UNIVERSITAT DE BARCELONA

FUNDAMENTALS OF DATA SCIENCE MASTER'S THESIS

Agent-based models for assessing the risk of default propagation in interconnected sectorial financial networks

---

Authors:
Philippe VAN AMERONGGEN ,
Ramon MIR MORA,
Sergi SÁNCHEZ DE LA BLANCA CONTRERAS

Supervisor:
Jordi NIN

A thesis submitted in partial fulfillment of the requirements for the degree of MSc in Fundamentals of Data Science

in the

Facultat de Matemàtiques i Informàtica

June 30, 2019

UNIVERSITAT DE BARCELONA

Abstract

Facultat de Matemàtiques i Informàtica

MSc

Agent-based models for assessing the risk of default propagation in interconnected sectorial financial networks

by Philippe VAN AMERONGGEN , Ramon MIR MORA,
Sergi SÁNCHEZ DE LA BLANCA CONTRERAS

Financial Institutions perform risk assessments continuously in order to judge if certain companies are viable and should receive funding or loans to prevent companies to go bankrupt (default). This task helps keeping the financial system healthy. However, risk assessment is a tremendously difficult task since there are many variables to take into account. This work is a continuation of Adrià Barja, 2019, in which a model is posed to simulate customer-supplier relationships. The model helps to explore the risk of default of companies under certain circumstances. We extended the model in several ways to make it more realistic. The main objective of the work is to gain better insights in how defaulted companies affect non-defaulted ones. This is analyzed by keeping track of the possible default cascades produced when a company goes bankrupt and stops paying. In addition, studying how financial networks behave, it is also possible shed some light about how the risk of specific companies or economical sectors can be tracked.

Acknowledgements

We would first like to thank our thesis advisor PhD. Jordi Nin of the Computer Science and Mathematics faculty at the University of Barcelona. Jordi Nin gave us freedom to experiment with new methods. He consistently allowed this paper to be our own work, but steered us in the right direction whenever he thought we needed it.

We would also like to thank MSc Martin Vogelaar of the Kapteyn Astronomical Institute at the University of Groningen for allowing us to use the norma4 computer cluster, a crucial part for obtaining results.

Lastly, we would like to acknowledge Dr. Santi Seguí, PhD. Lluis Garrido and Dr. Eloi Puertas of the Computer Science and Mathematics faculty at the University of Barcelona as the second readers of this thesis, showing much interest in the topic.

Philippe van Amerongen,
Ramon Mir Mora,
Sergi Sánchez de la blanca Contreras.

Contents

Chapter 1

Introduction

A longed for desire by banks, governments, business owners and, indirectly, citizens is to prevent the (partial) collapse of the financial system. Banks actively work on risk assessment of companies to decide if a company is healthy enough to receive a loan or credit, or on the contrary, its financial products must be reduced. This decision largely affects company's liquidity producing delays in its payments, and therefore, affecting other companies' liquidity.

Risk assessment is an incredibly difficult task due to the complex nature of economy; fluctuations in demand/offer, procurement expenses, labor market, etc. For these reasons, to mitigate this issue, client-supplier networks analytics are used. Those analysis can simulate a vast range of parameters, representing different market conditions and business relations.

This work is a continuation of a previous paper, based on transaction data between business clients from the Spanish bank BBVA Adrià Barja, 2019. The goal of this work is to create a set of simulations of different client-supplier relationships. This is done using a directional graph representation of companies, where every company is a node and the dependency of companies are weighted edges. The dependence is modelled based on the real money flows observed among companies. The larger the money flow, the higher the weight and therefore the dependency. Every company (or node) has a probability of going bankrupt (default). This is modelled as a node property and is updated every iteration of the simulation. This default probability solely depends on the default probability of the neighboring nodes, since the company only depends on transactions from the neighboring companies.

