic preferences in school friendship networks. *Physica A*, 379, 307–316.

Grady D., Thiemann C., Brockmann D. (2012). Robust classification of salient links in complex networks. *Nature Communication*, 3.

Grassberger P., Procaccia I. (1983). Characterization of strange attractors. *Physics Review Letters*, 50, 346–349.

Guimerá R., Amaral N. (2005). Functional cartography of complex metabolic networks *Nature*, 433, 895–900.

Halsey T.C., Jensen M.H., Kadanoff L.P., Procaccia I., Shraiman B.I. (1986). Fractal measures and their singularities: The characterization of strange sets., 33, 1141–1151.

Hanneman R.A., Riddle M. (2005). *Introduction to social network methods*. University of California Riverside, California.

Hopcroft J., Khan O., Kulis B., Selman B. (2004). Tracking evolving communities in large linked networks. *Proceedings of the National Academy of Sciences USA*, 110, 5249–5253.

Inaoka H., Nimoniya T., Taniguchi K., Shimizu T., Takayasu H. (2004). Fractal network derived from banking transactions – an analysis of network structures formed by financial institutions. *Bank of Japan Working papers*.

Iori G., Jafarey S. (2001) Criticality in a model of banking crisis. *Physica A*, 299, 205–212.

Iori G., De Masi G., Precup O.V., Gabbi G., Caldarelli G. (2007). A network analysis of the Italian overnight money market. *Journal of Economic Dynamics and Control,* 32, 259–278.

Jeong H., Mason S.P., Barabási A.L., Oltvai Z.N. (2001) Lethality and centrality in protein networks. *Nature*, 411 41–42.

Jeong H., Tombor B., Reka A., Zoltan N.O., Barabási A.L. (2000). The large-scale organization of metabolic networks. *Nature,* 407, 651–654.

Kaluza, P., Kölzsch, A., Gastner, M.T., Blasius, B. (2010). The complex network of global cargo ship movements. *Journal of the royal society interface*, 7, 1093–1103.

Kim D.H., Noh J.D., Jeong H. (2004). Scale-free trees: The skeletons of complex networks. *Physics Review E*, 70.

Kim J.S., Goh K.I., Salvi G., Oh E., Kahng B., Kim D. (2007). Fractality in complex networks: Critical and supercritical skeletons. *Physical Review* E, 75.

König M.D., Battiston S. (2009). From graph theory to models of economic networks, a tutorial in: Naimzada A.K., Stefani S. & Torriero A. (eds). Networks, topology and dynamics. *Lecture Notes in Economics and Mathematical Systems*. Springer, 613, 23–63.

Lewis A.C.F., Jones N.S., Porter M.A., Deane C.M. (2010) The function of communities in protein interaction networks at multiple scales. *BMC Systems Biology*, 4.

Li B.G., Yu Z.G., Zhou Y. (2014). Fractal and multifractal properties of a family of fractal networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2.

Liu J.L., Yu Z.G., Anh V. (2015). Determination of multifractal dimensions of complex networks by means of the sandbox algorithm. *Chaos*, 25.

Lu Y., Tang J. (2004). Fractal dimension of a transportation network and its relationship with urban growth: A study of the Dallas-Fort Worth area. *Environmental and Planning B*, 31, 895–911.

Lublóy A. (2006). Topology of the Hungarian large-value transfer system. Magyar Nemzeti Bank (Central Bank of Hungary). *MNB Occasional Papers*, 57.

Mandelbrot B. (1983). *The fractal geometry of nature*, NY Academic Press, New York.

Mantegna R. N., Stanley, H.E. (2000). *An Introduction to Econophysics: Correlations and complexity in finance*. Cambridge University Press, Cambridge.

Morone F., Makse H.A. (2015). Influence maximization in complex networks through optimal percolation. *Nature,* 524, 65–68.

Nakano T., White D. (2007). Network structures in industrial pricing: The effect of emergent roles in Tokyo supplier-chain hierarchies. *Structures and Dynamics*, 2, 130–154.

Newman M.E.J. (2004). Detecting community structure in networks. *European Physical. Journal B*, 38, 321–330.

Newman M.E.J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, 69.

Newman M.E.J. (2004) Analysis of weighted networks. *Physics Review E*, 70.

Newman M.E.J. (2010). *Networks: An introduction*. Oxford University Press, New York.

Palla G., Barabási A.L., Vicsek, T. (2007). Quantifying social group evolution. *Nature*, 446, 664–667.

Palla G., Derényi I., Farkas I., Vicsek T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435, 814–818.

Piccardi C., Calatroni L., Bertoni F. (2010). Communities in Italian corporate networks. *Physica A*, 389, 5247–5258.

Pollner P., Palla G., Vicsek T. (2006). Preferential attachment of communities: The same principle, but a higher level. *Europhysics Letters*, 73, 478–484.

Pont O., Turiel A., Pérez-Vicente C.J. (2009). Empirical evidences of a common multifractal signature in economic, biological and physical systems. *Physica A,* 388, 2025–2035.

Ravasz E., Somera A.L., Mongru D.A., Oltvai Z.N., Barabási A.L. (2002) Hierarchical organization of modularity in metabolic networks. *Science*, 297, 1551–1555.

Reka A., Barabási, A.L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74, 47–97.

Rendón de la T.S., Kalda J., Kitt R., Engelbrecht J. (2016). On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. *Chaos Solitons and Fractals*, 90, 18–27.

Reyes J., Schiavo S., Fagiolo G. (2008). Assessing the evolution of international economic integration using random-walk betweenness centrality: The cases of East Asia and Latin America. *Advances in Complex Systems*,11, 685–702.

Ripeanu M., Foster I., Iamnitchi I. (2002). Mapping the Gnutella network: properties of large-scale peer-to-peer systems and implications for system design. IEEE Internet Computing, 6, 50–57.

Schneider C.M, Kesselring T.A., Andrade J.S., Herrmann H.J. (2012). Box-covering algorithm for fractal dimension of complex networks. *Physical Review E*, 86.

Scott J. (2000). *Social network analysis: A handbook*. Sage Publications, UK.

Shen H.W. (2013). Community structure of complex networks. *Springer Science & Business Media.*

Song C., Havlin S., Makse H.A. (2005). Self-similarity of complex networks. *Nature*, 433, 392–395.

Song C., Gallos L.K., Havlin S., Makse H.A. (2007). How to calculate the fractal dimension of a complex network: the box covering algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 3.

Song Y.Q., Liu J.L., Yu Z.G., Li B.G. (2015). Multifractal analysis of weighted networks by a modified sandbox algorithm. *Scientific Reports*, 5.

Soramäki K., Bech M.L., Arnold J., Glass R.J., Beyeler W.E. (2007). The topology of interbank payment flows. *Physica A*, 379, 317–333.

Souma W., Fujiwara Y., Aoyama H. (2006). Heterogeneous economic networks in: Namatame A., Kaizouji T., Aruka Y. (eds). The complex networks of economic interactions. *Lecture Notes in Economics and Mathematical Systems*. Springer, 567, 79–92.

Standish R. K. (2008). Intelligent complex adaptive systems. *Australia: IGI Global*, 105–124.

Tél T., Fülöp A., Vicsek T. (1989). Determination of fractal dimensions for geometrical multifractals. *Physica A, 159*, 155–166.

Traud A.L., Mucha J.P., Porter M.A. (2012). Social structure of Facebook networks. *Physica A*, 391, 4165–4180.

Vitali S., Battiston B. (2014). The community structure of the global corporate network. *Plos,* 9.

Wang D.L., Yu Z.G., Anh V. (2012). Multifractal analysis of complex networks. *Chinese Physics B*, 21.

Watts D.J., Strogatz S.H. (1998). Collective dynamics of small-world networks. *Nature*, 393, 440–442.

Wei D.J. et al., (2013). Box-covering algorithm for fractal dimension of weighted networks. *Scientific Reports*, 3.

West D.B. (2003). *Introduction to graph theory.* Prentice Hall, Upper Saddle River, New Jersey.

Xie J., Kelley S., Boleslaw K.S. (2013). Overlapping community detection in networks: the state of the art and comparative study. *ACM Computing Surveys*, 45.

Yang Z., Algesheimer R., Tessone C.J. (2016) A comparative analysis of community detection algorithms on artificial networks*. Scienticif Reports*, 6.

Zemp D.C., Wiedermann M., Kurths J., Rammig A., J.F. Donges. (2014). Node-weighted measures for complex networks with directed and weighted edges for studying continental moisture recycling. *Europhysics Letters* 107.

Zhou W.X., Yiang Z.Q., Sornette D. (2007). Exploring self-similarity of complex cellular networks: The edge-covering method with simulated annealing and log-periodic sampling. *Physica A,* 375, 741–752.

## Acknowledgements

To stay focussed and work throughout these years of Doctoral Studies would have not been easy without the support of great friends and colleagues. Hoping that I do not forget anyone, this page is dedicated to all of them. I want to thank Prof. Jaan Kalda, my supervisor, thanks to whom I was able to move to Estonia for my Doctoral studies. The support he gave me during these years has been very valuable to me and *conditio sine qua non* for the success of my whole research project.

To Robert Kitt for his support in the bank, because I wouldn´t even be living and working in Estonia if it wasn't because of him and his initiatives and very important support through the projects in Swedbank. I thank Swedbank and him for allowing me to use their data for research purposes and providing me with very good job opportunities.

A grateful and due special thanks goes to Prof. Jüri Engelbrecht for his advice, directions and support since the very beginning of my studies in Estonia. Another, no lesser, goes to my parents, my sister and my grandmother.

To all my friends who have supported me in many different ways, with their kind words of encouragement and strength or with the perfect dinners and even through continental distances: Bibiana Garmendia, Pamela Moro, Sandra A. Olivera, Paola Rodriguez, Anette Väljamats and all my other national and international friends.

Big thanks go to Veronika Plotnikova, my direct boss in the bank. For her true understanding of what it means to do a Ph.D. while having a full-time job, and for her patience and support.

The financial support of the Dora Programme sponsored by the Archimedes Foundation, to the Institute of Cybernetics (CENS), to the School of Science and to the Tallinn University of Technology for my grant I would also like to acknowledge.

European Union
European Regional
Development Fund

Investing
in your future

DoRa

Eesti hariduse ja teaduse heaks
ARCHIMEDES

# Abstract

## From Econophysics to Networks: Structure of the Large-Scale Estonian Network of Payments

The thesis addresses the study of the structures and dynamics of economic complex networks through the exploration of different experiments on a unique, interesting and particular economic network: the large-scale Estonian network of payments. Mainly I focus on the analysis of global/local topology, community detection and fractal/multifractal properties.

This is the first study that analyses the economic development of a country during one year, from a complex network approach, through payments data. My data set is exclusive in its kind because around 80% of Estonia's bank transactions are done through only one bank and I obtained the payments data from that bank (Swedbank), hence, the economic structure of the whole country can be reconstructed and this data set reproduces fairly well the trends of money of the whole Estonian economy. In this network, the nodes represent Estonian companies and the links are established by payments between these nodes.

I explored the topology of this network by extracting the scaling-free and structural properties of this network. I show that this network has scale-free properties in its degree distributions. I also found that this network has small world characteristics, low clustering coefficient and is disassortative (degree). I performed simulations to reveal the resiliency of the network against random and targeted attacks of the nodes with two different approaches. In the first analysis, I used an approach based on centralities and the second analysis was based on a collective influencer method. The results of such analysis revealed the robustness of this economic network against random attacks but they also revealed its vulnerability towards smart attacks.

Revealing the community structure exhibited by real networks is a fundamental phase towards a comprehensive understanding of complex systems beyond the local organization of their components. I also studied the mesoscale structure of this network. I have analysed the community structure of the Estonian network of payments by using the Clique Percolation Method. I found that there are scale-free properties in the statistical distributions of the community structure. Size, overlap and membership distributions follow shapes that are compatible with power-laws.

I also presented the first multifractal analysis of a complex network of payments. In here, I studied specific fractal and multifractal properties. I found that the skeleton network had a slightly smaller fractal dimension than the original network. My results indicated that multifractality exists in the Estonian network of payments, and this suggests that the Estonian economy is multifractal (from the point of view of networks).

# Kokkuvõte

## Majandusfüüsikast võrgustikeni: Eesti suuremahulise maksevõrgustiku struktuur

Selles väitekirjas uuritakse majanduse kompleksvõrgustike struktuuri ja dünaamikat, viies selleks läbi mitmesuguseid eksperimente ühe ainulaadse, huvitava ja erilise majandusliku võrgustiku peal, milleks on Eesti suuremahuline maksete võrgustik. Analüüsis keskendume peamiselt globaalsele ja lokaalsele topoloogiale, kogukondade tuvastamisele ning fraktaalsete ja multifraktaalsete struktuuride tuvastamisele.

See on esimene teadustöö, milles analüüsitakse maksete andmestiku põhjal ühe riigi majanduse arengut tervikuna, kasutades selleks kompleksvõrgustiku meetodit. Meie andmestik on ainus omataoline, kuna umbes 80% pangatehingutest Eestis tehakse läbi ühe panga (Swedbanki) ja me saime kasutada nende maksete andmestikku. Seetõttu on võimalik rekonstrueerida terve riigi majanduse struktuur, sest meie andmestik kajastab üsna hästi terve Eesti majanduse rahatrende. Sõlmed selles võrgustikus tähistavad Eesti ettevõtteid ja ühendused on moodustunud nende sõlmede vahel toimuvatest maksetest.

Me uurisime saadud võrgustiku topoloogiat, võttes võrgustikust välja mitteskaleeritavad ja strukturaalsed tunnused. Näitasime, et võrgustiku valentside jaotusel on skaalata omadusi. Samuti leidsime, et meie võrgustikul on väikese maailma tunnuseid ja väike klasterdumiskoefitsient ning et see oli teatud määral mitteassortatiivne. Viisime kahte eri meetodit kasutades läbi simulatsioonid, mis paljastasid võrgustiku hea vastupidavuse juhuslike ja sihitud rünnakute korral sõlmedele. Esimeses analüüsis kasutasime tsentraalsusel põhinevat meetodit, teise analüüsi aluseks oli kollektiivsete mõjutajate meetod. Need tõid välja meie majandusvõrgustike tugevuse juhuslike rünnakute korral, kuid samuti võrgustike haavatavuse tarkade rünnakute korral ja läbiimbumise lävendi.

Kogukonna struktuuri esiletoomine reaalse võrgustiku näitel on oluline etapp teel komplekssüsteemide parema mõistmise poole, viies meid edasi komponentide lokaalsest struktuurist. Me uurisime oma võrgustikus ka keskmise mõõtkava struktuure. Oleme analüüsinud Eesti maksevõrgustike kogukondlikku struktuuri, kasutades selleks kogukonnatuvastuse meetodeid. Leidsime, et kogukonnastruktuuri jaotustel on skaalata tunnuseid. Suurus, kattuvus ja liikmete jaotus vastab kujunditele, mis on vastavuses võimsuse reeglitega.

Samuti teostasime maksete võrgustiku esimese multifraktaalanalüüsi. Uurisime selle võrgustiku konkreetseid fraktaalseid ja multifraktaalseid omadusi. Arvutasime välja tema fraktali mõõtmed ja võrdlesime neid algse võrgustiku fraktali mõõtmetega. Leidsime, et raamvõrgustiku fraktali mõõtmed olid veidi väiksemad kui algsel võrgustikul. Meie tulemused viitavad sellele, et Eesti maksete võrgustikus esineb multifraktaalsust, mis omakorda lubab oletada, et Eesti majandus on multifraktaalne (võrgustike seisukohalt). Empiiriliste uuringute kaudu lisanduvad tõendid, mis viitavad majandusvõrgustike fraktaalsusele ja multifraktaalsusele, on samm universaalsuse mõistmise poole ja avab meie ees majandussüsteemide keerukuse.

# Appendix

**Paper I**

**Rendón de la Torre S.**, Kalda J., Kitt R., Engelbrecht J. (2016). On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. Chaos, Solitons & Fractals, 90, 18–27. DOI:10.1016/j.chaos.2016.01.018.

Chaos, Solitons and Fractals 90 (2016) 18–27

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

# On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia

Stephanie Rendón de la Torre[a,*], Jaan Kalda[a], Robert Kitt[a,b], Jüri Engelbrecht[a]

[a] Institute of Cybernetics at Tallinn University of Technology, Akadeemia tee 21, 12618, Tallinn, Estonia
[b] Swedbank AS, Liivalaia 12, 15038 Tallinn, Estonia

ARTICLE INFO

*Article history:*
Received 29 September 2015
Revised 2 January 2016
Accepted 19 January 2016
Available online 15 February 2016

*Keywords:*
Complex systems
Network topology
Scale-free networks
Economic networks

ABSTRACT

This paper presents the first topological analysis of the economic structure of an entire country based on payments data obtained from Swedbank. This data set is exclusive in its kind because around 80% of Estonia's bank transactions are done through Swedbank; hence, the economic structure of the country can be reconstructed. Scale-free networks are commonly observed in a wide array of different contexts such as nature and society. In this paper, the nodes are comprised by customers of the bank (legal entities) and the links are established by payments between these nodes. We study the scaling-free and structural properties of this network. We also describe its topology, components and behaviors. We show that this network shares typical structural characteristics known in other complex networks: degree distributions follow a power law, low clustering coefficient and low average shortest path length. We identify the key nodes of the network and perform simulations of resiliency against random and targeted attacks of the nodes with two different approaches. With this, we find that by identifying and studying the links between the nodes is possible to perform vulnerability analysis of the Estonian economy with respect to economic shocks.



## 1. Introduction

The network approach applied to financial and economic systems has potential to go further on the frontiers of research; there are two currents of origin: one comes from finances, economics and sociology, and the second one comes from computer science, big data challenges, physics, and complex evolving network studies [1]. Both converge in how node representation is done and how the relationships and interactions across the nodes form, whatsoever the nature of these links are. This is an intuitive path that starts to follow the approach that fusions economy and complex systems studies.

Nowadays, networks are a central concept and they can be: biological, technological, economic, social, cultural, among other types. The physical approach has made significant effort during the recent years around the study of evolution and structure of networks [2–9] while some other works have been dedicated to certain network phenomena and specific properties [10,11].

Since the structure of a network has direct influence on the vulnerability and dynamic behavior of the underlying system, important network properties such as stability and robustness can be understood by analyzing the clustering coefficient, the degree distribution and by determining the

* Corresponding author. Tel.: +372 59062059.
*E-mail address:* stretomx@gmail.com (S. Rendón de la Torre).

http://dx.doi.org/10.1016/j.chaos.2016.01.018

average shortest path length between nodes in the network [12,13].

In networks, the degree distribution *P(k)* is the probability that a node links to *k* number of nodes. Complex networks can be separated into two classes based on their degree distributions:

(1) Homogeneous networks are identified by degree distributions that follow an exponential decay. The distribution spikes at an average *k* and then decays exponentially for large values of *k*, such as the random graph model [14,15] and the small-world model [4], both leading to an homogeneous network: in which each node has approximately the same number of links *k* and a normal distribution where the majority of the nodes has an average number of connections, and only some or none of the nodes have only some or lots of connections.
(2) Heterogeneous large networks or scale-free networks, are those for which *P(k)* decays as a power law with a characteristic scale. The degree distribution follows a Pareto form of distribution where many nodes have few links and few nodes have many links, therefore, highly connected nodes are statistically significant in scale-free networks.

Network topology gives a fair basis for investigating money flows of customer driven banking transactions. A few recent papers describe the actual topologies observed in different financial systems [13,16–19]. Other works have focused on shocks and robustness in economic complex networks [20,21–24].

Scale-free networks display a strong tolerance against random removal of nodes [14] whereas exponential networks not (this means an exponential network can break easily into isolated clusters). Scale-free networks are more resistant to random disconnection of nodes because one can eliminate a considerable number of nodes randomly and the network's structure is preserved and will not break into disconnected clusters. However, the error tolerance is acquired at the expense of survival attack capability. When the most connected nodes are targeted, the diameter of a scale-free network increases and the network breaks into isolated clusters. This occurs because when removing these nodes, the damage disturbs the heart of the system, whereas a random attack is most likely not. One way to entangle the interaction of the nodes is by taking a look to the heavy tail effects they produce and see the implications on their robustness. Heavy-tailed distributions are strong against random perturbations but are extremely sensitive to targeted attacks.

Unlike previous studies, we illustrate the topology of an unstudied complex system that can be analyzed as a particular case of a complex network: Estonia's network of payments. We study the full country economic development, found on Swedbank's data as a proxy. The main goal of our analysis is to study the structure of this economic network. Additionally, this data set is unique given the fact that around 80% of Estonia's bank transactions are done through Swedbank, hence it is expected to reproduce fairly well the structure of the Estonian economy.

This paper is organized as follows. In Section 2 we provide the description of the selected data and the methods utilized; Section 3 is devoted to the discussion of the results and Section 4 concludes the study.

## 2. Materials and methods

### 2.1. Data

Payment events data from Swedbank AS were used to create the network. Data and information related to identities of the nodes will remain confidential and cannot be disclosed. We believe the utilized data describes fairly well the tendencies of money transactions and is the best possible information available.

The considered dataset corresponds to year 2014. We analyze the network of the payment flows of Swedbank (Estonia), specifically: domestic payments transferred electronically from customer to customer (legal entities). There are 16,613 nodes and 2,617,478 payment transactions in the network. There are 43,375 links if we count them as undirected.

A network (or a graph) is a set of nodes connected by links. The links are the connections between the nodes. In our network, the nodes are the companies and a link is established from one node to another if at least 20 payments were executed, or more than 1000 money units were paid/received per year. When there is a link from a node to itself, it is called a loop. We eliminated loops resulting from parties making money transfers across their own bank accounts.

There are several ways to define the network of payments; in this study we consider three definitions. The first definition is to look at the structure as a weighted graph where the links have certain weights associated to them representing less or higher important relationships with the nodes. Transactions between any two parties add to the associated link weights in terms of value of payments settled. In this representation we built a payment adjacency matrix that represents the whole image of the network and each element represents the overall money flow traded between companies *i* and *j*. This non-symmetric matrix represents the weights of the volumes of money exchanged between the companies.

The second definition is to consider an undirected graph, ignoring directions and weights of the payments and considering that two parties are connected if they share at least one payment, then $a^u_{ij} = a^u_{ji}$ and $a^u_{ij} = 1$ if there is a transaction between company *i* and *j* or $a^u_{ij} = 0$ if there is no transaction between them. Diagonal elements are equal to 0 and non-diagonal elements are either 0 or 1.

The links can also represent directions on the flow of the relationship. They could be directed or undirected. The third definition is a un-weighted-directed graph where the links follow the flow of money, such that a link is incoming to the receiver and outgoing from the sender of the payment. For this case we have two more matrices, one for the in-degree case and another one for the out-degree case. The choice of the definition of the matrix representation depends on the focus of the analysis.

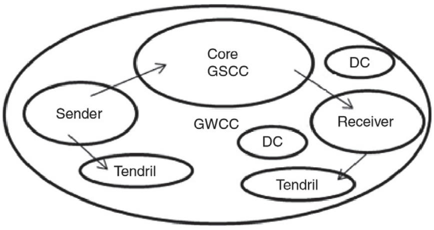



**Fig. 1.** Components of a directed network.

### 2.2. Components

Depending on how the nodes connect with each other, they can be partitioned into components. A component is a group of nodes such that any two nodes can be connected by a direct or indirect path. A path is a sequence of different nodes, each one connected to the next node. A component of an undirected network is a set of nodes such that for any pair of nodes $i$ and $j$ there is a path from $j$ to $i$; this means that two nodes share the same component if there is a path connecting them. Our analysis treats the network as both undirected and directed, and finds the components and their sizes.

In a directed network the largest component is known as the Giant Weakly Connected Component (GWCC) in which all nodes connect to each other via undirected paths. The core is the Giant Strongly Connected Component in which the nodes can reach each other through a directed path. The Giant Out-Component (GOUT) comprises the nodes that have a path from the GSCC and the Giant In-Component (GIN) comprises the nodes that have a path to the GSCC. The set of disconnected components (DC) are smaller components. Tendrils are nodes that have no directed path to or from the GSCC, but to GOUT and or the GIN [2]. These concepts are shown in Fig. 1.

In order to study the statistical properties and characterize the underlying structure of our network, we use specific useful network metrics [2,3,25].

## 3. Results

Fig. 2(b) displays a picture of the network as a weighted directed graph where each link is shaded by the corresponding weight: with darker shades indicating higher values on the cash flows. The bigger nodes represent those nodes with higher amount of values transferred. Fig. 2(a) shows the network as an undirected representation. This image includes 16,613 nodes and 43,374 links.

The high number of nodes and links makes difficult to have a detailed visualization of the graph's structure; therefore, we calculate topological and statistical measures that provide a clearer structure of the network.

### 3.1. Topology structure

We find all the components in the undirected graph. We obtained that the GCC is composed by 15,434 nodes which means that 92.8% of the nodes are reachable from one another by following either forward or backward links, suggesting it is a very well connected network. The remaining 7.2% nodes correspond to 508 DC. If we take a directed approach, the GSCC contains 24% of the nodes in the system.

A previous study of the structure of the WWW network components [25] focused on analyzing the robustness of the GCC against removal of nodes, and it was concluded that it is very difficult to destroy the structure of the WWW network by random elimination of links. (Table 3 displays the component sizes of the network of payments, among other statistics).

The degree of a node is defined as

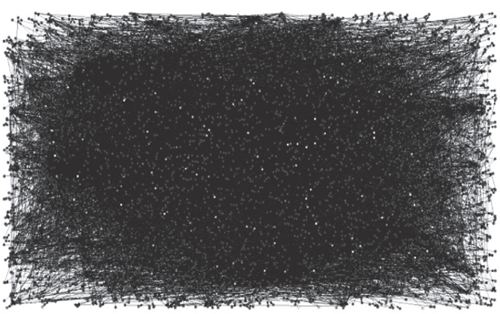
$$k_i = \sum_{j \in \zeta(i)} a_{ij} \ , \tag{1}$$

the sum goes over the set $\zeta(i)$ of neighbors of $i$. For example: $\zeta(i) = \{j | a_{ij} = 1\}$

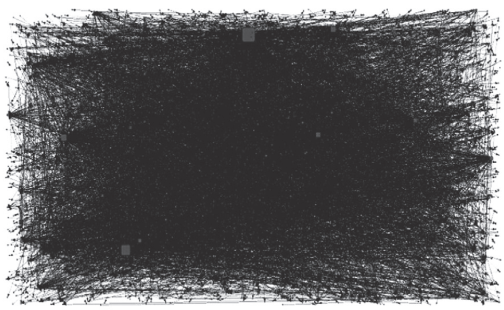
In a directed network there are two relevant characteristics of a node: the number of links that end at a node and the number of links that start from the node. These quantities are known as the out-degree $k^o$ and the in-degree $k^d$ of a node, and we define them as

$$k^d = \sum_{j \in \zeta(i)} a^d_{ij}, \ \ k^o = \sum_{j \in \zeta(i)} a^o_{ij}. \tag{2}$$

The average degree of a node in a network is the number of links divided by the number of nodes and is



(a) Connectivity network of payments.



(b) Weighted-directed representation of the network.

**Fig. 2.** (a) Connectivity network of payments. (b) Weighted-directed representation of the network.

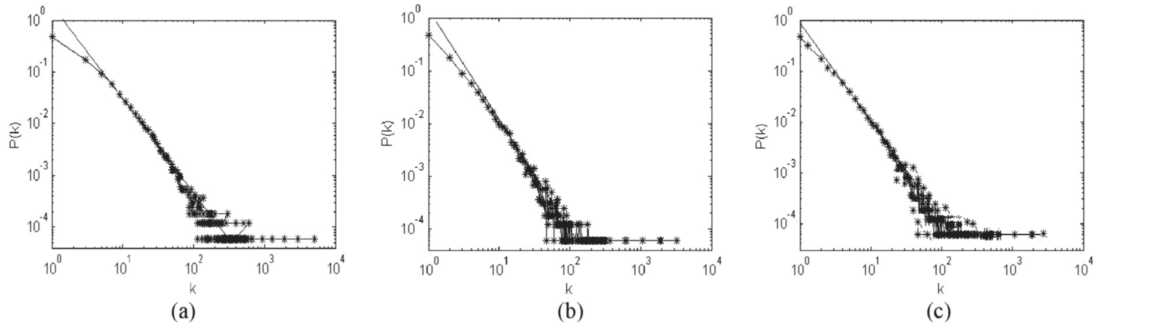



**Fig. 3.** *X*-axis corresponds to the number *k* of degrees and the *Y*-axis is $P(k)$. (a) Empirical degree distribution for the connectivity network. (b) Empirical in-degree distribution. (c) Empirical out-degree distribution. All the plots are log-log representations of histograms.

defined as

$$\langle k\rangle = \frac{1}{n}\sum k^{o} = \frac{1}{n}\sum k^{d} = \frac{m}{n}. \quad (3)$$

One can categorize networks by the degree distributions shown in the tails. In random networks it is very common to find Poisson distributed links, but in complex system networks it is common to find a distribution that follows a power law

$$P(k_i) \sim k_i^{-\gamma}, k \neq 0 \quad (4)$$

where $\gamma$ is the scaling exponent of the distribution. This distribution is called scale-free and networks with such a degree distribution are referred to as scale-free networks because have no natural scale and the distribution remains unchanged within a multiplicative factor under a rescaling of the random variables [26].

The average degree of our network is $\langle k\rangle = 20$. Most of the nodes have only 5 or less links, and 45% have only 1 link. Like other real networks, the degree distributions (undirected and directed) of the network of payments follow power laws. Fig. 3 displays the degree distributions. In all the distributions, we found regions that can be fitted by power laws, and this implies that the network has a scale-free structure. (We used the maximum likelihood estimation for obtaining the power law exponents [11]). The degree distribution in Fig. 3(a) follows a power law with a scaling exponent:

$$P(\geq k) \propto k^{-2.45}. \quad (5)$$

The in-degree distribution in Fig. 3(b) follows a power law defined as

$$P(k) \sim k^{-2.49}. \quad (6)$$

The out-degree distribution Fig. 3(c) follows a power law defined as

$$P(k) \sim k^{-2.39}. \quad (7)$$

In all the cases there is an area at the end of the tail that looks like a cut-off which can be explained by the fact that the system is finite and there is a maximum number of connections that a company could hold.

In a random network, the degree distributions follow a Poisson distribution. A degree distribution following a power law distribution appears to be a common feature in complex networks such as the World Wide Web, proteins interactions, phone calls and food webs, among others, but also shown in systems of payments of different banks [16–18]. The degree distributions obtained here are comparable to those obtained in the aforementioned studies. Table 2 includes a limited list of the power-law exponents obtained in different types of real networks.

**Table 1**
Network's characteristics.

| | |
|---|---|
| Companies analyzed | 16,613 |
| Total number of payments analyzed | 2,617,478 |
| Value of transactions | 3,803,462,026* |
| Average value of transaction per customer | 87,600* |
| Max value of a transaction | 121,533* |
| Min value of a transaction (aggregated in whole year) | 1000* |
| Average volume of transaction per company | 60 |
| Max volume of transaction per company | 24,859 |
| Min volume of transaction per company (aggregated in whole year) | 20 |

* All money amounts are expressed in monetary units.

*3.2. Weight, strength, size and diameter*

The basic properties of a network are the number of nodes *N* and the overall number of links *k* (Table 1 shows the general characteristics of our network). The number of nodes defines the size of the network while the number of links relative to the number of possible links defines the connectivity of a network.

Connectivity $(p)$ is the unconditional probability that two nodes are connected by a direct link. For a directed network, connectivity is defined as

$$(p) = \frac{k}{n(n-1)}. \quad (8)$$

In our case, the connectivity is 0.13 and this means the network is sparse because 87% of the potential connections are not used during the year.

