Master's Degree programme
Second Cycle (D.M. 270/2004) in

Models and Methods of Quantitative Economics

Final Thesis

MODELLING FINANCIAL NETWORKS WITH KRONECKER GRAPH

Supervisor
Ch. Prof. Roberto CASARIN

Graduand
Diana KARIMOVA
866296

Academic Year
2017/2018

Contents

List of Figures

1 Introduction

Modern financial system is characterized as a large complex system of different institutions. The main function of any financial system is to provide reliable and stable linkages between investors and depositors. As a consequence, financial institutions are highly interdependent within the system, the failure of one can cause a cascade failure of the whole system. This phenomenon is referred as systemic risk in the modern economic literature.

Various measures of systemic risk had been prosed in recent years. There is a class of measures addressed to individual characteristics of institutions. This measures can detect the financial firms, usually banks, that are *too-big-to-fail* by analysing its size, complexity, and interconnectedness. However, this approach does not provide insights how the systemic risk will spread and the status of the whole system.

A better approach consists in studying financial institutions as a whole *network*. Network theory has a huge variety of applications and is becoming more and more popular in the field of economics and finance. The scope of this thesis is to study networks of financial institutions and its application to systemic risk measurement.

There are two aspects to be analysed: extracting network and generating a proper mathematical model. For the first one, Granger-causality method for constructing network is discussed. While Kronecker graph model was chosen to fit the real network. An important issue in generating networks to fit real data is parameter estimation, which is intractable because of the complex structure of the network. Thus, sampling methods based on Bayesian inference and Monte Carlo simulation methods were implemented.

2 Systemic risk and financial networks

The economic motivation of studying financial networks is to measure and predict systemic risk. A definition of systemic risk is given in modern literature as follows: systemic risk can be defined as the probability that a series of correlated defaults among financial institutions, occurring over a short time span, will trigger a withdrawal of liquidity and widespread loss of confidence in the financial system as a whole [Billio et al., 2010]. This definition leads to an approach to study systemic risk within the complex system of financial institutions constructed as a network.

2.1 Financial networks and their role in financial crisis

There is a strong interest in studying financial networks among the modern economists, especially after the Financial Crisis of 2007-2009 since systemic risk plays a key role during the financial crises. The systemic risk corresponds to the probability of failure of the whole system due to the high level of connections between financial institutions, thus the more comprehensive way to measure it is to study the system of financial institutions as a whole system.

Within the framework of financial networks level of interconnectedness is a good representation of degree of systemic risk. Several econometric measures have been proposed recently. For example, [Billio et al., 2010] study the linkages between hedge funds, banks, brokers, and insurance companies based on principal components analysis.

As inter-temporal changes are observed in the financial statements of all the institution, these changes lead to different levels of connectedness between financial institutions, thus the level of systemic risk during the times of financial stability differs from its level during financial crisis. This fact provides the intuition to study not only static analysis, but also its dynamic analysis, at least in different periods, such as "stability" and "turbulence".

The most important implication of systemic risk measures is that they could be used to produce warning signals. The empirical study discovered that the most important financial institutions, those who suffered the most during the financial crisis, have the largest measures of interconnectedness. This indirect econometric measures can serve as signals of market dislocation and distress, or financial crisis.

A framework for studying the relationship between the financial network architecture and the likelihood of systemic failures due to contagion of risk was also provided in [Acemoglu et al., 2013]. The study in this paper focuses on financial networks based on the liabilities between financial institutions, having debt contracts representing the edges in the network. The finding of that work discovers that highly interconnected financial systems, beyond the certain threshold, cannot

guarantee stability of the whole system in the presence of large shocks.

2.2 Timeline of financial crises from 1997 to 2013

There are three big economic and financial crises within the observation period of the dataset used. They had affected and changed the whole financial system and the ways it operates. During a financial crisis, the value of financial institutions drops rapidly due to various possible reasons, such as irrational behaviour or unrealistic expectations of the future values. A financial crisis usually causes an economy to go into a recession or depression and leads to structural changes in market regulations.

The first one to be analysed is a so-called *Dot-com bubble* that occurred in the period of extreme growth in usage of Internet and information technologies. The rapid growth of internet based companies and unrealistically high hopes about their future rising prices caused the excessive speculation on the market. After the introduction of the first Internet browser *Mosaic* in 1993, the technology became widely accessible and less costly, that increased the level of its usage. Meanwhile, growing level of computer education enhanced industrial and domestic usage of various information technologies and personal computers.

At the same time, low interest rates increased the availability of capital for the technological firms, encouraging big investment flows into the industry. The procedures of IPO for Internet companies, or *dotcoms*, were extremely efficient in term of attracting capital. The technological revolution around information technology, Internet and telephony, as well as the personal computers, helped to increase productivity in every other industry. This led to optimistic and often exaggerated expectations about technology companies in general, and Internet companies in particular. Investors were willing to take the advantage and invest in any company related to Internet, especially if it had *".com"* suffix in its name. In this environment, many investors neglected the traditional metrics of profitability, such as price/earnings ratio, and base confidence on technological advancements, leading to a stock market bubble. New technologies made it easy to operate private investment schemes. The popularization of the personal investment had affected American job market as well, as people were quitting their jobs to be engaged in full-time day trading.

The business models of the Internet companies were concentrated on attracting the capital and building customer awareness, that led to an aggressive marketing strategies and inefficient spendings. Weak business strategies caused the failure of dotcoms to turn a profit. Investors had hight expectations for short-run returns, however the companies failed to earn sufficient profits due to the inadequate business models. Between 1995 and 2000, the NASDAQ Composite stock market index, which included many Internet-based companies, rose 400% reaching its

peak on March 10, 2000 at 5048, and had lost 78% against its peak by October 2002. The burst of the bubble forced investors and economists to examine the measure of profitability for technology companies and come up with *"a new realism to the internet economy."*

Despite of the fact that many businesses were unable to survive the market crash, a few companies managed to get through the times of market instability. Companies such as Amazon.com, eBay, Priceline.com, Shutterfly are a few that could stay in the marker since their foundation in early 1990s.

The United States being the leader in the technological development, was also the source of the bubble. Most of the flagman technological companies were located in the USA and trades on American stock exchanges, while in Europe the number of such companies was significantly smaller in the 1990s, thus the speculation activity was not that strong. However, the American Dot-come crisis had affected world economy and European financial sector.

After a few years of recovery, there was another financial crisis caused by high default rate in subprime mortgage in the American financial market. It had started in the USA, yet shortly had developed into an international banking crisis. Global economic downturn followed right after that is considered to be the the worst economic crisis since the Great Depression of the 1930s.

Subprime mortgage bubble appeared as a result of accumulation of a risky loans in the federal financial institutions. First of all, low interest rates encouraged mortgage lending. As the number of these loans was quite large, many mortgages were bundled together and formed into new financial instruments called *mortgage-backed securities.* The procedure of *securitization* was supposed to create low-risk rate financial products, that were traded between banks often without the thorough check of the real risk evaluation. In fact, the rate of default of the original mortgage loans was very high, and the bundles that were created and traded with hight credit ratings were extremely risky. The lax regulation of such loans enlarged the scale of predatory lending, allowing banks and credit agencies issue mortgages with floating credit rate, that later caused high default rated. Bundles of subprime loans were sold, finally accruing to American quasi-government agencies, such as *Fannie Mae* and *Freddie Mac*, that provided an implicit guarantee by the US federal government that in its turn contributed to an excess of risky lending.

Manipulating the interest rate on mortgages resulted in high mortgage approval rates and drove up housing prices. This "bubble" burst by a rising single-family residential mortgages delinquency rate beginning in August 2006. The high delinquency rates led to a rapid devaluation of financial instruments, based on the mortgage loans, i.e. mortgage-backed securities including bundled loan portfolios, derivatives and credit default swaps.

As the value of these assets decreased, the market for these securities evap-

orated and banks who were heavily invested in these assets began to experience a liquidity crisis. A number of bailouts of quasi-government agencies had taken place. There is a huge discussion of how the American government had chosen financial agencies to bail out and to provide the federal support, as well as was its strategy successful. However, two biggest holders of subprime loans *Freddie Mac* and *Fannie Mae* were taken over by the federal government on September 7, 2008. Considerable amount of federal support for other financial institution, such as *Merrill Lynch*, *AIG*, *HBOS*, *Royal Bank of Scotland*, *Bradford & Bingley*, *Fortis*, *Hypo Real Estate*, and *Alliance & Leicester* followed shortly in 2009.

The active phase of the crisis, which manifested as a liquidity crisis, can be dated from August 9, 2007, when *BNP Paribas* terminated withdrawals from three hedge funds citing "a complete evaporation of liquidity." The most dramatic moment of the crisis was indeed the collapse of *Lehman Brothers* investment bank on September 15, 2008.

The consequences of the banking crisis in the USA were observed in many other countries resulting in global banking crisis. While the collapse of large financial institutions was prevented by the bailout of banks by national governments, stock markets still dropped worldwide. The housing market was the first to suffer, resulting in evictions, foreclosures, and prolonged unemployment. The crisis played a significant role in the failure of key businesses, declines in consumer wealth, and a downturn in economic activity and contributing to the *European sovereign-debt crisis.*

The bursting of the US housing bubble caused the values of securities tied to US real estate pricing to go down, damaging financial institutions globally. By the time when the mortgage bubble burst, the global financial market had developed to a level of high connectedness across the countries. New technologies of trading financial sequesters made it possible and less costly to invest overseas, leading to interdependence of financial markets in different countries.

One of the preliminary causes of the financial crisis of 2008 was easy credit conditions for loans. On one hand, it accelerated the speed of business growth and development, however rising the level of risk for financial institutions. It was clear that the standards of lending and borrowing should be reconsidered. In 2012 OECD realised a study, [Slovik, 2012], that suggests that bank regulation based on the Basel accords encourage unconventional business practices and contributed to or even reinforced the financial crisis. As a response, a new standards were issued in *Basel III*. These standards aim to strengthen the regulation, supervision and risk management of banks and were adopted by countries around the world.