This work focuses on extending the model to better cope with reality as well as giving more insights for risk assessment. The first improvement with regards to previous contributions is making it a stochastic system. This change is explained by the fact that default of companies is binary; they are either bankrupt or not. Intermediary probabilities are possible, but after recalculating the default probability this gets either translated to defaulted or not, 1 or 0. The second change is implementing a so-called default delay. This delay comes forth from forgiveness; a company will not immediately cut ties with another company if they do not deliver. In addition to these two contributions and to enhance the usefulness of the simulation, we have also analyzed default cascades. Such cascades are chains of defaulted companies, stemming from a single defaulted company. Using cascades we can inspect the health of the entire supply chain. This information is extremely useful in risk assessment of companies, since it accounts for the start and end of its value chain, because we can assess the economic strength of the neighborhood of the different companies recursively.

To assess the inner workings of the system, we pose two different scenarios: a liberal market network and a network representing sharing economy. In the liberal market network there will be some nodes with many outgoing nodes, which also likely have many incoming nodes. These nodes represent big corporations that are very resilient to default. On the other hand, this network mostly consists of sparsely connected nodes, which are much more prone to default. In the sharing economy networks, nodes will have more equally distributed number of connections. The hypothesis is that the liberal market network will be much more resilient to large cascades or even to the entire collapse of the system. However, there will be more small cascades and defaulted companies than the sharing economy network. This hypothesis is supported by the fact that the liberal market network has some very large nodes. These large nodes provide a buffer for smaller companies to recover if they go default. However, not all small nodes are connected to such big nodes and therefore are very prone to get 'infected', hence going default due to the poor economic strength of its economic neighborhood. Whenever a large node in the liberal market network defaults, it has disastrous consequences to the entire system, but the probability of a situation like such arising is extremely small due to the abundance of suppliers (outgoing connections). In a sharing economy, nodes have much more equal numbers of suppliers and customers. Therefore, any location in the network is of medium economic strength, causing the probability of companies to default to be smaller than in the liberal market network. However, because of the same reason, in this network the default 'infection' can spread rather rapidly, making the system in general less resilient to cascades than the liberal market system.

The objective is to simulate client-supplier sectorial networks as closely to reality as possible with feasible computation times and usable results for risk assessment. We simulate a reduced sectorial network based on statistics of transaction data provided by the BBVA. We expect to see different behaviours in each sector regarding the number of defaulted companies as well as cascade sizes. Some sectors have much higher numbers of connections than other sectors and should therefore be more resilient against infection.

This document is organized as follows. First, in Section 2 an introduction is made explaining which data is used, how it is used, how it is structured and a reference to previous work made by the BBVA. In Section 3 we elaborate on the fundamentals of graph theory used in this project. In Sections 4, 5 and 6, the theory of the workings and use of financial client-supplier networks is illustrated. After explaining the theory, in Section 7, we show the results and hint towards our hypothesis. Finally, the document ends with a discussion about obtained results, suggestions for future work and our conclusions, including comments towards the confirmation and deviations regarding our hypothesis.

Chapter 2

Database

The data used in this work is provided by the BBVA bank. The data is collected using the transaction data the bank has between its business clients. Their business clientele is mainly composed of a vast amount Spanish companies. This data set contains the following information.

- Name of the company.
- Economical sector of the company.
- Companies that receive money from other companies.
- Relative importance of the relation.
- Defaulted probability of each company [0 or 1] where 0 is not defaulted and 1 is defaulted.
- Weights that follow the direction of the money flow (Normalized for all the outgoing edges of a node)

We treat this data set as a financial client-supplier network, we also have the information of each node and their edges. For more information about the real data see Table 2.1. To get a better grasp at complex networks and to better understand this data, an introduction into the matter is given in 3.

2.1 Metadata

Due to confidentially issues with the data, we ran our simulations on a synthetic network designed to mimic the BBVA data. Besides confidentially, another important issue was the lack of hardware to store and analyze the real network. Therefore, we decided to create a smaller synthetic network. More information about the generation of the network can be found in Subsection 4.1.1.

The sectors that have been analyzed during the study are as follows:

- Financial Institutions: A financial institution is responsible for the supply of money to the market through the transfer of funds from investors to the companies in the form of loans, deposits, and investments. Large financial institutions such as JP Morgan Chase, HSBC, Goldman Sachs or Morgan Stanley can even control the flow of money in an economy.
- Energy: The energy sector is a category of stocks that relate to producing or supplying energy. This sector includes companies involved in the exploration and development of oil or gas reserves, oil and gas drilling and refining, or integrated power utility companies including renewable energy and coal.