The diameter is the maximum distance between two companies (measured by the number of links) and in our network this distance is equal to 29. This number is substantially higher when compared to the diameter of

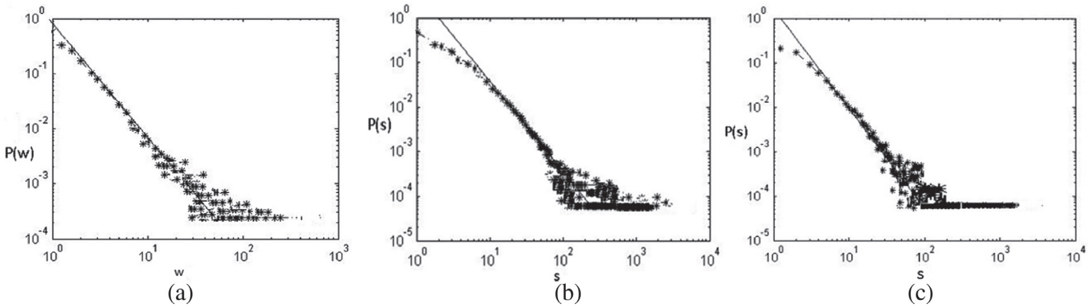



**Fig. 4.** (a) Link weight distribution by volume. (b) Node in-degree distribution by strength. (c) Node out-degree distribution by strength.

a random network of comparable characteristics (19) and this big difference in the diameter points to the existence of certain companies that send or receive money to other specific companies, and this contours specific and preferred paths for transactions. Intuitively this makes sense, because for companies in general, it is important to choose carefully their trading partners, clients, service providers or suppliers based on geographical location, affinity in the goals of the companies, cost policies, future joint ventures, agreements or any other reasons.

The strength of the nodes is the sum of the weights of all the links. In this case, the strength measures the overall transaction volume for any given node. The node-weighted strength is defined as

$$s_i = \sum_{j\in\zeta(i)} w_{ij}, \tag{9}$$

where $w_{ij}$ is the weight of the link between nodes $i$ and $j$ and the sum runs over the set $\zeta(i)$ of neighbors of $i$. The average strength can be calculated as a function of the $k$ number of links of a node to examine the bond between the strength and the degree.

Fig. 4(a) displays the distribution of link weights weighted by the number of payments transacted. This distribution follows a power law. The same power law relationship occurs between the strength and the degree of a node Fig. 4(b) and Fig. 4(c). These results were fitted by power laws with the following scaling exponents:

The volume link weight distribution

$$P(w) \sim w^{-1.98}, \tag{10}$$

where the scaling exponent is 1.98.

The volume out-degree strength distribution

$$P(s) \sim s^{-2.21}, \tag{11}$$

where the scaling exponent is 2.21.

The volume out-degree strength distribution

$$P(s) \sim s^{-2.32}, \tag{12}$$

where the scaling exponent is 2.32. There are some deviations from the power law behavior but they are sufficiently small.

### 3.3. Clustering, betweenness and average shortest path length

The clustering coefficient of a node is the tendency to cluster; is the density around node $i$. It represents the proportion of the closest nodes of a node which are linked to each other.

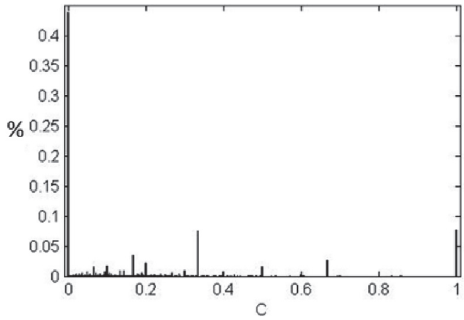



**Fig. 5.** Distribution of the clustering coefficient.

$$C(i) = \frac{1}{k_i(k_i-1)} \sum_{j\neq k} a_{ij}a_{jk}a_{ik}. \tag{13}$$

The overall clustering coefficient is the mean of the clustering coefficients $\langle C\rangle$ of all the nodes. It indicates if there is a link between two companies who have a common trading partner. In our case, the average clustering coefficient is 0.183, suggestive of cliquishness in our network. This means that two companies that are trading partners with a third one, have an average probability of 18.3% to be trading partners with one another, than will any two other companies randomly chosen. The clustering coefficient across nodes is highly spread, as seen in Fig. 5.

Fig. 5 shows that more than 52% of the nodes have a clustering coefficient of 0 or 1; therefore, the network is dispersed. There is ∼9% probability that two neighbors of a node are linked whereas around 45% are not linked at all. This high level percentage of unlinked neighbor nodes can be explained by the high number of nodes with degrees equal to 1 which are very frequent in scale-free networks.

Compared with other real networks this average clustering coefficient is low (See Table 2 for comparison). In this study, such a coefficient is fair. Business relationships between companies are commonly settled through medium or long term contracts. A company would like to remain doing business with the same parties (suppliers, clients, service providers, institutions, etcetera)

**Table 2**
Scaling exponents and clustering coefficients for different types of reported networks.

| Type | Network | Exponent | Clustering coefficient* | References |
|---|---|---|---|---|
| Economical | Bank of Japan payments | $\gamma=2.1$ | – | [16] |
| | US Federal Reserve Bank | $\gamma^i=2.11$ | 0.53 | [17] |
| | | $\gamma^o=2.15$ | | |
| | Austrian Interbank Market payments | $\gamma^i=1.7$ | 0.12 | [18] |
| | | $\gamma^o=3.1$ | | |
| Technological | WWW | $\gamma^o=2.4$ | – | [36] |
| | | $\gamma^i=2.1$ | | |
| | Peer-to-peer network | $\gamma=2.1$ | 0.012 | [37] |
| | Digital electronic circuits | $\gamma=3$ | 0.03 | [38] |
| Social | Film actors | $\gamma=2.3$ | 0.78 | [4] |
| | Email messages | $\gamma^i=1.5$ | 0.16 | [39] |
| | | $\gamma^o=2.0$ | | |
| | Telephone call | $\gamma=2.1$ | – | [40] |
| Biological | Protein interactions (yeast) | $\gamma=2.4$ | 0.022 | [41] |
| | Metabolism reactions | $\gamma^i=2.2$ | | |
| | | $\gamma^o=2.2$ | 0.32 | [42] |
| | Energy lancscape for a 14-atom cluster | $\gamma=2.78$ | 0.073 | [43] |

$\gamma^i$ = scaling exponent for in-degree distribution. $\gamma^o$ = scaling exponent for the out-degree distribution. $\gamma$ = scaling exponent for the connectivity distribution.
* Refers to average clustering coefficient.

**Table 3**
Summary of statistics.

| Statistic | Value | *Components* | # of nodes |
|---|---|---|---|
| $N$ | 16,613 | GCC | 15,434 |
| # Payment | 2,617,478 | DC | 1179 |
| Undirected Links | 43,375 | GSCC | 3987 |
| $< k >$ | 20 | GOUT | 6054 |
| $\gamma^o$ | 2.39 | GIN | 6172 |
| $\gamma^i$ | 2.49 | Tendrils | 400 |
| $\gamma$ | 2.45 | Cutpoints | 1401 |
| $< C >$ | 0.183 | Bi-component | 4404 |
| $< l >$ | 7.1 | *k*-core | 1081 |
| $C$ | 0.13 | | |
| $D$ | 29 | | |
| $< \sigma >$ (nodes) | 110 | | |
| $< \sigma >$ (links) | 40 | | |

$N$ = number of nodes. $<k>$ = average degree. $\gamma^o$ = scaling exponent for the out-degree empirical distribution. $\gamma^i$ = scaling exponent for the in-degree empirical distribution. $\gamma$ = scaling exponent for the connectivity degree distribution. $<C>$ = average clustering coefficient. $< l >$ = average shortest path length. $C$ = connectivity per cent. $D$ = Diameter. $< \sigma >$ = average betweenness.

because in general, it is easier and cheaper than changing them time after time. A change on a trading party could mean a decrease on profits or an increase on costs. A low clustering coefficient in our payment network reflects this perspective.

In our case, the clustering coefficient is higher than the connectivity, therefore, the network is not random (in a random network the clustering coefficient is equal to the connectivity; a random network is built by randomly adding links to a given set of nodes). A random network of a comparable size has a clustering coefficient around 70 times lower than our network.

Betweenness $\sigma(m)$ of a node $m$ is the total number of shortest paths between all possible pairs of nodes that pass through this node; it is a measure of the number of paths between other nodes that run through the node $i$; the more paths this node has, the more central is the node $i$ in the network. It indicates whether or not a node is important in the traffic of the network.

$$\sigma(m) \equiv \sum_{i \neq j} \frac{B(i, m, j)}{B(i, j)}, \tag{14}$$

where $B(i, j)$ is the total number of shortest paths between nodes $i$ and $j$ and the sum goes over all the pairs of nodes for which at least one path exists, with $B(i, j) > 0$. The nodes with high betweenness control the network.

The results in Table 2 show that the average betweenness for the links is 40 and for the nodes is 110, meaning that each company handles in average 110 shortest paths, and the higher is the number of shortest paths the more central the company is for the network.

The average shortest path length $< l >$ was calculated with Dijkstra's algorithm [27]. In the connectivity network this value is equal to 7.1. The network is a small world with 7.1 degrees of separation, so in average any company can be reached by other one only in a few steps. Our network has low connectivity but it is densely connected. This characteristic is in line with the fact that there are companies that act as hubs or key nodes and lead to short distances between the other companies.

93% of the nodes are within 7 links of distance from each other and this suggests that the network of payments is comprised of a core of nodes with whom the other companies interact with. There is a smaller group of 1081 nodes (6.5% of the total number of nodes in the network) connected by high value links. This group contains weighted links that comprise 75% of the overall value of the funds transferred. Fig. 6 shows the graph of the k-core. A k-core in an undirected graph is a connected maximal induced sub-graph which has minimum degree greater than or equal to $k$. Alternatively, the k-core is the (unique) result of iteratively deleting nodes that have degree less than $k$, in any order.

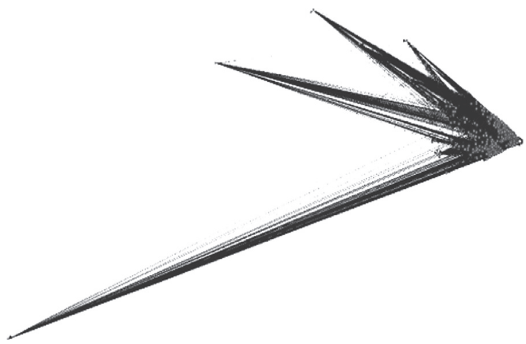


**Fig. 6.** Graph representation of the *k*-core.

### 3.4. Robustness simulation and degree correlations

One of the characteristics that makes a hub or a key node an important node is its high betweenness not just its high degree. Hubs often connect groups of clusters of sub-areas of the graph that are not connected to one another directly. These nodes are important because they shorten path lengths making for high reachability and fast movement of information. But they may also be important as brokers and key-players that connect the graph because of their betweenness [28].

In order to gain more understanding on how the network is likely to behave as a whole, let us address the question: if a node were removed, would the structure of the network become divided into disconnected clusters? One can consider several approaches to find the key nodes in the network which may act as enablers among otherwise disconnected groups and we find the nodes that connect the network by locating the vulnerable parts (see Hanneman and Riddle [29]).

A total of 1401 cut-points or key nodes were found, this means that around 8% of the nodes are relevant to keep the structure connected as it is, or in other terms, if we remove these nodes then the number of components and the average path lengths between the nodes would increase, leaving the network vulnerable to break.

We run a simulation for the GCC that shows random removal of a fraction of nodes and another simulation considering strategically chosen nodes. Then we measure the average shortest path length $< l >$ and the relative size of the GCC as functions of the percentage $d$ of deleted nodes [2,30,31]. The results are displayed in Fig. 7. The effect of the targeted removal of nodes causes a quick growth in the average shortest path length until the GCC disappears, $GCC(p_c) = 0$ at a very low level of targeted damage (less than 10%). We will call this level the percolation threshold $p_c$. It is noticeable that a weak but smart attack destroys the network. In the random removal of nodes the damage is less than in the targeted damage. We established that our network of payments has shown scale-free properties, and this kind of networks are resilient to random damage, so it is barely possible to destroy the network of payments by random removal, but if we remove the exact portion of particularly selected nodes, it breaks completely. This effect has been seen in financial systems in economic crisis before: companies or banks may declare in bankruptcy and the whole system stays healthy, but if certain organizations declare in bankruptcy then the whole system collapses.

It is not rare that the GCC in heavy-tailed networks is resilient against random removal of nodes. If the degree distribution of the network is fat-tailed, then this fact determines the topology of the system. However, it might be possible that when removing nodes in a random way, the tail of the degree distribution changes and then the GCC structure would be damaged [2].

There are other heuristic methods available in literature to calculate the optimal percolation threshold of nodes that breaks the network into disconnected clusters (such as high degree node, *k*-core, closeness and eigenvector centralities). However, a common characteristic in these approaches is that they do not necessarily optimize a measure that reflects the collective influence arising from considering the entire influential nodes at once. Under a collective influence approach, nodes' inherent strength and weakness arises collectively from the configuration of interactions they have with the other components.

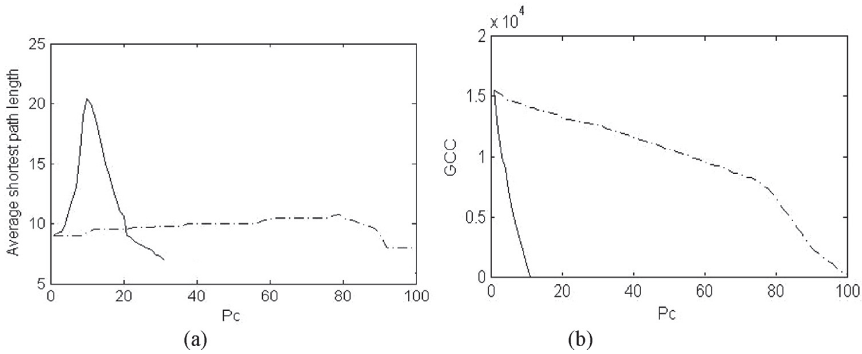



**Fig. 7.** Plots of the targeted and random damage over the network of payments. (a) The average shortest-path length $< l >$ in the GCC plotted against the percentage of removed nodes. (b) The GCC plotted against the percentage of removed nodes. Continuous lines display the effect of the targeted removal and the dashed lines display the effect of the random removal of nodes. $p_c$ are the percolation thresholds, for each case.

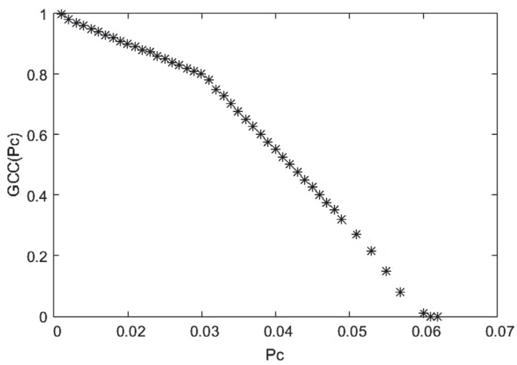



**Fig. 8.** GCC of the network of payments as a function of the percolation threshold $p_c$ by using the collective influence algorithm.

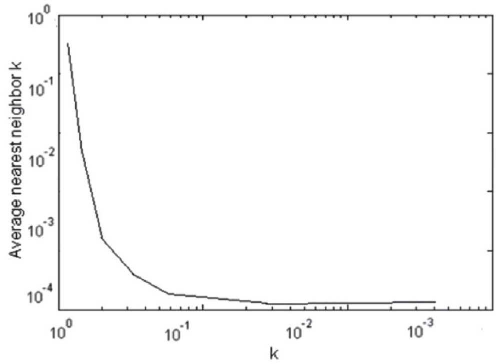



**Fig. 9.** Affinity of the connectivity network.

Morone and Makse [32] designed an algorithm suitable for large networks that has proven to perform better than other empirical methods because it predicts a smaller set of optimal influencer nodes (the nodes that destroy the network if removed).

The collective influence of a node is defined as the product of the node's reduced degree (the number of its nearest connections, $k_i$, minus one), and the total reduced degree of all nodes $k_j$ at a distance $\ell$ from it ($\ell$ is defined as the shortest path)

$$CI_\ell(i) = (k_i - 1) \sum_{j \in \partial \mathrm{Ball}(i,\ell)} (k_j - 1), \tag{15}$$

where Ball($i$, $\ell$) is the set of nodes inside a ball of radius $\ell$ around node $i$. $\partial$Ball($i$, $\ell$) is the frontier of the ball and comprises the nodes $j$ that are at a distance $\ell$ from $i$. By computing the *CI* for each node, it is possible to find the ones with the highest collective influence and remove them.

The collective influence algorithm maps the problem of optimal influence on the computation of the minimum structural amount of nodes that reduces the largest eigenvalue of the non-backtracking matrix of the network (see Morone and Makse [32]).

We performed a simulation using the *CI* approach, and the results are shown in Fig. 8. We measure the collective influence of a group of nodes as the fall in the size of the GCC which would occur if the nodes in question were eliminated. The figure shows the GCC when a fraction of its nodes is eliminated. The optimal percolation threshold occurs when removing around 6% of the nodes because that is the point where $GCC(p_c) = 0$. This result also implies that there are a huge number of companies with a large number of payments which in fact have a minor influence in the whole economic network.

Fig. 8 shows a better performance than previous strategy used on Fig. 7 (which is based on a betweenness centrality and where the optimal percolation threshold is higher than in collective influence method).

A practical measure of correlations is the average nearest neighbor degree function. A network is called assortative if its nodes with a certain degree are more likely to have links with nodes of similar degree, and it is called disassortative when the contrary occurs. For example, when low degree nodes are more likely connected with nodes of higher degrees, or when high degree nodes are more likely connected with low degree nodes.

A method that calculates these aforementioned correlation measures is the average nearest neighbor degree function (see Serrano et al. [33,10]). The conditional probability of a node with degree $k$ to be connected to a node of degree $k'$ is defined as

$$P\left(k'|k\right) = \frac{P\left(k', k\right)}{P(k)}, \tag{16}$$

where $P(k',k)$ is the probability of two nodes, with degree $k'$ and $k$ to be connected by a link. $P(k)$ is the degree distribution. The average nearest neighbor degree function is defined as

$$\langle k_{nn} \rangle (k) = \sum_{k'} k' P(k'|k) \tag{17}$$

Previous studies [2] have shown that social networks usually have significant assortative mixing; biological, technological and other financial networks have shown disassortative mixing [17,24]. Fig. 9 shows the affinity of the connectivity network. The correlation is –0.18, which means there is a negative dependence between the degree of a node and the degrees of its neighbors; therefore the system exhibits disassortative mixing. The function $\langle k_{nn} \rangle(k)$ decreases with $k$ suggesting that the high degree nodes, which are represented by companies who have many business partners such as service providers, clients or suppliers, usually have a large number of links to companies which have only one link (or just few), then the high degree nodes tend to connect with the low degree ones.

Disassortative mixing has implications for network resilience. For example, when this type of mixing is found, the attacks to the highest degree nodes are effective when trying to destroy the network quickly because these nodes are being approximately distributed over the network and forming links on different paths between other nodes, hence, this characteristic makes our network particularly vulnerable to targeted attacks.

## 4. Conclusion

We studied the structure of the economic network of an entire country using Swedbank's payments database. After extracting the network's topology, characteristics and statistics we conclude that this economic network shares many of the features found in empirical complex networks, such as scale-free degree distributions, small world characteristic and low clustering coefficient.

Our results show that this economic network is disassortative in terms of degree. The system shows topological heterogeneity due to its heavy tails and scale-free structure in the degree distributions. This scale-free structure indicates that few companies in Estonia trade with many parties while the majority trade with only few.

In our network, the clustering coefficient is low and disperse (more than 52% of the nodes have either a clustering coefficient of 0 or 1). A low coefficient is a fair result because it shows how companies perceive business partners change as an avoidable expense. A company might prefer to keep working with regular trading partners (for example: service providers, clients or suppliers) for a medium or long term instead of changing them often, in order to save money and time.

The network is a small world with just 7 degrees of separation: in average any company can be reached by other only in a few steps. The connectivity is smaller than the overall clustering coefficient; therefore, the network cannot be classified as random.

Regarding the diameter size of our network: it is high when compared with that of a random network (1.5 times higher). The diameter in our results suggests a preference for specific paths of money flows between companies. This preference refers to companies that trade more with specific parties over others based on decisions relative to costs saving, geographical location, convenience, or any other type of decision.

We performed two separate analyses to reveal the robustness of our economic network. The first one is based on centralities and the second one is based on an approach focused on collective influencer nodes. First, we found the key nodes that prevent the network to break into disconnected components. The simulation for the GCC assuming a targeted removal of key nodes causes a quick growth in the average shortest path length until the GCC disappears at a percolation threshold of 8%, while in the random removal the damage is extremely small. This revealed the robustness of our economic network against random attacks but also revealed its vulnerability to smart attacks.

In the second analysis we followed the collective influence strategy. The percolation threshold is close to 6%; therefore, the performance of the optimal percolation threshold is better when following this method because it reduces the percentage substantially. The interpretation for this low level of optimal percolation threshold is that there are a lot of companies with enormous amounts of payments that have a weak influence in the economic network as a whole, and this also reveals that the most influential companies in the network are not necessarily the most connected ones or those having more economic activity. Both results agree on the fact that a small portion of economic entities maintains the whole network unified.

## Acknowledgments

Algorithms in UCINET [34] and Pajek [35] were used to calculate network statistics and images. We thank Swedbank AS for enabling us with the data set for the analysis. This research was supported by the European Union through the European Regional Development Fund (project TK 124).

## References

[1] Schweitzer F, Fagiolo G, Sornette D, Vega-Redondo F, White DR. Economic networks: what do we know and what do we need to know? Adv Comp Syst 2009;12:407–22. doi:10.1142/S0219525909002337.
[2] Dorogovtsev SN, Mendes JFF. Evolution of networks. New York: Oxford University Press; 2003.
[3] Albert R, Barabási AL. Statistical mechanics of complex networks. Rev Mod Phys 2002;74:47–97. doi:10.1103/RevModPhys.74.47.
[4] Watts DJ, Strogatz SH. Collective dynamics of small-world networks. Nature 1998;393:440–2. doi:10.1038/30918.
[5] Watts DJ. Six degrees: the science of a connected age. New York: WW Norton & Company; 2003.
[6] Newman MEJ. Assortative mixing in networks. Phys Rev Lett 2002;89:208701. doi:10.1103/PhysRevLett.89.208701.
[7] Barrat A, Barthélemy M, Vespigniani A. Dynamical processes on complex networks. Cambridge, UK: Cambridge University Press; 2008.
[8] Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, et al. Graph structure in the web. Comput Netw 2000;33:309–20. doi:10.1016/S1389-1286(00)00083-9.
[9] Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the internet topology. Comput Commun Rev 1999;29:251–62. doi:10.1145/316194.316229.
[10] Catanzaro M, Boguña M, Pastor-Satorras R. Generation of uncorrelated random scale-free networks. Phys Rev E 2005;71(2):027103. doi:10.1103/PhysRevE.71.027103.
[11] Newman MEJ. Power laws, Pareto distributions and Zipf's law. Contemp Phys 2005;46(5):323–51. doi:10.1016/j.cities.2012.03.001.
[12] Barabási AL, Albert R. Emergence of scaling in random networks. Science 1999;286:509–12. doi:10.1126/science.286.5439.509.
[13] Lublóy A. Topology of the Hungarian large-value transfer system. Magyar Nemzti Bank (Central Bank of Hungary) MNB Occasional Papers 2006;57.
[14] Barabási AL, Albert R, Jeong H. Mean-field theory of scale-free random networks. Physica A 1999;272:173–87. doi:10.1016/s0378-4371(99)00291-S.
[15] Erdös P, Rényi A. On random graphs, 6. Publicationes Mathematicae Debrecen; 1959. p. 290–7.
[16] Inaoka H, Nimoniya T, Taniguchi K, Shimizu T, Takayasu H. Fractal network derived from banking transactions – An analysis of network structures formed by financial institutions. Bank of Japan; 2004. 04-E-04, Working papers.
[17] Soramäki K, Bech ML, Arnold J, Glass RJ, Beyeler WE. The topology of interbank payment flows. Physica A 2007;379(1):317–33. doi:10.1016/j.physa.2006.11.09.
[18] Boss M, Helsinger H, Summer M, Thurner S. The network topology of the Interbank Market. Quant Financ 2004;4(6):677–84.
[19] Rordam KB, Bech ML. The topology of Danish Interbank money flows. Finance Research Unit; 2008. Working paper 59.
[20] Deschatres F, Sornette D. The dynamics of book sales: endogenous versus exogenous shocks in complex networks. Phys Rev E 2005;72(1):016112. doi:10.1103/PhysRevE.72.016112.
[21] Iori G, Jafarey S. Criticality in a model of banking crisis. Physica A 2001;299(1):205–12. doi:10.1016/S0378-4371(01)00297-7.
[22] Crucitti P, Latora V, Marchiori M, Rapisarda A. Error and attack tolerance of complex networks. Physica A 2005;340(1):388–94. doi:10.1371/journal.pone.0059613.
[23] Kitt R. Economic decision making: application of the theory of complex systems. In: Banerjee S, Ercetin SS, Tekin A, editors. Chaos theory in politics. Dordrecht: Springer; 2014. Ch. Part II. Politics, Complex Systems, Basin of Attractions p. 51–73. doi:10.1007/978-94-017-8691-1_4.

[24] Iori G, De Masi G, Precup OV, Gabbi G, Caldarelli G. A network analysis of the Italian overnight money market. J Econ Dyn Control 2007;32(1):259–78. doi:10.1016/j.jedc.2007.01.032.
[25] Adamic LA, Huberman BA, Barabási AL, Albert R, Jeong H, Bianconi G. Power-law distribution of the World Wide Web. Science 2000;287:2115a. doi:10.1126/science.287.5461.2115a.
[26] Newman MEJ. The structure and function of complex networks. SIAM Rev 2003;45(2):167–256. doi:10.1137/S003614450342480.
[27] Gibbons A. Algorithmic graph theory. Cambridge, UK: Cambridge University Press; 1985.
[28] Borgatti SP, Everett MG, Johnson JC. Analyzing social networks. London, UK: SAGE Publications Limited; 2013.
[29] Hanneman RA, Riddle M. Introduction to social network methods. Riverside, CA: University of California Riverside; 2005.
[30] Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. Nature 1999;406(6794):378–82. doi:10.1038/35019019.
[31] Cohen K, Erez D, Avraham D, Havlin S. Resilience of the internet to random breakdowns. Phys Rev Lett 2000;85:4625–30. doi:10.1103/PhysRevLett.86.3682.
[32] Morone F, Makse HA. Influence maximization in complex networks through optimal percolation. Nature 2016;524:65–8. doi:10.1038/nature14604.
[33] Serrano M.A., Maguitman M.A., Bogunan A., Fortunato M., Vespignani A. Decoding the structure of the WWW: Facts versus sampling biases. ArXiv:cond-mat, 0511035, 2005.
[34] Borgatti SP, Everett MG, Freeman LC. Ucinet 6 for Windows: software for social network analysis. Harvard, MA: Analytic Technologies; 2002.
[35] Batagelj V, Mrvar A. Exploratory social network analysis with Pajek. Cambridge, UK: Cambridge University Press; 2005.
[36] Albert R, Jeong H, Barabási AL. Diameter of the World Wide Web. Nature 1999;401:130–1. doi:10.1038/43601.
[37] Ripeanu M, Foster I, Iamnitchi I. Mapping the Gnutella network: properties of large-scale peer-to-peer systems and implications for system design. IEEE Internet Comput 2002;6(1):50–7.
[38] Ferrer R, Cancho RV, Solé RV. Optimization in complex networks. Statistical mechanics of complex networks. Lecture Notes in Physics 2003(625):114–25. doi:10.1007/b12331.
[39] Ebel H, Mielsch LI, Bornholdt S. Scale-free topology of e-mail networks. Phys Rev E 2002;66:35103(R). doi:10.1103/PhysRevE.66.035103.
[40] Aiello W, Chung F, Lu L. A random graph model for massive graphs. In: Proceedings of the thirty-second annual ACM symposium on theory of computing; 2000. p. 171–80. doi:10.1145/335305.335326.
[41] Jeong H, Mason SP, Barabási AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–2. doi:10.1038/35075138.
[42] Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. Nature 2000;407:651–4.
[43] Doye JPK. The network topology of a potential energy landscape: A static scale-free network. Phys Rev Lett 2002;88:238701. doi:10.1103/PhysRevLett.88.238701.

**Paper II**

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2017). Fractal and multifractal analysis of complex networks: Estonian network of payments. The European Physical Journal B, 90. DOI: 10.1140/epjb/e2017-80214-5.

Eur. Phys. J. B (2017) 90: 234
DOI: 10.1140/epjb/e2017-80214-5

THE EUROPEAN PHYSICAL JOURNAL B

Regular Article

# Fractal and multifractal analysis of complex networks: Estonian network of payments

Stephanie Rendón de la Torre[1,a], Jaan Kalda[1], Robert Kitt[1,2], and Jüri Engelbrecht[1]

[1] School of Science, Department of Cybernetics, Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia
[2] Swedbank AS, Liivalaia 12, 15038 Tallinn, Estonia

Received 11 April 2017 / Received in final form 11 June 2017
Published online 4 December 2017 – 

**Abstract.** Complex networks have gained much attention from different areas of knowledge in recent years. Particularly, the structures and dynamics of such systems have attracted considerable interest. Complex networks may have characteristics of multifractality. In this study, we analyze fractal and multifractal properties of a novel network: the large scale economic network of payments of Estonia, where companies are represented by nodes and the payments done between companies are represented by links. We present a fractal scaling analysis and examine the multifractal behavior of this network by using a sandbox algorithm. Our results indicate the existence of multifractality in this network and consequently, the existence of multifractality in the Estonian economy. To the best of our knowledge, this is the first study that analyzes multifractality of a complex network of payments.

## 1 Introduction

In recent years, complex networks have been studied extensively and have attracted much attention from researchers belonging to different fields of knowledge and science. Complex networks theory is developing at a fast pace and has already made significant progress toward designing the framework for unraveling the organizing principles that govern complex networks and their evolution. In fact, several topological characteristics and a variety of dynamical aspects of complex networks have been the center of extensive research and studies in the last years.

A fractal is a quantity or a fragmented geometric object which can be split into parts, each of which is a reduced-size copy of the whole and has the same statistical character as the whole. A fractal displays self-similarity on all scales. Fractals have infinitely complex patterns that are self-similar across different scales; these objects do not need to display exactly the same structures at all scales, but the same "type" of structures must appear on all scales. Fractals can be created by repeating simple recursive processes. Another definition of a fractal states that is a set whose Hausdorff-Besicovitch dimension strictly exceeds its topological dimension. The "fractal geometry" of nature was first labeled as a term by Benoit Mandelbrot in the late 60s, and after that, the fractal approach has been widely used to gain insight into the fundamental scaling of numerous complex structures. Fractal analysis helps to distinguish global features of complex networks, such as the fractal dimension. However, the fractal formalism is insufficient to characterize the complexity of many real networks which cannot simply be described by a single fractal dimension. Furuya and Yakubo [1] demonstrated analytically and numerically that fractal scale-free networks may have multifractal structures in which the fractal dimension is not sufficient to describe the multiple fractal patterns of such networks, therefore, multifractal analysis rises as a natural step after fractal analysis.

Multifractal structures are abundant in social systems and are plentiful in a variety of physical phenomena. Multifractal analysis is a systematic approach and a generalization of fractal analysis that is useful when describing spatial heterogeneity of fractal patterns [2]. It has proven to be a useful tool for studying turbulence phenomena [3,4], time series analysis [5,6], economic and financial modeling [7], medical pattern recognition [8], biological and geophysical systems [9–18].

Fractal and multifractal analysis can help to uncover the structure of all kinds of systems in order to have a better understanding of such systems. In particular, both approaches have many different interesting applications in economy. An interesting line of research is related with the relevance and applicability of fractal and multifractal analysis in social and economic topics. Inaoka et al. [19] showed that the study of the structure of a banking network provides useful insight from practical points of view. By knowing and understanding the network structure and characteristics of banking networks (in terms of transactions and their patterns), a systemic contagion could potentially be prevented. In their study, these authors showed that the network of financial transactions of Japanese financial institutions has a fractal structure. Regarding social studies, Lu and Tang [20] showed

[a] e-mail: stretomx@gmail.com

the importance of road patterns for urban transportation capacity based on fractal analysis of such network. In this study, the authors were able to link the fractal measurement with city mass measurements.

Another direction of these studies has tilted toward the development of multifractal models for financial networks [21]. A few recent studies have focused on the analysis of the changes of multifractal spectra across time to assess changes in economy during crisis periods [22,23]. Some other studies have focused on gathering empirical evidence of the common multifractal signature in economic, biological and physical systems [24].