After the US crisis the world economy experienced slow growth. Level of tax revenues stayed very low making high budget deficits unsustainable. Thus, in the end of 2009 the next wave of financial crisis hit the European countries. A few eurozone member states, that is Greece, Portugal, Ireland, Spain, Italy

and Cyprus, were unable to repay or refinance their government debt or to bail out over-indebted banks under their national supervision without the assistance of third parties. Events related to this are known as *Eurozone crisis* or the *European sovereign debt crisis.*

A complex combination of factors caused the crisis. In general, private debts were transferred to sovereign debt as a result of banking system bailouts and government responses to slowing economies. The fast recovery measures were impossible to implement due to the *currency union* union structure without fiscal union across the EU member states. Different tax and public pension rules resulted to the crisis and limited the ability of European leaders to respond to the arising difficulties. In addition, availability of complex financial instruments, currency and credit derivatives in combination with inconsistent accounting, off-balance-sheet transactions, made it possible to mask budget debts and deficit.

In late 2009 the new Prime Minister of Greece announced the true realistic size of the nation's deficits, that previous governments were hiding. Amount of Greece's debts was extremely large that actually exceed the size of the nation's entire economy, accounting to approximately 120% of the country's GDP. This can be marked as a beginning of the crisis. The market reacted by demanding higher yields on Greece's bonds, which raised the cost of the country's debt burden. anticipating problems similar to what occurred in Greece, investors acted the same towards other highly indebted countries in the region. Such "contagion" had spread across the region, as investors lost their confidence in government bonds causing the excessive selling.

As a measure to stop the crisis, countries Greece, as well as Ireland, and Portugal had received bailouts in 2010-2011. The *European Financial Stability Facility (EFSF),* a legal instrument financed by members of the eurozone, was proposed in May, 2010 to provide emergency lending to countries in financial difficulty. European Central Bank was involved in the process of restructuring the debts of the countries in need. In August 2011 ECB announced a plan, according to which it will purchase government bonds if necessary in order to keep yields on the optimal level. This measures had helped countries like Italy and Spain, that were too big to bailout by ECB or any other institution.

The major part of governments' debt was owned by European banks. They are required to keep a certain amount of assets on their balance sheets relative to the amount of debt they hold. A default of a government could lead to a reduction of their assets on balance sheet, and as a result a possible insolvency. The high level of interconnectedness in financial system plays a significant role in situations like this. A failure of a number of small banks can cause, *by domino effect,* further failures of other bigger financial institutions, as it had happened to Lehman Brothers. Its collapse was provoked by series of collapses by smaller financial institutions on US market.

In 2012, ECB authorities had once again confirmed bank's strong commitment to preserve eurozone. Some troubled European countries went down during the second half of the year and bond prices rose out of critical level. However, that did not solve all the existing problems, lower yields, have bought time for the high-debt countries to address their broader issues. In spite of the measures that ECB had taken together with IMF, the countries of the region continued to experience financial difficulties, resulting into banking crisis in Cyprus.

3 Mathematical models for generating real world networks

The idea of investigation networks in the real world is highly discussed in the modern literature. The use of these models is surprisingly wide - from medicine and biology, to economics, management and social sciences. The level of development of information technologies provides a field for empirical experiments. Availability of big datasets in various fields allows scientists use empirical examples to explore the properties among real world networks.

Though the first examples of networks originally come from neuroscience, it was recently discovered that many social and economic processes can be explained using network models. In the field of economics network analysis helps to analyse such problems as failures of financial institutions, contagion, international trade patterns, importance of social connections in the labour market, risk sharing across the individuals and social effects as immigration and aid transmission.

3.1 Properties of real world networks

Empirical studies of networks in real world have proved the particular features of networks, such as small diameters, heavy tailed distributions for degrees as well as specific temporal evolution patterns. As in [Barabási and Albert, 1999] it was first discovered that complex networks such as World Wide Web evolve continuously over time by adding new vertex, and new vertex have a property of *preferential attachment*, that is a new node tends to be connected to a similar set of existing nodes.

One of the key characteristics of any network is *a degree of a vertex*. The degree of a vertex is number of the other nodes to which one has a connection. *Degree distribution* is the probability distribution of these degrees over the entire set of nodes in network. It was proved that degree distribution follows the power law: $P(k) \sim k^{\gamma}$, where $P(k)$ is probability of having k links in a node. These features were discovered not only in WWW, but also in the networks of a quite different nature, such as scientific collaborations and actors playing in one film, see [Barabási, 2009].

Another important characteristic of a network is its topology or its geometrical form. In particular, the level of *connectivity* of a network reflects its shape. It is referred as a minimum number of elements (nodes or edges) that need to be removed to disconnect the remaining nodes from each other. Networks have different degrees of connectivity, that is how many vertexes are connected one to another by sequence of existing links. It is important to understand a topology of the network, as it is explains the patterns of spreading the contagion through the

network. The aspect of contagion comes from biological networks, and was later applied to financial networks to study the spreading of risk.

The simplest geometrical characteristics of a network is its *diameter*, that represents a linear size of the network. It is calculated as a he shortest distance between the two most distant vertexes in the network. The diameter in different networks is indeed different, however there is a surprising similarity across various networks as most of the real world networks have a small diameter. This property is often called "small world", was studied first in mailing experiment of Stanley Milgram and further has been observed in social, biological and technological networks.

3.2 Evolution of the graph modelling

The problem of finding a mathematically well defined model for generating graphs is of a big interest. The scientific approach requires a convenient mathematical model that will allow the researcher perform statistical tests and deliver forecasts, investigate various "what-if" scenarios in order to provide powerful insights, as well as study abnormalities in social and financial networks. Meanwhile, models should reflect all the properties of a real network.

The most intuitive mathematical representation of a network is a *graph*. In the classic literature the graph is defined as an ordered pair of disjoint sets $G = (V, E)$, where V is the set of nodes and E is the set of edges. The set $E \subset V \times V$ and defines the edge between to nodes as follows: if $x, y \in V$ and $x, y \in E$ then there is and edge between x and y. The intuition of representing real world networks as a graph is straightforward: nodes of a graph correspond to vertex, edges to links. One can expand the relationship between nodes by adding direction and weight, as to fit better to the real world network, depending on the research questions. This gives a certain flexibility for mathematical graphs and makes them so appealing for research.

The most common and convenient representation of graph is *adjacency matrix*. It is defined as $n \times n$ matrix A, where $n = |V|$, is the cardinality of the vertex set of G. The elements of matrix A are defined as follows:

$$a_{ij} = \begin{cases} 1, \text{if } ij \in E \\ 0, \text{if } ij \notin E. \end{cases}$$

This representation allows to use the theoretical tools of linear algebra and matrix operations to model specific features of networks. Symmetry of adjacency matrix A indicates that graph is indirected, if node x is connected to y, then node y is simultaneously connected to x. Otherwise, there is a separation between

the edges and the starting and ending point define their direction. Some useful characteristics of graphs must be mentioned, as they are the key factors that describe its structure.

Assume $i_1, ..., i_k$ is a sequence of nodes in some graph G. A walk from node i_1 to i_k is a sequence of links $\{i_1i_2, i_2i_3, ..., i_{k-1}i_k\}$, such that $i_{k-1}i_k \in G(E)$ for each k. A walk in which every node is distinct is a *path* and a shortest path is *geodesic*. It is easy to measure the scale of a graph by looking at its diameter – the largest geodesic (largest shortest path) or, as an alternative, at average path length. A graph can also contain a *cycle,* that is a path in which a node is reachable from itself.

The graph is called connected when there is a path between every node of the graph. This is not always the case, the degree of connectedness of nodes can vary across the nodes. Define $N_i(G) = \{j | ij \in E(G)\}$ – neighbourhood of the node $i \in V(G)$. The degree of the node i is defined as $d_i = |N_i(G)|$. Maximum and minimum degree of graph G defined as $\Delta(G)$ and $\delta(G)$ give the preliminary information about the topology of a given graph. The degree of connectedness can be interpreted as a community structure of a network and can serve for solving the problem of clustering.

The mathematical framework described above has provided rich theory, yet it is not consistent with real world networks. In reality, the links between the nodes are not always observed. Moreover, it is interesting to look further to evolution of network and build the predictive model.

This issues were partly solved by introducing random graphs by Erdös and Rényi [Erdös and Rényi, 1959]. In their model defined by $G(n, p)$ each node from $V(G)$, where $|V(G)| = n$, has a probability p to be connected to any other node independently. It was proved that degree distribution within the frames of this model is normal, which is a serious obstacle for fitting the real world data. Another issue of this model is its simplicity: the probability of having the edge between the nodes is equal across all the nodes, which is an unrealistic assumption. Since the first introduction of the random graph model a lot of alternative graph generating methods have been proposed, but there are still a lot of unsolved questions in this direction.

In 1998 Watts and Strogatz proposed a model that allowed random clusters in networks. Their model is more realistic compared to the random graph. It was observed from empirical data that networks with large number of nodes tend to have clusters and community structure, thus a new model was introduced to capture this properties, see [Watts and Strogatz, 1998]. Networks generated according to this model have low path length on average and high clusterisation level, that correspond to community structure of networks. The complexity of graph generation models consists of keeping both small diameters and community structure, as these two properties may contradict each other.

Three parameters are required to generate a network: N number of nodes, K as a mean degree, and $\beta \in [0, 1]$ as a probability parameter. The model generates a graph with N nodes and $NK/2$ edges in the following way:

- Construct a *ring lattice,* that is a set of ordered nodes arranged in a line where the last node is connected to the first one, with N nodes of mean degree $2K$, thus each node will be connected to its K nearest neighbours on either side.
- For each edge in the graph, rewire the target node with probability β.

There is a straightforward relationship between Watts-Strogatz model and a classic random graph: with $p = K/(N-1)$ and setting $\beta = 1$ the two models are equivalent.