- Financial Services: Financial services are the economic services provided by the finance industry, which encompasses a broad range of businesses that manage money, including credit unions, banks, credit-card companies, insurance companies, accountancy companies, consumer-finance companies, stock brokerages, investment funds, individual managers and some government-sponsored enterprises.

- Utilities: Is an economic term introduced by the noted 18th century Swiss mathematician Daniel Bernoulli referring to the total satisfaction received from consuming a good or service. The economic utility of a good or service is important to understand because it will directly influence the demand, and therefore price, of that good or service. A consumer's utility is hard to measure, however, but it can be determined indirectly with consumer behavior theories, which assume that consumers will strive to maximize their utility.

- Telecom: Information Communication & Telecommunication Economics refers to a broad range approach to the micro and macro economics of data consumption and management, voice or data. This application of micro cum macro economic principles to the subject matter here is referring to three clear strategies vis-a-vis information, communication and telecommunication. Information refers to data that is accurately organized and timely presented so as to affect the end user's behavior.

- Basic Materials: Companies included in the basic materials sector are involved in the physical acquisition, development, and initial processing of the many products commonly referred to as raw materials. Oil, gold, and stone are examples. Raw materials, for the most part, are naturally occurring substances and resources. Some are finite. Others are reusable but are not available in infinite quantities at any given point in time.

- Transportation: Transport Economics is the study of the movement of people and goods over space and time. It is a branch of economics that deals with the allocation of resources within the transport sector. Historically, it has been thought of as the intersection of microeconomics and civil engineering, as shown on the right.

- Retail: Retail sales is the purchases of finished goods and services by consumers and businesses. These goods and services have made it to the end of the supply chain.

- Retailers: The middle of the supply chain is wholesale sales. They distribute the goods and services to retailers. The retailers sell them to the consumer.

- Capital Goods: also called complex products and systems (CoPS) is a durable good that is used in the production of goods or services. Capital goods are one of the three types of producer goods, the other two being land and labour. The three are also known collectively as "primary factors of production"

- Auto: The automotive industry comprises a wide range of companies and organizations involved in the design, development, manufacturing, marketing, and selling of motor vehicles. It is one of the world's largest economic sectors by revenue. The automotive industry does not include industries dedicated to the maintenance of automobiles following delivery to the end-user, such as automobile repair shops and motor fuel filling stations.

- Consumer and Healthcare: The healthcare sector consists of companies that provide medical services, manufacture medical equipment or drugs, provide medical insurance, or otherwise facilitate the provision of healthcare to patients.

- Construction and Infrastructure: The construction industry plays an important role in the economy, and the activities of the industry are also vital to the achievement of national socio-economic development goals of providing shelter, infrastructure and employment.

- Real Estate: Real estate is "property consisting of land and the buildings on it, along with its natural resources such as crops, minerals or water; immovable property of this nature; an interest vested in this (also) an item of real property, (more generally) buildings or housing in general. Also: the business of real estate; the profession of buying, selling, or renting land, buildings, or housing."

- Leisure: the leisure industry is the segment of business focused on recreation, entertainment, sports, and tourism (REST)-related products and services

- Institutions: organizations founded for a religious, educational, professional, or social purpose.

- Unknown: The sector is unknown, most likely due to the instance not being a client of the bank.