In the last years, numerous algorithms for calculating the fractal dimension and studying self-similar properties of complex networks have been developed and tested extensively [25–29]. Song et al. [30] developed a method for calculating the fractal dimension of a complex network by using a box-covering algorithm and identified self-similarity as a property of complex networks [31]. Moreover, a myriad of algorithms and studies on networks' multifractal analysis have been proposed and developed lately [32–37].

The main objective of this study is to analyze the fractality and multifractality of a novel and unstudied network: the large scale Estonian network of payments. We present a study that contributes to the field of complex networks (particularly to economic complex networks studies) by adding empirical evidence in favor of fractalism and multifractalism with a new case of study. The study is done thanks to the application of known network methods. Also, the goal is to expand the knowledge of the structure of this network of payments by analyzing its fractal and multifractal structures anticipating that this analysis could be useful in the future for further studies. Multifractal analysis of a payment network could be the starting point for developing economical and financial future studies related with: opportune detection of key factors driving the multifractal spectrum changes across time, money-flows forecasting and risk-pattern recognition (during turbulent financial times, for example). To the extent of our knowledge, this is the first study that examines multifractality of a complex network of payments.

We present a fractal scaling analysis by calculating the fractal dimension of our network and its skeleton. Then, we use a sandbox algorithm for complex networks [33] to calculate the spectrum of the generalized fractal dimensions $D(q)$ and mass exponents $\tau(q)$ in order to study the multifractal behavior of the network. Section 1 presents an introduction and an overview of literature related with fractal and multifractal network studies. Section 2 is devoted to a detailed description of the data set and the methods used. Section 3 presents our main results and Section 4 concludes with a discussion of the results.

## 2 Data and methodology

In this section we describe the nature and the scope of our data set. Then, we introduce the box-covering algorithm used to calculate the fractal dimension of our network and its skeleton. To conclude, we introduce the sandbox algorithm used for multifractal analysis in this study.

### 2.1 Data

To create the network of payments we used payment transactions data from Swedbank. At present, Swedbank is one of the leading banks in the Nordic and Baltic regions of Europe and operates in the following countries: Latvia, Lithuania, Estonia and Sweden. The data and all the information related with the identities of the nodes are very sensitive and will remain confidential. Our data set is unique and very interesting because ~80% of Estonia's electronic bank transactions are executed through Swedbank's system of payments, hence, this data set reproduces fairly well the trends of money of the whole Estonian economy. Our Estonian network of payments focuses exclusively on domestic payments transferred electronically from customer to customer (company-to-company) during the year 2014. There are 16 613 nodes, 2 617 478 payments and 43 375 undirected links in the selected data set.

A network is a set of nodes connected by links. In this study, the nodes represent companies and the links represent the payments between the companies. We mapped a symmetric payments adjacency matrix $A_{N\times N}$ where $N$ is the total number of nodes in the network. The payments adjacency matrix $A_{N\times N}$ represents the whole image of the network. For simplicity, we considered an undirected graph approach where two nodes have a link if they share one or more payments, then each element represents a link if there is a transaction between companies $i$ and $j$ as follows: $a^u_{ij} = a^u_{ji}$ and $a^u_{ij} = 1$; otherwise, $a^u_{ij} = 0$ if there is no transaction between companies $i$ and $j$.

### 2.2 Fractal scaling analysis

Fractal analysis assists on the calculation and the understanding of the fractal dimension of complex networks. In general, fractal analysis consists of several methods to measure complexity by using the fractal dimension and other fractal characteristics. According to Song et al. [31] complex networks may have self-similar structures. According to these authors, the box-counting algorithm is an appropriate method to examine global properties of complex networks. The fundamental relation of fractal scaling is based on the box-covering method which counts the total number of boxes that are needed to cover a network with boxes of certain size. The box-covering method is analogous to the box-counting method widely used in fractal geometry and is a basic tool to measure the fractal dimension of fractal objects embedded in Euclidean space [38]. However, the Euclidean metric is not well defined for networks, thus we use the adaptation for networks developed by Wang et al. [35] of the random sequential box-covering algorithm created by Kim et al. [39], to determine the fractal dimension of our network and its skeleton. The aforementioned method contains a random process for selecting the position of the center of each box. We let $N_B(r_B)$ be the minimum number of boxes needed to tile the whole network, where the lateral size of the boxes is the measure of radius $r_B$ as follows:

$$N_B(r_B) \sim r_B^{-d_B}, \quad (1)$$

where $d_B$ is the fractal dimension. If we measure the number $N_B$ for different box sizes, then it is possible to obtain the fractal dimension $d_B$ by obtaining the power law fitting of the distribution. The algorithm selects a random node at each step, and this node is the seed that will be the center of a box. Then, we search the network by distance $r_B$ from the seed node and cover all the nodes that are located within that distance, but only if they have not been covered yet. Then, we assign the newly covered nodes to the new box; if there are no more newly covered nodes then the box is removed. This process is repeated until all the nodes of the network belong to boxes.

Before using the algorithm, we calculate the skeleton of the network. The concept of skeleton was first introduced by Kim et al. [40]. The skeleton is a particular type of spanning tree based on the link betweenness centrality (a simplified quantity to measure the traffic of networks) that is entrenched beneath the original network. The skeleton delivers a shell for the fractality of the network and is formed by links with the highest betweenness centralities. Only the links that do not form loops are included. The remaining links from the original network which are not included in the skeleton are local shortcuts that contribute to loop formation, meaning that the distance between any two nodes in the original network may increase in the skeleton. A fractal network has a fractal skeleton beneath it which is distressed by these local shortcuts, but preserving its fractality. For a scale-free network, its skeleton also follows a power-law degree distribution where its degree exponent might differ slightly from that of the original network. When studying the origin of fractality in networks, actually the skeleton is more useful than the original network itself due to its unsophisticated tree structure [41]. In general, the skeleton preferentially collects the sections of the network where betweenness is high, and this preserves the structure and simplifies its complexity. Consequently, by looking at the properties of the skeleton it is easier to appreciate the topological organization of the original network.

To calculate the skeleton of a complex network, the link betweenness of all the links in the network has to be calculated. The betweenness centrality of a network (for a link or a node), is defined as follows:

$$b_i = \sum_{j,k \in N, j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \quad (2)$$

where $N$ is the total number of nodes, $n_{jk}$ is the total number of shortest-paths between nodes $j$ and $k$, $n_{jk}(i)$ is the total number of shortest-paths linking nodes $j$ and $k$ that passes through the node $i$.

In order to perform the fractal scaling analysis, we used Dijkstra's algorithm [42]; then we used the box-covering algorithm to calculate the fractal dimension of the network and the skeleton to compare both values.

### 2.3 Sandbox algorithm for multifractal analysis of complex networks

Scale-free networks are commonly observed in a wide array of different contexts of nature and society. Previous studies [43,44] have shown that in scale-free networks, independently of the system and the identity of their components, the probability $P(k)$ that a node in the network interacts with $k$ other links decays as a power-law, following that $P(k) \sim k^{-\gamma}$, $k \neq 0$; this points to a tendency for large networks to self-organize into a scale-free state. We found scale-free properties characterized by power-law degree distributions in our previous study on the Estonian network of payments [45] (see Tabs. 1 and 2 for details of the main features and statistics of this network).

In general, multifractality is expected to appear in scale-free networks due to the fluctuations that occur in the density of local nodes. Multifractal analysis requires taking into account a physical measure, like the number of nodes within a box of specific size in order to analyze how the distribution of such number of nodes scales in a network as the size of the box grows. Tél et al. [46] introduced a sandbox algorithm based on the fixed-size box-counting algorithm [47] which was used and adapted for multifractal analysis of complex networks by Liu et al. [33]. In order to determine the multifractal dimensions of our complex network, we chose this adapted sandbox algorithm because it is precise, efficient and practical. Moreover, a study by Song et al. [2] has shown that this algorithm gives better results when it is used in unweighted networks, and this is our case.

The fixed-size box-counting algorithm is one of the most known and efficient algorithms for multifractal analysis. For a given probability measure $0 \leq \mu \leq 1$ in a metric space $\Omega$ with a support set $E$, we consider the following partition sum:

$$Z_\varepsilon(q) = \sum_{\mu(B) \neq 0} [\mu(B)]^q, \quad (3)$$

where the parameter $q \in R$, and it describes the moment of the measure. The sum runs over all different non-overlapping (or non-empty) boxes $B$ of a given size $\epsilon$ that covers the support set $E$. From this definition, it is easy to obtain $Z_\varepsilon(q) \geq 0$ and $Z_\varepsilon(0) = 1$. The function of the mass exponents $\tau(q)$ of the measure $\mu$ is defined by:

$$\tau(q) = \lim_{\varepsilon \to 0} \left( \frac{\ln Z_\varepsilon(q)}{\ln \varepsilon} \right). \quad (4)$$

Then, the generalized fractal dimensions $D(q)$ of the measure $\mu$ are defined as follows:

$$D(q) = \frac{\tau(q)}{q-1}, \quad q \neq 1, \quad (5)$$

and

$$D(1) = \lim_{\varepsilon \to 0} \frac{Z_{(1,\varepsilon)}}{\ln \varepsilon}, \quad q = 1, \quad (6)$$

**Table 1.** Network's characteristics.

| | |
|---|---|
| Companies analyzed | 16 613 |
| Total number of payments analyzed | 2 617 478 |
| Value of transactions | 3 803 462 026* |
| Average value of transaction per customer | 87 600* |
| Max value of transaction | 121 533* |
| Min value of transaction (aggregated in whole year) | 1000* |
| Average volume of transaction per company | 60 |
| Max volume of transaction per company | 24 859 |
| Min volume of transaction per company (aggregated in whole year) | 20 |

* All money amounts are expressed in monetary units and not in currencies in order to protect the confidentiality of the data set. The purpose of showing monetary units is to provide a notion of the proportion of quantities and not to show exact amounts of money.

**Table 2.** Summary of statistics.

| Statistic | Value | Components | No. of nodes |
|---|---|---|---|
| Undirected links | 43 375 | GCC | 15 434 |
| $\langle k \rangle$ | 20 | DC | 1179 |
| $\gamma^o$ | 2.39 | GSCC | 3987 |
| $\gamma^i$ | 2.49 | GOUT | 6054 |
| $\gamma$ | 2.45 | GIN | 6172 |
| $\langle C \rangle$ | 0.183 | Tendrils | 400 |
| $\langle l \rangle$ | 7.1 | Cutpoints | 1401 |
| $T$ | 0.13 | Bi-component | 4404 |
| $D$ | 29 | $k$-core | 1081 |
| $\langle \sigma \rangle$ (nodes) | 110 | | |
| $\langle \sigma \rangle$ (links) | 40 | | |

$N$, number of nodes; $\langle k \rangle$, average degree; $\gamma^o$, scaling exponent of the out-degree empirical distribution; $\gamma^i$, scaling exponent of the in-degree empirical distribution; $\gamma$, scaling exponent of the connectivity degree distribution; $\langle C \rangle$, average clustering coefficient; $\langle l \rangle$, average shortest path length; $T$, connectivity %; $D$, diameter; $\langle \sigma \rangle$, average betweenness; GCC, giant connected component; DC, disconnected component; GSCC, giant strong connected component; GOUT, giant out component; GIN, giant in component.

where

$$Z_{1,\varepsilon} = \sum_{\mu(B)\neq 0} \mu(B) \ln \mu(B). \tag{7}$$

The generalized fractal dimensions $D(q)$ can be estimated with linear regression of $[\ln Z_\varepsilon(q)]/[q-1]$ against $\ln \varepsilon$ for $q \neq 1$, and similarly a linear regression of $Z_{1,\varepsilon}$ against $\ln \varepsilon$ for $q = 1$. $D(0)$ is the fractal dimension or the box-counting dimension of the support set $E$ of the measure $\mu$, $D(1)$ is the information dimension and $D(2)$ is the correlation dimension.

For a complex network, a box of size $B$ can be defined in terms of the distance $l_B$, which corresponds to the number of links in the shortest-path between two nodes. This means that every node is less than $l_B$ links away from another node in the same box. The measure $\mu$ of each box is defined as the ratio of the number of nodes that are covered by the box and the total number of nodes in the whole network.

Multifractality of a complex network can be determined by the shape of $\tau(q)$ or $D(q)$ curves. If $\tau(q)$ is a straight line or $D(q)$ is a constant, then the network is monofractal; similarly if $D(q)$ or $\tau(q)$ have convex shapes, then the network is multifractal. A multifractal structure can be identified by the following signs [48]: (a) multiple slopes of $\tau(q)$ vs $q$; (b) non-constant $D(q)$ vs $(q)$ values and (c) $f(\alpha)$ vs $\alpha$ value covers a broad range (not accumulated at nearby non-integer values of $\alpha$).

First, we calculate the shortest-path distance between any two nodes in the network and map the shortest-path adjacency matrix $B_{N\times N}$ using the payments adjacency matrix $A_{N\times N}$. We use the shortest-path adjacency matrix $B_{N\times N}$ as input for multifractal analysis. The central idea of the sandbox algorithm is simply to select a node of the network in a random fashion as the center of a sandbox and then count the number of nodes that are inside the sandbox. Initially, none of the nodes has been chosen as a center of a box or as a seed. We set the radius $r$ of the sandbox which will be used to cover the nodes in the range $r \in [1, d]$, where $d$ (diameter) is the longest distance between nodes in the network and radii $r$ are integer numbers. We ensure that the nodes are chosen randomly as center nodes by reordering the nodes randomly in the whole network. Depending on the size $N$ of the network, we choose $T$ nodes in random order as centers of $T$ sandboxes; then we find all the neighboring nodes within radius $r$ from the center of each box. We count the number of nodes contained in each sandbox of radius $r$, and denote it by $S(r)$. We calculate the statistical averages $\langle [S(r)^{q-1}] \rangle$ of $[S(r)^{q-1}]$ over all the sandboxes $T$ of radius $r$. The previous steps will be repeated for each of the different values of radius $r$ to obtain the statistical average $\langle [S(r)^{q-1}] \rangle$ and then use it for calculating linear regression.

The generalized fractal dimensions $D(q)$ of the measure $\mu$ are defined by

$$D(q) = \lim_{r\to 0} \frac{\ln\langle [S(r)/S(0)]^{q-1}\rangle}{\ln(r/d)} \frac{1}{q-1}, \quad q \in R, \tag{8}$$

or rewritten as

$$\ln(\langle [S(r)]^{q-1}\rangle) \propto D(q)(q-1)\ln(r/d) + (q-1)\ln(S_0), \tag{9}$$

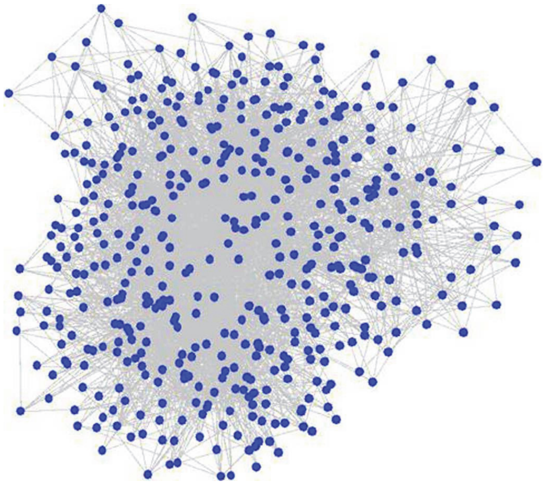


Fig. 1. Visual graph representation of the skeleton of the Estonian network of payments.

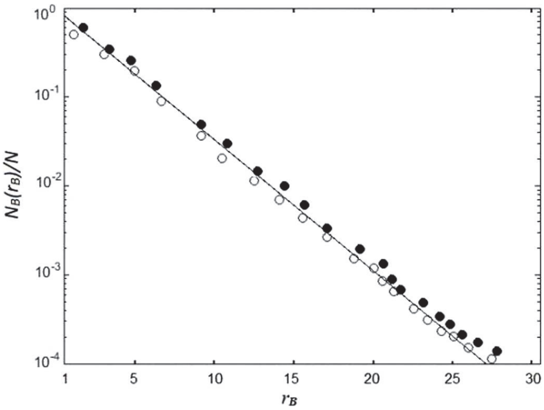



Fig. 2. Fractal scaling representation of our network. The original network (○) and the skeleton network (●). The straight line is included for guidance and has a slope of 2.3. The analysis includes only the giant connected cluster of the network.

where $S(0)$ is the size of the network and the brackets mean taking statistical average over the random selection of the sandbox centers.

We run the linear regression of $\ln(\langle[S(r)]^{q-1}\rangle)$ against $(q-1)\ln(r/d)$ to obtain the generalized fractal dimensions and similarly, calculate the linear regression of $\ln(\langle[S(r)]^{q-1}\rangle)$ against $\ln(r/d)$ to obtain the mass exponents $\tau(q)$. From the shapes of the generalized fractal dimensions curves, we can conclude if multifractality exists or not in our network.

## 3 Results

### 3.1 Fractal scaling analysis

The general characteristics and statistics of the Estonian network of payments are listed in Tables 1 and 2.

We present a fractal scaling analysis by using the box-counting algorithm expressed in equation (1). We calculated the fractal dimension of our network and its skeleton (see Figs. 1 and 2), where the fractal dimension is the absolute value of the slope of the linear fit. Figure 1 depicts a visualization of the graph representation of the skeleton of our network. The box-covering method yields a fractal dimension $d_{Bs} = 2.32 \pm 0.07$ for the skeleton network and for the original network the fractal dimension is $d_{Bo} = 2.39 \pm 0.05$.

The comparison of the fractal scaling in our network and its skeleton structure revealed its own patterns according to the fractality of the network. Figure 2 depicts the fractal scaling representation of our network. As seen in Figure 2, the respective number of boxes needed to cover both networks is similar but not identical: more boxes were needed for covering the skeleton. The largest distance between any two nodes in the network of payments is 29, while the largest distance between any two nodes in the skeleton network is 34.

### 3.2 Multifractal characterization

Linear regression is an important step for obtaining the correct range of radius $r \in [r_{\min}, r_{\max}]$ that is needed to calculate the generalized fractal dimensions (defined by Eqs. (8) and (9)) and the mass exponents (defined by Eq. (4)). We found an appropriate range of radii $r$ within the range of the interval located between 2 and 29 for linear regression. Thus, we selected this linear fit scaling range to perform multifractal analysis. We set the range of $q$ values from $-7$ to 12.

We calculated $\tau(q)$ and the $D(q)$ curves using the sandbox algorithm by Liu et al. [33] and based upon the shapes obtained from the spectrum in Figure 3, it can be seen that the curves are non-linear, suggesting that the network is multifractal.

In Figure 3b, the $D(q)$ function decreases sharply after the peak reaches its end when $q$ is close to $-4.5$. This trait could be interpreted as high densities around the hubs in the network. The hubs have a high number of links connected to them; therefore the density of links around the sections near the hubs is higher than in other parts of the network. These hub nodes or important companies have a noticeable larger amount of business partners (for example: their own customers, or their suppliers or any other business parties that interact financially with them) than the rest of the companies in the network have, and it is interesting to observe that this characteristic can be explored and identified by looking at the values of $D(q)$ spectra. The multifractality seen in our network reveals that the system cannot be described by a single fractal dimension, suggesting that the multifractal approach provides a better characterization than the fractal approach; hence, this means that the Estonian economy is multifractal.

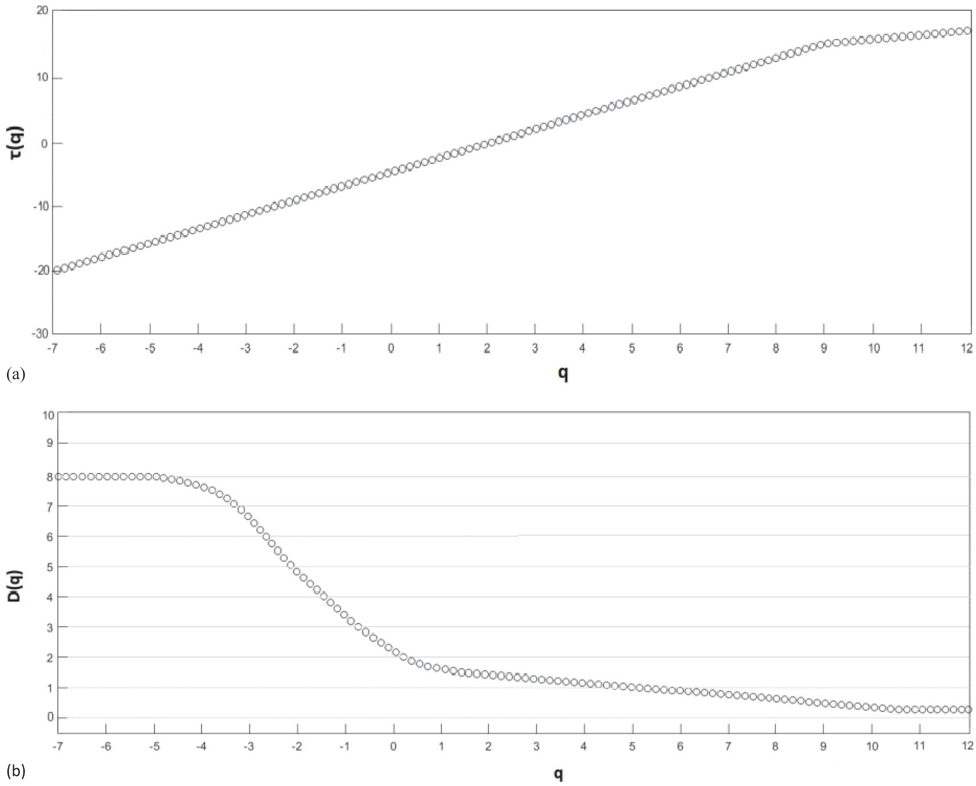



**Fig. 3.** (a) Plot of mass exponents $\tau(q)$ as function of $q$. (b) Plot of generalized fractal dimensions $D(q)$ as function of $q$. Curves indicated by circles represent numerical estimations of the mass exponents and generalized fractal dimensions, respectively.

The quantity $\Delta D(q)$ describes density changes of links in our network. We use $\Delta D(q) = D(q)_{\max} - \lim D(q)$ to observe how the values of $D(q)$ change along the spectrum. From Figure 3b, we found that $\lim D(q) = 0.37$ and $D(q)_{\max} = 7.8$ and this means that $\Delta D(q) = 7.43$. A large value of $D(q)$ means that the distribution of links is very irregular in our network, suggesting that there are areas of hubs where the links are very densely grouped contrasting with areas where the nodes are connected with only just few links. For the economy of a country, this makes sense because not all the companies have the role of hubs in the network of payments; many companies are just small participants. For a comparison of the maximum values of $D(q)$ of different networks, please see Table 3.

## 4 Conclusions

We presented the first multifractal analysis of a complex network of payments. We studied specific fractal and multifractal properties of a novel and unique network: the Estonian network of payments. In this study, we presented a fractal scaling analysis where we identified the underlying skeleton structure of the network. We calculated its fractal dimension and compared it with the fractal dimension of the original network. We found that the skeleton network had a slightly smaller fractal dimension than the original network. This comparison, between the fractal scaling in our original network and the corresponding skeleton network reveals that there are only slightly distinct patterns according to the fractality in the network. This means that the skeleton network preserves the structure very well while simplifying the complexity of the network. Then, the skeleton network captures the general structure of the network and by observing the properties of the skeleton, an easier visualization of the topological organization of the network can be achieved.

Fractal analysis helps to calculate and understand the fractal dimension of complex networks. However, it is necessary to describe and characterize the multiple fractal patterns which cannot be described by a single fractal dimension, thus we also performed a multifractal analysis to our network. Multifractal analysis allows the calculation of a set of fractal dimensions, particularly the generalized fractal dimensions. We examined the general multifractal structure and explored some statistical features of our network. In order to study the multifractal

**Table 3.** Comparison of the maximum values of $D(q)$ in different networks.

| Network | Number of nodes | Highest $D(q)$ | Reference |
|---|---|---|---|
| Pure fractal network | 6222 | 2.8 | [32] |
| Small world network | 6222 | 6.6 | [32] |
| Semi fractal network | 6222 | 3.1 | [32] |
| Sierpinski weighted fractal network | 9841 | 2.0 | [2] |
| Cantor dust weighted fractal network | 9841 | 3.2 | [2] |
| High-energy theory collaboration weighted network | 8361 | 6.0 | [2] |
| Astrophysics collaboration weighted network | 16 706 | 6.2 | [2] |
| Computational geometry collaboration weighted network | 7343 | 5.1 | [2] |
| Barabási & Albert model scale-free network | 10 000 | 3.6 | [33] |
| Newman and Watts model small-world network | 10 000 | 4.8 | [33] |
| Erdös-Rényi random graph model | 10 000 | 3.9 | [33] |
| Barabási & Albert model scale-free network | 7000 | 3.4 | [35] |
| Random network | 5620 | 3.5 | [35] |
| Random network | 449 | 2.4 | [35] |
| Protein–protein interaction network: Human | 8934 | 4.9 | [35] |
| Protein–protein interaction network: *Arabidopsis thaliana* | 1298 | 2.5 | [35] |
| Protein–protein interaction network: *C. elegans* | 3343 | 4.5 | [35] |
| Protein–protein interaction network: *E. coli* | 2516 | 4.1 | [35] |
| Small world network | 5000 | 3.0 | [35] |
| Estonian network of payments | 16 613 | 7.8 | [45] |

structure, we calculated the spectrum of the mass exponents $\tau(q)$ and the generalized fractal dimensions $D(q)$ curves, using a sandbox algorithm for multifractal analysis of complex networks adapted by Liu et al. [33]. This algorithm is based on the fixed-size box-counting algorithm developed by Tél et al. [46]. The sandbox algorithm utilized in this study could also be used to explore and characterize other similar kinds of economic networks.

Our results indicated that multifractality exists in the Estonian network of payments, and this suggests that the Estonian economy is multifractal (from the point of view of networks). We found large values of $D(q)$spectra and this means that the distribution of links is quite irregular in the network, suggesting there are specific nodes which hold densely connected links, meanwhile other nodes hold just few links. This type of structure could be relevant when specific critical events occur in the economy that could threaten the whole network. It is important to continue observing, describing and analyzing the structures and characteristics of economic complex networks in order to be able to understand their underlying processes or to be able to detect patterns that could be useful for predicting or forecasting events and trends. The addition of evidence through empirical studies in favor of fractality and multifractality of economic networks represents a step forward toward the knowledge on the universality and the unraveling of the complexity of economic systems.

Further applications and studies could extend this topic by examining the potential factors that drive the strength of the multifractal spectrum. Some applications could involve studying the origin of such factors. Another interesting line of research would be to study the patterns and the changes of the multifractal spectrum across different periods of time. Particularly, it would be interesting to analyze such patterns during financial crisis periods for risk pattern recognition purposes. Also, it would be interesting to take into account different probability measures for such kind of multifractal analysis. Other direction of the studies could focus on building network models that attempt to forecast country money flows or potential industry growth trends based on transactions data.

We thank Swedbank (Estonia) for allowing us to retrieve the data set for the analysis. This research was supported by the European Union through the European Regional Development Fund (Project TK 124).

## Author contribution statement

SR: data analysis and interpretation of results. JK: research hypotheses formulation, results interpretation. RK, JE: discussions and ideas, validation.

## References

1. S. Furuya, K. Yakubo, Phys. Rev. E **84**, 036118 (2011)
2. Y.Q. Song, J.L. Liu, Z.G. Yu, B.G. Li, Sci. Rep. **5**, 17628 (2015)
3. G. Boffetta, A. Mazzino, A. Vulpiani, J. Phys. A: Math. Theor. **41**, 363001 (2008)
4. B. Lashermes, S. Roux, P. Abry, S. Jaffard, Eur. Phys. J. B **61**, 201 (2008)
5. J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde, H.E. Stanley, Physica A **316**, 87 (2002)
6. E. Bacry, J. Delour, J.F. Muzy, Physica A **299**, 84 (2001)
7. B.B. Mandelbrot, L. Calvet, A. Fisher, Cowles Foundation, Discussion paper 1164, 1997
8. R. Lopes, N. Betrouni, Med. Image Anal. **13**, 634 (2009)
9. Z.G. Yu, V.V Anh, K.S. Lau, Phys. Rev. E **64**, 31903 (2001)
10. J.S. Gagnon, S. Lovejoy, D. Schertzer, Nonlinear Process. Geophys. **13**, 541 (2006)

11. D. Schertzer, S. Lovejoy, J. Appl. Meteorol. **36**, 1296 (1997)
12. V.V. Anh, K.S. Lau, Z.G. Yu, Phys. Rev. E **66**, 031910 (2002)
13. Z.G. Yu, V.V. Anh, K.S. Lau, Phys. Rev. E **68**, 021913 (2003)
14. L.Q. Zhou, Z.G. Yu, J.Q. Deng, V.V. Anh, S.C. Long, J. Theor. Biol. **232**, 559 (2005)
15. J.W. Kantelhardt, E.K. Bunde, D. Rybski, P. Braun, A. Bunde, S. Havlin, J. Geophys. Res. **111**, D01106 (2006)
16. D. Veneziano, A. Langousis, P. Furcolo, Water Resour. Res. **42**, W06D15 (2006)
17. V. Venugopal, S.G. Roux, E.F. Georgiou, A. Arneodo, Water Resour. Res. **42**, W06D14 (2006)
18. Z.G. Yu, V.V. Anh, Y. Wang, D. Mao, J. Wanliss, J. Geophys. Res. **115**, A10219 (2010)
19. H. Inaoka, T. Nimoniya, K. Taniguchi, T. Shimizu, H. Takayasu, Bank of Japan, Working papers, 2004
20. Y. Lu, J. Tang, Environ. Plan. B: Plan. Des. **31**, 6 (2004)
21. Y. Zhang, X. Li, Eur. Phys. J. B **88**, 61 (2015)
22. F.M. Siokis, Physica A **395**, 283 (2014)
23. A.I. Olemskoi, V.N. Borisyuk, I.A. Shuda, A.A. Bagdasaryan, J. Nano-Electron. Phys. **1**, 3 (2009)
24. O. Pont, A. Turiel, C.J. Pérez-Vicente, Physica A **388**, 10 (2009)
25. W.X. Zhou, Z.Q. Yiang, D. Sornette, Physica A **375**, 741 (2007)
26. G. Palla, I. Derényi, I. Farkas, T. Vicsek, Nature **435**, 814 (2005)
27. L.K. Gallos, C. Song, S. Havlin, H.A. Makse, Physica A **386**, 686 (2007)
28. C.M. Schneider, T.A. Kesselring, J.S. Andrade, H.J. Herrmann, Phys. Rev. E **86**, 016707 (2012)
29. V.M. Eguiluz, E. Hernández-García, O. Piro, K. Klemm, Phys. Rev. E **68**, 055102 (2003)
30. C. Song, L.K. Gallos, S. Havlin, H.A. Makse, J. Stat. Mech.: Theor. Exp. **3**, 4673 (2007)
31. C. Song, S. Havlin, H.A. Makse, Nature **433**, 392 (2005)
32. B.G. Li, Z.G. Yu, Y. Zhou, J. Stat. Mech.: Theor. Exp. **2**, P02020 (2014)
33. J.L. Liu, Z.G. Yu, V. Anh, Chaos **25**, 023103 (2015)
34. D.J. Wei et al., Sci. Rep. **3**, 3049 (2013)
35. D.L. Wang, Z.G. Yu, V. Anh, Chin. Phys. B **21**, 080504 (2012)
36. Z.G. Yu, H. Zhang, D.W. Huang, Y. Lin, V. Anh, J. Stat. Mech.: Theor. Exp. **2016**, 033206 (2016)
37. J.L. Liu, J. Wang, Z.G. Yu, X.H. Xie, Sci. Rep. **7**, 45588 (2017)
38. J. Feder, *Fractals* (Plenum, New York, 1988)
39. J.S. Kim, K.I. Goh, G. Salvi, E. Oh, B. Kahng, D. Kim, Phys. Rev. E **75**, 016110 (2007)
40. D.H. Kim, J.D. Noh, H. Jeong, Phys. Rev. E **70**, 046126 (2004)
41. K.-I. Goh, G. Salvi, B. Kahng, D. Kim, Phys. Rev. Lett. **96**, 018701 (2006)
42. A. Gibbons, *Algorithmic graph theory* (Cambridge University Press, UK, 1985)
43. H. Jeong, B. Tombor, R. Albert, Z.N. Oltvai, A.L. Barabási, Nature **407**, 651 (2000)
44. B.B. Mandelbrot, *The fractal geometry of nature* (Academic Press, New York, 1983)
45. S. Rendón de la Torre, J. Kalda, R. Kitt, J. Engelbrecht, Chaos Solitons Fractals **90**, 18 (2016)
46. T. Tél, A. Fülöp, A. Vicsek, Physica A **159**, 155 (1989)
47. T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia, B.I. Shraiman, Phys. Rev. A **33**, 1141 (1986)
48. P. Grassberger, I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983)

**Paper III**

**Rendón de la Torre S.**, Kalda J. (2018) Review of structures and dynamics of economic complex networks: Large scale payment network of Estonia. In: Zengqjang C., Dehmer M., Emmert-Streib F., Shi Y. (eds.), Modern and interdisciplinary problems in network science. Taylor & Francis CRC Group, USA, 193–226. https://www.crcpress.com/Modern-and-Interdisciplinary-Problems-in-Network-Science-A-Translational/Chen-Dehmer-Emmert-Streib-Shi/p/book/9780815376583.