The authors also introduced a measure of clusterization which determines whether a graph is a small-world network. Assume that i is a node and N_i is its neighbourhood in some graph G, $k_i = |N_i|$. The *local clustering coefficient* is then defined as a ratio:

$$C_i = \frac{|\{e_{jl} : j, l \in G(N), jl \in G(E)\}|}{k_i(k_i - 1)},$$

that is a proportion of links between the vertices inside its neighbourhood over the number of links that could possibly exist between them. For the undirected graph there is no distinction between links jl and lj, therefore the coefficient C_i must be multiplied by 2. The clustering coefficient over the network is defined as an average over the nodes: $\bar{C} = \sum_{i=1}^{N} C_i$.

For the case when $\beta = 0$, that corresponds to a simple ring lattice of the first stage of the generating algorithm, that is a single cluster. In this case the clustering coefficient is equal to $\bar{C} = \frac{3(K-2)}{4(K-1)}$ and converges to $3/4$ when for large K, while for $\beta = 1$ it is inversely proportional to network size: $\bar{C} = K/(N-1)$. The clustering coefficient remains close to the value $\frac{3(K-2)}{4(K-1)}$ for small β and decreases to $3/4$ only for β close to 1, thus the model is able to capture the local clusters in the network.

There is a class of models for generating complex networks that exploit the property of preferential attachment, the most famous of them was described in [Albert and Barabási, 2002]. In such models new nodes in the network are added one by one and are linked to the existing nodes with probability proportional to their degrees. While the degree distribution has heavy tails, the diameter of the generated network becomes larger with bigger number of nodes, which violates the small world property.

Another popular and rather simple model for community structured networks is *Stochastic Block Model* (SBM). It was initially developed for social networks as a particular case of a stochastic multigraph in [Holland et al., 1983] and later became a popular alternative to a classic random graph.

In the general case the SBM model is defined by three components: a number of nodes N, a partition of node labels into K disjoint sets $C_1, C_2, ..., C_K$ and a symmetric $K \times K$ probability matrix P. The edges then are generated at random respecting the given partition: any two nodes $u \in C_i$ and $v \in C_j$ are connected with probability p_{ij} which is (i, j) element of a matrix P. It is easy to notice that with all the entries of matrix P being identical the model is equivalent to Erdös-Rényi random graph. This definition requires a predetermined partition of nodes $C_1, C_2, ..., C_K$ that is usually unknown. Thus a more convenient definition is following. Denote p as a probability vector of dimension K, and W as $K \times K$ matrix with elements in $[0, 1]$ interval. Each node has a community label from the set $\{1, 2, ..., k\}$ according to a given probability vector p, while a node with a label i is connected to a node with a label j with probability $(W)_{ij}$.

SBM model is a convenient instrument in machine learning and computer science. The algorithms that analyse parameters' combination allow to detect whether there is a latent community structure or clusters in the observed network by comparing models' fit. These algorithm can be either for *detection* of the community structure or its *recovery*. An extended version of this model named Weighted Stochastic Block Model was used in [Casarin et al., 2017] to capture the community structure network of financial institutions.

To sum up, all the mentioned models can capture some properties of the real world networks while neglecting the others. However, the model based on Kronecker multiplication is able to match multiple properties, this model will be discussed in details in the next chapter.

4 Kronecker graphs

The aim of studying different modes of network creation is to model real life features as close as possible. Some models can capture either small diameter property, or heavy tailed degree distribution, but never all of them together. However, the model based on the *Kronecker multiplication* of matrices manages to satisfy all those features. At first, consider the two matrices A and B, their Kronecker product is defined as follows:

Definition. For two given matrices $\mathbf{A}$ and $\mathbf{B}$ with corresponding dimensions $n_1 \times m_1$ and $n_2 \times m_2$ the Kronecker product $\mathbf{C}$ of dimension $(n_1 n_2) \times (m_1 m_2)$ is defined by

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{1,1}\mathbf{B} & a_{1,2}\mathbf{B} & \dots & a_{1,m_1}\mathbf{B} \\ a_{2,1}\mathbf{B} & a_{2,2}\mathbf{B} & \dots & a_{2,m_1}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n_1,1}\mathbf{B} & a_{n_1,2}\mathbf{B} & \dots & a_{n_1,m_1}\mathbf{B} \end{pmatrix}$$

The operation was named after a German mathematician *Leopold Kronecker* (1823 - 1891), though it was another scientist, *Georg Zehfuss* (1832-1901) who was first to introduce this operation and its properties in 1858. He worked on linear algebra as well as physics and thermodynamics and exploited a tensor matrix multiplication a lot in his works. He was also the first to prove the determinant relation $|A \otimes B| = |A|^n |B|^m$ for square matrices A and B of dimension n and m correspondingly.

Being a particular case of a tensor product, Kronecker multiplication has a set of simple and handy properties, like bilinearity and associativity:

$$\text{(i)} \;\; A \otimes (B + C) = A \otimes B + A \otimes C$$

$$\text{(ii)} \;\; (A + B) \otimes C = A \otimes C + B \otimes C$$

$$\text{(iii)} \;\; (A \otimes B) \otimes C = A \otimes (B \otimes C)$$

$$\text{(iv)} \;\; \lambda A \otimes B = A \otimes \lambda B = \lambda (A \otimes B), \text{for some scalar } \lambda$$

As the adjacency matrix of some graph, or network, in the deterministic case consists of 0 and 1, the Kronecker product of two adjacency matrices is as well an adjacency matrix. Therefore, *Kronecker product of two graphs* with matrices A and B can be defined as a graph with adjacency matrix $A \otimes B$. In the context of this definition, it is easy to look at the *Kronecker power* $G^{[k]}$ of some graph G, which model will be described later.

4.1 Properties of Kronecker graph

It is interesting to look at the properties of the Kronecker graph, that makes it handy in modelling real world networks. Geometric characteristics of Kronecker product of graphs were well studied and proved in [Weichsel, 1962]. For instance, if both two components of the multiplication are connected graphs, the product will be connected if and only if the either of them contains an odd cycle. If at least one of two graphs is a disconnected graph, then the Kronecker product is also disconnected.

Another remarkable property of the Kronecker product of two matrices is its invariability under the process of permutation. A *permutation matrix* $\mathcal{P}$ is a square binary matrix that has one entry of 1 in each row and each column with 0s elsewhere. Permutation matrix is orthogonal: $\mathcal{P}\mathcal{P}^T = I$, where I is identity matrix. Pre-multiplication of a matrix A by permutation matrix $\mathcal{P}$ of the corresponding size represents permutation of rows, while post-multiplying represents permutation of columns.

Notice that, for any permutation matrices $\mathcal{P}_1$ and $\mathcal{P}_2$ there exists a permutation matrix $\mathcal{P}$, such that

$$\mathcal{P}(A \otimes B)\mathcal{P}^{-1} = (\mathcal{P}_1 A \mathcal{P}_1^{-1}) \otimes \mathcal{P}_2 B \mathcal{P}_2^{-1},$$

i.e. regardless of the node order of A and B their Kronecker products will be the same up to isomorphism. This condition is useful in the modelling, as it is not always possible to track the correspondence between vertices in the graphs and nodes in the real network.

The hierarchical nature of the Kronecker product allows to track the nodes in $A \otimes B$ by decomposing it two parts. An edge (X_{ij}, X_{kl}) lies in $A \otimes B$, if and only if $X_{ik} = (X_i, X_k)$ is an edge in A and $X_{jl} = (X_j, X_l)$ is a node in B. The result of Kronecker product of two matrices is a block matrix, that consists of the blocks similar to the first matrix in the multiplication. This property is useful when describing real networks, as they were proven to have a block, or community, structure as well, therefore with a proper initiator matrix, its Kronecker power will serve as a good model for real networks.

The model of Kronecker power graph, developed in [Leskovec et al., 2010] is designed as an iterative multiplication of an *initiator matrix*. Denote the initiator as G_1, the $k-$th Kronecker power is a matrix G_k, such that

$$G_k = G_1^{[k]} = \underbrace{G_1 \otimes G_1 \otimes \cdots \otimes G_1}_{\text{k times}} = G_{k-1} \otimes G_1, k = 1, 2, ...$$

From now firther the term *Kronecker graph* will refer to the $k-$th Kronecker

power of some initiator matrix. The defined matrix G_k is used as an adjacency matrix to represent some real world graph. It was proven in [Leskovec et al., 2010] that Kronecker graph model performs a good fit for a number of large real-world networks, such as citation graphs, Internet Autonomous Systems, web and blog graphs, collaboration networks of co-authorships, internet and peer-to-peer and the like. However, there is no published literature of modelling financial networks with Kronecker graph.

The choice of Kronecker power k depends on both number of nodes in the real network of interest and the dimension of the initiator matrix G_1 itself. It was proven that it is better to have a model with more nodes than it is in the real network. To achieve the balance of nodes number, a few isolated nodes can be added to the real network. It will not affect the goodness of fit, while deleting nodes from the real network can affect the degree distribution.

Kronecker power preserves the diameter of the graph. If the initiator graph G_1 has a diameter d, then its Kronecker power of any order k $G_1^{[k]}$ will also have diameter d. It follows from the edge transitions in the Kronecker product. This property is a common feature for the real world networks, sometimes referred as *small world* property. Moreover, in some real world networks, the diameter is shrinking, the distance between nodes decreases as number of nodes gets bigger. This is true for world wide web network and citations.

Each Kronecker multiplication exponentially increases the size of the graph, that is G_1^k has N^k nodes and E^k edges, where N is number of nodes in initiator adjacency matrix and E is its number of edges. Empirical observations over different networks tell that number of edges grows when new edges appear, therefore the network becomes denser over time. Such phenomena is called *densification power law*: $E(t) \propto N(t)^a$, where a is some constant. According to [Leskovec et al., 2005b] the constant a lies between 1 and 2 for different real world networks, $a = 1$ corresponding to constant average degree over time, while $a = 2$ corresponds to an extremely dense network. In this sense, Kronecker graph modes is able to capture the behaviour of real network.