The data has been simulated using the following statistics:

sector	size(%)	$\overline{k_{in}}$	$\overline{k_{out}}$	default (%)
Financial Institutions	0.046	39.613	45.529	3.650
Energy	0.083	12.844	8.666	1.111
Financial Services	1.165	6.300	20.265	0.786
Utilities	1.529	5.589	5.903	1.264
Telecom	3.299	5.960	5.194	1.1776
Basic Materials	2.745	5.789	5.350	2.782
Transportation	4.023	5.411	4.336	1.868
Retail	4.064	3.973	3.233	1.217
Retailers	4.273	5.001	3.613	1.885
Capital Goods	8.698	4.528	3.098	1.866
Auto	1.470	4.454	2.991	1.786
Consumer and Healthcare	7.055	3.259	3.770	1.539
Construction and Infrastructure	8.907	3.067	3.270	1.539
Unknown	10.159	0.930	1.413	1942
Real Estate	6.843	1.517	1.844	3.603
Leisure	12.861	2.509	2.512	1.511
Institutions	3.219	5.547	10.764	0.535

TABLE 2.1: Network properties of all financial sectors of the BBVA data (Real one)

Table 2.2 represents the metadata of the simulated BBVA graph. The graph is simulated using random statistics based on the original BBVA metadata, but applied on much less nodes. In total 10^4 nodes have been generated for the BBVA simulation. The generation procedure starts with generating the nodes per sector

sector	size(%)	k_{in}	k_{out}	default (%)
Financial Institutions	0.050	36.800	46.200	0.000
Energy	0.080	11.125	8.750	0.000
Financial Services	1.160	6.147	20.500	0.000
Utilities	1.530	5.797	5.889	0.000
Telecom	3.300	5.758	5.188	0.303
Basic Materials	2.740	6.077	5.401	2.555
Transportation	4.060	5.308	4.374	0.985
Retail	23.592	3.956	3.179	1.314
Retailers	4.270	5.178	3.628	1.171
Capital Goods	8.691	4.525	3.075	2.877
Auto	1.470	4.415	3.027	2.041
Consumer and Healthcare	7.061	3.360	3.796	1.133
Construction and Infrastructure	8.911	2.975	3.226	3.030
Unknown	10.161	0.928	1.260	2.067
Real Estate	6.841	1.456	1.971	3.216
Leisure	12.861	2.495	2.526	1.555
Institutions	3.220	5.326	10.745	5.901

TABLE 2.2: Network properties of all financial sectors of the BBVA data (simulated)

according to the relative frequency. For each node we randomly generate a number of connections the node should have. This number is drawn from a $N(n, n/10)$ distribution. Here, n is the average outgoing number of connection for the sector of the node. Afterwards, we pick a selection of sectors, equal to the number of connections previously generated. For each selected sector a random node is picked to connect to. The sector selection probability relies on the relative size of the sector multiplied by the number of average incoming connections. Comparing the metadata of the simulated graph, Table 2.2, with the original metadata, Table 2.1 some deviance can be noticed, especially in the percentage of defaulted companies. These differences can be explained by the fact that the simulated graph is simply a smaller projection of the statistical data of the original graph. For example, Financial Institutions takes up 0.046% of the entire original network and the percentage of defaulted companies is 3.650%. Lets calculate, $0.046\% \times 10000 \sim 5$ companies, which is 0.05% of the simulated network. However, if even one Financial Institution in the simulated network is defaulted, a default percentage of $1/5 = 20\%$ is obtained. Now it should be clear how the simulation results in these differences.

2.2 Previous work done by BBVA

The main goal of this work is to improve and expand the previous work carried out by BBVA Data science team (Adrià Barja, 2019). The purpose now is compare the results obtained from BBVA, and expand including default cascades analysis. The BBVA team had proposed a computational model, based on the probabilities of default contagion, to study the default diffusion at individual and aggregated levels. They had been performed massive experiments based on this model by varying several parameters such as the initial default rate, the contagion rate β and the recovery rate μ. And the results that they have been obtained shown the relationship between dynamical and topological properties for more than 140,000 BBVA firms aggregated using their economic sector. This allow us to create a ranking of sectors which can be used in specific scenarios.