# Chapter 8

## Review of Structures and Dynamics of Economic Complex Networks: Large-Scale Payment Network of Estonia

**Stephanie Rendón de la Torre and Jaan Kalda**

**CONTENTS**

| | | | |
|---|---|---|---|
| 8.1 | Introduction | | 193 |
| 8.2 | Summary | | 195 |
| 8.3 | Topologic Structure and Components: Analytic Metrics | | 196 |
| | 8.3.1 | Robustness of the network | 202 |
| 8.4 | Community Detection | | 204 |
| | 8.4.1 | Structure of communities | 209 |
| 8.5 | Multifractal Networks | | 211 |
| | 8.5.1 | Fractal network analysis | 212 |
| | 8.5.2 | Multifractal network analysis | 215 |
| 8.6 | Conclusions | | 220 |
| | 8.6.1 | Further applications | 221 |
| References | | | 222 |



## 8.1 Introduction

The neologism *econophysics* was first coined by H. Eugene Stanley in a Statphys conference in 1995 held in Kolkata, India. Mantegna and Stanley [1] defined *econophysics* as a multidisciplinary field that denotes the activities of physicists who work on economic problems in order to test a variety of new conceptual approaches derived from physical sciences. Much has been studied and developed in this area since then and even before then, mainly originated from models of statistical mechanics. Similarly, problems related with distributions of income, wealth, and economic returns in financial markets have been already addressed in research papers, and mostly these topics are related with the insufficiency to explain non-Gaussian distributions and scaling properties empirically detected by means of traditional economic theoretic approaches. Some of the most relevant outcomes of the research accomplished in topics of econophysics are related with detection and explanation of power-law tails in the distribution of different types of financial data, the existence of certain underlying universalities in the behavior of individual market agents, and the detection of similarities between financial time series and natural phenomena.

In recent years, a part of the main focus of research has tilted towards the discovery and understanding of the underlying financial, social, and economic systems' structures through the use of the tools of complex networks science. In this context, the network approach has two sources of origin: one source originates from economics, finance, and sociology, while the second source originates from computer science, physics, complexity, and mathematics. The convergence point of both sources of origin attempts to combine economy and complex systems studies, and this approach can be translated into a graph representation of economic systems in order to study how interactions among the components of the graph occur whatsoever the nature of the relations between the components is.

Network science is an interdisciplinary active field of research that originates from the mathematics branch of graph theory, and it has been extended into different directions including towards economics, statistical mechanics, computer science, neuroscience, sociology, transportation, ecological systems, and biology. With complex networks, it is possible to describe the structure of any system, when the system is suitable to be represented as a graph.

"Complexity" may refer to the quality of a system or to a quantitative characterization of a system. As a quality of the system, it refers to what makes the system complex and it has something to do with the ability to understand a system; it refers to the existence of emergent properties, which appear as a consequence of the interactions of the components of the system [2]. An example of a property that emerges as a consequence of global organizational structure of a network is the "small-world" property, which is characterized by small average path length and a high number of triangles in the network. In the second definition of complexity, this term is used as a quantity when referring to something that is more complicated than other thing; it refers to the quantity of information needed to specify the system. For real-world networks, a huge amount of information is needed to describe a system, such as the number of nodes, links, degree correlations, degree distributions, clustering coefficients, diameter, betweenness, centralities, community structure, average or shortest paths, communication patterns, and other quantities. In the case of random networks, the only information needed to describe their structure is the number of nodes and the probabilities for linking pairs of nodes. The network representation of real networks is called "complex networks" because of two reasons. Firstly, because there are characteristics that arise as a consequence of the global topological organization of the system, and secondly because these structures cannot be trivially described like in the cases of random or regular graphs [3].



The theoretical framework behind complex networks is continuously developing, advancing at a fast pace and has already made significant progress towards unraveling the organizing principles governing complex networks structures and their dynamics. Studies related with: topological features, dynamical aspects, community detection, network phenomena, and particular properties of networks have been the focus of attention of extensive research in the last couple of decades [4–9].

Networks play an important role in a wide range of economic and social phenomena, and the use of techniques and methods from graph theory has permitted economic network theory to expand the knowledge and insights into such phenomena in which the embeddedness of individuals or agents in their social or economic interrelations cannot be neglected [10]. For example, Souma et al. [11] studied a shareholder network of Japanese companies where the authors analyzed the companies' growth through economic networks. Other examples of interesting applications of complex networks in economics are provided by the regional investment or ownership networks where European company-to-company investment stocks show power-law distributions that allow predicting the investments that will be received or made in specific regions, based on the connectivity and transactional activity of the companies [12,13]. Nakano and White [14] have shown that analytic concepts and methods related with complex networks can help to uncover structural factors that may influence the price formation for empirical market-link formations of economic agents. Reyes et al. [15] used a weighted network analysis focused on using random walk betweenness centrality to study why high-performing Asian economies have higher economic growth than Latin-American economies in the last years. Network-based approaches are very useful and provide a means by which to monitor complex economic systems and may help on providing better control in managing and governing these systems. Other interesting line of research is related with network topology as a basis for investigating money flows of customer-driven banking transactions. A few recent papers describe the actual topologies observed in different financial systems [16–19]. Some other works have focused on economic shocks and robustness in economic complex networks [20,21].



## 8.2 Summary

Networks can be studied from different points of view, for example, from a local, global, or mesoscale perspective. The contribution of this chapter is to explore these approaches using different methodologies with the goal of studying general and particular properties of networks through the analysis that consists of different experiments on a unique, interesting, and particular economic network. This is a review of our research on the structures and characteristics of the large-scale Estonian network of payments [22–24]. In this novel and unique economic network, the nodes represent Estonian companies and the links represent payments done between the companies. Mainly, we focus on the analysis of:

a. Global and local topology

b. Community detection and structure

c. Fractal and multifractal properties

Our data set was obtained from Swedbank's databases. Swedbank is one of the leading banks in the Nordic and Baltic regions of Europe. The bank operates actively in Estonia, Latvia, Lithuania, and Sweden. All the information related to the identities of the nodes is very sensitive and thus will remain confidential and unfortunately cannot be disclosed. The data set is unique in its kind and very interesting since ~80% of Estonia's bank transactions are executed through Swedbank's system of payments, therefore, this data set reproduces fairly well the transactional trends of the whole Estonian economy, and hence we use this data set as a proxy of the economy of Estonia. Such data set comprises domestic payments (company-to-company electronic transactions) of year 2014. The network consists of 16,613 nodes, 2,617,478 payment transactions, and 43,375 links.

## 8.3 Topologic Structure and Components: Analytic Metrics

In this sub-section of the chapter, we focus on analyzing some interesting structural properties of our network, with a special focus on topologic components. Graph theory definitions not introduced in this chapter can be found in [25,5].

A random network is the most basic model of all network formations, and it is based on the assumption that a fully random process is responsible for the structure of the links in a network. The properties of random network models [26] provide rich insight of the characteristics and features that many economic and social networks share. Such models are useful benchmarks to compare empirical networks in order to be able to identify the elements that are a result of randomness and the ones that can be rooted to other factors. Some properties of random networks that are useful for studying general networks are, for example, the distribution of links across nodes, connectivity in terms of paths, distances within networks, shortest-average paths, diameter, and etcetera.



A graph is a mathematical and symbolic representation of a network and of its connectivity. A simple undirected graph $G$ is a set of vertices $V$ connected with edges $E$, therefore, $G = (V, E)$. $A$ graph is defined by the structural information contained in its adjacency matrix. A network may have an arbitrary large amount of additional information on top of it: for example, edges can have attributes such as capacity or weight, or it may be a function of other variables. Also, in a network, the vertices are called nodes and the edges are called links. Network terminology is generally used when the links transport or send something between the nodes (like in social, computer, biological, transport, or economic networks).

There are several ways to define our network of payments and in this study we consider more than one definition. In the first definition, we mapped an undirected graph, a symmetric payment adjacency matrix $A_{N\times N}$, where $N$ is the total number of nodes in the network, then $a^u_{ij} = a^u_{ji}$ and $a^u_{ij} = 1$. Otherwise, $a^u_{ij} = 0$ if there is no transaction between companies $i$ and $j$.

The links can also represent directions, where the links follow the flow of money. The second definition is a directed graph where the links follow the flow of money, such that a link is incoming to the receiver and outgoing from the sender of the payment. For this case, we have two more matrices, one for the in-degree case and another one for the out-degree case. The choice of the definition of the matrix representation depends on the focus of the analysis.

The most basic properties of a network are the number of nodes $N$ and the overall number of links $k$. The number of nodes defines the size of the network, while the number of links relative to the number of possible links defines the connectivity of a network. Connectivity $(p)$ is the unconditional probability that two nodes are connected by a direct link. For a directed network, connectivity is defined as follows:

$$(p) = \frac{k}{n(n-1)}. \tag{8.1}$$

The connectivity of our network is 0.13, meaning that the network is sparse and 87% of the potential connections are disabled. Diameter $d$ is the maximum distance between two nodes (measured by the number of links), and this distance is equal to 29; this number is substantially higher when compared to the diameter of a random network of comparable characteristics ($d \sim 19$). The difference between the diameter number of our network and a comparable random network is substantially high, and it could be explained by the preferred money paths that nodes have in our network. Preferred money paths means that some companies have specific preferences when considering the counterparties they transact with. Intuitively, this makes sense because for a company it is important to choose carefully which counterparties become trading partners, clients, service providers, or suppliers and which ones not. Usually, this decision is based upon determined factors such as geographical location, goals affinity, cost policies, future joint ventures, legal agreements, nature of the business, or any other reasons, and it is interesting to notice how this particular feature can be observed through the comparison of the connectivity of our network and a random network.



A path is a sequence of nodes such that each node is linked to the next one along the path by a link. A path consists of $n+1$ nodes and $n$ links. A path between nodes $i$ and $j$ is an ordered list of $n$ links. The length of this path is $n$. The path length of all node pairs could be represented in the form of a distance matrix. The average path length is the average of the shortest path lengths across all node pairs in the network.

Other simple quantity that can be observed in a network is the number of nodes of a given degree. The degree of a node is the number of neighbors of that node and is defined as

$$k_i = \sum_{j \in \zeta(i)} a_{ij}\,, \tag{8.2}$$

the sum runs over the set $\zeta(i)$ of neighbors of $i$. For example: $\zeta(i) = \{j | a_{ij} = 1\}$.

The average degree of a network is the number of links divided by the number of nodes and is defined as

$$\langle k\rangle = \frac{1}{n}\sum k^0 = \frac{1}{n}\sum k^d = \frac{m}{n}, \tag{8.3}$$

where $m$ is the number of links and $n$ is the number of nodes. The average degree of our network is 20.

In a directed network, there are two important characteristics of a node: the number of links that end at a node and the number of links that start from the node. These quantities are known as the out-degree $k_o$ and the in-degree $k_d$ of a node, and we define them as

$$k_d = \sum_{j\in\zeta(i)} a_{ij}^d, \quad k_o = \sum_{j\in\zeta(i)} a_{ij}^o. \tag{8.4}$$

Also, it is possible to categorize networks by the degree distributions shown in their tails. In general, real-world networks are very different compared with random networks, when referring to their degree distributions. Random networks commonly show Poisson distributions, while real-world networks might have long tails in the right part of the distribution with values that are far above the mean. Measuring the tail of the distribution of the degree data could be achieved by building a plot of the cumulative distribution function. In real-world networks, it is common to find distributions that follow power laws in their tails:



$$P(k) \sim \sum_{k'=k}^{\infty} k'^{-\gamma} \sim k^{-(\gamma-1)}, \tag{8.5}$$

where $\gamma$ is the scaling exponent of the distribution and the degree distribution $P(k)$ is the probability that the degree of a node is equal to $k$. This type of distribution is called scale-free and networks with such degree distributions are referred to as scale-free networks. Such distributions have no natural scale and the functional form of the distribution remains unchanged within a multiplicative factor under a rescaling of the random variables. Previous studies [27,28] have shown that in large-scale-free networks, independently of the system and the origin of the components, the probability $P(k)$ that a node in the network interacts with $k$ other links decays as a power law, suggesting that there is a tendency for large networks to self-organize into a scale-free state. A degree distribution with power laws is a characteristic commonly seen in complex networks such as in the World Wide Web network, protein interaction networks, phone calls networks, food webs networks, citation networks, actors-movies networks, and it also appears in systems of payments from different banks around the world [17–19].

Complex networks can be classified as homogeneous or heterogeneous depending on their degree distributions. Homogeneous networks are identified by degree distributions that follow an exponential decay. In these networks, the distribution peaks at an average $k$ and then decays exponentially for large values of $k$, such as the distributions formed in the random graph model [26] and the small-world model [9] where each node has approximately the same number of links $k$, a normal distribution and the majority of the nodes has an average number of connections and only few or none of the nodes have either some or lots of connections. In heterogeneous large networks or scale-free networks, the degree distribution decays as a power law with a characteristic scale. The degree distribution follows a Pareto form of distribution where many nodes have few links and few nodes have many links, therefore, highly connected nodes are statistically significant in scale-free networks.

Figure 8.1a shows the cumulative degree distribution of the Estonian network of payments (undirected). A straight line was added as eye guideline.

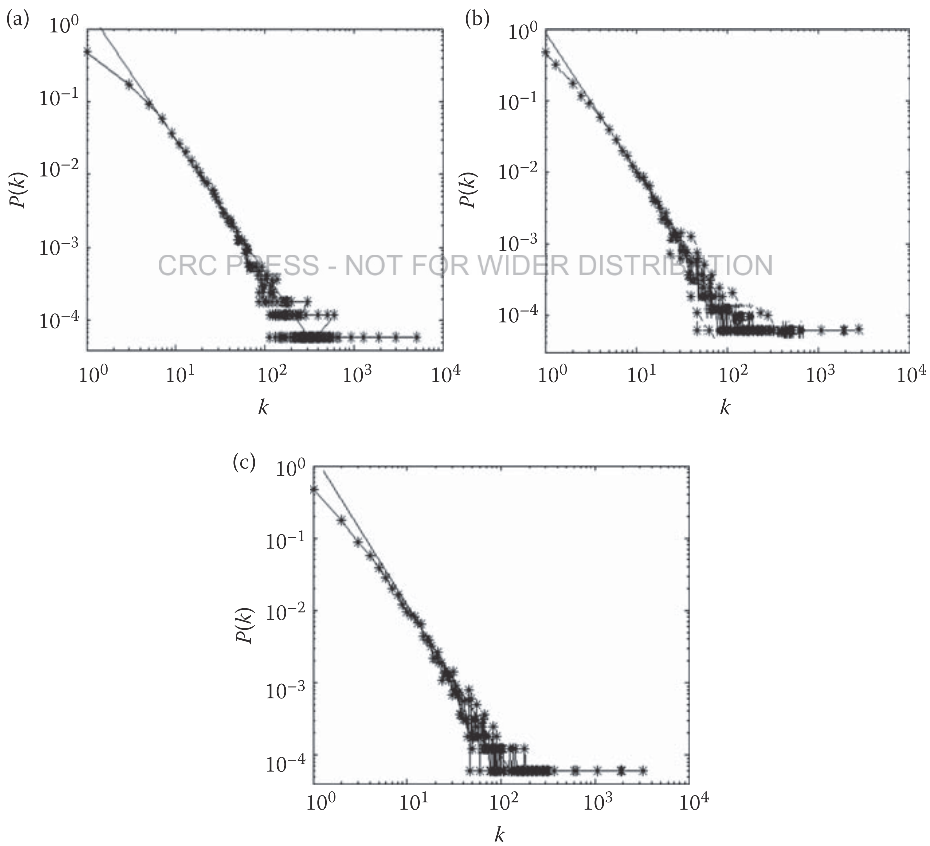





**FIGURE 8.1:** (a) Empirical degree distribution for the connectivity network of the Estonian network of payments. $X$-axis is the number of $k$ degrees and $Y$-axis is $P(k)$. (b) out-degree distribution of the network, $P(k) \sim k^{-2.39}$. (c) Empirical in-degree distribution $P(k) \sim k^{-2.49}$.

The distribution in Figure 8.1a follows a power law with the following scaling exponent:

$$P(\geq k) \propto k^{-2.4}. \tag{8.6}$$

Figure 8.1b shows the out-degree distribution and Figure 8.1c shows the in-degree distribution of our network. In all the distributions, we found regions that can be fitted by power laws, and this implies that the network has a scale-free structure.

Another interesting and fundamental metric of complex networks is the clustering coefficient of a node. It represents the probability that any two neighbors of a node are connected; it is the density around a node. In our study, it indicates whether or not there is a link between two companies that have a common third business party.

$$C(i) = \frac{1}{k_i(k_i - 1)} \sum_{j \neq k} a_{ij} a_{jk} a_{ik}. \tag{8.7}$$

The average clustering coefficient is the mean of the clustering coefficients $\langle C \rangle$ of all the nodes. In our network, the average clustering coefficient is 0.18, and this suggests there is cliquishness in the network. This means that two companies that are trading partners with a third one, have an average probability of 18% of being trading partners with another than the probability than any two other companies randomly chosen have. For visualization purposes, Figure 8.2 displays the distribution of the clustering coefficient of our network. As seen in the plot, there are a high number of unlinked neighbor nodes (45% of the nodes) that might be explained by the large number of nodes with degrees equal to 1 which appear frequently in scale-free networks.



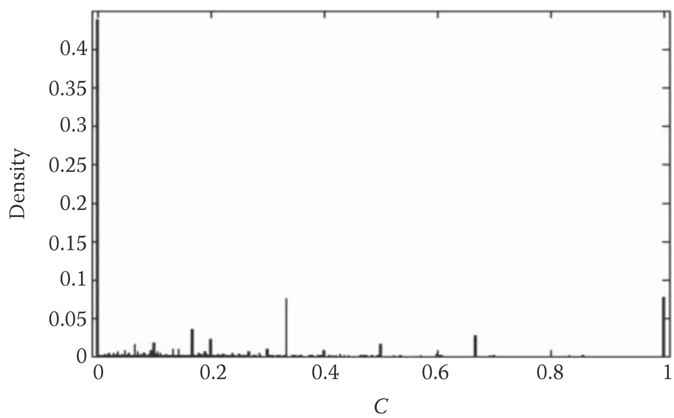



**FIGURE 8.2:** Distribution of the clustering coefficient of the Estonian network of payments.

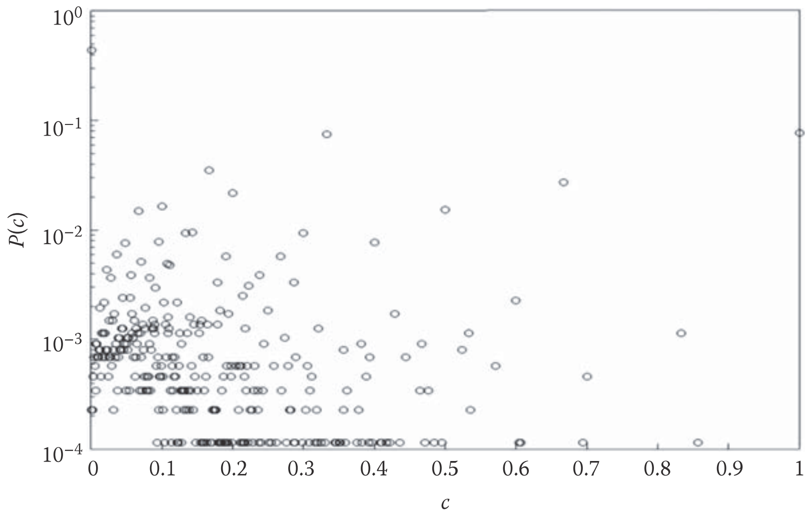



**FIGURE 8.3:** Probability distribution of the clustering coefficient of the Estonian network of payments.

We use the set of clustering coefficients of node $i$ to construct a probability distribution. Figure 8.3 shows the probability distribution of clustering coefficients of our network. As observed in the plot, the irregularity of the clustering coefficients is noticeable.



Compared to other real-world networks, such as the U.S. Federal Reserve Bank network of payments [18], the film-actor network [9], or the metabolism reactions network [28,29], the average clustering coefficient in our network is low.

As it was mentioned earlier, the most basic model of networks is the random network model $G(n, p)$ developed by Erdös and Rényi [26] and this model has two parameters: $n$ and $p$ ($n$ is the number of nodes of the graph and $p$ is the probability to link). The model works under the assumption that there could be a link $i-j$ between two nodes $i$ and $j$, and this assumption holds no matter if the nodes had a common neighbor node before the link was formed. The outcome of the model is the generation of random network graphs with a low clustering coefficient and a low variation in the degrees of the nodes. A random network cannot capture the decreasing nature of the clustering coefficient of the nodes with increase in the node degree, because the clustering coefficient of the nodes in this type of network is totally independent of the node degree and is equal to the probability of a link between any two nodes [7].

The general characteristics and statistics of the Estonian network of payments are listed in Tables 8.1 and 8.2. Regarding other statistical measures of the Estonian network of payments, as per Table 8.2, the average shortest path length $l$ is equal to 7.1 (calculated with Dijkstra's algorithm). Our network is a "small world" with 7.1° of separation, meaning that in average any company

**TABLE 8.1:** Network's characteristics

| | |
|---|---|
| Companies analyzed | 16,613 |
| Total number of payments analyzed | 2,617,478 |
| Value of transactions | 3,803,462,026[a] |
| Average value of transaction per customer | 87,600[a] |
| Max value of a transaction | 121,533[a] |
| Min value of a transaction (aggregated in whole year) | 1000[a] |
| Average volume of transaction per company | 60 |
| Max volume of transaction per company | 24,859 |
| Min volume of transaction per company (aggregated in whole year) | 20 |

[a]All money quantities are expressed in monetary units and not in real currencies in order to protect the confidentiality of the data set. The purpose of showing monetary units is to provide a notion of the proportions of quantities and not to show exact amounts of money.

**TABLE 8.2:** Summary of statistics

| **Statistic** | **Value** | **Components** | **# Nodes** |
|---|---|---|---|
| $N$ | 16,613 | GCC | 15,434 |
| Number of payments | 2,617,478 | DC | 1179 |
| Undirected Links | 43,375 | GSCC | 3987 |
| $\langle k \rangle$ | 20 | GOUT | 6054 |
| $\gamma^o$ | 2.39 | GIN | 6172 |
| $\gamma^i$ | 2.49 | Tendrils | 400 |
| $\gamma$ | 2.45 | Cutpoints | 1401 |
| $\langle C \rangle$ | 0.183 | Bi-component | 4404 |
| $\langle l \rangle$ | 7.1 | $k$-core | 1081 |
| $T$ | 0.13 | | |
| $D$ | 29 | | |
| $\langle \sigma \rangle$ (nodes) | 110 | | |
| $\langle \sigma \rangle$ (links) | 40 | | |



*Abbreviations:* $N$ = number of nodes. $\langle k \rangle$ = average degree. $\gamma^o$ = scaling exponent of the out-degree empirical distribution. $\gamma^i$ = scaling exponent of the in-degree empirical distribution. $\gamma$ = scaling exponent of the connectivity degree distribution. $\langle C \rangle$ = average clustering coefficient. $\langle l \rangle$ = average shortest path length. $T$ = connectivity %. $D$ = Diameter. $\langle \sigma \rangle$ = average betweenness. GCC = Giant Connected Component. DC = Disconnected Component. GSCC = Giant Strongly Connected Component. GOUT = Giant Out-Component. GIN = Giant In-Component.

can be reached by other company in just a few links. Also, our network showed low connectivity ($C = 0.13$) but at the same time the network is densely connected. This characteristic is in line with the fact that there are companies that act as hubs and lead to short distances between the other companies.

### 8.3.1 Robustness of the network

In complex networks, some nodes are essential while others are not, and identifying them is a critical task for many situations. The most essential nodes are those which if removed from the network, would cause the whole system to collapse. In order to have a deeper understanding on how the network is likely to behave as a whole in the presence of perturbations, we will address the next question: if a portion of nodes were removed, would the structure of the network become divided into disconnected clusters? How will the network respond to an actual removal of nodes? There are many approaches on how to tackle this problem and locate the "key nodes" in the network, or on how to calculate the optimal percolation threshold of nodes that would break the network into disconnected clusters. Morone and Makse [30] designed an approach that has proven to perform better than other heuristic methods (such as high degree node, $k$-core, closeness, and eigenvector centralities). Morone and Makse's algorithm optimizes a measure that can reflect the collective influence effect that arises when taking into account the entire influential set of nodes at once. This algorithm predicts a smaller set of optimal influencer nodes (the group of nodes that destroy the network if they are removed).

The collective influence of a node $CI$ is defined as the product of the node's reduced degree (the number of its nearest connections $k_i - 1$), and the total reduced degree of all nodes $k_j$ at a distance $\ell$ from it, and is represented as follows:

$$CI_\ell(i) = (k_i - 1) \sum_{j \in \partial \mathrm{Ball}(i,\ell)} (k_j - 1), \tag{8.8}$$

where $\ell$ is defined as the shortest path.  Ball$(i, \ell)$ is the set of nodes inside a ball of radius $\ell$ around node $i$. $\partial$Ball$(i, \ell)$ is the frontier of the ball and comprises the nodes $j$ that are at a distance $\ell$ from $i$. By computing $CI$ for each node, it is possible to locate the nodes with the highest collective influence. The collective influence algorithm addresses the problem of optimal influence on the computation of the minimum structural total number of nodes that reduces the largest eigenvalue of the nonbacktracking matrix of the network.

We performed a simulation using the $CI$, where we calculate the collective influence of a group of nodes as the fall in the size of the Giant Connected Component (GCC) which would occur if the nodes of the GCC were eliminated. The GCC contains 15,434 nodes, and this quantity represents 92.8% of the nodes of the whole network. These results are displayed in Figure 8.4. The plot shows the GCC when a fraction of its nodes has been removed. The optimal percolation threshold occurs when 6.0% of the nodes are removed and that is the point where GCC$(Pc) = 0$. This result implies that there are many companies that execute a large number of payments which, in fact, have a weak influence in the economic network as a whole. The most influential companies in the network are not necessarily the most connected ones, neither those having more economic activity. A weak but smart node attack where only 6.0% of the nodes are removed destroys the whole network of payments, meaning that a few nodes maintain unified the whole network.

Scale-free networks are resilient to random removal of nodes, but are vulnerable to smart attacks. Our network is a scale-free network (with power laws in

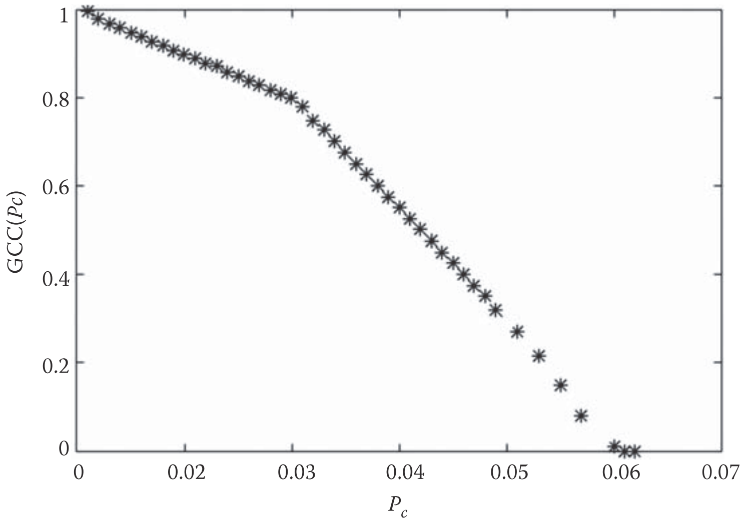



**FIGURE 8.4:** GCC of the network of payments as a function of the percolation threshold $P_c$.

the degree distribution) and its own scale-free nature makes hardly possible to destroy the network by a random removal of nodes, but if the exact portion of particularly selected nodes are removed, then the network collapses completely. This "collapse" effect has been already observed in financial systems when severe economic crisis occur and specific companies or banks declare themselves in bankruptcy and the whole system breaks down. An example is the global financial crisis of 2008 that started with the collapse of the famous investment bank Lehman Brothers, followed by Bear Sterns, UBS, and other financial entities that dragged the whole global financial system into severe liquidity problems.



## 8.4 Community Detection

Community detection analysis is essential for understanding the structure and functionality of large networks, and it also helps to expand the knowledge of the local organization of their components. Networks have sub-sections in which the nodes are more densely connected to each other than to the rest of the nodes in the network, and such sub-sections are called communities. Community detection is a graph partitioning process that provides valuable insight of the organizational principles of networks and is essential for exploring and predicting connections that are not yet observed. Thus far, recent advances of the underlying mechanisms that rule dynamics of communities in networks are limited, and this is why the achievement of an extensive and wider understanding on communities is important. Locating the underlying community structure in a network allows studying the network more easily and could provide insights into the function of the system represented by the network, as communities often correspond to functional units of systems. The study of communities and their properties also helps on revealing relevant groups of nodes, creating meaningful classifications, discovering similarities, or revealing unknown linkages between nodes. Communities have a strong impact in the behavior of a network as a whole and studying them is fundamental in order to expand the knowledge of the community structure beyond the local organization of the components of networks.

In this sub-section of the chapter, we study the overlapping community structure of our network by examining its characteristics and scale-free properties through the Clique Percolation Method (CPM) [31,32]. First, we detect communities and then we analyze the global structure of the whole network through the distribution functions of four basic quantities. In this analysis, our data set included ~3.4 million payments from the period of October 2013 to December 2014.

The majority of previous studies on communities have essentially been devoted to the description of structures inside the communities and their applications: communities representing real social groupings [33–35] communities in a co-authorship network representing related publications of specific topics [36], protein–protein interaction networks [37], communities in a metabolic network representing cycles and functional units in biology [38], and communities in the World Wide Web representing web pages with related contents [39]. Regarding community studies on economic networks and their applications, Vitali and Battiston [40] studied the community structure of a global corporate network and found that geography is the major driver of organization within that network. Fenn et al. [41] studied the evolution of communities of a foreign exchange market network in which each node represents an exchange rate and each link represents a time-dependent correlation between the rates. By using community detection, they were able to uncover major trading changes that occurred in the market during the credit crisis of 2008. Other related economic studies have focused on the overlapping feature of communities, such as in [42,43].



Most of the algorithms for community detection can be classified as divisive, agglomerative, or optimization-based methods, and each method has specific strengths and weaknesses. Previous studies on communities based on divisive and agglomerative methods consider that structures of communities can be expressed in terms of separated groups of clusters [44], but most of the real networks are characterized by well-defined statistics of overlapping communities. An important limitation of the popular node partitioning methods is that a node must be in one single community, whereas it is often more appropriate to attribute a node to several different communities, particularly in real-world networks. An example where community overlapping is commonly observed is in social networks where individuals typically belong to many communities such as: work teams, religious groups, friendship groups, hobby clubs, family, or other similar social communities. Moreover, members of social communities have their own communities and this, in turn, results in a very complex web of communities [32]. The phenomenon of community overlapping has been already noticed by sociologists but has barely been studied systematically for large-scale networks [31,45].