It is clear to see that Kronecker graphs have multinomial degree distributions. Consider some node of a degree d in the initial graph G_1. With every multiplication this node will expand to a sequence $\{d*d_1, d*d_2, \ldots, d*d_N\}$ where $\{d_1, d_2, \ldots, d_N\}$ is a sequence of degrees of a corresponding multiplier. Graph $G_1^{[k]}$ will then have a degrees of the form $d_{i_1} * d_{i_2} * \cdots * d_{i_k}$ for every combination of indices $i_1, i_2, \ldots, i_k \in (1, ..., N)$ and N being a number of nodes in G_1. This results to the multinomial distribution of the degrees for every Kronecker power.

Multinomial distribution is in character for the spectral values of Kronecker powers. Eigenvalues as well as the components of eigenvectors follow multinomial distribution.

4.2 Extension to the random graph

In a stochastic adjacency matrix every entry represents a probability of an edge between the corresponding nodes. Kronecker graphs can also be transformed into a stochastic graph.

Construction of a stochastic Kronecker graph faces some challenges, such as the mechanism of introducing the randomness and choosing the appropriate parametrization. The easiest way to make a graph stochastic is to generate deterministic Kronecker graph and then choose uniformly at random the elements of the adjacency matrix and replace 0 to 1 and visa versa. However, it is not the optimal way, as this procedure will affect the degree distribution, making it binomial as in the classic random graph.

A better approach to introduce probabilities in the model is to transform all the entries of the initiator matrix by values in $[0, 1]$ interval as a probability of an edge, by that transforming in into a *probability matrix*. Notice that sum by columns or by row will not necessarily be equal to 1, as the elements of the matrix are independent. The result will be a generalization of Erdös-Rényi model. The Kronecker power of such a probability initiator matrix will be a probability matrix as well.

The second step is to choose an appropriate parameters in the model. First, all the elements in the initiator matrix can be considered as independent parameters. This approach is not efficient as the number of parameters will be N^2 and the problem of overfitting arise. The level of complexity of calculations will be too high as well. Another extreme case is to introduce one single parameter for all the elements in the initiator matrix. In this case model will be equivalent to a classic Erdös-Rényi random graph, which is shown below.

Proposition When the initiator matrix of the stochastic Kronecker graph has identical probability p over the edges,$\Theta = (\theta_{ij}) = p$, model is identical to classic random graph model $G(n, \tilde{p})$ with $\tilde{p} = p^k$.

Proof The structure of Kronecker multiplication allows to explicitly write the resulting matrix in the case when $\Theta = (\theta_{ij}) = p$. Consider the Kronecker multiplication of two $n \times n$ matrices of the stated form $\Theta \otimes \Theta$:

$$\Theta \otimes \Theta = \begin{pmatrix} p\Theta & p\Theta & ... & p\Theta \\ p\Theta & p\Theta & ... & p\Theta \\ \vdots & \vdots & \ddots & \vdots \\ p\Theta & p\Theta & ... & p\Theta \end{pmatrix} = (p^2)_{i,j=1...n^2} = \Theta^{[2]}$$

Furthermore,

$$\Theta^{[3]} = \Theta^{[2]} \otimes \Theta = \begin{pmatrix} p^2\Theta & p^2\Theta & ... & p^2\Theta \\ p^2\Theta & p^2\Theta & ... & p^2\Theta \\ \vdots & \vdots & \ddots & \vdots \\ p^2\Theta & p^2\Theta & ... & p^2\Theta \end{pmatrix} = (p^3)_{i,j=1...n^2} = \Theta^{[3]}$$

It is obvious that the k-th multiplication matrix the $\Theta^{[k]}$ will have the entries $(\theta_{ij}) = p^k$. $\square$

In this context, two-parameter initiator matrix is appealing to be tested. In the original paper of Kronecker graph [Leskovec et al., 2010] the authors use only two parameters α and β both being in the interval $[0, 1]$ to reshape the deterministic initiator matrix. All the elements of value 1 are replaced by α, while all the elements of value 0 are replaced by β. This model was then fitted to the citation network with 4×4 initiator matrix chosen as a star graph, which is one node connected to all the other nodes without any other edges. The stochastic Kronecker graph managed to match the qualitative structure of the real data having a similar degree distribution as in the real network, therefore the approach is proven to be reliable.

The comparison of the proposed Kronecker graph and the real network can be done by matching the adjacency matrices. Having a probability matrix in $G_1^{[k]}$ the adjacency matrix is straightforward. Then the likelihood function can be calculated as a measure of similarity of two matrices and the value of the likelihood will serve as a goodness of fit.

5 Network of Financial Institutions

In the case of the network of financial institutions the procedure of establishing the connections between node is not obvious. There are different ways to construct financial networks, one of the most promising one is based on Granger causality relation between institutions.

5.1 Granger causality networks

The approach of extracting networks using Granger-causality relations was introduced in [Billio et al., 2012], where it was implemented to capture the connectedness between fnancial institutions of four types, which are hedge funds, publicly traded banks, broker and insurance companies. The 25 largest financial institutions in each category were used for the analysis. According to the authors, development of more complex financial instruments and derivative, together with the practice of more sophisticated insurance strategies in the latest years made it reasonable to include other types of the financial institutions into the models of estimating systemic risk. The main idea to extracting a network from the available financial data was to model Granger-causality relations between institutions based on monthly equity returns using the rolling window approach. The advantage of using market returns over other indicators, such as accounting variables, lies in the faster reflection of the market changes. Moreover, application of Granger causality approach helped to discover the unusual asymmetry in the connections: the returns of banks and insurers seem to have more significant impact on the returns of hedge funds and broker/dealers than vice versa. That asymmetry can serve as an evidence of additional risk that banks and insurers may have taken on, which cannot be managed by traditional regulatory instruments, and potentially be the source, as well as an indicator, of the financial crisis. To establish the connections, linear Granger-causality tests were applied over 36-month rolling sub-periods of monthly data returns. Furthermore, Granger causalities for daily returns with the rolling window approach were used in [Casarin et al., 2017], as well as in [Billio et al., 2016], for obtaining a network of financial institutions in Europe. The dynamic network constructed in such a way, was then used to study contagion patterns and community structure of the system.

The notion of so-called Granger causality for time series data lies in the linear dependencies, that can be detected by linear regression coefficients. Suppose that vectors X and Y are two different time series, then X is said to *Granger-cause* Y if X values provide statistically significant information about future values of Y and it can be confirmed by t-tests. Formally, for the case of two time series of market returns, the definition can be written as follows:

Definition. Let $\{X_t\}$ and $\{Y_t\}$ be the time series of market returns. Consider

linear regression models of the form

$$\begin{cases} X_{t+1} = a_1 X_t + a_2 Y_t + \epsilon_{t+1} \\ Y_{t+1} = b_1 Y_t + b_2 X_t + \eta_{t+1} \end{cases}$$

where ϵ_{t+1}, η_{t+1} - uncorrelated normal error terms, a_1, a_2; b_1, b_2 - coefficients of the model. Then, if a_2 is significantly different from 0, then Y_t Granger causes X_t. Correspondingly, if b_2 is significantly different from 0, then X_t Granger causes Y_t.

The dataset that was used in the current estimation experiment consists of the networks (for 210 time points) of European financial firms classified under the ICB code class 8000. Daily returns obtained from source, DataStream were used to estimate Granger causalities, summing up for monthly periods. The time period considered is from 8 January, 1997 till 16 January, 2013.

6 Bayesian inference for parameters of network

6.1 Metropolis-Hastings sampling algorithm

Estimating the parameters of the graph most of the time is intractable because of the complex structure of the network. The standard methods based on sampling a number of values of the parameter do not work in this case. To provide the proper statistical inference for the parameters of the stochastic graphs models in general, and Kronecker graph model in particular, it's better to use sampling methods based on Bayesian inference and Monte Carlo simulation methods such as Markov Chain Monte Carlo sampling.

In Bayesian inference difficulty arise from the fact that the true type of distribution of parameter is unknown. For such situations algorithms, that come from so-called Markov Chain Monte Carlo (MCMC) family, allow to obtain samples of the parameter of interest that converge to the true distribution as the sample size grows. The simplest version of MCMC algorithm of sampling a target distribution results in a Markov chain that converges to that target distribution.

Definition. Let $X_1, X_2, ..., X_t$ be a sequence of random variables, $x_1, x_2, ..., x_t$ - its realization, $\mathcal{X}$ - sample space for X_t and $A \subset \mathcal{X}$. Then if for all t

$$\mathbb{P}(X_{t+1} \in A | x_0, x_1, x_2, ..., x_t) = \mathbb{P}(X_{t+1})$$

for all such A., the sequence is called *Markov chain*. The probability distribution of any state of a Markov chain given all the preceding states depends only on the previous realization on the chain.

The nature of Markov chain makes it convenient to describe its evolution only by providing the rule of transition from state X_t to the state X_{t+1}, that is called the *transition kernel*. However, the first values of the chain are highly dependent on the starting value X_1. Therefore, when Markov chains are used for creating samples of some distribution, the first steps of chain are usually removed from the sample as *burn-in* or *warm-up* to achieve the independent sample in the end.

Modern MCMC algorithms evolved from the original Monte Carlo simulation methods at the early age of computers development. In many cases the analytical expression of the quantity of interest is not available, but can be expressed as an integral, which is an expected value of some function. For example, let I be the quantity of interest, such that it can be written as

$$I = \int h(x) f(x) dx = \mathbb{E}_f[h(x)], \text{where} f(x) \text{ is a density function.}$$

If the iid sample $X_1, X_2, ..., X_n$ from the distribution $f(\cdot)$ is available, then the

quantity of interest I can be approximated by an empirical mean value of the function h evaluated at sample:

$$\hat{I} = \frac{1}{n} \sum_{i=1}^{n} h(X_i).$$

The result is an unbiased estimator, its accuracy depends on the number of observations in a sample n. Having large enough samples, Monte Carlo method of simulation can produce estimation of the required accuracy. In particular, by the standard central limit theorem, the accuracy is inversely proportional to the square root of the sample size. Being first developed by Stanislav Ulam and John von Neumann in 1940s to resolve complicated calculations in nuclear physics, the method is now widely used to calculate the deterministic values using the statistical sampling and has applications in a wide range of disciplines, from physics and biology to finance. In spite of its relatively simple application, Monte Carlo method relies on simulation of large samples of random or pseudo-random numbers, that could be an intractable problem in the early era of computing machines. Therefore, the method needed further development.