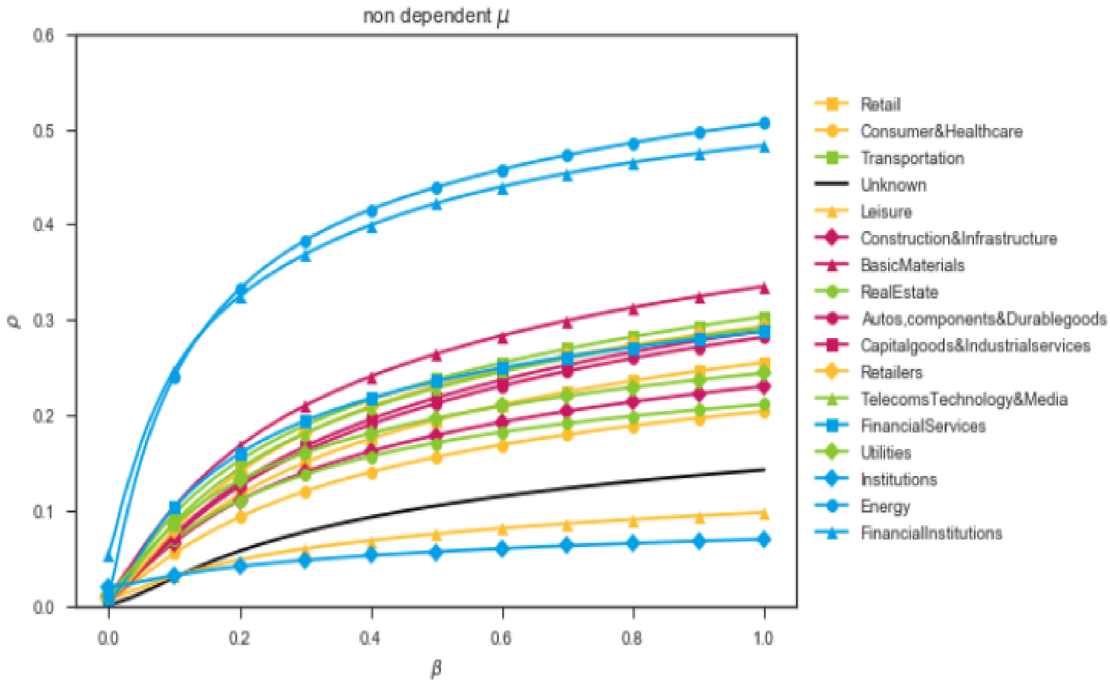
2.3 Results of previous work

what they did on the paper (Adrià Barja, 2019) was provide a simple mechanistic model to assess the impact of a particular diffusion process on financial networks, that of default contagion. To this end, they take advantage of a probabilistic computational framework named microscopic Markov chain approach (MMCA) to compute the probability of the states of individual agents in contagion processes in complex networks, and adapted its formulation to the understanding of the default propagation in financial networks.

To analyze the behaviour of our proposed model, they use real data from the anonymized database of BBVA from December 2015 to December 2016, covering about 140000 public and private Spanish firms. With this data they had access to the real network of interactions and to the default endogenous propagation dynamics.

The results of the model that was applied to the real data reveal which sectors are more at risk in the propagation of default, which sectors are more resilient to the default avalanches, and what are the expectations for the cascades of default under different stochastic conditions. Below is shown some of the results obtained for a baseline model with no edge rewiring.

Plot 2.3 shows $\rho(\beta, \mu)$ for each economical sector. Clearly, not all sectors behave in the same way regarding default dynamics. Broadly speaking, economical sectors can be grouped in three blocks given its response to default propagation. On one hand, Institutions, Leisure and unknown sector show a low propensity to default propagation, where the sector default probability density range from approximately 0.04 to 0.10 for high infectious rate β. On the other hand, Financial Institutions and Energy evince a high propensity to default contagion, with ρ reaching almost 0.50. In other words, on average, each company of these two sectors has a 0.50 probability of being in default for extreme parameter conditions. In-between, we find the rest of the sectors with density variations ranging from 0.15 to almost 0.30. Interestingly, the exposure of the economical sectors is quite different to one another. This result, may allow current risk assessment models (ex. Generalized Linear Models) to include a quantification of sectorial risk and rank accordingly.


non dependent μ
ρ
β
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.0
0.2
0.4
0.6
0.8
1.0
Retail
Consumer&Healthcare
Transportation
Unknown
Leisure
Construction&Infrastructure
BasicMaterials
RealEstate
Autos,components&Durablegoods
Capitalgoods&Industrialservices
Retailers
TelecomsTechnology&Media
FinancialServices
Utilities
Institutions
Energy
FinancialInstitutions

Chapter 3

Complex Networks

In this section one can find a general overview into complex networks focusing on the necessary knowledge to understand the original and simulated data, Tables 2.1,2.2, as well as the model networks (BA and ER) used in section 7.