Networks have sections in which the nodes are more densely connected to each other than to the rest of the nodes in the network, and such sub-sections are called communities. Communities might exist in networked systems of different nature, such as economics, sociology, biology, engineering, politics, and computer science. There is no unique definition of community in the existing literature. Definitions change depending on the author and the type of study, and precisely one of the core issues in community detection is the lack of a unified definition of what is a community. We use the CPM definition because such algorithm allows overlapping nodes among communities, a condition that arises when a node is a member of more than one community. In economic systems, the nodes could frequently belong to multiple communities; therefore, forcing each node to belong to a single community could result into a misleading characterization of the underlying community structure.

An overlapping community graph is a network that represents links between communities. In our study, the nodes represent communities and the links represent shared nodes between communities. CPM is based on the density of links and the definition of community for this algorithm is local and it is not too restrictive. Overlapping communities arise when a node is a member of more than one community. CPM is based on the assumption that a community comprises overlapping sets of fully connected sub-graphs and detects communities by searching for adjacent cliques. A clique is a complete (fully connected) sub-graph. A $k$-clique is a complete subgraph of size $k$ (the number of nodes in the sub-graph). Two nodes are connected if the $k$-cliques that represent them share $k-1$ members. The method begins by identifying all cliques of size $k$ in a network. When all the cliques are identified, then a $N_c \times N_c$ clique–clique overlapping symmetric matrix $\boldsymbol{O}$ can be built, where $N_c$ is the number of cliques and $\boldsymbol{O}_{ij}$ is the number of nodes shared by cliques $i$ and $j$ [46]. This overlapping matrix $\boldsymbol{O}$ encodes all the important information needed to extract the $k$-clique communities for any value of $k$. In the overlapping matrix $\boldsymbol{O}$, rows and columns represent cliques and the elements are the number of shared nodes between the corresponding two cliques. Diagonal elements represent the size of the clique and when two cliques intersect they form a community. For certain $k$-values, the $k$-clique communities form such connected clique components in which the nearby cliques are linked to each other by at least $k-1$ adjacent nodes. In order to find these components in the overlapping matrix $\boldsymbol{O}$, one should keep the entries of the overlapping matrix which are larger than or equal to $k-1$, set the others to zero and finally locate the connected components of the overlapping matrix $\boldsymbol{O}$. The formed communities are the identified separated components.



For our method, it is important to select a proper parameter $k$. This parameter affects the constituents of the overlapping regions between communities. The larger the parameter $k$ is, the less the number of nodes which can arise in the overlapping regions. When $k \to \infty$, the maximal clique network is identical to the original network and no overlap is identified. The choice of $k$ depends on the network. It is observed from many real-world networks that the typical value of $k$ is often between 3 and 6 [47].

Figure 8.5 shows a plot of the number of communities and the average size of the communities at different $k$-values. As $k$ increases the number of communities decreases while the size of communities increases fast. When $k$ decreases the number of communities increases fast and the size of the communities remains low. In order to obtain the optimal value $k$ for our network, we tested different values ranging from 3 to 10 and *a posteriori* we found that the optimal number is $k = 5$. When $k < 5$ a high number of communities arises and the partitions become very low; when $k > 5$ a lower number of communities arises and the partitions become unreal. At the level of $k = 5$, we obtain the richest partition with the most widely distributed cluster sizes set for which no giant community appears.

For visualization purposes and in order to draw a readable map of the network, Figure 8.6 shows a graphic view of a representative section of the overlapping network of communities where big and small communities can easily be distinguished. This image depicts 25 overlapping communities and each colored circle represents a node which, in turn, represents an overlapping community. The links represent the shared nodes between the communities. The size of the nodes characterizes the size of each community. For example, the big node in the middle represents a community with 61 companies.



The usefulness of identifying the communities within this network lies in how this information could be used in a practical scenario. The output of the

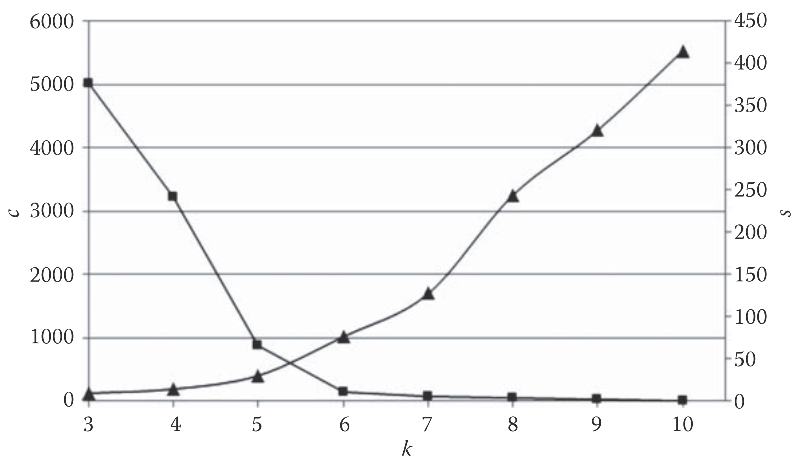



**FIGURE 8.5:** Plot of average community sizes "*s*" and number of communities "*c*" as $k$ increases. Squares represent the number of communities and triangles represent the size of the communities.

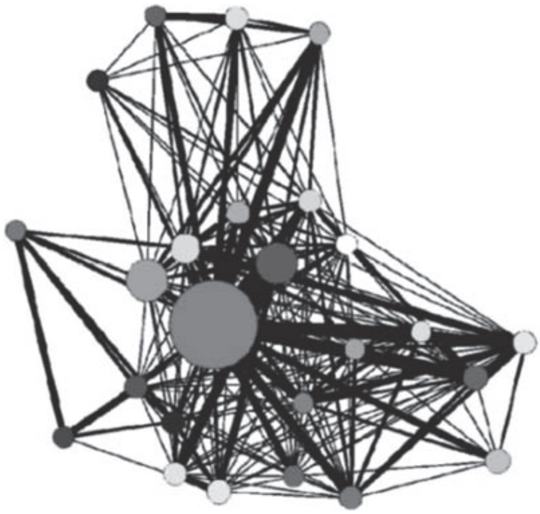


**FIGURE 8.6:** Visual representation of a section of the overlapping network of communities (Estonian network of payments). Nodes represent communities and the links represent shared nodes between communities.



community analysis could be used for targeted marketing. For example, it could be useful when integrating criteria for creating target groups of companies or customers to whom certain products or lines of products would be offered. Companies included in the same community would be located in the same target group and later on after a product offer is made it would be possible and interesting to assess the contagion effect of the product acquisition among companies of the same communities who received the offer. Another useful application is that the output of the analysis could help on creating customer-level segmentations or marketing profiles. Knowing the community (or communities) where a company or customer belongs to could be one of the drivers for creating a customer profile or grouping level. An alternative usage of the results of the community detection analysis is in predictive analytics for building churn models. Churn models usually define a measure of the potential risk of a customer cancelling a product or service and provide awareness and metrics to execute retention efforts against churning. Additionally, community detection analysis could be used as input for product affinity analysis and recommender systems. Affinity analysis is a data mining technique that helps to group customers based on historical data of purchased products and is used for cross-selling product recommendations. Another useful and immediate application is in product acquisition propensity models. These models calculate customers' likelihood to acquire a product based on a myriad of variables and the output of the overlapping community analysis could be input for such propensity models and support efficiency in sales processes.

### 8.4.1 Structure of communities

We studied the global community structure of our network by inspecting the distribution functions of four elemental quantities: community size $P(s)$, overlap size $P(s_o)$ community degree $P(d)$, and membership number $P(m)$. The distributions of such quantities are shown in Figure 8.7a–d which show important statistics that describe the community structure of our network. In general, nodes in a network can be characterized by a membership number which represents the number of communities a node belongs to. This means that, for example, any two communities may share some of their nodes which correspond to the overlap size between those communities. There is also a network of communities where the overlaps are the links and the communities are the nodes, and the number of such links is called community degree. The size of any of those communities is defined by the number of its nodes.

Figure 8.7a displays the cumulative distribution function of the community size $P(s)$. This is the probability for a community of having a community size higher or equal to $s$, calculated over different points in time $t$ (where $t$ is the time expressed in months, $t = 1$ is October, $t = 2$ is November, etc.). The overall distribution of community sizes resembles a power law $P(s) \propto s^a$, where $a$ is the scaling exponent, and a power law is valid nearly over all times $t$, suggesting

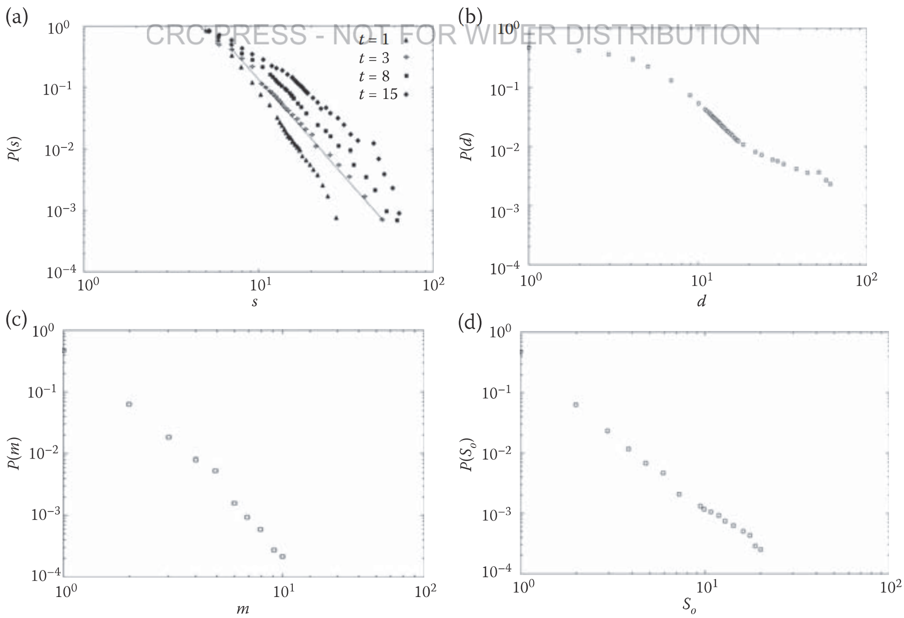




**FIGURE 8.7:** (a) Cumulative community size distribution at different times $t$. (b) Cumulative distribution of community degrees $d$. (c) Cumulative distribution function of the membership number $m_i$. (d) Cumulative distribution function of the overlap size $s_o$.

there is no characteristic community size in the network. The sizes of the communities on $t = 1$ are smaller than in the rest of the months; as time increases the size increases, particularly the size of the largest communities. The distribution at different moments in time follows similar decaying patterns, but in general, the scaling tail is higher as $t$ increases. The shapes of the power laws observed in the community size distributions of Figure 8.7a suggest there is no characteristic community size in the network. A fat tail distribution implies that there are numerous small communities coexisting with few large communities [48,49]. The scaling exponent when $t = 3$ is $-2.8$ (included for eye guideline) and Equation 8.9 is

$$P(s) \propto s^{-2.8}. \tag{8.9}$$

In a network of overlapping communities, the overlaps are represented by the links and the number of those links is represented by the community degree $d$. Then, the degree $d$ is the number of communities another community overlaps with. Figure 8.7b shows the cumulative distribution of the community degrees in the network. There are some outstanding community degrees by the end of the tail and these include communities that cluster the majority of the biggest customers in the network. The central part of the distribution decays faster than the rest of the distribution. There is an observable curvature in the log–log plot, however no approximation method fitted the distribution. This plot shows that the maximum number of degrees $d$ is 63 and corresponds to a relatively small quantity of nodes.



The membership number $m_i$ represents the number of communities a node $i$ belongs to while the overlap size $s_o$ is the number of nodes that two communities share. Figure 8.7c and d show the distributions of both measures, respectively. Both distributions seem to have a power-law behavior and indicate that there is no characteristic scale in the overlapping size or in the membership sizes. Regarding the overlap size, the range to which the communities overlap with each other is also an important property of our network.

As shown in Figure 8.7c, the largest membership number found in the network was 10, meaning that a company can belong to maximum 10 different communities simultaneously. This plot shows that the fraction of nodes that belong to many different communities is quite small, while the fraction of nodes belonging to at least 1 community is high. For example, when $m = 1$ the percentage of nodes that belong to at least one community is 50%, while the percentage of nodes that belong simultaneously to 10 communities ($m = 10$) is extremely small. However the rest of the communities belong to at least 2 or more communities.

In our previous study [22], we found scale-free properties in the degree distributions of the Estonian network of payments and it is very interesting to observe that the scale-free property is also preserved at a higher level of organization where overlapping communities are present.

## 8.5 Multifractal Networks

In the late 1960s Benoit Mandelbrot was the first to coin the term "fractal" and he also was the first one in describing the fractal geometry of nature [27], and since then the fractal approach has been widely spread and used in extensive research studies related with the underlying scaling of different complex structures, including networks.

Whether if a single fractal scaling spans or not all the constituents or areas of a system, is a fundamental issue that helps on distinguishing when a system is multifractal or just fractal. One scaling exponent is enough to characterize completely a monofractal process. Monofractals are considered as homogeneous objects because they have the same scaling properties branded by one singularity exponent. Instead, a multifractal object requires several exponents to characterize its scaling properties. Multifractals are inherently more complex and inhomogeneous than monofractals and portray systems with high variations or fluctuations that originate from specific characteristics.

Fractal and multifractal analysis can help to reveal the structure of all kinds of systems in order to have a better understanding of them. In particular, both approaches have many different interesting applications in economy. An interesting line of research is related with the relevance and applicability of fractal and multifractal analysis in social and economic topics. Inaoka et al. [17] showed that the study of the structure of a banking network provides useful insight from practical points of view. By knowing and understanding the structure and characteristics of banking networks (in terms of transactions and their patterns), a systemic contagion could potentially be prevented. In their study, these authors showed that the network of financial transactions of Japanese financial institutions has a fractal structure. Regarding social studies, Lu et al. [50] showed the importance of road patterns for urban transportation capacity based on fractal analysis of such network. In this study, the authors were able to link the fractal measurement with city mass measurements. A few recent studies have focused on the analysis of the changes of multifractal spectra across time to assess changes in economy during crisis periods [51]. Some other studies have focused on gathering empirical evidence of the common multifractal signature in economic, biological, and physical systems [52].



Fractal analysis helps to distinguish global features of complex networks, such as the fractal dimension. However, the fractal formalism is insufficient to characterize the complexity of many real networks which cannot be described by a single fractal dimension. Furuya and Yakubo [6] demonstrated analytically and numerically that fractal scale-free networks may have multifractal structures in which the fractal dimension is not sufficient to describe the multiple fractal patterns of such networks, therefore, multifractal analysis rises as a natural step after fractal analysis.

Multifractal structures are abundant in social systems and in a variety of physical phenomena. Inhomogeneous systems which do not follow a self-similar scaling law with a sole exponent could be multifractal if they are characterized by many interweaved fractal sets with a spectrum of various fractal dimensions. Multifractal analysis is a systematic approach and a generalization of fractal analysis that is useful when describing spatial heterogeneity of fractal patterns [53]. Multifractal network analysis requires taking into account a physical measure, like the number of nodes within a box of specific size in order to analyze how the distribution of such number of nodes scales in a network as the size of the box grows or reduces. In the last years, numerous algorithms for calculating the fractal dimension and studying self-similar properties of complex networks have been developed and tested extensively [31,54–57]. Song et al. [58] developed a method for calculating the fractal dimension of a complex network by using a box-covering algorithm and identified self-similarity as a property of complex networks [59]. Additionally, several algorithms and studies on multifractal analysis of networks have been proposed and developed recently [60–63].

In this sub-section of the chapter, we analyze fractal and multifractal properties of the large-scale economic network of payments of Estonia. We perform a fractal scaling analysis by estimating the fractal dimension of our network and its skeleton. Then, we study the multifractal behavior of the network by using a sandbox algorithm for complex networks to calculate the spectrum of the generalized fractal dimensions $D(q)$ and mass exponents $\tau(q)$.



### 8.5.1 Fractal network analysis

According to Song et al. [59], the box-counting algorithm is an appropriate method to study global properties of complex networks. The fundamental relation of fractal scaling is based on the box-covering method which counts the total number of boxes that are needed to cover a network with boxes of certain size. The box-covering method is equivalent to the box-counting method widely used in fractal geometry and is a basic tool for measuring the fractal dimension of fractal objects embedded in Euclidean space [64]. However, the Euclidean metric is not well defined for networks, thus we use the networks' adaptation [63] of the random sequential box-covering algorithm [65] in order to calculate the fractal dimension of our network and its skeleton. This method involves a random process for selecting the position of the center of each box. We let $N_B(r_B)$ be the minimum number of boxes needed to tile the whole network, where the lateral size of the boxes is the measure of radius $r_B$ as follows:

$$N_B(r_B) \sim r_B^{-d_B}, \tag{8.10}$$

where $d_B$ is the fractal dimension. If we measure the number of $N_B$ for different box sizes, then it is possible to obtain the fractal dimension $d_B$ by obtaining the power-law fitting of the distribution. The algorithm selects a random node at each step, and this node is the seed that will be the center of a box. Then we search the network by distance $r_B$ from the seed node and cover all the nodes that are located within that distance, but only if they have not been covered yet. Later we assign the newly covered nodes to the new box; if there are no more newly covered nodes then the box is removed. This process is repeated until all the nodes of the network belong to boxes. Before using the algorithm we calculate the skeleton of our network.

One of the main challenges of complex network studies is the identification of critical structural features that are underneath the network's complexity. This is related with the basic concept of: the distinctive character of a whole is inside just a few of its parts, for example, in specific colors and shapes of a painting, particular notes or tunes in a song or certain keywords in a text or speech. This basic concept is also true for complex networks, where only a few parts of the whole network reflect the most important properties of it. For example, in large-scale networks, only a small number of links are critical for the network to exist as a whole. A skeleton network is generally smaller than the original and it reproduces all the fundamental properties of the whole because it contains the essence of the network. Grady et al. [66] analyzed the network of international flight connections and discovered that the skeleton network consists of just 6.76% of the original network. The skeleton network concept can be used to detect epidemic propagations of disease when indicating which individuals are key participants in a social network or it can be useful when describing ecosystems to identify the species that should not be damaged at all to avoid jeopardizing the whole network.



The concept of skeleton was first introduced by Kim et al. [67]. The skeleton is a particular type of spanning tree based on the link betweenness centrality (a simplified quantity to measure the traffic of networks) that is entrenched beneath the original network. The skeleton provides a shell for the fractality of the network and is formed by links with the highest betweenness centralities. Only the links that do not form loops are included. The remaining links from the original network which are not included in the skeleton are local shortcuts that contribute to loop formation, meaning that the distance between any two nodes in the original network may increase in the skeleton. A fractal network has a fractal skeleton beneath which is distressed by these local shortcuts but it preserves fractality. For a scale-free network the skeleton also follows a power-law degree distribution where the degree exponent might differ slightly from that of the original network. When studying the origin of fractality in networks, actually the skeleton is more useful than the original network itself due to its unsophisticated and simplistic tree structure [68]. In general, the skeleton preferentially collects the sections of the network where betweenness is high and this preserves the structure and simplifies its complexity. Therefore, by looking at the properties of the skeleton it is easier to appreciate the topological organization of the original network.

In order to calculate the skeleton of a complex network, the link betweenness of all the links in the network has to be calculated. The betweenness centrality of

**FIGURE 8.8:** Graph representation of the skeleton of the Estonian network of payments.

a network (for a link or a node), is defined as follows:



$$b_i = \sum_{j,k \in N, j \neq k} \frac{n_{jk}(i)}{n_{jk}}, \tag{8.11}$$

where $N$ is the total number of nodes, $n_{jk}$ is the total number of shortest paths between nodes $j$ and $k$, $n_{jk}(i)$ is the total number of shortest-paths linking nodes $j$ and $k$ that passes through the node $i$. In order to perform the fractal scaling analysis, we used Dijkstra's algorithm [69]; then we used the box-covering algorithm to calculate the fractal dimension of the network and the skeleton to compare both values.

We present a fractal scaling analysis by using the box-counting algorithm expressed in Equation 8.10 and we calculated the fractal dimension of our network and its skeleton. Figure 8.8 shows a visualization of the graph representation of the skeleton of our network. The box-covering method yields a fractal dimension $d_{Bs} = 2.32 \pm 0.07$ for the skeleton network and for the original network the fractal dimension is $d_{Bo} = 2.39 \pm 0.05$.

The comparison of the fractal scaling in our network and its skeleton structure revealed its own patterns according to the fractality of the network. Figure 8.9 shows a fractal scaling representation of our network and its skeleton, where the fractal dimension is the absolute value of the slope of the linear fit.

As seen in the plot of Figure 8.9, the respective number of boxes needed to cover both networks is very similar but not identical, actually more boxes were needed for covering the skeleton. The largest distance between any two

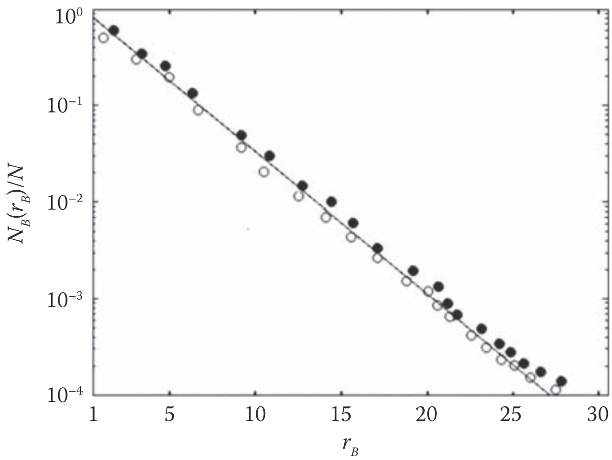



**FIGURE 8.9:** Fractal scaling representation of our network. The original network (○) and the skeleton network (●). The straight line is included for guidance and has a slope of 2.3. The analysis includes only the giant connected cluster of the network.

nodes in the network of payments is 29, while the largest distance between any two nodes in the skeleton network is 34.



## 8.5.2 Multifractal network analysis

Scale-free networks are commonly observed in a wide array of different contexts of nature and society. In the first sub-section of this chapter, we have shown that the Estonian network of payments has scale-free properties characterized by power-law degree distributions.

In general, multifractality is expected to appear in scale-free networks due to the fluctuations that occur in the density of local nodes. Tél et al. [70] introduced a sandbox algorithm based on the fixed-size box-counting algorithm [71] which was used and adapted for multifractal analysis of complex networks by Liu et al. [61]. In order to determine the multifractal dimensions of our complex network, we chose this adapted sandbox algorithm because it is precise, efficient, and practical. Moreover, a study by Song et al. [53] has shown that this algorithm gives better results when it is used in unweighted networks, and this is our case.

The fixed-size box-counting algorithm is one of the most known and efficient algorithms for multifractal analysis. For a given probability measure $0 \leq \mu \leq 1$ in a metric space $\boldsymbol{\Omega}$ with a support set $E$, we consider the following partition sum:

$$Z_{\varepsilon}(q) = \sum_{\mu(B)\neq 0} [\mu(B)]^{q}, \tag{8.12}$$

where the parameter $q \in R$, and describes the moment of the measure. The sum runs over all different non-overlapping (or non-empty) boxes $B$ of a given size $\varepsilon$ that covers the support set $E$. From this definition, it is easy to obtain $Z_\varepsilon(q) \geq 0$ and $Z_\varepsilon(0) = 1$. The function of the mass exponents $\tau(q)$ of the measure $\mu$ is defined by

$$\tau(q) = \lim_{\varepsilon \to 0}\left(\frac{\ln Z_\varepsilon(q)}{\ln \varepsilon}\right). \tag{8.13}$$

Then, the generalized fractal dimensions $D(q)$ of the measure $\mu$ are defined as follows:

$$D(q) = \frac{\tau(q)}{q-1}, \quad q \neq 1, \tag{8.14}$$

and

$$D(1) = \lim_{\varepsilon \to 0} \frac{Z_{(1,\varepsilon)}}{\ln \varepsilon}, \quad q = 1, \tag{8.15}$$

where

$$Z_{1,\varepsilon} = \sum_{\mu(B) \neq 0} \mu(B) \ln \mu(B). \tag{8.16}$$



The generalized fractal dimensions $D(q)$ can be estimated with linear regression of $[\ln Z_\varepsilon(q)]/[q-1]$ against $\ln \varepsilon$ for $q \neq 1$, and similarly a linear regression of $Z_{1,\varepsilon}$ against $\ln\ \varepsilon$ for $q = 1$. $D(0)$ is the fractal dimension or the box-counting dimension of the support set $E$ of the measure $\mu$, $D(1)$ is the information dimension and $D(2)$ is the correlation dimension.

For a complex network, a box of size $B$ can be defined in terms of the distance $l_B$, which corresponds to the number of links in the shortest path between two nodes. This means that every node is less than $l_B$ links away from another node in the same box. The measure $\mu$ of each box is defined as the ratio of the number of nodes that are covered by the box and the total number of nodes in the whole network.

Multifractality of a complex network can be determined by the shape of $\tau(q)$ or $D(q)$ curves. If $\tau(q)$ is a straight line or $D(q)$ is a constant, then the network is monofractal; similarly if $D(q)$ or $\tau(q)$ have convex shapes, then the network is multifractal. A multifractal structure can be identified by the following signs [72]: multiple slopes of $\tau(q)$ versus $q$, non-constant $D(q)$ versus $(q)$ values and $f(a)$ versus $a$ value covers a broad range (not accumulated at nearby non-integer values of $a$).

Firstly, we calculate the shortest-path distance between any two nodes in the network and map the shortest-path adjacency matrix $B_{N\times N}$ using the payments adjacency matrix $A_{N\times N}$. Then we use the shortest-path adjacency matrix $B_{N\times N}$ as input for multifractal analysis. The central idea of the sandbox algorithm is simply to select a node of the network in a random fashion as the center of a sandbox and then count the number of nodes that are inside the sandbox. Initially, none of the nodes has been chosen as a center of a box or as a seed. We set the radius $r$ of the sandbox which will be used to cover the nodes in the range $r \in [1, D]$, where $D$(diameter) is the longest distance between nodes in the network and radii $r$ are integer numbers. We ensure that the nodes are chosen randomly as center nodes by reordering the nodes randomly in the whole network. Depending on the size $N$ of the network, we choose $T$ nodes in random order as centers of $T$ sandboxes; then we find all the neighboring nodes within radius $r$ from the center of each box. We count the number of nodes contained in each sandbox of radius $r$, and denote that quantity by $S(r)$. We calculate the statistical averages $[S(r)^{q-1}]$ of $[S(r)^{q-1}]$ over all the sandboxes $T$ of radius $r$. The previous steps are repeated for each of the different values of radius $r$ in order to obtain the statistical average $[S(r)^{q-1}]$ and use it for calculating linear regression.

The generalized fractal dimensions $D(q)$ of the measure $\mu$ are defined by

$$D(q) = \lim_{r \to 0} \frac{\ln [S(r)/S(0)]^{q-1}}{\ln (r/d)} \frac{1}{q-1}, \quad q \in R \tag{8.17}$$

or rewritten as



$$\ln ([S(r)]^{q-1}) \propto D(q)(q-1) \ln\left(\frac{r}{d}\right) + (q-1)\ln(S_0), \tag{8.18}$$

where $S(0)$ is the size of the network and the brackets mean taking statistical average over the random selection of the sandbox centers. We run the linear regression of $\ln([S(r)]^{q-1})$ against $(q-1)\ln(r/d)$ to obtain the generalized fractal dimensions and similarly, calculate the linear regression of $\ln([S(r)]^{q-1})$ against $\ln(r/d)$ to obtain the mass exponents $\tau(q)$. From the shapes of the generalized fractal dimension curves, we can conclude if multifractality exists or not in our network.

Linear regression is an important step to obtain the correct range of radius $r \in [r_{min}, r_{max}]$ that is needed to calculate the generalized fractal dimensions (defined by Equations 8.17 and 8.18) and the mass exponents (defined by Equation 8.13). We found an appropriate range of radii $r$ within the range of the interval located between 2 and 29 for linear regression, thus we selected this linear fit scaling range to perform multifractal analysis (we set the range of $q$-values from $-7$ to 12).

We calculated $\tau(q)$ and the $D(q)$ curves using the sandbox algorithm by Liu et al. [61] and based upon the shapes obtained from the spectrum in Figure 8.10a and b, it can be seen that the curves are non-linear, suggesting that the network is multifractal.

In Figure 8.10b, the $D(q)$ function decreases sharply after the peak reaches its end when $q$ is $-4$. This could be interpreted as the high densities around

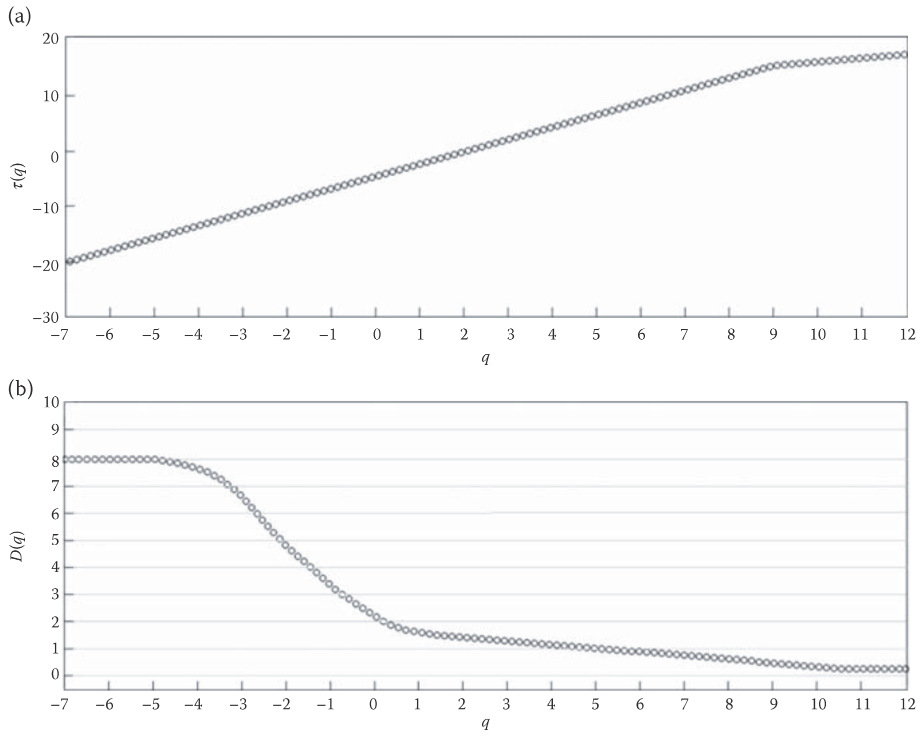





**FIGURE 8.10:** (a) Plot of mass exponents $\tau(q)$ as function of $q$. (b) Plot of generalized fractal dimensions $D(q)$ as function of $q$. Curves indicated by circles represent numerical estimations of the mass exponents and generalized fractal dimensions, respectively.

the hubs in the network. The hubs have a high number of links connected to them; therefore, the density of links around the sections near the hubs is higher than in other parts of the network. These hub nodes or important companies have a noticeable larger amount of business partners (for example: customers, suppliers, or any other business parties that interact financially) than the rest of the companies in the network have, and it is interesting to observe that this characteristic can be explored and identified by looking at the values of $D(q)$ spectra. The multifractality seen in our network reveals that the system cannot be described by a single fractal dimension suggesting that the multifractal approach provides a better characterization; hence, this means that the Estonian economy is multifractal.

The quantity $\Delta D(q)$ describes the changes on link density in our network. We use $\Delta D(q) = D(q)_{max} - \lim D(q)$ to observe how the values of $D(q)$ change along the spectrum. From Figure 8.10b, we found that $\lim D(q) = 0.37$ and $D(q)_{max} = 7.8$ and this means that $\Delta D(q) = 7.43$. A large $D(q)$ value means that the link distribution is very irregular, suggesting that there are areas of hubs where the links are densely grouped contrasting with areas where the nodes are connected with only just few links. In our network, this means that just a few companies have the role of hubs, while the rest are just small participants of the payments network. Table 8.3 shows a comparison of the maximum values of $D(q)$ in different networks.