According to [Robert and Casella, 2011], after the approbation of the original version of the Monte Carlo method on the on of the first computers, that had certain limitations in large calculations, further evolution of the technique continued in the following way. The availlabilty of the first computers and techliques for generation of pseudorandom numbers made it possible to test the proposed way of calculations. In order to improve the efficiency of the existing Monte Carlo method, Metropolis proposed the random walk modifications, that was publidhed in [Metropolis et al., 1953]. Metropolis algorithm generates the sequence of observations, in which at each step the next observation is accepted with a certain probability, otherwise the previous observation is replicated. With a large number of iterations, the sequence converges to the stationary distribution.

The Metropolis algorithm was later generalized by Hastings in [Hastings, 1970], who used the Markov process and its transition kernels to construct the sequence of states that converges to the desirable distribution. The motivation to use Markov chains is that it is uniquely defined by its transition probabilities, i.e. transition kernels, that simplifies the calculation.

The formal description of Metropolis-Hastings algorithm should be started with the conditions, that will guarantee the convergence of the algorithm to a unique desirable distribution. The *existence* and *uniqueness* of such a distribution π is guaranteed by the mathematical properties of the Markov process, therefore the process converges to a desirables distribution. The distribution π is called stationary for a Markov chain if it is invariant to state changes, i.e. it is preserved at each state. For instance, if π is a distribution for X_1, then X_2 has distribution π and so on. Formally, if we have two different states x and x', the following must

be true: $\pi(x)\mathbb{P}(x'|x) = \pi(x')\mathbb{P}(x|x')$. The existence of distribution of a sequence is given by construction of the algorithm, such that the chain in MCMC algorithms is always constructed to have a specific distribution.

The steps of Metropolis-Hastings algorithm can be described as follows. Assume that π is a target density; $x_1, x_2, x_3, \ldots, x_n$ is a sequence to be constructed; x_j is a current state of the sequence and $q(x|x_j)$ is the proposal distribution. At each iteration the following steps are realized:

Metropolis-Hastings algorithm

- Sample x^* from the proposal distribution, $x^* \sim q(x|x_j)$
- Calculate the acceptance probability $\rho = \min\{1, \frac{\pi(x^*)}{\pi(x_j)} \frac{q(x_j|x^*)}{q(x^*|x_j)}\}$
- For the proceeding state of the sequence x_{j+1}:

$$x_{j+1} = \begin{cases} x^* & \text{with probability } \rho \\ x_j & \text{with probability } 1 - \rho \end{cases}$$

There exist other variations of accept-reject algorithms, but the particularity of the Metropolis-Hastings algorithm is that it can remain at the previous state. Another special case when a proposal density satisfy $q(x|y) = q(y|x)$ is called the *Metropolis update*.

The further improvement of sampling can be achieved by adding a random walk transformation for the proposal. Random walk jumps make the sequence visit all the possible values of the support of the target distribution and therefore result in a more realistic samples. at every iteration the random walk modification can be added as: $\tilde{x} = x^* + \epsilon$, where $\epsilon \sim \mathcal{N}(0, \tau), \tau > 0$. It will be shown later in the empirical experiment that variance τ plays important role. Its optimal value depends on the scale of the support. Larger values of τ correspond to bigger jumps of the sequence, while very small values can cause the sequence to stuck near the initial point.

6.2 Sampling parameters of random graph $G(n, p)$

To begin with, MCMC method can be applied to a trivial case of Kronecker graph model, where the edges have identical probability p over all pair of nodes across the network. Hence, the initiator matrix will be of the following type: $\Theta = (\theta_{ij}) = p$. and Kronecker graph is identical to Erdös-Rényi random graph model $G(n, \tilde{p})$ with $\tilde{p} = p^k$.

Assume that the stochastic initiator matrix of the Kronecker graph is assumed to be a $2 \times 2-$matrix with one parameter p:

$$\Theta = \begin{pmatrix} p & p \\ p & p \end{pmatrix}$$

As it was mentioned in the previous section, the trivial version of Kronecker graph is then constructed to fit the given adjacency matrix. Since the dimension of the initiator matrix is 2×2, the dimension of the $k - th$ Kronecker power of Θ will be $2^k \times 2^k$. Denote the dimension of the Kronecker graph as N, then $N = 2^k$, which leads to $k = \log(N)/\log(2)$. The resulting Kronecker power of Θ, which is denoted as $\Theta^{[k]}$ should have enough nodes to fit the real adjacency matrix, but the number of nodes in the adjacency matrix is not always an even number, or more general, the value $\log(N)/\log(n)$ can be a fraction if n is the dimension of the Kronecker initiator. Thus, the power k would be calculated as

$$k = \lfloor \log(N)/\log(n) \rfloor + 1.$$

If the number of the nodes in the Kronecker power $\Theta^{[k]}$ is greater than the one of the real adjacency matrix A, the lacking rows and columns will be added having values 0 in all the cells. That corresponds to the procedure of adding isolated nodes to the existing network, which will not affect the results of the fit, while the removal of random nodes corrupts the degree distribution.

The symmetrized version of the adjacency matrix A was used to simplify the further calculations. The symmetric adjacency matrix corresponds to undirected network and the direction of the links can be neglected in this experiment. A simple procedure was used to modify the initial matrices: the new version of the matrix A^{symm} is obtained from the original as a sum of a matrix with its transposition $A^{symm} = A + A^T$, in which all the zeros stayed unchanged, while all the other numbers, which is either 1 or 2, changed to ones. In the further calculation the notation A refers to a symmetrized version of the original adjacency matrix.

The aim of the empirical experiment is to estimate the unknown parameter p by applying Metropolis-Hastings algorithm. The adjacency matrix A contains the information of the links between the financial institutions at a fixed time

point t, which lies between 8 January, 1997 and 16 January, 2013. As prior *beta*-distribution $q \sim \mathcal{B}(a, b)$ has been used, with parameters $a = b = 4$, that gives a "flat" density function. The posterior distribution of the parameter is unknown, thus Metropolis-Hastings algorithm is used to approximate it.

In the case of the general Kronecker graph model likelihood function $\mathcal{L}$ depends on the latent graph $\mathbb{G}$, the observed adjacency matrix A and the nodes permutation σ:

$$\mathcal{L}(\mathbb{G}, \sigma, A) = \prod_{(u,v)\in A} \mathbb{G}(\sigma_u, \sigma_v) \prod_{(u,v)\notin A} (1 - \mathbb{G}(\sigma_u, \sigma_v)),$$

where $\mathbb{G}(i, j)$is the (i, j) element of the graph $\mathbb{G}$ (u, v) is an edge between nodes with numbers u and v, while σ_u, σ_v denote the node permutation. It is calculated as similarity measure between the real data and the fitted matrix. If the link is presented or not presented in both real matrix and the fit, the value of the product is equal to 1. In the general case of Kronecker graph permutation of nodes is necessary, as the likelihood function is calculated by comparing elements of adjacency matrix to Kronecker graph without predetermined order of nodes. However, in the case of a random graph model, it can be neglected as all the cells in the fitted graph are identical.

The algorithm generates the sequence of values p_j as was proposed by the Metropolis-Hastings method. The initial value of the parameter can be chosen randomly, as it does not affect the resulting sample. At each iteration p_j is generated according to the following algorithm:

Metropolis-Hastings algorithm of sampling parameter p

- Set the current value of the sequence as equal to the previous values:

  $$p_j = p_{j-1}$$

  To be consistent with the indexes, the initial value of p_1 is chosen randomly and the iteration process starts with $j = 2$.

- Use the logarithmic transformation of p_j :

  $$\tilde{p}_j = \phi(p_j) = \log\left(\frac{p_j}{1 - p_j}\right)$$

- Propose $\tilde{p}_j^*$ as a random walk centred at $\tilde{p}_{j-1}$, calculated at the previous step:

  $$\tilde{p}_j^* = \tilde{p}_{j-1} + \gamma$$

where $\gamma \sim \mathcal{N}(0, \tau)$ is the standard normal distribution and $\tau > 0$ is the variance of the random walk transformation, that is taken to be equal to $\tau = 0.1$

- Transform the proposed value to: $p^* = \frac{1}{1+\exp(-\tilde{p}_j^*)}$
- Generate the Kronecker graph $\mathbb{G}$ as the k^{th} power of the initiator matrix for the current parameter p_j, $\mathbb{G} = \Theta^{[k]}$, as well as for the proposal parameter p^*, $\mathbb{G}^* = \Theta^{*[k]}$
- Calculate the Metropolis-Hastings acceptance rate ρ :

$$\rho = min\left\{1, \frac{\mathcal{L}(\mathbb{G}^*, A)}{\mathcal{L}(\mathbb{G}, A)} \frac{q(p^*; a, b)}{q(p_j; a, b)} \frac{J(\phi^{-1}(p^*))}{J(\phi^{-1}(p_j))}\right\}$$

where $\phi(p) = \log(\frac{p}{1-p})$, $q(\cdot; a, b)-$ density function of beta distribution with parameters a, b and $J(\phi^{-1})$ is a Jacobian of the transformation ϕ^{-1}, that is equal to

$$J(\phi^{-1}(p)) = \frac{\exp(-p)}{(1-\exp(-p))^2}$$

- Accept the proposal p^* as a next iteration of the sequence p_{j+1} with probability ρ and set $p_{j+1} = p_j$ with probability $(1-\rho)$

$$p_{j+1} = \begin{cases} p^*, & \text{with probability } \rho \\ p_j, & \text{with probability } 1-\rho \end{cases}$$

Similarly, update the intermediate values of $\tilde{p}_{j+1}$ as follows:

$$\tilde{p}_{j+1} = \begin{cases} \tilde{p}^*, & \text{with probability } \rho \\ \tilde{p}_j, & \text{with probability } 1-\rho \end{cases}$$

In practice to realize the accept-reject mechanism a random number u is drawn $u \sim \mathcal{U}_{[0,1]}$ and the proposal p^* is accepted when $u \leq \rho$ and rejected otherwise:

$$p_{j+1} = \begin{cases} p^* & \text{if } u \leq \rho \\ p_j & \text{if } u > \rho \end{cases}$$

The proposed algorithm was implemented in MATLAB and applied to the described dataset. The results of the algorithm applied to adjacency matrix of the first available data point, that is January, 8, 1997, is the following. The sample has a unimodal density function, the kernel estimation of the posterior distribution is shown on Figure 1. The number of iterations used in the algorithm was set to 1000, and the first 200 iteration were neglected to eliminate the dependence of the sample from the initial value.