The main difference between simple graph/network theory and complex networks lies in the fact that complex networks are based on real networks and present several non-trivial topological features and unexpected dynamics.

3.1 General definitions

The two main components of a complex networks are the **nodes** and the **edges**. A node represents the actual individuals, companies, elements that are part of the system to analyze and the edges are the relations between those individuals. In that sense, if two nodes are related with each other then, an edge is established between them. For example, in a social network the edges usually symbolize friendship between two individuals, in an airport network they symbolize flights and in a financial network transactions between two companies. This allows us to draw the network as seen in Figure 3.1. Notice that the nodes are drawn in two distinct colours. In general one can use the colour of the nodes to represent some properties of the node, in that case it is done to symbolize the two main factions that co-exist inside the club.

Given a network structure with nodes and edges one can differentiate between several networks,

- **Undirected/Directed:**If the edge (the relation) between the two nodes is independent of the direction of such relation we have an undirected complex network with edges that resemble lines. However, if the direction matters, like is the case in a financial network, where the edges symbolize monetary transaction flowing from one company to the other, then, we have a directed network with edges that resemble arrows.

- **Weighted/Unweighted complex network:** In an unweighted complex network all edges have the same value for the involved nodes (again, is the case of a social network). However, if the edges are different ones from the others a number (weight) is assigned to symbolize that difference. For instance, in a financial network the weight is the value of the monetary transaction between the companies. Clearly the higher the value of the weight the more important is that edge over the rest for a given node.

To sum up, a financial network is a directed weighted complex network.

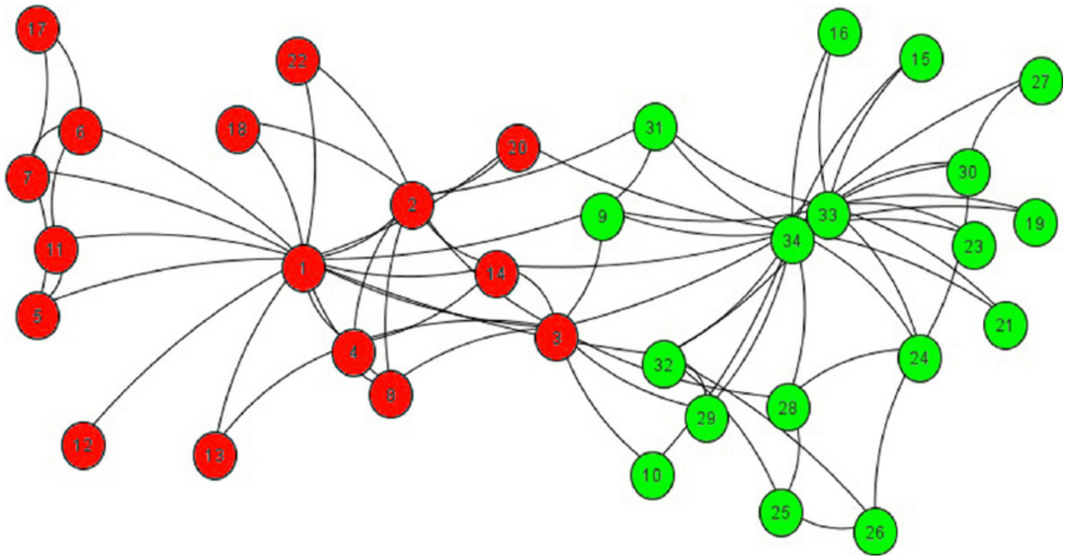


FIGURE 3.1: Prototypical complex network where the nodes represent the members of a karate club and the edges (lines) friendship between them.

Given a node we can define:

- **Degree (k):** The number of edges that connect the node with the rest.

In the case of a directed network we need to differentiate between in-degree and out-degree.