**TABLE 8.3:** Comparison of the maximum values of $D(q)$ in different networks

| Network | Number of nodes | Highest $D(q)$ | Reference |
|---|---|---|---|
| Pure fractal network | 6222 | 2.8 | [60] |
| Small-world network | 6222 | 6.6 | [60] |
| Semi fractal network | 6222 | 3.1 | [60] |
| Sierpinski weighted fractal network | 9841 | 2.0 | [53] |
| Cantor dust weighted fractal network | 9841 | 3.2 | [53] |
| High-energy theory collaboration weighted network | 8361 | 6.0 | [53] |
| Astrophysics collaboration weighted network | 16,706 | 6.2 | [53] |
| Computational geometry collaboration weighted network | 7343 | 5.1 | [53] |
| Barabási and Albert model scale-free network | 10,000 | 3.6 | [61] |
| Newman and Watts model small-world network | 10,000 | 4.8 | [61] |
| Erdös–Rényi random graph model | 10,000 | 3.9 | [61] |
| Barabási and Albert model scale-free network | 7000 | 3.4 | [63] |
| Random network | 5620 | 3.5 | [63] |
| Random network | 449 | 2.4 | [63] |
| Protein–Protein interaction network: Human | 8934 | 4.9 | [63] |
| Protein–Protein interaction network: *Arabidopsis thaliana* | 1298 | 2.5 | [63] |
| Protein–Protein interaction network: *C. elegans* | 3343 | 4.5 | [63] |
| Protein–Protein interaction network: *E. coli* | 2516 | 4.1 | [63] |
| Small-world network | 5000 | 3.0 | [63] |
| Estonian network of payments | 16,613 | 7.8 | [22] |

## 8.6 Conclusions

Complex networks can be considered as the skeleton of complex systems and they are present in many kinds of social, economic, biological, chemical, physical, and technological systems. In this chapter, we have reviewed global properties and statistics related with the topological structure of the large-scale payments network of an entire country (Estonia) by using payments data. Additionally, we have reviewed some topics related with its community structure and moreover, we have analyzed some aspects related with multifractal and fractal properties.

In the network of Estonian payments, we found scale-free degree distributions, small-world property, low clustering coefficient, disassortative degree, and heterogeneity. Its scale-free structure indicates that a low number of companies in Estonia trade with a high number of companies, while the majority of the companies trade with only few. The clustering coefficient distribution suggests the existence of a hierarchic structure in the network. Our network is a small world with just $7^\circ$ of separation. The connectivity is smaller than the overall clustering coefficient therefore, our network is not random. The diameter value suggests there is a preference among companies for particular paths of money.

We tested the robustness of the network with an approach that focuses on the collective influencer nodes. First, we located the key nodes that prevent the network of breaking into disconnected components. The simulation assumed a targeted removal of key nodes which cause a quick growth in the average shortest path length until the network was destroyed at an optimal percolation threshold of 6%, while in the random removal the damage was extremely small. This revealed the robustness of our economic network against random attacks but also revealed its vulnerability to smart attacks. The low percentage of the optimal percolation threshold reveals that the most influential companies in the network are not necessarily the most connected ones or those having more economic activity and that a small quantity of companies maintains the whole network unified.



Later, we analyzed the community structure of our network by using the CPM. We found that there are scale-free properties in the statistical distributions of the community structure, too. Size, overlap, and membership distributions follow shapes that are compatible with power laws. Power-law distributions have already appeared in this network at a global scale in the level of nodes, and in this community study we have shown that power laws are present at the level of overlapping communities, too.

An immediate application for the community detection output is that it can be used in targeted marketing activities, as input for predictive analytical models such as in product acquisition propensities, churn, product affinity analyses, for creating marketing profiles or customer segmentations and for creating customer target lists for product offering (in an effort to propagate consumer buzz effects). Further applications for community detection in similar economic networks could involve identification of patterns between companies, tracking suspicious business activities, and strengthening relationships between companies of the same community for improving performance of the whole network.

In the last part of the chapter, we presented a fractal and multifractal analysis of the network. We identified the underlying structure of the network (its skeleton) and measured the fractal dimension of the skeleton to compare it with the fractal dimension of the original network. Both fractal dimensions were similar but the fractal dimension of the skeleton was slightly smaller. We also analyzed the general multifractal structure by calculating the spectrum of the mass exponents $(q)$ and the generalized fractal dimension $D(q)$ curves, through a sandbox algorithm for multifractal analysis of complex networks. Our results indicated that multifractality exists in the Estonian network of payments, and this suggests that the Estonian economy is multifractal (from the point of view of networks). We found large values of $D(q)$ spectra, which mean that the distribution of links is quite irregular in the network, suggesting there are specific nodes which hold densely connected links while other nodes hold just a few links. This type of structure could be relevant when critical events occur in the economy that could threaten the whole network.

It is important to continue studying the structures and characteristics of economic complex networks in order to be able to understand their underlying processes and to be able to detect patterns that could be useful for predicting or forecasting events and trends. The addition of evidence through empirical studies in favor of fractality, multifractality, communities' detection, and structural properties of economic networks represents a step forward towards the knowledge on the unraveling of the complexity of economic systems.



### 8.6.1 Further applications

Regarding community structure in economic networks, a question that remains open for future research is to investigate if the similarities in communities' features amongst different complex networks arise randomly or if there are any unknown properties shared by all of them. Another interesting open line of research is to study the plausibility of predicting changes in a payment network through community detection analysis. Further applications in economic networks could involve strengthening relationships between companies of the same community to improve the performance of the whole network, targeted marketing, identification of patterns between companies, and tracking of suspicious business activities.

Further applications of multifractal studies in economic networks might involve examining the potential factors that drive the strength of the multifractal spectrum. Some applications could involve studying the origin of such factors. Another interesting line of research would be to study the patterns and the changes of the multifractal spectrum across different periods of time.

Particularly, it would be interesting to analyze such patterns during financial crisis periods for risk pattern recognition purposes. Also, it would be interesting to take into account different probability measures for such kind of multifractal analysis. Other direction of the studies could focus on building network models that attempt to forecast country money flows or potential industry growth trends based on data of transactions.

## References

1. Mantegna R.N., & Stanley H.E. 2000. *An Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge, UK: Cambridge University Press.

2. Standish R.K. 2008. *Intelligent Complex Adaptive Systems*. New South Wales, Australia: IGI Global, 105–124.

3. Estrada E. 2011. *The Structure of Complex Networks*. New York, NY: Oxford University Press.

4. Reka A., & Barabási A.L. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74(1), 47–97.

5. Dorogovtsev S.N., & Mendes J.F.F. 2003. *Evolution of Networks*. New York, NY: Oxford University Press.



6. Furuya S., & Yakubo K. 2011. Multifractality of complex networks. *Physical Review E*, 84(3), 036118.

7. Newman M.E.J. 2010. *Networks: An Introduction*. New York, NY: Oxford University Press.

8. Palla G., Barabási A.L., & Vicsek T. 2007. Quantifying social group evolution. *Nature*, 446, 664–667.

9. Watts D.J., & Strogatz S.H. 1998. Collective dynamics of small-world networks. *Nature*, 393(6563), 440–442.

10. König M.D., & Battiston S. 2009. From graph theory to models of economic networks, a tutorial in: Naimzada A.K., Stefani S., & Torriero A. (eds). *Networks, Topology and Dynamics*. Lecture Notes in Economics and Mathematical Systems. Germany: Springer, 613, 23–63.

11. Souma W., Fujiwara Y., & Aoyama H. 2006. Heterogeneous economic networks. in: Namatame A., Kaizouji T., & Aruka Y. (eds). *The Complex Networks of Economic Interactions*. Lecture Notes in Economics and Mathematical Systems. Germany: Springer, 567, 79–92.

12. Battiston S., Rodrigues J.F., & Zeytinoglu H. 2007. The network of inter-regional direct investment stocks across. *Advances in Complex Systems*, 10(1), 29–51.

13. Glattfelder J.B., & Battiston S. 2009. Backbone of complex networks of corporations: The flow of control. *Physical Review E*, 80(3), 036104. https://doi.org/10.1103/PhysRevE.80.036104

14. Nakano T., & White D. 2007. Network structures in industrial pricing: The effect of emergent roles in Tokyo supplier-chain hierarchies. *Structures and Dynamics*, 2(3), 130–154.

15. Reyes J., Schiavo S., & Fagiolo G. 2008. Assessing the evolution of international economic integration using random-walk betweenness centrality: The cases of East Asia and Latin America. *Advances in Complex Systems*, 11(5), 685–702.

16. Lublóy A. 2006. Topology of the Hungarian large-value transfer system. *Magyar Nemzeti Bank (Central Bank of Hungary) MNB Occasional Papers*, 57.

17. Inaoka H., Nimoniya T., Taniguchi K., Shimizu T., & Takayasu H. 2004. Fractal network derived from banking transactions—An analysis of network structures formed by financial institutions. *Bank of Japan Working Papers Series*, Bank of Japan, 04(4).

18. Soramäki K., Bech M.L., Arnold J., Glass R.J., & Beyeler W.E. 2007. The topology of interbank payment flows. *Physica A*, 379(1), 317–333.

19. Boss M., Helsinger H., Summer M., & Thurner S. 2004. The network topology of the interbank market. *Quantitative Finance*, 4(6), 677–684.



20. Iori G., & Jafarey S. 2001. Criticality in a model of banking crisis. *Physica A*, 299(1), 205–212.

21. Iori G., De Masi G., Precup O.V., Gabbi G., & Caldarelli G. 2007. A network analysis of the Italian overnight money market. *Journal of Economic Dynamics and Control*, 32(1), 259–278.

22. Rendón de la T.S., Kalda J., Kitt R., & Engelbrecht J. 2016. On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. *Chaos Solitons and Fractals*, 90, 18–27.

23. Rendón de la T.S., Kalda J., Kitt R., & Engelbrecht J. 2017. Detecting overlapping community structure: Estonian network of payments (submitted).

24. Rendón de la T.S., Kalda J., Kitt R., & Engelbrecht J. 2017. Fractal and multifractal analysis of complex networks: Estonian network of payments, *European Journal of Physics B*, 90(12), 234–241.

25. West D.B. 2003. *Introduction to Graph Theory*. Upper Saddle River, NJ: Prentice Hall.

26. Erdös P., & Rényi A. 1959. On random graphs. *Publicationes Mathematicae Debrecen*, 6, 290–297.

27. Mandelbrot B. 1983. *The Fractal Geometry of Nature*. New York, NY: Academic Press.

28. Jeong H., Tombor B., Reka A., Zoltan N.O., & Barabási A.L. 2000. The large-scale organization of metabolic networks. *Nature*, 407, 651–654.

29. Barabási A.L., & Zoltan N.O. 2004. Network biology: Understanding the cell's functional organization. *Nature Review Genetics*, 5(2), 101–103.

30. Morone F., & Makse H.A. 2015. Influence maximization in complex networks through optimal percolation. *Nature*, 524(7563), 65–68.

31. Palla G., Derényi I., Farkas I., & Vicsek T. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435, 814–818.

32. Derényi I., Palla G., & Vicsek T. 2005. Clique percolation in random networks. *Physics Review Letters*, 94(16), 160–202.

33. Traud A.L., Mucha J.P., & Porter M.A. 2012. Social structure of Facebook networks. *Physica A*, 391(16), 4165–4180.

34. González M.C., Herrmann H.J., Kertész J., & Vicsek T. 2007. Community structure and ethnic preferences in school friendship networks. *Physica A*, 379(1), 307–316.

35. Palla G., Barabási A.L., & Vicsek T. 2007. Quantifying social group evolution. *Nature*, 446(7136), 664–667.

36. Pollner P., Palla G., & Vicsek T. 2006. Preferential attachment of communities: The same principle, but a higher level. *Europhyics Letters*, 73(3), 478–484.



37. Lewis A.C.F., Jones N.S., Porter M.A., & Deane C.M. 2010. The function of communities in protein interaction networks at multiple scales. *BMC Systems Biolology*, 4, 100.

38. Guimerá R., & Amaral N. 2005. Functional cartography of complex metabolic networks. *Nature*, 433, 895–900.

39. Dourisboure Y., Geraci F., & Pellegrini M. 2007. Extraction and classification of dense communities in the web. *Proceedings of the 16th International Conference on the World Wide Web*, 1, 461–470.

40. Vitali S., & Battiston B. 2014. The community structure of the global corporate network. *Plos ONE*, 9(8), 104655. https://doi.org/10.1371/journal.pone.0104655

41. Fenn D. et al. 2009. Dynamic communities in multichannel data: An application to the foreign exchange market during the 2007–2008 credit crisis. *Chaos*, 19(3), 033119.

42. Piccardi C., Calatroni L., & Bertoni F. 2010. Communities in Italian corporate networks. *Physica A*, 389(22), 5247–5258.

43. Bóta A., & Krész M. 2015. A high resolution clique-based overlapping community detection algorithm for small-world networks. *Informatica*, 39, 177–187.

44. Yang Z., Algesheimer R., & Tessone C.J. 2016. A comparative analysis of community detection algorithms on artificial networks. *Scientific Reports*, 6, 30750.

45. Devi J.C., & Poovammal E. 2016. An analysis of overlapping community detection algorithms in social networks. *Proceedia Computer Sciences*, 89, 349–358.

46. Everett M.G., & Borgatti S.P. 1998. Analyzing clique overlap. *Connections*, 21(1), 49–61.

47. Shen H.W. 2013. *Community Structure of Complex Networks*. Germany: Springer Science & Business Media.

48. Newman M.E.J. 2004. Fast algorithm for detecting community structure in networks. *Phyical Review E*, 69, 066133.

49. Clauset A., Newman M.E.J., & Moore C. 2004. Finding community structure in very large networks. *Physical Review E*, 70(6), 066111. http://dx.doi.org/10.1103/PhysRevE.70.066111

50. Lu Y., & Tang J. 2004. Fractal dimension of a transportation network and its relationship with urban growth: A study of the Dallas-Fort Worth area. *Environmental and Planning B*, 31(6), 895–911.

51. Fotios M., & Siokis M. 2014. European economies in crisis: A multifractal analysis of disruptive economic events and the effects of financial assistance. *Physica A*, 395, 283–292.



52. Pont O., Turiel A., & Pérez-Vicente C.J. 2009. Empirical evidences of a common multifractal signature in economic, biological and physical systems. *Physica A*, 388(10), 2025–2035.

53. Song Y.Q., Liu J.L., Yu Z.G., & Li B.G. 2015. Multifractal analysis of weighted networks by a modified sandbox algorithm. *Scientific Reports*, 5, 17628.

54. Zhou W.X., Yiang Z.Q., & Sornette D. 2007. Exploring self-similarity of complex cellular networks: The edge-covering method with simulated annealing and log-periodic sampling. *Physica A*, 375(2), 741–752.

55. Gallos L.K., Song C., Havlin S., & Makse H.A. 2007. A review of fractality and self-similarity in complex networks. *Physica A*, 386, 686–691.

56. Schneider C.M., Kesselring T.A., Andrade J.S., & Herrmann H.J. 2012. Box-covering algorithm for fractal dimension of complex networks. *Physical Review E*, 86, 016707.

57. Eguiluz V.M., Hernández-García E., Piro O., & Klemm K. 2003. Effective dimensions and percolation in hierarchically structured scale-free networks. *Physical Review E*, 68(5), 551021–551024.

58. Song C., Gallos L.K., Havlin S., & Makse H.A. 2007. How to calculate the fractal dimension of a complex network: The box covering algorithm. *Journal of Statistical Mechanics: Theory and Experiment*, 3, P03006.

59. Song C., Havlin S., & Makse H.A. 2005. Self-similarity of complex networks. *Nature*, 433, 392–395.

60. Li B.G., Yu Z.G., & Zhou Y. 2014. Fractal and multifractal properties of a family of fractal networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2, P02020.

61. Liu J.L., Yu Z.G., & Anh V. 2015. Determination of multifractal dimensions of complex networks by means of the sandbox algorithm. *Chaos*, 25, 023103.

62. Wei D.J. et al. 2013. Box-covering algorithm for fractal dimension of weighted networks. *Scientific Reports*, 3, 3049.

63. Wang D.L., Yu Z.G., & Anh V. 2012. Multifractal analysis of complex networks. *Chinese Physics B*, 21(8), 080504.

64. Feder J. 1988. *Fractals*. New York, NY: Plenum.

65. Kim J.S., Goh K.I., Salvi G., Oh E., Kahng B., & Kim D. 2007. Fractality in complex networks: Critical and supercritical skeletons. *Physical Review E*, 75, 177.

66. Grady D., Thiemann C., & Brockmann D. 2012. Robust classification of salient links in complex networks. *Nature Communication*, 3(864), 864.

67. Kim D.H., Noh J.D., & Jeong H. 2004. Scale-free trees: The skeletons of complex networks. *Physics Review E*, 70, 046126.



68. Goh K.I., Salvi G., Kahng B., & Kim D. 2006. Skeleton and fractal scaling in complex networks. *Physics Review Letters*, 96(1), 018701.

69. Gibbons A. 1985. *Algorithmic Graph Theory*. Cambridge, UK: Cambridge University Press.

70. Tél T., Fülöp A., & Vicsek T. 1989. Determination of fractal dimensions for geometrical multifractals. *Physica A*, 159(2), 155–166.

71. Halsey T.C., Jensen M.H., Kadanoff L.P., Procaccia I., & Shraiman B.I. 1986. Fractal measures and their singularities: The characterization of strange sets. *Physics Review A*, 33(2), 1141–1151.

72. Grassberger P., & Procaccia I. 1983. Characterization of strange attractors. *Physics Review Letters*, 50(5), 346–349.

**Paper IV**

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2019) Detecting overlapping community structure: Estonian network of payments. Proceedings of the Estonian Academy of Sciences, 68(1) 79-88. DOI:10.3176/proc.2019.1.08.

Proceedings of the Estonian Academy of Sciences,
2019, **68**, 1, 79–88
https://doi.org/10.3176/proc.2019.1.08
Available online at www.eap.ee/proceedings

NETWORK SCIENCE
STUDIES

# Detecting overlapping community structure: Estonian network of payments

Stephanie Rendón de la Torre[a*], Jaan Kalda[a], Robert Kitt[a,b], and Jüri Engelbrecht[a]

[a] Department of Cybernetics, School of Science, Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia
[b] Swedbank AS, Liivalaia 12, 15038 Tallinn, Estonia

Received 28 September 2018, accepted 16 November 2018, available online 24 January 2019



**Abstract.** Revealing the community structure exhibited by real networks is a fundamental phase towards a comprehensive understanding of complex systems beyond the local organization of their components. Community detection techniques help in providing insights into understanding the local organization of the components of networks. We identified and investigated the overlapping community structure of an interesting and unique case of study: the Estonian network of payments. In order to perform the study, we used the Clique Percolation Method and explored statistical distribution functions of the communities, where in most cases we found scale-free properties. In this network the nodes represent Estonian companies and the links represent payments made between the companies. Our study adds to the literature of complex networks by presenting the first overlapping community detection analysis of a country's network of payments.

**Key words:** complex networks, economic networks, overlapping communities, scale-free networks.

## 1. INTRODUCTION

A network is a set of nodes connected by links. A complex network has nontrivial topological features and most of the real-world networks are complex. Complex networks can be described by a combination of local, global, and mesoscale approaches. The exploration of intermediate-sized structures that are responsible for "coupling" local properties demands partitioning networks into useful groups of nodes [1]. Networks have sections in which the nodes are more densely connected to each other than to the rest of the nodes in the network, and such sub-sections are called communities. Communities may exist in networked systems of different nature, such as economics, sociology, biology, engineering, politics, and computer science.

Community detection is a graph partitioning process that provides valuable insight into the organizational principles of networks and is essential for exploring and predicting connections that are not yet observed. Thus far, recent advances in the underlying mechanisms that rule the dynamics of communities in networks are limited, and this is why the achievement of an extensive and wider understanding of communities is important. Locating the underlying community structure in a network allows studying the network more easily and can provide insights into the function of the system represented by the network, as communities often correspond to functional units of systems. The study of communities and their properties also helps in revealing relevant groups of nodes, creating meaningful classifications, discovering similarities, or revealing unknown linkages between nodes. Communities have a strong impact on the behaviour of a network as a whole and studying them is

[*] Corresponding author, stretomx@gmail.com

fundamental in order to expand the knowledge of the community structure beyond the local organization of the components of networks.

The usefulness of identifying the communities within networks lies in how this information could be used in a practical scenario. Particularly, in the context of the bank industry the output of our community analysis (based on payments between companies which are customers of a bank) could be used for targeted marketing. For example, it could be used at the moment of integrating criteria for creating target groups of customers to whom certain products or lines of products would be offered. Customers in the same community would be included in the same target group and later on, after one offer is made to them, it would be possible and interesting to assess the contagion effects of the product acquisition among customers of the same communities who received the same offer. Another useful application is when helping to create customer-level segmentations or marketing profiles. To know the community (or communities) where a customer belongs to could be one of the main features for creating customer profiles or clustering levels. An alternative usage of the output of community analysis is in predictive analytics, for example, when building churn models. Churn models usually define a measure of the potential risk of a customer cancelling a product or service and provide awareness and metrics to execute retention efforts against churning. The communities to which the companies/customers belong to could be used as variables or features when using logistic regression, random forest, or neural network models. Additionally, community detection analysis could be used as input for product affinity and recommender systems. Affinity analysis is a data mining technique that helps group customers based on historical data of purchased products and is used for cross-selling product recommendations. Another useful and immediate application is in product acquisition propensity models. These models calculate customers' likelihood to acquire a product after an offer is made based on a myriad of variables and with this evidence the sales process can become more efficient.

The objective of this study is to detect the overlapping community structure of the large-scale payments network of Estonia by examining its characteristics and scale-free properties through the Clique Percolation Method [2,3]. First, we detect communities and then we analyse the global structure of the network through the distribution functions of four basic quantities.

The research questions for this study are the following: Which is the community structure of the Estonian network of payments? Are there scale-free properties in the community structure?

Section 1 provides a general introduction and an overview of the objectives. In Section 2 we deliver a description of the data set used in this study. Section 3 provides a literature review of studies related to similar networks and their applications. In Section 4 we present the method used to develop this study, while Section 5 presents our main results and findings. Finally, Section 6 concludes with a discussion of our results.

## 2. ANALYSED DATA

Our data set was obtained from Swedbank's databases. Swedbank is one of the leading banks in the Nordic and Baltic regions of Europe. The bank operates actively in Estonia, Latvia, Lithuania, and Sweden. All the information related to the identities of the nodes is very sensitive and thus will remain confidential and unfortunately cannot be disclosed. The data set is unique in its kind and very interesting since ~80% of Estonia's bank transactions are executed through Swedbank's system of payments. Hence, this data set reproduces well the transactional trends of the whole Estonian economy, so we use it as a proxy of the Estonian economy.

The data set consists of electronic company-to-company domestic payments, including data of 16 613 companies and 3.4 million payment transactions (October 2013–December 2014). In this study, the nodes represent companies and the links represent the payments between the companies. For simplicity, we focus on the basic case where the network of payments is defined by a symmetric payment adjacency matrix that represents the whole image of the network. We consider an undirected graph approach where two nodes have a link if they share one or more payments. Then each element represents a link as follows: $a_{ij}^u = a_{ji}^u$, where $a_{ij}^u = 1$ if there is a transaction between companies $i$ and $j$, and $a_{ij}^u = 0$ if there is no transaction between $i$ and $j$.

Tables 1 and 2 show main measures and statistics of our network of payments. The average degree of our network is $\langle k \rangle = 21$ while the diameter is 29. The average betweenness of links is 41, while it is 112 for nodes. The average shortest path length $\langle l \rangle = 7.3$ Our network is a "small world" with 7 degrees of separation, so on average any company can be reached by another within seven steps. An average degree of separation of 7 is a very small value for a network of size $N = 16\,613$. The network displays scale-free properties in the degree distribution. The degree distribution follows a power-law where the scaling exponent is $P(\geq k) \propto k^\wedge(-2.46)$. The network has a low average clustering coefficient of 0.19 and displays disassortative mixing behaviour, where high-degree nodes, represented by companies who have

**Table 1.** Network characteristics

| | |
|---|---|
| Number of companies analysed | 16 613 |
| Total number of payments analysed | 3 406 651 |
| Total value of transactions | 4 342 109 265* |
| Average value of transaction per customer | 99 904* |
| Maximum value of a transaction | 135 736* |
| Minimum value of a transaction (aggregated) | 1000* |
| Average volume of transaction per company | 76 |
| Maximum volume of transaction per company | 34 665 |
| Minimum volume of transaction per company (aggregated) | 20 |

* All monetary quantities are expressed in monetary units and not in real currencies in order to protect the confidentiality of the data set. The purpose of showing monetary units is to provide a notion of the proportions of quantities and not to show exact amounts of money.

**Table 2.** Summary of statistics

| Statistic | Value |
|---|---|
| $\langle k \rangle$ | 21 |
| $\gamma^o$ | 2.41 |
| $\gamma^i$ | 2.50 |
| $\gamma$ | 2.46 |
| $\langle C \rangle$ | 0.19 |
| $\langle l \rangle$ | 7.3 |
| $T$ | 0.13 |
| $D$ | 29 |
| $\langle \sigma \rangle$ (nodes) | 112 |
| $\langle \sigma \rangle$ (links) | 41 |

$\langle k \rangle$ = average degree, $\gamma^o$ = scaling exponent of the out-degree distribution, $\gamma^i$ = scaling exponent of the in-degree distribution, $\gamma$ = scaling exponent of the connectivity degree distribution, $\langle C \rangle$ = average clustering coefficient, $\langle l \rangle$ = average shortest path length, $T$ = connectivity %, $D$ = diameter, $\langle \sigma \rangle$ = average betweenness.

many counterparties such as business partners, service providers, clients, or suppliers, have a large number of links to companies which have only one link, or just few links.

## 3. LITERATURE DISCUSSION

Networks play an important role in a wide range of economic and social phenomena. The use of techniques and methods from graph theory has permitted economic network theory to expand the knowledge and understanding of economic phenomena in which the embeddedness of individuals or agents in their social or economic interrelations cannot be ignored [4]. For example, Souma et al. [5] studied a shareholder network of Japanese companies by analysing the companies' growth through economic networks dynamics. Other examples of interesting applications of complex networks in economics are provided by the regional investment or ownership networks where European company-to-company investment stocks show power-law distributions that allow predicting the investments that will be received or made in specific regions, based on the connectivity and transactional activity of the companies [6,7]. Nakano and White [8] showed that analytic concepts and methods related to complex networks can help to uncover structural factors that may influence the price formation for empirical market-link formations of economic agents. Reyes et al. [9] used a weighted network analysis focused on using random walk betweenness centrality to study why high-performing Asian economies had higher economic growth than Latin American economies between 1980 and 2005. Network-based approaches are very useful serving as a means for monitoring complex economic systems and may help in providing better control in managing and governing these systems. Another interesting line of research is related to network topology as a basis for investigating money flows of customer-driven banking transactions. A few recent papers describe the actual topologies observed in different financial systems [10–13]. Other works have focused on economic shocks and robustness in economic complex networks [14,15].

Regarding community studies on economic networks and their applications, Vitali and Battiston [16] studied the community structure of a global corporate network and found that geography is the major driver of organization within that network. In this study they also assessed the role of the financial sector in the architecture of the global corporate network by analysing centralities of communities. Fenn et al. [17] studied the evolution of communities of a foreign exchange market network in which each node represents an exchange rate and each link represents a time-dependent correlation between the rates. By using community detection, they were able to uncover major trading changes that occurred in the market during the credit crisis of 2008. Other economic communities' studies have focused on the overlapping feature of communities (e.g. [18,19]).

General community detection studies on other types of networks deal with communities representing real social groupings [20–22], communities in a co-authorship network representing related publications of specific topics [23], protein–protein interaction networks [24], communities in a metabolic network representing cycles and functional units in biology [25,26], and communities in the World Wide Web representing web pages with related contents [27].

Most algorithms for community detection can be distinguished in divisive, agglomerative, and optimization-based methods and each one has specific strengths and weaknesses. Previous studies on network communities based on divisive and agglomerative methods consider that structures of communities can be expressed in terms of separated groups of clusters [28–31], but most of the real networks are characterized by well-defined statistics of overlapping communities. An important limitation of the popular node partitioning methods is that a node must be in one single community, whereas it is often more appropriate to attribute a node to several different communities, particularly in real-world networks. An example where community overlapping is commonly observed is in social networks where individuals typically belong to many communities such as work teams, religious groups, friendship groups, hobby clubs, family, or other similar social communities. Moreover, members of social communities have their own communities and this in turn results in a very complex web of communities [3]. The phenomenon of community overlapping has already been noticed by sociologists but has been barely studied systematically for large-scale networks [2,32–35].

## 4. METHOD

Overlapping communities arise when a node is a member of more than one community. In economic systems the nodes can frequently belong to multiple communities, therefore, forcing each node to belong into a single community might result in a misleading characterization of the underlying community structure. The Clique Percolation Method [2,3] is based on the assumption that a community comprises overlapping sets of fully connected subgraphs and detects communities by searching for adjacent cliques. A clique is a complete (fully connected) subgraph. A $k$-clique is a complete subgraph of size $k$ (the number of nodes in the subgraph). Two nodes are connected if the $k$-cliques that represent them share $k-1$ members. The method begins by identifying all cliques of size $k$ in a network. When all the cliques are identified, then an $N_C \times N_C$ clique–clique overlapping symmetric matrix $\boldsymbol{O}$ can be constructed, where $N_C$ is the number of cliques and $\boldsymbol{O}_{ij}$ is the number of nodes shared by cliques $i$ and $j$ [36]. This overlapping matrix $\boldsymbol{O}$ encodes all the important information needed to extract the $k$-clique communities for any value of $k$. In the overlapping matrix $\boldsymbol{O}$, rows and columns represent cliques and the elements are the number of shared nodes between the corresponding two cliques. Diagonal elements represent the size of the clique and when two cliques intersect, they form a community.

For certain $k$ values, the $k$-clique communities form such connected clique components in which their nearby cliques are linked to each other by at least $k-1$ adjacent nodes. In order to find these components in the overlapping matrix $\boldsymbol{O}$, one should keep the entries of the overlapping matrix which are larger than or equal to $k-1$, set the others to zero, and finally locate the connected components of the overlapping matrix $\boldsymbol{O}$. Communities correspond to each one of the identified separated components [2].

## 5. RESULTS

### 5.1. Parameter $k$

For the Clique Percolation Method it is important to choose a parameter $k$. The parameter $k$ affects the constituents of the overlapping regions between communities. The larger the parameter $k$, the smaller the number of nodes which can arise in the overlapping regions. When $k \to \infty$, the maximal clique network is identical to the original network and no overlap is identified. The choice of $k$ will depend on the network. It is observed from many real-world networks that the typical value of $k$ is often between 3 and 6 [37].

Figure 1 shows a plot of the number of communities and the average size of the communities at different $k$ values. As $k$ increases, the number of communities decreases, while the size of the communities increases rapidly. When $k$ decreases, the number of communities increases rapidly, while the size of the communities remains low. In order to obtain the optimal value of $k$, we tested different values ranging from 3 to 10 and *a posteriori* we chose $k = 5$ because when $k < 5$, a large number of communities arise and the partitions become very low and giant communities appear (with sizes of more than 3200); at the level $k = 5$ we obtain a rich partition with the most widely distributed cluster sizes set for which no giant community appears.

An overlapping community graph is a representation of a network that denotes links between communities, where the nodes represent the communities and the links are represented by the shared nodes between communities. For visualization purposes and in order to draw a readable map of the network, Fig. 2 shows a graphic view of a representative section of the overlapping network of communities where big and small communities can easily be distinguished. Figure 2 depicts 25 overlapping

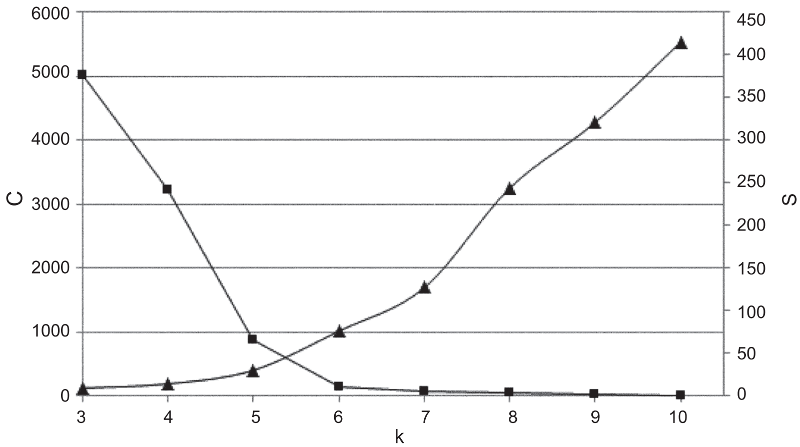



**Fig. 1.** Plot of the average size of community ($s$) and number of communities ($c$) as $k$ increases. Squares represent the number of communities and triangles represent the size of the communities.