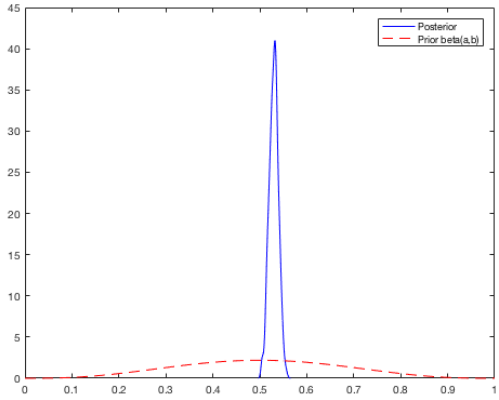
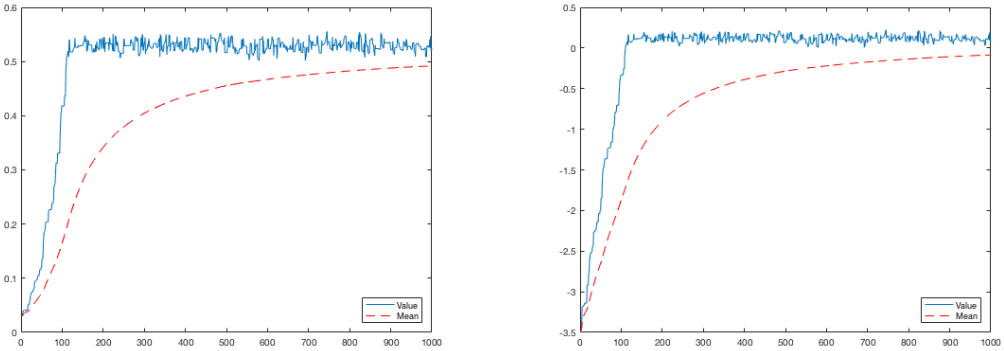


Figure 1: Prior *(dashed line)* and posterior *(solid line)* distributions of the initiator matrix parameter p, fitted to the adjacency matrix of the date *January, 8, 1997*

As the algorithm goes on, the values of p and the intermediate parameter $\tilde{p}$ visits various values in the support of the distribution, but concentrates on the "true" value, that is most likely to be accepted. It can be also seen from the Figure 2 that the sequence stabilizes. In addition, it is clear to see why the first observations in the sequence is necessary to be deleted as they are highly dependent on the initial value due to the Markov nature of the algorithm.



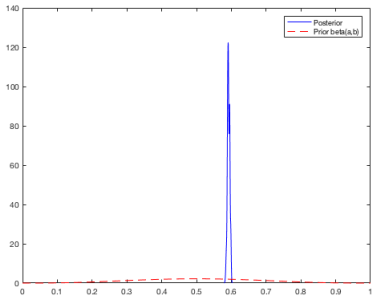
(a) Sample of the parameter p, basic values

(b) Sample of the parameter $\tilde{p}$, random walk

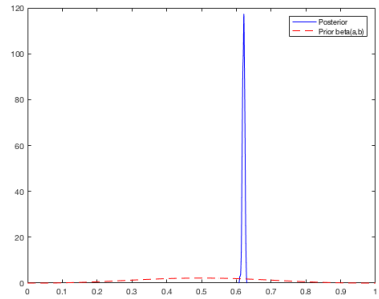
Figure 2: Sample plots of initiator matrix parameter p for the adjacency matrix of the date *8, January, 8 1997:* value of the parameter *(solid line)* and mean value *(dashed line)*.

Running the algorithm for different adjacency matrices of the given dataset result in the same types of the densities, as in the Figure 3.

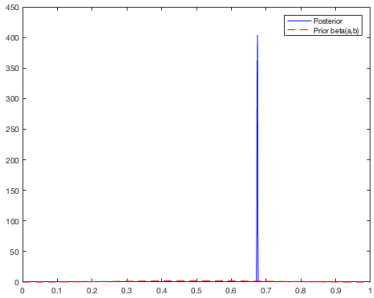
Another interesting observation that can be concluded from studying sample is that the shape of the posterior distribution depends on the variance of the



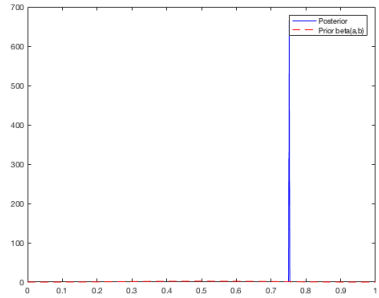
(a) Date: *11 October 2000*



(b) Date: *8 September 2004*



(c) Date: *06 August 2008*



(d) Date: *11 April 2012*

Figure 3: Prior *(dashed line)* and posterior (*solid line)* distributions of the initiator matrix parameter p at different points of time

random walk transformation τ. It can be seen from the Figure 4 that small values of random walk variance give more flat distribution, while bigger values of the variance result in the univariate densities with high peaks. All the densities in the figure 4 are calculated for the 1000 iterations with 200 burn-in. When the variance of the random walk is too small, the algorithm does not converge to one value, as the jumps are too small to visit all the support.

Figure 5 shows that sample with very low variance (5a) moves slowly and does not converge within 1000 iterations, meanwhile the sample with large variance (5b) jumps too far away from the support of the true parameter distribution and then stays at one value for the proceeding iterations. It can also bee seen from the graph that acceptance rate is lower in the case of $\tau = 0.5$.

To proceed further, all the static estimations from available time points can be combined in one graph to see the changes over time. Parameter p estimate reflects the density of the network, that is the number of edges over the possible number of edges in the complete graph. In the classic mode of random graph $G(n, p)$ the entries of the matrix refer to a probability of an edge. However, in the setup of Kronecker graph, while using parameter p in the initiator matrix, the final probability of an edge is in fact equal to p^k, where the k is the Kronecker power

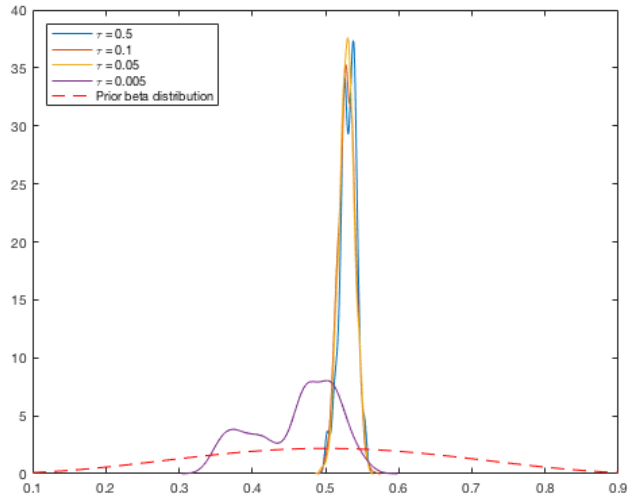
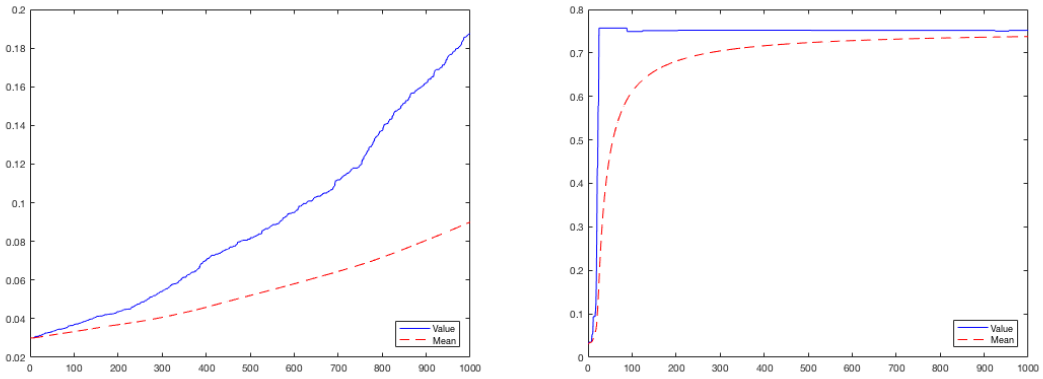


Figure 4: Kernel density estimations of the posterior distribution of the initiator matrix parameter p with different values of the random walk variance τ *(solid lines)*



(a) Sample pf parameter p for $\tau = 0.005$ (b) Sample pf parameter p for $\tau = 0.5$

Figure 5: Samples of initiator matrix parameter p generated with different values of random walk variance τ

of the initiator matrix, that depends on the size of the real adjacency matrix A. Therefore, for the actual comparison of edge probabilities values of p^k should be taken. As the power function is monotonically increasing, the relative comparison of values of p instead of p^k at different time point will lead to the same conclusion.

The obtained results can be compared to the empirical evidence of financial crisis hat affected European financial system from 1997 till 2013. The hypothesis that during the times of financial instability the network gets denser as the insti-

tutions are becoming more interconnected was confirmed in the latest literature and the described model of the random graph does not contradict it. It can be seen in Figure 6, that value of the parameter p is higher during the peaks of the crises.

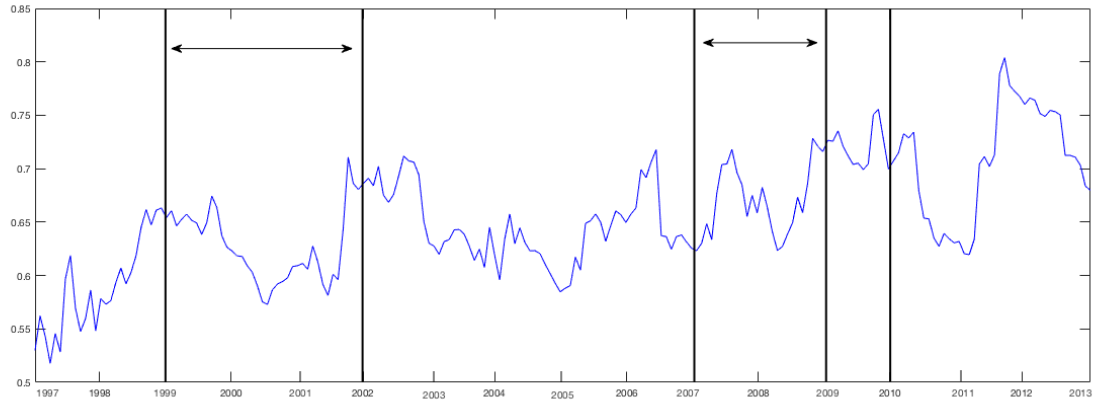
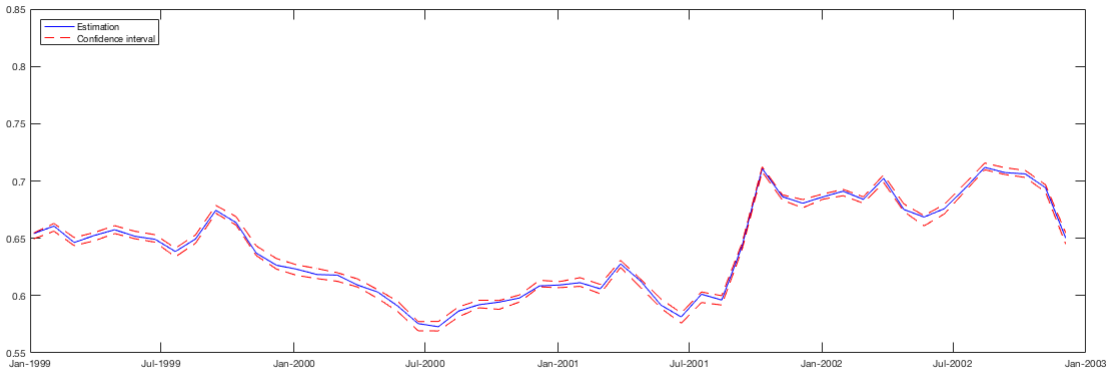


Figure 6: Time varying the parameter fit

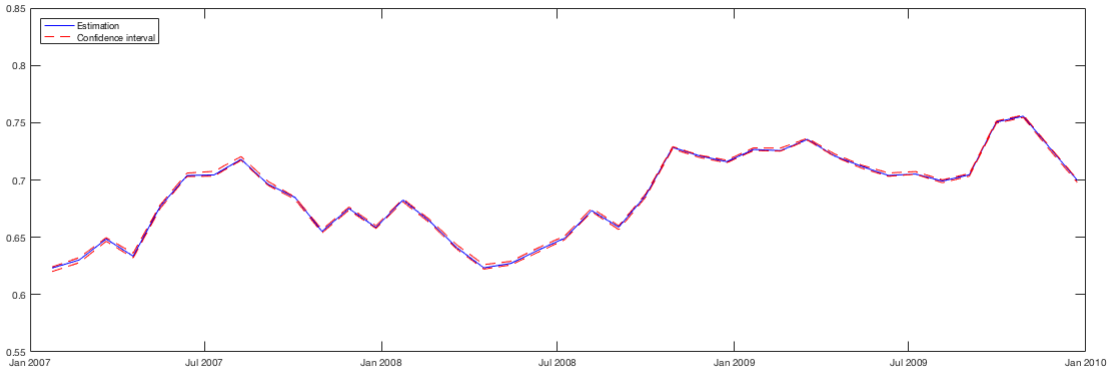
The first peak of first observed *Dot-com* crisis were between 1999 and 2000, when the markets started to question the credibility of investments into internet companies. This was followed by the sharp drop of the NASDAQ index in March 2002. The graph on of the estimation in Figure 7a reflects the same evolution, keeping high values throughout the year of 2002.

High values of p can be also observed during 2009, when the series of bailouts of biggest American banks had taken place. The rapid rise of the parameter between 2008 and 2009 coincides with the worst period of the subprime mortgage crisis.

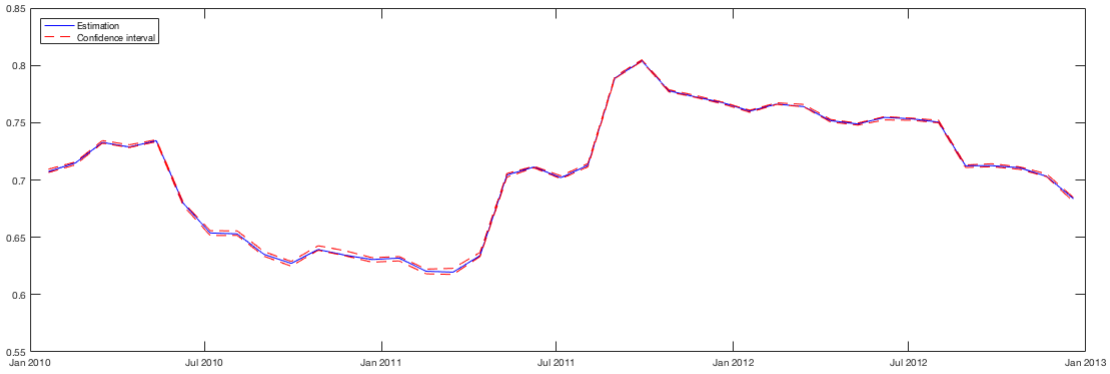
For the period of Eurozone crisis, the lasting period from the second half of 2011 till 2013 the value of parameter p stays between 0.6 and 0.8 that is on average the largest of all the observed period. This reflects the sovereign debt crisis on Eurozone that had begun in 2010. It can also be explained as the condition of the system in general: financial institutions in Europe operating in a monetary union became more interconnected since the introduction of unified currency.



(a) Parameter fit *(solid line)* and 95% credible interval *(dashed line)* during the Dotcom crisis 1999-2002.



(b) Parameter fit *(solid line)* and 95% credible interval *(dashed line)* during the US mortgage crisis 2007-2009.



(c) Parameter fit *(solid line)* and 95% credible interval *(dashed line)* during the US mortgage crisis 2007-2009.

6.3 Sampling parameters of Kronecker graph $G = [\Theta]^k$

The simplest case of the random Kronecker graphs can be represented by a 2×2 initiator matrix Θ with 2 parameters, α and β:

$$\Theta = \begin{pmatrix} \alpha & \beta \\ \beta & \alpha \end{pmatrix}$$

The parameters $\alpha, \beta \in [0, 1]$ serve as preliminary probabilities of having a link between the corresponded nodes in the graph. The sample of these two parameters can be created using the proposed Metropolis-Hastings algorithm with a proper extension. The algorithm will simultaneously choose two parameters at every step of iteration, calculating Metropolis-Hastings acceptance rate as a ratio of likelihood functions for current and proposed value. As in the previous case, a flat prior $\mathcal{B}(a, b)$ distribution was used, with parameters $a = b = 4$. It is important to mention that in two parameter Kronecker graph it is necessary to include permutations in the evaluation of the likelihood. It was shown previously that different permutation of the matrices - multipliers in the Kronecker product end in a isomorphic graph. To pick a better fit from the family of isomorphic graphs, the sample of nodes permutations will be included in the algorithm.

Equivalently to the one parameter case, the initial values of α and β does not affect estimation and are chosen randomly. The same applies to the permutation vector σ, which initial value is set as a random permutation of $\{1, 2, \ldots, n\}$ where n is a number on nodes in the adjacency matrix.

Metropolis-Hastings algorithm of sampling the parameters α, β

- Set the current value of the sequence as equal to the previous values:

$$\alpha_j = \alpha_{j-1},\ \beta_j = \beta_{j-1}$$

- Use the logarithmic transformation of the parameters:

$$\tilde{\alpha}_j = \phi(\alpha_j) = \log\left(\frac{\alpha_j}{1 - \alpha_j}\right);$$
$$\tilde{\beta}_j = \phi(\beta_j) = \log\left(\frac{\beta_j}{1 - \beta_j}\right);$$

- Propose a new values of $\tilde{\alpha}_j^*$ and $\tilde{\beta}_*$ as a random walk centred at $\tilde{\alpha}_{j-1}$ and $\tilde{\beta}_{j-1}$ correspondingly, calculated at the previous step:

$$\tilde{\alpha}_j^* = \tilde{\alpha}_{j-1} + \gamma,$$
$$\tilde{\beta}_j^* = \tilde{\beta}_{j-1} + \eta,$$

  where $\gamma, \eta \sim \mathcal{N}(0, \tau)$ is the standard normal distribution and $\tau > 0$ is the variance of the random walk transformation.