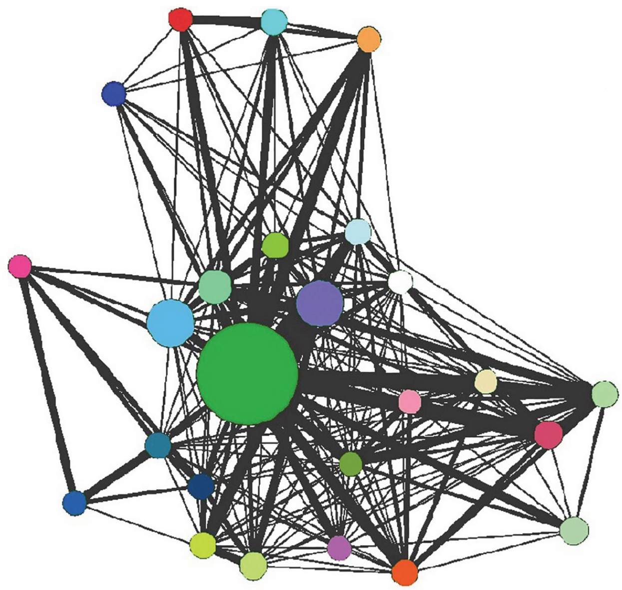


**Fig. 2.** Visual representation of a section of the overlapping network of communities (Estonian network of payments). The circles (nodes) represent communities and the black lines between them represent shared nodes between communities.

communities and each coloured circle represents a node which in turn represents an overlapping community. The links represent the shared nodes between the communities. The size of the nodes characterizes the size of each community. For example, the big node in the middle represents a community with 61 companies.

### 5.2. Structure of communities

In order to study and characterize the global community structure of our network, we investigated the distribution functions of the following four elementary quantities: community size $P(s)$, overlap size $P(s_o)$, community degree $P(d)$, and membership number $P(m)$. The aforementioned distributions are shown in Figs 3–7. In general, nodes in a network can be characterized by a membership number which is the number of communities a node belongs to. This means that, for example, any two communities may share some of their nodes which correspond to the overlap size between those communities. There is also a network of communities where the overlaps are represented by the links and the communities are represented by the nodes, and the number of such links is called community degree. The size of any of those communities is defined by the number of nodes it has.

The community size distribution is an important statistic that describes partially the system of communities. Figure 3 displays the cumulative distribution function of the community size $P(s)$ and shows the probability of a community to have a size higher than or equal to $s$ calculated over different points in time, where $t$ is the time in months. The overall distribution of community sizes resembles a power-law $P(s) \propto s^{\alpha}$, where $\alpha$ is the scaling exponent, and a power-law is valid nearly over all times $t$. The scaling exponent (calculated by maximum likelihood estimators) when $t = 3$ is –2.8 (included for eye guideline) and its corresponding equation is as follows:

$$P(s) \propto s^{-2.8}.$$

The sizes of the communities at $t = 1$ are smaller than in the rest of the months; as time increases, the size increases, particularly the size of the largest communities. The shapes of the power-laws observed in the community size distributions of Fig. 3 suggest there is no characteristic community size in the network. The distribution at different moments in time follows similar decaying patterns, but in general, the scaling tail is higher as $t$ increases. A fat tail distribution implies that there are numerous small communities coexisting with few large communities [38,39]. Figure 4 shows statistics of the community sizes across time and according to the plot, both the standard deviation of community sizes and the average size of communities increased with time.

In a network of overlapping communities, the overlaps are represented by the links and the number of those links is represented by the community degree $d$. Then, the degree $d$ is the number of communities another community overlaps. Figure 5 shows the cumulative distribution of the community degrees in the network. Some outstanding community degrees occur by the end of the tail and these include communities that cluster the majority of the biggest customers in the network. The central part of the distribution decays faster than the rest of the distribution. There is an observable curvature in the log–log plot, however, no approximation method fitted the distribution. Figure 5 shows that the maximum

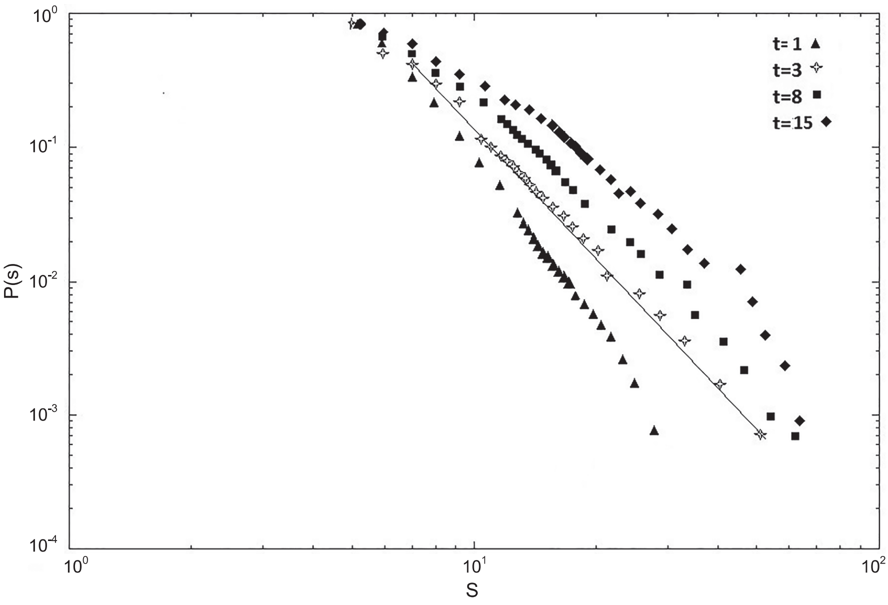



**Fig. 3.** Cumulative community size distribution at different times $t$ (log–log scale).

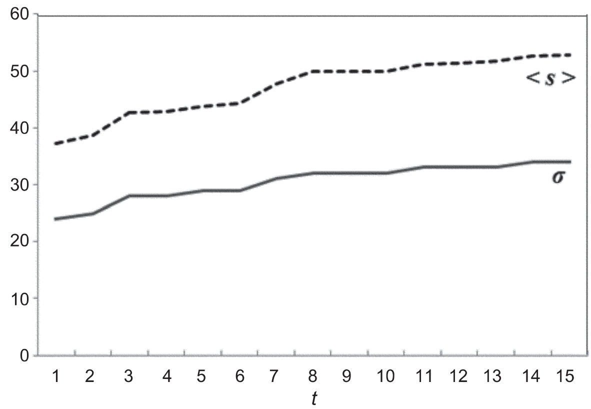



**Fig. 4.** Statistics of community size; $\langle s \rangle$ is the average community size, $\sigma$ is the standard deviation of the size of communities at different times $t$.

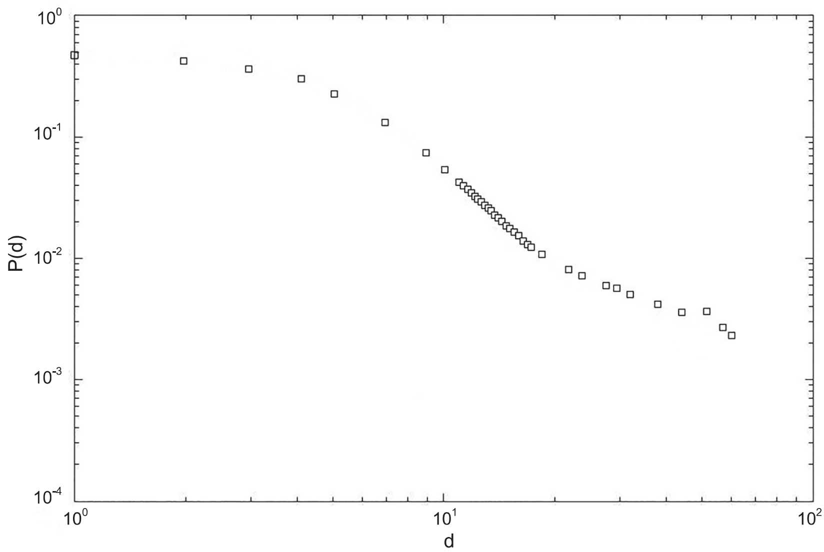



**Fig. 5.** Cumulative distribution of community degrees $d$ (log–log scale).

number of degrees $d$ is 63 and corresponds to a relatively small quantity of nodes.

A node $i$ of a network can be characterized by a membership number $m_i$, which is the number of communities where the node $i$ belongs to. Figure 6 shows the cumulative distribution of the membership number $m_i$. The distribution follows a power-law where no characteristic scale exists. The largest membership number found in the network was 10, meaning that a company can belong to a maximum of 10 different communities simultaneously. Figure 6 shows that the fraction of nodes that belong to many different communities is quite small, while the fraction of nodes belonging to at least one community is high. For example, when $m = 1$, the percentage of nodes that belong to at least one community is 50%, while the percentage of nodes that belong simultaneously to 10 communities $(m = 10)$ is extremely small. However, the rest of the

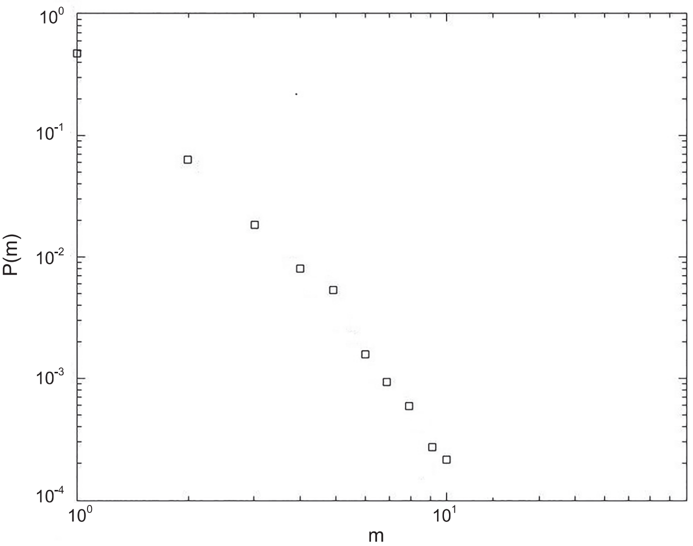



**Fig. 6.** Cumulative distribution function of the membership number $m_i$ (log–log scale).

communities belong to at least two or more communities. The companies that overlap 10 communities belong to the energy and water services. The majority of the nodes that have $m \neq 1$ have a degree that is less than $k-1$, meaning they are weakly connected.

The range to which the communities overlap each other is also an important property of our network. The overlap size is defined as the number of nodes that two communities share. $P(s_o)$ is the proportion of overlaps larger than $s_o$. Figure 7 shows the cumulative distribution

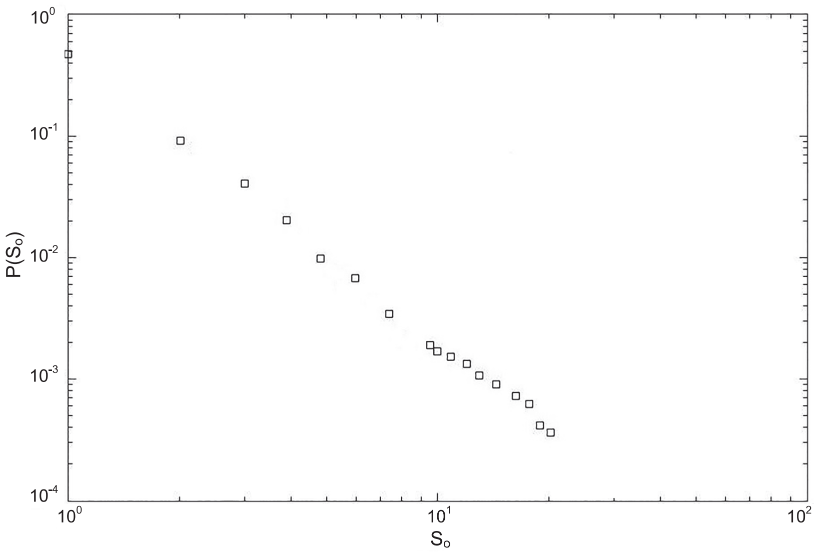



**Fig. 7.** Cumulative distribution function of the overlap size $s_o$ (log–log scale).

function of the overlap size. In general, although the extent of overlap sizes is limited, the data is close to power-law dependence, meaning there is no characteristic overlap size. The largest overlap size is 22, however, at $s_o \geq 9$ the number of overlapping nodes becomes small.

In our previous study [40] we found scale-free properties in the degree distributions of the Estonian network of payments and it is interesting to observe that the scale-free property is also preserved at a higher level of organization where overlapping communities are present. Scale-free networks are resilient against random removal of nodes. This means that it is difficult to destroy a complex network by random mechanisms, but if the exact portion of particularly selected nodes is removed, the network breaks easily. When the degree distributions of networks present scale-free structure, then this fact determines the topology of the system. Scale-free networks are robust against random damages but vulnerable against targeted attacks of nodes.

## 6. CONCLUSIONS

In this study we analysed the community structure of the Estonian network of payments by using the Clique Percolation Method. We found that there were scale-free properties in the statistical distributions of the community structure. The size, overlap, and membership distributions follow shapes that are compatible with power-laws. Power-law distributions have already appeared in this network at a global scale in the level of nodes [40], and in this community structure study we have shown that power-laws are present at the level of overlapping communities as well. This study adds to the existing literature on complex networks by presenting the first overlapping community analysis of a country's network of payments.

An immediate application and usefulness of the community detection output is that it can be used in targeted marketing activities, as input for predictive analytical models such as product acquisition propensities, churn propensities, product affinity analyses, for creating marketing profiles or customer segmentations, and for creating customer target lists for product offering (in an effort to propagate consumer buzz effects). Further applications of community detection in similar economic networks could involve strengthening relationships between companies of the same community for improving the performance of the whole network, identification of patterns between companies, and tracking suspicious business activities.

A question that remains open for future research is to investigate if the similarities in communities' features amongst different complex networks arise randomly or if there are any unknown properties shared by all of them. Another line of research that remains open is the plausibility of forecasting changes in a payment network through communities' detection analysis.

## ACKNOWLEDGEMENTS

The program Cfinder was used to calculate some network statistics and images. We thank Swedbank AS for allowing us to study this data set. This research was supported by the European Union through the European Regional Development Fund (Project TK 124). The publication costs of this article were covered by the Estonian Academy of Sciences.

## REFERENCES

1. Newman, M. E. J. *Networks: An introduction*. Oxford University Press, 2010.
2. Palla, G., Derényi, I., Farkas, I., and Vicsek, T. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 2005, **435**, 814–818.
3. Derényi, I., Palla, G., and Vicsek, T. Clique percolation in random networks. *Phys. Rev. Lett.*, 2005, **94**(16), 160202.
4. König, M. D. and Battiston, S. From graph theory to models of economic networks, a tutorial. In *Networks, Topology and Dynamics* (Naimzada, A. K., Stefani, S., and Torriero, A., eds), *Lecture Notes in Econom. and Math. Systems*, 2009, **613**, 23–63.
5. Souma, W., Fujiwara, Y., and Aoyama, H. Heterogeneous economic networks. In *The Complex Networks of Economic Interactions* (Namatame, A., Kaizouji, T., and Aruka, Y., eds), *Lecture Notes in Econom. and Math. Systems*, 2006, **567**, 79–92.
6. Battiston, S., Rodrigues, J. F., and Zeytinoglu, H. The network of inter-regional direct investment stocks across Europe. *Advs. Complex Syst.*, 2007, **10**(1), 29–51.
7. Glattfelder, J. B. and Battiston, S. Backbone of complex networks of corporations: the flow of control. *Phys. Rev. E.*, 2009, **80**(3).
8. Nakano, T. and White, D. Network structures in industrial pricing: the effect of emergent roles in Tokyo supplier-chain hierarchies. *Struct. and Dyn.*, 2007, **2**(3), 130–154.
9. Reyes, J., Schiavo, S., and Fagiolo, G. Assessing the evolution of international economic integration using random-walk betweenness centrality: the cases of East Asia and Latin America. *Advs. Complex Syst.*, 2007, **11**(5), 685–702.
10. Lublóy, A. Topology of the Hungarian large-value transfer system. *Magyar Nemzeti Bank (Central Bank of Hungary) MNB Occasional Papers*, 2006, 57.
11. Inaoka, H., Nimoniya, T., Taniguchi, K., Shimizu, T., and Takayasu, H. Fractal network derived from banking

transactions – an analysis of network structures formed by financial institutions. *Bank of Japan Working Papers*, 2004.
12. Soramäki, K., Bech, M. L., Arnold, J., Glass, R. J., and Beyeler, W. E. The topology of interbank payment flows. *Physica A*, 2007, **379**(1), 317–333.
13. Boss, M., Helsinger, H., Summer, M., and Thurner, S. The network topology of the interbank market. *Quant. Finance*, 2004, **4**(6), 677–684.
14. Iori, G. and Jafarey, S. Criticality in a model of banking crisis. *Physica A*, 2001, **299**(1), 205–212.
15. Iori, G., De Masi, G., Precup, O. V., Gabbi, G., and Caldarelli, G. A network analysis of the Italian overnight money market. *J. Econ. Dyn. Control*., 2007, **32**(1), 259–278.
16. Vitali, S. and Battiston, B. The community structure of the global corporate network. *PLoS ONE*, 2014, **9**(8), 0104655.
17. Fenn, D., Porter, M., McDonald, M., Williams, S., Johnson, N., et al. Dynamic communities in multichannel data: an application to the foreign exchange market during the 2007–2008 credit crisis. *Chaos*, 2009, **19**(3), 3184538.
18. Piccardi, C., Calatroni, L., and Bertoni, F. Communities in Italian corporate networks. *Physica A*, 2010, **389**(22), 5247–5258.
19. Bóta, A. and Kresz, M. A high resolution clique-based overlapping community detection algorithm for small-world networks. *Informatica*, 2015, **39**(2), 177–187.
20. Traud, A. L., Mucha, J. P., and Porter, M. A. Social structure of Facebook networks. *Physica A*, 2012, **391**(16), 4165–4180.
21. González, M. C., Herrmann, H. J., Kertész, J., and Vicsek, T. Community structure and ethnic preferences in school friendship networks. *Physica A*, 2007, **379**(1), 307–316.
22. Palla, G., Barabási, A. L., and Vicsek, T. Quantifying social group evolution. *Nature*, 2007, **446**(7136), 664–667.
23. Pollner, P., Palla, G., and Vicsek, T. Preferential attachment of communities: the same principle, but a higher level. *Europhys. Lett*., 2006, **73**(3), 478–484.
24. Lewis, A. C. F., Jones, N. S., Porter, M. A., and Deane, C. M. The function of communities in protein interaction networks at multiple scales. *BMC Syst. Biol*., 2010, **4**, 100.
25. Guimerá, R. and Amaral, N. Functional cartography of complex metabolic networks. *Nature*, 2005, **433**, 895–900.
26. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabási, A. L. Hierarchical organization of modularity in metabolic networks. *Science*, 2002, **297**(5586), 1551–1555.
27. Dourisboure, Y., Geraci, F., and Pellegrini, M. Extraction and classification of dense communities in the web. In *Proceedings of the 16th International Conference on the World Wide Web*, 2007, 1, 461–470.
28. Newman, M. E. J. Detecting community structure in networks. *Eur. Phys. J. B*, 2004, **38**(2), 321–330.
29. Yang, Z., Algesheimer, R., and Tessone, C. J. A comparative analysis of community detection algorithms on artificial networks. *Sci. Rep.*, 2016, **6**, 30750.
30. Hopcroft, J., Khan, O., Kulis, B., and Selman, B. Tracking evolving communities in large linked networks. *Proc. Natl. Acad. Sci. USA*, 2004, **110**, 5249–5253.
31. Scott, J. *Social Network Analysis: A Handbook*. Sage Publications, UK, 2000.
32. Gavin, A. C., Bösche, M., Krause, R., Grandi, P., Marzioch, M., et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, 2002, **6868**, 141–147.
33. Devi, J. C. and Poovammal, E. An analysis of overlapping community detection algorithms in social networks. *Proc. Comp. Sci.*, 2016, **89**, 349–358.
34. Xie, J., Kelley, S., and Boleslaw, K. S. Overlapping community detection in networks: the-state-of-the-art and comparative study. *ACM Comput. Surv.*, 2013, **45**(4), Art. 43, 1–35.
35. Ding, Z., Zhang, X., Sun, D., and Luo, B. Overlapping community detection based on network decomposition. *Sci. Rep.*, 2016, **6**, 24115.
36. Everett, M. G. and Borgatti, S. P. Analyzing clique overlap. *Connections*, 1998, **21**(1), 49–61.
37. Shen, H. W. *Community Structure of Complex Networks*. Springer Science & Business Media, 2013.
38. Newman, M. E. J. Fast algorithm for detecting community structure in networks. *Phys. Rev. E*, 2004, **69**, 066133.
39. Clauset, A., Newman, M. E. J., and Moore, C. Finding community structure in very large networks. *Phys. Rev. E*, 2004, **70**(6), 066111.
40. Rendón de la Torre, S., Kalda, J., Kitt, R., and Engelbrecht, J. On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia. *Chaos Soliton. Fract*., 2016, **90**, 18–27.

## Kattuvate kommuunide tuvastamine Eesti maksevõrgustikus

Stephanie Rendón de la Torre, Jaan Kalda, Robert Kitt ja Jüri Engelbrecht

Reaalselt eksisteerivates võrgustikes sisalduvate kommuunide tuvastamine on üheks põhietapiks teel komplekssüsteemide selliste seaduspärasuste mõistmise poole, mis lähevad üksikelementide lokaalsete interaktsioonide käsitlemisest sügavamale. Kommuunide tuvastamise meetodid aitavad võrgustike komponentide lokaalstruktuuridele valgust heita. Käesolevas uurimuses identifitseerime ja uurime kattuvate kommuunide struktuure olulises unikaalses võrgustikus: Eesti maksevõrgustikus. Selleks otstarbeks kasutame nn klikk-perkolatsiooni meetodit ja uurime kommuunide jaotusfunktsioone ning kommuunide mastaabi-invariantseid omadusi. Antud võrgustikus on sõlmpunktideks Eesti ettevõtted ja sidemeteks maksed erinevate ettevõtete vahel. Tegemist on esmakordse uurimusega, kus tuvastatakse kattuvad kommuunid ühe riigi ettevõtete vaheliste maksete võrgustikus.

**Paper V**

**Rendón de la Torre S.**, Kalda J., Kitt R. (2019) Specific statistical properties of the strength of links and nodes of the Estonian network of payments. Proceedings of the Estonian Academy of Sciences, 68(3). Manuscript (in press). DOI:10.3176/proc.2019.3.02.

Proceedings of the Estonian Academy of Sciences,
2019, **68**, 3,
https://doi.org/10.3176/proc.2019.3.02
Available online at www.eap.ee/proceedings

NETWORK SCIENCE STUDIES

# Specific statistical properties of the strength of links and nodes of the Estonian network of payments

Stephanie Rendón de la Torre[a*], Jaan Kalda[a], and Robert Kitt[b]

[a] Department of Cybernetics, School of Science, Tallinn University of Technology, Akadeemia tee 21, 12618 Tallinn, Estonia
[b] Swedbank AS, Liivalaia 10, 15040 Tallinn, Estonia

Received 10 October 2018, accepted 14 December 2018, available online 28 March 2019



**Abstract.** We investigated the strength of the interactions of the elements of the Estonian network of payments (link weight of payments and volume of payments) by the realization of particular experiments. Specific statistical measures of this network, which combine the topology of the relations of the strength of links and nodes and their specific weights, were studied with the purpose of discovering beyond the topological architecture of our network and revealing aspects of its complex structure. Moreover, scale-free properties between the strengths and the degree values were found. We also identified clear patterns of structural changes in such a network over the analysed period.

**Key words:** economic networks, complex systems, scale-free networks, weighted networks, strength of nodes.

## 1. INTRODUCTION

Complex network systems have been studied across many fields of science [1–4]. Undoubtedly, many ystems in nature can be modelled as networks where the elements of the nodes are the elements of the system and the links represent the interactions between these elements. Some examples of such systems are technological networks such as the internet [5] (a network of routers or domains connected via cables), the World Wide Web [6] (where nodes are HTML documents connected by links pointing from one page to another), power grids [7] (electricity networks), social networks [8,9] (such as acquaintance networks and collaboration networks), biological and metabolic networks [10,11], transport networks [12] (worldwide airport network), and economic networks [13] (Japanese bank transaction network). In the last decades, networks have received a great deal of attention and a great portion of recent research has focused on statistical and topological properties, for example, the small-world property [14] and scale-free behaviours [15].

Most of the real networks, alongside with their complex topological structure, display a gradation of interaction regarding connections and their intensities; this is commonly quantified by the link weight. The link weight reveals significant functional properties such as the concentration of friendships between people in social networks, the quantities of flows of money between banks, and the capacity to pass information in a network of communication or transport.

The research questions of this study are the following: Which is the characterization of the strength of the links of the nodes of the Estonian network of payments? Is there any relevant relationship between the weighted quantities and the underlying network structure?

[*] Corresponding author, stretomx@gmail.com

In this study, we characterize the links by investigating the strength of the interactions of the elements of our network (link weight of payments and volume of payments). We analyse specific statistical measures of the weighted Estonian network of payments that combine the topology of the relations of the strength of links and nodes and their specific weights with the purpose of discovering beyond the topological architecture of the network and revealing aspects of its complex structure.

Section 1 provides a general introduction and an overview of the objectives. In Section 2, we deliver a description of the data set used in this study. Section 3 provides a literature review of related studies. Section 4 presents the methods used to develop this study, while Section 5 gives our main results. Section 6 concludes with a discussion of our results.

## 2. DATA

Our data set was obtained from the databases of Swedbank, one of the leading banks in the Nordic and Baltic regions of Europe operating actively in Estonia, Latvia, Lithuania, and Sweden. As all the information related to the identities of the nodes is very sensitive, it will remain confidential and cannot be disclosed. Since ~80% of Estonia's bank transactions are executed through the Swedbank's system of payments, the data set is unique in its kind and very interesting. Hence it reproduces fairly well the trends of flows of money of the whole Estonian economy, and we use this dataset as a proxy of the economy of Estonia.

A network is a set of nodes connected by links. In this study, the nodes represent companies and the links represent the payments between the companies. The network of payments is defined by three matrices that map the whole image of the network: $A$ is an undirected connectivity symmetric adjacency matrix $A_{N\times N}$, where $N$ is the total number of nodes in the network and two nodes have a link if they share one or more payments; then each element represents a link as follows: $A_{ij}^{u} = a_{ji}^{u}$, where $A_{ij}^{u} = 1$ if there is a transaction between companies $i$ and $j$ or $A_{ij}^{u} = 0$ if there is no transaction between companies $i$ and $j$. The weighted connectivity matrix $B$ contains the number of transactions between companies $i$ and $j$. This definition allows looking at the structure of the network as a weighted graph where the links have certain weights associated with them, representing less or more important relationships. Transactions between any two parties add to the associated link weights in terms of volume. The elements $w_{ij}$ of the weighted connectivity matrix $B$ denote the overall number of transactions between companies $i$ and $j$.

Additionally, our data set allows us to construct directed graphs where the links follow the flow of the money, such that a link is incoming to the receiver $i$ and outgoing from the payer $j$. The matrix $C$ is a weighted-directed graph where the links follow the flow of the money, such that a link is incoming to the receiver and outgoing from the sender of the payment. For this case we have two more matrices: in-degree and out-degree. The choice of usage of the matrix representation depends on the focus of the analysis.

The data set contains 3.4 million electronic company-to-company domestic payments of the full calendar year 2014, including data of 16 613 companies. This network shares typical structural characteristics known in other complex networks: degree distributions follow a power law, low clustering coefficient, and low average shortest path length. The average degree of the Estonian network of payments is $\langle k \rangle = 20$, the maximal degree is 345, and the diameter is 29. The average betweenness of the links is 40 while it is 110 for the nodes. The average shortest path length $\langle l \rangle = 7.1$. The network is a small world with 7.1 degrees of separation, which means that on average any company can be reached by another only in a few steps. Our network has low connectivity but is densely connected. The network also displays scale-free properties in its degree distributions. This scale-free structure indicates that few companies in Estonia trade with many parties while the majority trade with only few.

The network has a low average clustering coefficient of 0.18 and shows disassortative mixing. The detailed information of our network's full topologic structure and its basic characteristics can be found in our previous paper [16]. In that study, we performed an analysis to reveal the robustness of our network derived from its scale-free structure. We found that the network is resilient to random removal of the nodes but is vulnerable to targeted removal of nodes. The percolation threshold is 6%, and this means that a small portion of economic entities maintains the whole network unified. We found that most of the influential companies in the network are not necessarily the most connected ones and that a considerable number of companies who have high transactional activities have weak influence on the economic network as a whole.

## 3. LITERATURE DISCUSSION

Real networks, which are organized in a complex topological structure, show a large heterogeneity in the capacity and intensity of the connections (the weight of the links) [17]. Recently, many features of weighted networks have been studied, for example, the relationship between the node degree and node strength [18], degree correlations and perturbations [19], node correlations [20], and dynamical properties of nodes and degrees [21]. The study of highly interconnected systems became an important area of multidisciplinary research in network science involving physics, mathematics, biology, and social sciences, but recently the interest has shifted towards weighted networks. In the last 20 years, a large set of measures and metrics which combine topological and weighted observables has been proposed to characterize the statistical properties of nodes and links and to investigate the relationships between the weighted quantities and underlying network structures. Barrat et al. [17] presented a quantitative and general approach to understanding the complex architecture of weighted networks. They studied representative examples of social and large infrastructure systems, and defined specific metrics considering the weights of nodes and links in order to investigate the correlations among the weighted quantities and the underlying topological structure of these networks. They showed that a more complete view of complex networks is provided by the study of the interactions defining the links of these systems.

Zemp et al. [22] developed new versions of some measures for directed and/or weighted networks in order to take the importance of nodes into account. They showed that the use of their measures avoids systematic biases created by a higher node density and larger weights of the links. Newman [23] showed that weighted networks could be analysed by using a simple mapping from a weighted network to an unweighted multigraph, which allows using standard techniques for unweighted and weighted networks.

Network-based approaches are very useful and provide means for monitoring complex economic systems. They may also help in ensuring better control in managing and governing these systems. Regarding applications of economic networks and other recent studies, Souma et al. [24] studied a shareholder network of Japanese companies. In this study, the authors analysed the growth of companies through the analysis of the network's dynamics. A similar work by Rotundo and D'Arcangelis [25] dealt with the relationships of shareholders in the Italian stock market. Reyes et al. [26] made a weighted network analysis focused on using random walk betweenness centrality to study why high-performing Asian economies had higher economic growth than Latin American economies in 1980–2005. Other relevant studies on economic networks concentrate on the regional investment and ownership networks [27,28]. In these networks, European company-to-company foreign direct investment stocks show a power-law distribution with the number of employees in the investing company and in the company invested in, and with the volume of in- and outgoing investments of both companies. This power-law feature allows predicting the investments that will be received or made in specific regions, based on the connectivity and transactional activity of the companies. Nakano and White [29] showed that analytic concepts and methods related to complex networks can help to uncover structural factors that may influence the price formation for empirical market-link formations of economic agents.

Another interesting line of research is related to network topology as a basis for investigating money flows of customer-driven banking transactions. A few papers describe the actual topologies observed in different financial systems [30–32]. Other similar studies focus on economic shocks, robustness, and growth in economic or social complex networks [33–35]. Interesting reviews of complex network models and methods present the applications to socioeconomic issues [36,37].