- Transform the proposed value backwards using function

$$\phi^{-1}(p) = 1/(1 + \exp(-p)) :$$

$$\alpha^* = \frac{1}{1 + \exp(-\tilde{\alpha}_j^*)}$$
$$\beta^* = \frac{1}{1 + \exp(-\tilde{\beta}_j^*)}$$

- Generate the Kronecker graph G as the kth power of the initiator matrix for the current values of α and β, as well as for the proposal parameter α^* and β^*, denoted as G^*

- Calculate the Metropolis-Hastings acceptance rate ρ, taking into account the transformation of the parameters, but keeping the permutation σ fixed:

$$\rho = min\left\{1, \frac{\mathcal{L}(\mathbb{G}^*, \sigma, A)}{\mathcal{L}(\mathbb{G}, \sigma, A)} \frac{q(\alpha^*; a, b)}{q(\alpha; a, b)} \frac{J(\phi^{-1}(\alpha^*))}{J(\phi^{-1}(\alpha)} \frac{q(\beta^*; a, b)}{q(\beta; a, b)} \frac{J(\phi^{-1}(\beta^*))}{J(\phi^{-1}(\beta)}\right\}$$

  where the notation is adopted from one p parameter case.

- Accept the proposal values of α^* and β^* as a next stage of the sequence with probability ρ and set the following values to the parameters:

$$\alpha_{j+1} = \begin{cases} \alpha^*, & \text{with probability } \rho \\ \alpha, & \text{with probability } 1 - \rho \end{cases}$$

$$\beta_{j+1} = \begin{cases} \beta^*, & \text{with probability } \rho \\ \beta_j, & \text{with probability } 1 - \rho \end{cases}$$

  Update the intermediate values of $\tilde{\alpha}_{j+1}$, $\tilde{\beta}_{j+1}$ as follows:

$$\tilde{\alpha}_{j+1} = \begin{cases} \tilde{\alpha}^*, & \text{with probability } \rho \\ \tilde{\alpha}_j, & \text{with probability } 1 - \rho \end{cases}$$

$$\tilde{\beta}_{j+1} = \begin{cases} \tilde{\beta}^*, & \text{with probability } \rho \\ \tilde{\beta}_j, & \text{with probability } 1-\rho \end{cases}$$

- Propose the new value of the permutation vector σ as some independent random permutation of the set $\{1, 2, \ldots, n\}$, denote as σ^*
- Calculate Metropolis-Hastings acceptance rate for the new proposal of σ :

$$\tilde{\rho} = min\left\{1, \frac{\mathcal{L}(G, \sigma^*, A)}{\mathcal{L}(G, \sigma, A)}\right\}$$

- Accept the proposal vector σ^* probability $\tilde{\rho}$:

$$\sigma_{j+1} = \begin{cases} \sigma^*, & \text{with probability } \tilde{\rho} \\ \sigma, & \text{with probability } 1-\tilde{\rho} \end{cases}$$

There is another way to choose the permutation vector, apart from the proposed random permutation. As an alternative, at each iteration a new σ^* can be obtained by swapping two nodes. Two node indexes can be chosen uniformly at random from the set of indexes and replaced one by another. For the large networks this alternative version of node permutations takes calculating time.

The variance of the random walk has the same effect as in the case of one parameter initiator matrix. It was set to 0.1 as an optimal scale. To guarantee the convergence 2000 iterations were calculated. The algorithm was tested on the first data point, the result obtained is shown on the Figure 8.

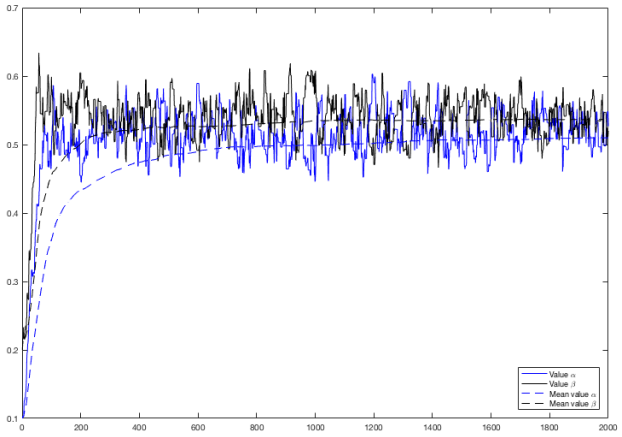


Figure 8: Samples of the initiator matrix parameters α and β for the adjacency matrix of the date *8 January, 1997:* parameters values (solid lines) and mean values (dashed lines)

It is interesting to see that estimated values for both parameters α and β are close to the value of the previously estimated parameter p for the corresponding adjacency matrix, that is around 0.5. Though the estimation values are similar, the interpretation for these two models is distinctive. The form of the 2 parameter Kronecker graph is no longer identical entries, as in the previous model. Elements of the Kronecker graph $G = \Theta(\alpha, \beta)^{[k]}$ are interpreted as the probabilities, though their numerical values are different from the initial parameters α and β.

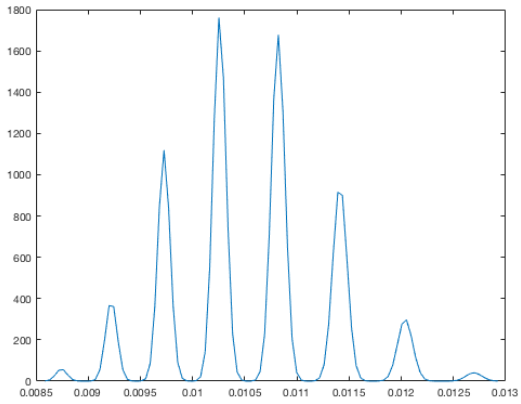


Figure 9: Kernel density estimation for the values of Kronecker graph $G = \Theta(\alpha, \beta)^{[k]}$

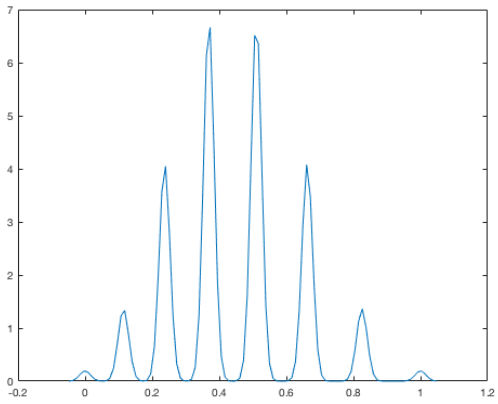


Figure 10: Kernel density estimation for the values of Kronecker graph $G = \Theta(\alpha, \beta)^{[k]}$, linear rescaling

The computed values are too small to be considered as a realistic probabilities of edges. Their distribution can serve as a approximate shape of the graph.

7 Conclusion

The implementation of the Markov Chain Monte Carlo sampling method for the parameters of random graph and Kronecker graph models provided useful results for fitting the model to the real world data. These results can be explained by the empirical events of financial crisis.

The parameter of the random graph p that serves as an estimated measure of connectedness among nodes demonstrates high values during the times of financial instability, that goes in line with the existing results of financial networks studies. The estimated parameters of the Kronecker graph model provide an abstract topological structure of the observed network. The shape of the parameters' distribution confirms imbalanced structure in the connectivity level of nodes that can be interpreted as a community structure of the financial system.

References

[Acemoglu et al., 2013] Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2013). Systemic risk and stability in financial networks. Working Paper 18727, National Bureau of Economic Research.

[Aicher et al., 2014] Aicher, C., Jacobs, A. Z., and Clauset, A. (2014). Learning latent block structure in weighted networks. *Journal of Complex Networks*, 3(2):221–248.

[Albert and Barabási, 2002] Albert, R. and Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47.

[Barabási, 2009] Barabási, A.-L. (2009). Scale-free networks: a decade and beyond. *Science*, 325(5939):412–413.

[Barabási and Albert, 1999] Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.

[Billio et al., 2016] Billio, M., Casarin, R., Costola, M., and Pasqualini, A. (2016). An entropy-based early warning indicator for systemic risk. *Journal of International Financial Markets, Institutions and Money*, 45:42–59.

[Billio et al., 2010] Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2010). Econometric measures of systemic risk in the finance and insurance sectors. Working Paper 16223, National Bureau of Economic Research.

[Billio et al., 2012] Billio, M., Getmansky, M., Lo, A. W., and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559.

[Bollobás, 2013] Bollobás, B. (2013). *Modern graph theory*, volume 184. Springer Science & Business Media.

[Casarin et al., 2017] Casarin, R., Costola, M., and Yenerdag, E. (2017). Financial bridges and network communities.

[Diestel, 2000] Diestel, R. (2000). *Graph theory* Springer-Verlag Berlin and Heidelberg GmbH & amp.

[Elliott et al., 2014] Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–53.

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

[Hastings, 1970] Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109.

[Holland et al., 1983] Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137.

[Hurd, 2016] Hurd, T. R. (2016). *Contagion! Systemic Risk in Financial Networks*. Springer.

[Lancaster, 2004] Lancaster, T. (2004). *An introduction to modern Bayesian econometrics*. Blackwell Oxford.

[Leskovec et al., 2005a] Leskovec, J., Chakrabarti, D., Kleinberg, J., and Faloutsos, C. (2005a). Realistic, mathematically tractable graph generation and evolution, using kronecker multiplication. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 133–145. Springer.

[Leskovec et al., 2010] Leskovec, J., Chakrabarti, D., Kleinberg, J., Faloutsos, C., and Ghahramani, Z. (2010). Kronecker graphs: An approach to modeling networks. *Journal of Machine Learning Research*, 11(Feb):985–1042.

[Leskovec et al., 2005b] Leskovec, J., Kleinberg, J., and Faloutsos, C. (2005b). Graphs over time: densification laws, shrinking diameters and possible explanations. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 177–187. ACM.

[Metropolis et al., 1953] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of Chemical Physics*, 21(6):1087–1092.

[Robert and Casella, 2011] Robert, C. and Casella, G. (2011). A short history of markov chain monte carlo: Subjective recollections from incomplete data. *Statistical Science*, pages 102–115.

[Slovik, 2012] Slovik, P. (2012). Systemically important banks and capital regulation challenges. *OECD Economics Department Working Papers, No. 916, OECD Publishing*

[Watts and Strogatz, 1998] Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. *nature*, 393(6684):440.

[Weichsel, 1962] Weichsel, P. M. (1962). The kronecker product of graphs. *Proceedings of the American mathematical society*, 13(1):47–52.