## 4. METHODS

The degree of a node is defined as

$$k_i = \sum_{j \in \zeta(i)} a_{ij}, \tag{1}$$

where the sum goes over the set $\zeta(i)$ of neighbours of *i*. For example, $\zeta(i) = \{ j \mid a_{ij} = 1 \}$. The degree of a node (company) refers to the number of payments linked to it.

Two relevant characteristics of a node occur in a directed network: the number of links that end at a node and the number of links that start from a node. These quantities are known as the in-degree $k^d$ and out-degree $k^o$ of a node, and we define them as

$$k^d = \sum_{j \in \zeta(i)} a_{ij}^d \, , \; k^o = \sum_{j \in \zeta(i)} a_{ij}^o. \tag{2}$$

Weights $w_{ij}$ of the links $i$ and $j$ in a network show the importance of each link. The strength $s_i$ of the nodes is the sum of the weights of all the links. In our network, the strength measures the overall transaction value/volume for any given node, and is defined by the formula

$$s_i = \sum_{j \in \zeta(i)} w_{ij}, \tag{3}$$

where the sum runs over the set $\zeta(i)$ of neighbours of $i$.

For a given node $i$ with connectivity $k_i$ and strength $s_i$, the weights of the links might be of the same order of magnitude $s_i / k_i$, or they can be distributed heterogeneously with some links predominating over others. Then, the participation ratio is defined as follows:

$$H_2^w(i) = \sum_{j \in \zeta(i)} \left[ \frac{w_{i,j}}{s_i^w} \right]^2, \quad (4)$$

or, equivalently,

$$H_2^c(i) = \sum_{j \in \zeta(i)} \left[ \frac{c_{i,j}}{s_i^c} \right]^2. \quad (5)$$

We define the participation rates to separate outgoing and incoming links. Then, the average participation ratio is calculated as

$$H_2^w = \frac{1}{N} \sum_i H_2^w(i), \quad (6)$$

and

$$H_2^c = \frac{1}{N} \sum_i H_2^c(i), \quad (7)$$

respectively.

We calculate the participation ratio as a function of a company's inverse degree, where the objective is to identify the links that are used more often. If a low number of weights are dominant, then $H_2^w$ is close to 1, but if all the weights are of the same order of magnitude, then $H_2^w \sim 1/k_i$. The value of $H_2^w$ close to 1 indicates the existence of preferential interactions between the nodes, meaning that companies prefer to transact with certain companies.

## 5. RESULTS

### 5.1. Patterns of payments

The general characteristics and statistics of the Estonian network of payments are listed in Tables 1 and 2.

In this section, we will focus on the structure of our network and its time evolution. The objective of this effort is to identify structural changes and compare the emerging patterns. Figure 1a shows the monthly volume of transactions during 2014 while Fig. 1b displays the total number of transactions. Figure 1c shows the monthly average number of active links as a function of time. Figure 1b,c show that the number of transactions decreases dramatically in the third quarter of the year, while the number of active links decreases already in the second quarter. Also, these plots show that the number of transactions and the active links increase in the last quarter of the year, suggesting that liquidity in the Estonian network of payments increases by the end of the year through increased transaction volumes and payments, and higher than the usual number of active counterparties. It is interesting that the concentration in the volume of payments is high from August till the end of the year, while the number of payments diminishes dramatically in the same period of time. These observations indicate that the average number of active companies has decreased 20%, while the volume of transactions has increased 14% and the number of transactions has decreased 66% by the end of the year (compared with the beginning of the year). This indicates that companies in Estonia manage higher volumes of money at the end of the year than at the beginning of the year, while not all the companies remain active by the end of the year. A full explanation of this pattern of financial liquidity is not possible due to the lack of complete information about the overall financial and commercial activities of the companies in this network. Nonetheless, there are some possible explanations for these patterns. For example, these patterns could be highly affected by business cycles of payments, or by seasonal effects on the liquidity of companies, or by macroeconomic variations such as changes in the monetary policy of the Euro area and Estonia. Another explanation for the increased volume of transactions and

**Table 1.** Network characteristics

| | |
|---|---|
| Companies analysed | 16 613 |
| Total number of payments analysed | 2 617 478 |
| Value of transactions | 3 803 462 026* |
| Average value of transaction per customer | 87 600* |
| Maximum value of transaction | 121 533* |
| Minimum value of transaction (aggregated in the whole year) | 1000* |
| Average volume of transaction per company | 60 |
| Maximum volume of transaction per company | 24 859 |
| Minimum volume of transaction per company (aggregated in the whole year) | 20 |

* All the money amounts are expressed in monetary units and not in currencies in order to protect the confidentiality of the data set. The purpose of showing monetary units is to provide a notion of the proportion of quantities and not to show exact amounts of money.

**Table 2.** Summary of statistics

| Statistic | Value | Components | Number of nodes |
|---|---|---|---|
| Undirected links | 43 375 | GCC | 15 434 |
| $\langle k \rangle$ | 20 | DC | 1179 |
| $\gamma^o$ | 2.39 | GSCC | 3987 |
| $\gamma^i$ | 2.49 | GOUT | 6054 |
| $\gamma$ | 2.45 | GIN | 6172 |
| $\langle C \rangle$ | 0.183 | Tendrils | 400 |
| $\langle l \rangle$ | 7.1 | Cutpoints | 1401 |
| $T$ | 0.13 | Bi-component | 4404 |
| $D$ | 29 | $k$-core | 1081 |
| $\langle \sigma \rangle$ (nodes) | 110 | | |
| $\langle \sigma \rangle$ (links) | 40 | | |

$\langle k \rangle$ = average degree, $\gamma^o$ = scaling exponent of the out-degree empirical distribution, $\gamma^i$ = scaling exponent of the in-degree empirical distribution, $\gamma$ = scaling exponent of the connectivity degree distribution, $\langle C \rangle$ = average clustering coefficient, $\langle l \rangle$ = average shortest path length, $T$ = connectivity %, $D$ = diameter, $\langle \sigma \rangle$ = average betweenness, GCC = Giant Connected Component, DC = Disconnected Component, GSCC = Giant Strong Connected Component, GOUT = Giant Out Component, GIN = Giant In Component.

increased liquidity at the end of the year is that there might be a generalized release of delayed payments, like when companies try to spend the remaining balances of their annual budgets.

### 5.2. Strength and degree of nodes

We calculated the probability $P(s)$ that a company has $k$ outgoing and incoming links. As per Fig. 2, the distribution of the out-degree volume (strength) follows a power-law decay

$$P(s) \sim s^{-2.32}, \tag{8}$$

where the scaling exponent is 2.32. There are some deviations from the power-law behaviour but they are sufficiently small. A similar distribution was also found in the in-degree volume (strength) distribution [16]. The power-law tail signals that the probability of finding companies paying out very large quantities of money is small. Moreover, while the companies have an absolute freedom in choosing how much money to pay or the counterparties who they interact with, the overall system obeys a scaling law, which is a particular property observed in critical phenomena and in highly interactive self-organized systems.

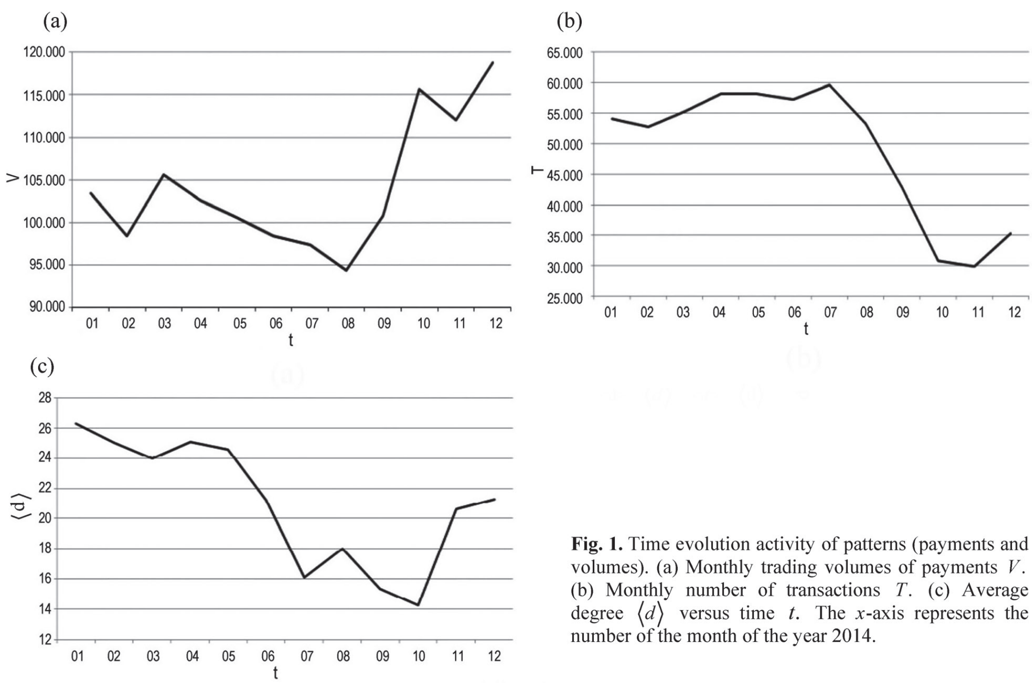



**Fig. 1.** Time evolution activity of patterns (payments and volumes). (a) Monthly trading volumes of payments $V$. (b) Monthly number of transactions $T$. (c) Average degree $\langle d \rangle$ versus time $t$. The $x$-axis represents the number of the month of the year 2014.

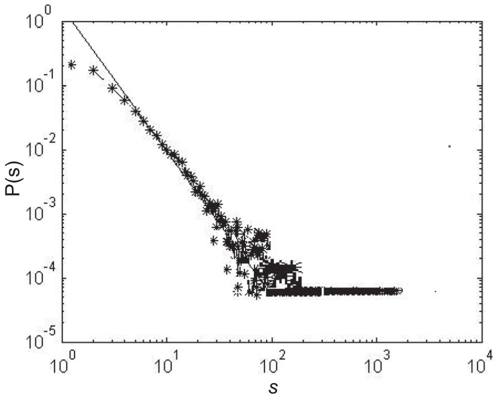



**Fig. 2.** Volume out-strength distribution.

We also analyse the bond between the strength and degree of a node. Figure 3a,b depict the volume and value (in and out) strengths as functions of degree of both outgoing and incoming links (in-degree and out-degree). The strength $s$ is normalized by dividing it over the average link weight $\langle w_{ij} \rangle$. The following power-law relationship exists between the strength and the degree:

$$s(k) \sim k^{\alpha}, \tag{9}$$

where $\alpha$ is the coefficient of this scaling distribution. The power-law fit of Fig. 3a has an exponent $\alpha_{\text{vol}} = 1.5$, when volume is used as the weight, and $\alpha_{\text{val}} = 2.4$ when the value is used instead. These values imply that the out-strength of nodes $s_{\text{no}}$ and in-strength of nodes $s_{\text{ni}}$ grow faster than the degree $k$ of a node, as seen in Fig 3a. It means that the most connected companies execute a higher number of payments with higher values of money than would be suggested only by their degree. This indicates that if a company has twice as many payments (out links) as another company, it could be expected that this company sends three times the number of payments, and almost five times the total value of payments. Figure 3b indicates that the relationships between in-degree and in-strength show similar trends to the out-degree and out-strength cases seen in Fig. 3a.

The strength of a node scales with its degree $k$, indicating that highly connected companies have payments of higher weight. The strength of a company grows generally faster than its degree. In other words, highly connected companies not only have many payments, but their payments also have a higher than average weight. This observation agrees with the fact that big companies are better equipped to handle large amounts of payments with higher amounts of money. Comparable results were found in the cargo ship movements network [38] and in the airport network [17], which may hint at a generic pattern in such large-scale networks.

### 5.3. Participation ratio

Figure 4b shows a plot of the participation ratio $H_2^c$ as a function of the inverse degree of the nodes. The plot shows the links that are used more often than others. For example, for a degree up to 10 $H_2^c(i) \sim 1/k_i$ and for higher degrees the participation ratio is higher than the inverse degree, suggesting there is a disposition in the direction of preferential trading with specific counterparties. Figure 4b shows the average participation ratio during the whole year for outgoing payments and in-

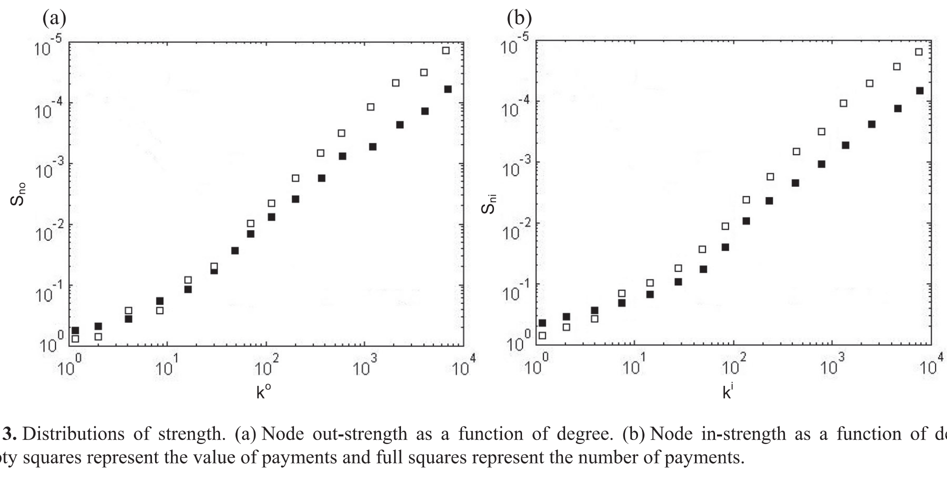
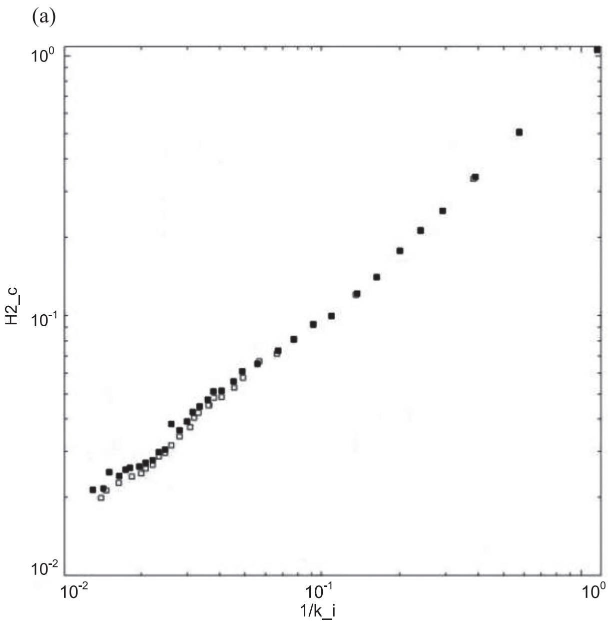



**Fig. 3.** Distributions of strength. (a) Node out-strength as a function of degree. (b) Node in-strength as a function of degree. Empty squares represent the value of payments and full squares represent the number of payments.



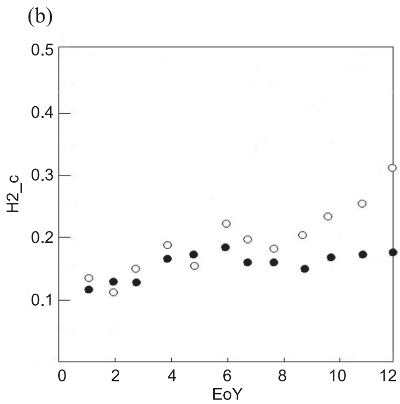




**Fig. 4.** Participation ratio plots. (a) Participation ratio as a function of the inverse degree for outgoing payments and incoming payments. Black squares represent outgoing payments, white squares incoming payments. (b) Participation ratio as a function of the distance from the end of the year expressed in months for incoming (black circles) and outgoing (white circles) payments.

coming payments. By the end of the year the participation ratio for all the payments decreases. Particularly, the participation ratio of the outgoing payments decreases dramatically. This reveals that the preferential linking is limited. By the end of the year, the preference for trading with only certain counterparties becomes less important. This could be caused by an increased payments/liquidity tendency that could potentially be driven by generalized unspent company annual budgets or delayed payments that were completed before the year ended.

## 6. CONCLUSIONS

In this study, we explored the relations between weighted quantities and their network underlying structures. We investigated the strength of interactions (number of payments and the volumes of payments) and the interconnectivities among these interactions in the Estonian network of payments by the realization of particular experiments, calculating specific metrics, and revealed interesting microstructural features.

We detected a clear pattern of structural changes over the analysed period in the network degree and number of payments decreasing by the end of the year, while the volume of payments increased. This indicates that Estonian companies handle higher volumes of cash flows at the end of the year than at the beginning of the year, while not all the companies remained active by the end of the year.

Scale-free properties were determined between the strengths and the degree values. We found that the most connected companies executed a higher number of payments with higher values of money than what would be suggested only by their degree (the out-strength of nodes and in-strength of nodes grow faster than the degree of a node).

It is important to continue observing, describing, and studying the structures and characteristics of economic complex networks in order to be able to understand their underlying processes and to detect patterns that could be useful for predicting or forecasting events and trends. The addition of evidence through empirical studies in favour of economic networks represents an important step towards the knowledge on the universality and the understanding of the complexity of economic systems.

## ACKNOWLEDGEMENTS

The data set used in this study was provided by Swedbank AS. Financial support for the research was granted by the European Regional Development Fund (Project TK 124). The publication costs of this article were covered by the Estonian Academy of Sciences.

## REFERENCES


1. Newman, M. E. J. *Networks: An introduction.* Oxford University Press, 2010.
2. Strogatz, S. H. Exploring complex networks. *Nature*, 2001, **410**, 268–276.
3. Albert, R. and Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 2002, **74**, 47–91.
4. Liu, Y. Y. and Barabási, A. L. Control principles of complex systems. *Rev. Mod. Phys.*, 2016, **88(**3), 035006–035064.
5. Faloutsos, M., Faloutsos, P., and Faloutsos, C. On power-law relationships of the internet topology. *ACM SIGCOMM Comput. Commun*. *Rev*., 1999, **9**(4), 251–262.
6. Albert, R., Jeong, H., and Barabási, A. L. Diameter of the world wide web. *Nature*, 1999, **401**, 130–131.
7. Pagani, G. A. and Aiello, M. The power grid as a complex network: a survey. *Physica A*, 2013, **392**(11), 2688–2700.
8. Newman, M. E. J. Scientific collaboration networks. I. Network construction and fundamental results. *Phys. Rev. E*, 2001, **64**(1), 016131.
9. Davidsen, J., Ebel, H., and Bornholdt, S. Emergence of a small world from local interactions: modeling acquaintance networks. *Phys. Rev. Lett*., 2002, **88**(12), 128701.
10. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A. L. The large-scale organization of metabolic networks. *Nature*, 2000, **407**(6804), 651–654.
11. Doye, J. Network topology of a potential energy landscape: a static scale-free network. *Phys. Rev. Lett*., 2002, **88**(23), 238701.
12. Guimerà, R. and Amaral, L. A. N. Modeling the world-wide airport network. *Eur. J. Phys. B*, 2004, **38**(2), 381–385.
13. Inaoka, H., Nimoniya, T., Taniguchi, K., Shimizu, T., and Takayasu, H. Fractal network derived from banking transactions – an analysis of network structures formed by financial institutions. *Bank of Japan Working Papers*, 2004.
14. Watts, D. J. *Small Worlds: The Dynamics of Networks Between Order and Randomness*. Princeton University Press, NJ, 2003.
15. Barabási, A. L. Scale-free networks: a decade and beyond science. *Science*, 2009, **325**(5939), 412–413.
16. Rendón de la Torre, S., Kalda, J., Kitt, R., and Engelbrecht, J. On the topologic structure of economic complex networks: empirical evidence from large scale payment network of Estonia. *Chaos Solitons Fractals*, 2016, **90**, 18–27.
17. Barrat, A., Barthelemy, M., Pastor-Satorras, R., and Vespignani, A. The architecture of complex weighted networks. *Proc. Natl Acad. Sci. USA*, 2004, **101**(11), 3747–3752.
18. Opsahl, T., Agneessens, F., and Skvoretz, J. Node centrality in weighted networks: generalizing degree and shortest paths. *Soc. Networks*, 2010, **32**(3), 245–251.
19. Xiang, J., Hu, K., Zhang, Y., Hu, T., and Li, J. M. Analysis and perturbation of degree correlation in complex networks. *Europhys. Lett.*, 2015, **111**(4), 48003.
20. Boguña, M. and Pastor-Satorras, R. Epidemic spreading in correlated complex networks. *Phys. Rev. E*, 2002, **66**(4), 047104.
21. Nie, T., Guo, Z., Zhao, K., and Zhe-Ming, Lu. The dynamic correlation between degree and betweenness of complex network under attack. *Physica A*, 2016, **457**(1), 129–137.
22. Zemp, D. C., Wiedermann, M., Kurths, J., Rammig, A., and Donges, J. F. Node-weighted measures for complex networks with directed and weighted edges for studying continental moisture recycling. *Europhys. Lett.*, 2014, **107**(5), 58005.
23. Newman, M. E. J. Analysis of weighted networks. *Phys. Rev. E*, 2004, **70**(5), 056131.
24. Souma, W., Fujiwara, Y., and Aoyama, H. Heterogeneous economic networks. In *The Complex Networks of Economic Interactions* (Namatame, A., Kaizouji, T., and Aruka, Y., eds), *Lecture Notes in Economics and Mathematical Systems*, 2006, **567**, 79–92. Springer, Berlin, Heidelberg.
25. Rotundo, G. and D'Arcangelis, A. M. Ownership and control in shareholding networks. *J. Econ. Interact Coord*., 2010, **5**(2), 191–219.
26. Reyes, J., Schiavo, S., and Fagiolo, G. Assessing the evolution of international economic integration using random-walk betweenness centrality: the cases of East Asia and Latin America. *Advs. Complex Syst*., 2007, **11**(5), 685–702.
27. Battiston, S., Rodrigues, J. F., and Zeytinoglu, H. The network of inter-regional direct investment stocks across Europe. *Advs. Complex Syst.*, 2007, **10**(1), 29–51.
28. Glattfelder, J. B. and Battiston, S. Backbone of complex networks of corporations: the flow of control. *Phys. Rev. E*, 2009, **80**(3), 036104.
29. Nakano, T. and White, D. Network structures in industrial pricing: the effect of emergent roles in Tokyo supplier-chain hierarchies. *Struct. and Dyn*., 2007, **2**(3), 130–154.
30. Lublóy, A. Topology of the Hungarian large-value transfer system. *Magyar Nemzeti Bank (Central Bank of Hungary) MNB Occasional Papers*, 2006, **57**.
31. Soramäki, K., Bech, M. L., Arnold, J., Glass, R. J., and Beyeler, W. E. The topology of interbank payment flows. *Physica A*, 2007, **379**(1), 317–333.
32. Boss, M., Helsinger, H., Summer, M., and Thurner, S. The network topology of the interbank market. *Quant. Finance*, 2004, **4**(6), 677–684.
33. Iori, G. and Jafarey, S. Criticality in a model of banking crisis. *Physica A*, 2001, **299**(1), 205–212.
34. Iori, G., De Masi, G., Precup, O. V., Gabbi, G., and Caldarelli, G. A network analysis of the Italian overnight money market. *J. Econ. Dyn. Control*., 2007, **32**(1), 259–278.
35. Safdari, H., Zare Kamali, M., Shirazi, A., Khalighi, M., Jafari, G., and Ausloos, M. Fractional dynamics of network growth constrained by aging node interactions. *PLOS ONE*, 2016, **11**(5), e0154983.
36. Varela, L. M., Rotundo, G., Ausloos, M., and Carrete Montana, J. Complex network analysis in socioeconomic models. In *Complexity and Geographical Economics*

(Commendatore, P., Kayam, S., and Kubin, I., eds), *Dynamic Modeling and Econometrics in Economics and Finance*, 2015, **19**. Springer, Cham.
37. Chapelle, A. and Szafarz, A. Controlling firms through the majority voting rule. *Physica A*, 2005, **355**(2–4), 509–529.
38. Kaluza, P., Kölzsch, A., Gastner, M. T., and Blasius, B. The complex network of global cargo ship movements. *J. R. Soc. Interface*, 2010, **7**(48), 1093–1103.

## Eesti maksete võrgustiku sidemete ja sõlmpunktide statistilised eriomadused

Stephanie Rendón de la Torre, Jaan Kalda ja Robert Kitt

Selles töös uurisime elementidevaheliste seoste tugevust (maksete seoste kaalu ja maksete mahtu) Eesti maksete võrgustikus, viies selleks läbi spetsiaalseid eksperimente. Uurisime võrgustiku konkreetseid statistilisi näitajaid, mis ühendavad seoste tugevuse suhete topoloogia sõlmede ja nende erikaaluga, eesmärgiga minna sügavamale võrgustiku topoloogilisest arhitektuurist ning välja tuua selle kompleksstruktuuri aspekte. Lisaks leidsime skaalata omadusi tugevuse ja järkude väärtuste vahel. Samuti täheldasime analüüsitud perioodi jooksul strukturaalsete muutuste selget mustrit võrgustikus.

# Curriculum vitae

**Personal data**
Name: Stephanie Rendón de la Torre
Date of birth: 09/03/85
Place of birth: Mexico City, Mexico
Citizenship: Mexican

**Contact data**
E-mail: stretomx@gmail.com

**Education**
2012–2013, MSC, National Autonomous University (UNAM Mexico City), Mexico (Cum laude and winner of Medal Antonio Caso (academic merit: best of 2012 class)
2004–2008 BSC, Accounting and Finances, Technology Institute of Superior Studies of Monterrey, Mexico City Campus, Mexico
2001–2003 High school, Liberty Heroes High School, Mexico City, Mexico

**Language competence**
Spanish: Native language
English: Fluent
French: Basic skills
Estonian: Basic skills

**Professional employment**
2015–to date, Tallinn University of Technology, Institute of Cybernetics (from 01.01.2017: School of Science, Department of Cybernetics). Engineer
September 2016–to date, Swedbank AS, Customer Intelligence, Data Scientist Sr.
September 2015–September 2016, Swedbank AS, Asset Management, Research analyst
September 2014–September 2015, Swedbank AS, Corporate Offering & Strategies, Data Analyst
2009–August 2014, AM Advisors (AM Capital Hedge Fund), CFO
2008–2009, Rabobank, Operations Officer
2007–2008, GBM (stockbrokerage house), Back Office International Treasury analyst
2005–2007, Ernst & Young, Taxes Staff

**Special courses and further training**
May 2017: Python Coursera certifications:

Python for everybody
Python Data Structures
Using databases with Python
Using Python to Access Web Data
Capstone: Retrieving, Processing, and Visualizing Data with Python

July, 2015, 5th Ph.D. School-Conference on Mathematical Modeling of Complex Systems (Patras, Greece)

July, 2017 SIGMAPHI 2017 School of Statistical Physics) (Corfu, Greece).

October, 2017, Predict, Europe'e Leading Data Conference / Data Modelling Workshop (Dublin, Ireland)

October 2018, PyCon and PyData Conference / Karlsruhe, Germany

October 2018, Data and Machine Learning Conference / Stockholm, Sweden

October 2018, Spark Programming certificate - by Instructor Ali Benamara, insite, Tallinn

**Research activity: Publications**

*Articles indexed by the Web of Science and Scopus database (1.1):*

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2016). On the topologic structure of economic complex networks: Empirical evidence from large scale payment network of Estonia. *Chaos, Solitons & Fractals*, 90, 18−27 DOI:10.1016/j.chaos.2016.01.018.

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2017). Fractal and multifractal analysis of complex networks: Estonian network of payments. *The European Physical Journal B*, 90. DOI: 10.1140/epjb/e2017-80214-5.

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht J. (2019) Detecting overlapping community structure: Estonian network of payments. *Proceedings of the Estonian Academy of Sciences*, 68(1) 79-88. DOI:10.3176/proc.2019.1.08

**Rendón de la Torre S., Kalda J., Kitt R. (2019)** Specific statistical properties of the strength of links and nodes of the Estonian network of payments. *Proceedings of the Estonian Academy of Sciences,* 68(3). Manuscript (in press). DOI:10.3176/proc.2019.3.02

*Peer-reviewed articles in other international journals (1.2) and collections (3.1):*

**Rendón de la Torre S**., Kalda J. (2018) Review of structures and dynamics of economic complex networks: Large scale payment network of Estonia. In: Zengqjang C., Dehmer M., Emmert-Streib F., Shi Y. (eds.), *Modern and interdisciplinary problems in network science.* Taylor & Francis CRC Group, USA, 193-226 https://www.crcpress.com/Modern-and-Interdisciplinary-Problems-in-Network-Science-A-Translational/Chen-Dehmer-Emmert-Streib-Shi/p/book/9780815376583.

**International conferences presentations:**

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. Estonian network of payments. Poster presentation at the *5th Ph.D. School-Conference on Mathematical Modeling of Complex Systems* (20-30 July, 2015, Patras, Greece).

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. On the topologic analysis of economic complex networks: Swedbank's network of payments in Estonia. Oral presentation at *Data Science Challenges* (14-17 October, 2015, Torino, Italy).

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. Communities' detection and evolution: Estonian economic network of payments. Poster presentation at the *26$^{th}$ STATPHYS IUPAP International Conference on Statistical Physics* (18-22 July, 2016, Lyon, France).

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. Communities detection and dynamics: Estonian economic network of payments. Poster presentation at *International Conference Statistical Physics SIGMAPHI 2017 (and School of Statistical Physics)* (6-14 July , 2017, Corfu, Greece).

**Rendón de la Torre S**., Kalda J., Kitt R., Engelbrecht, J. Fractal and multifractal analysis of complex networks: Estonian network of payments. Poster presentation at *NetSci-X 2018 International School and Conference on Network Science* (5-8 January, 2018 Hangzhou, China).

# Elulookirjeldus

**Isikuandmed**
Nimi: Stephanie Rendón de la Torre
Sünniaeg: 09/03/1985
Sünnikoht: México, Mehhiko
Kodakondsus: Mehhiko

**Kontaktandmed**
E-post: stretomx@gmail.com

**Hariduskäik**
2012–2013 National Autonomous University (UNAM México City), Mehhiko, MSc
2004–2008 Monterrey Kõrgem Tehnoloogiainstituut, Mehhiko, BSc rahanduses

**Keelteoskus**
Hispaania keel – emakeel
Inglise keel – kõrgtase
Prantsuse keel – algtase
Eesti keel – algtase

**Teenistuskäik**

2015–praeguseni Tallinna Tehnikaülikooli küberneetika osakond (School of Science, Department of Cybernetics), insener
2016–praeguseni Swedbank AS, klienditeabe osakond, vanemanalüütik
2015–2016 Swedbank AS, varahalduse osakond, teadur-analüütik
2014–2015 Swedbank AS, äriklientide haldus ja strateegiad, analüütik
2009–2014 AM Advisors (AM Capital Hedge Fund), finantsjuht
2008–2009 Rabobank, tootmisjuht
2007–2008 GBM (börsimaaklerite firma), keskkontori rahvusvahelise vara analüütik
2005–2007 Ernst & Young, maksupersonal

**Täiendõpe**

oktoober 2018: andmeteaduse ja masinõppe konverents (Stockholm, Rootsi)

oktober 2018 Spark Programming (õpetaja Ali Benamara, Tallinn, Eesti)

PyCon ja PyData konverents (Karlsruhe, Saksamaa)

oktoober 2017: Predict, Euroopa juhtiv andmeteaduse konverents, modelleerimise töötuba (Dublin, Iirimaa)

juuli 2017: SIGMAPHI 2017, statistilise füüsika kursus (Korfu, Kreeka)

mai 2017: Python Coursera kursused:

Python kõigile
Pythoni andmestruktuurid
andmebaaside kasutamine Pythoniga
Pythoni veebiandmestiku kasutamine
lõpukursus: andmete hankimine, töötlemine ja visualiseerimine Pythoniga

juuli 2015: 5. doktorikool/konverents komplekssüsteemide matemaatilisest modelleerimisest (Patras, Kreeka)

**Teadustegevus**
Avaldatud teadusartiklite ja konverentsiteeside ning peetud konverentsiettekannete loetelu on toodud ingliskeelse CV juures.