- **In-degree (k_{in}):** The number of edges whose ending point is the node.
- **Out-degree (k_{out}):** The number of edges whose starting point is the node.

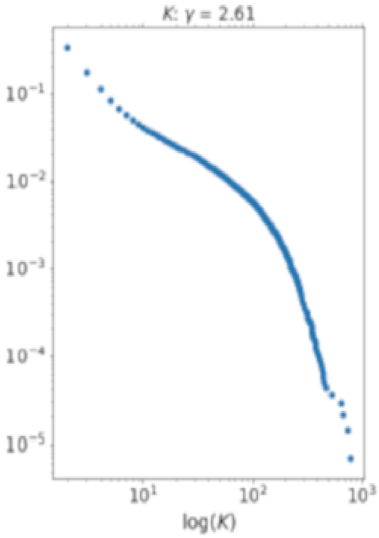
For a financial network the in-degree symbolizes for how many companies the node is a supplier, and the out-degree symbolizes how many suppliers the node has.

Obs 3.1 Since default contagion goes from suppliers up, the higher the in-degree the higher the chances the company spreads the default and the higher the out-degree the less susceptible is a company of getting infected when one of his suppliers goes into default. With this information now tables 2.2 and 2.1 should be clear. Looking at them we can start to think which sectors will be more susceptible to default contagion and which sectors have the higher risk for the network. More information about default contagion, failure cascades and risk will be introduced and developed in sections 5,6.

Most complex networks, as is the case for a financial network, are what is called scale-free networks. A scale-free network is such that the degree distribution of the nodes follows a power law.

$$P(k) = k^{-\gamma} \tag{3.1}$$

Meaning, nodes whose degree is at several different scales can be found and thus, there is no characteristic scale. That is why they are called scale-free networks. The nodes with the higher degree scale in this type of networks are what we call Hubs. Most of the strange and unexpected dynamical behaviours and topological features have their root in the Hubs. Figure 3.2 shows that both the BBVA and the generated data follow a similar power law distribution. The first values of the curve are the only ones that seem to differ, suggesting our generated data has fewer companies with degrees 1,2,3 than it should. Otherwise it is a pretty accurate fit. Also, notice that the Hubs have a much lower degree in this case, as a result of having a much smaller network. However, the fact that they follow a similar power law gives us confidence that the relative magnitude of the Hubs remains equivalent.



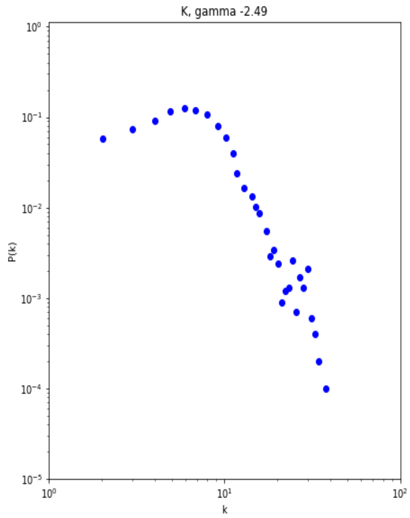


FIGURE 3.2: Degree distribution of the BBVA data (left) and our simulated metadata (right) in log-log scale.

One of the main models that generates a scale-free network is the proposed by Barabási-Albert. On the other hand, a model such as the Erdós-Rènyi one, for which all nodes have closely the same degree, is an example of a single scale network. In section 7 a comparison between a financial network following both models is performed.

3.2 Barabási–Albert (BA) model

The BA model generates a scale-free network using what is called preferential attachment, which essentially is the common rule of the rich gets richer and the poor, poorer.

The hypothesis of preferential attachment are that there is a simple initial random graph and the true complex network arises from introducing more nodes to the preexisting graph. Every new node enters the network with x edges and the probability of connecting those edges to each node depends on the actual degree of every node. Thus, nodes with higher degree have a higher probability of gaining even more connections, generating the so cold Hubs, while the lower degree nodes will remain mostly untouched.

The algorithm goes as follows, given a connected random graph with n_0 nodes and m_0 edges. Every new node enters with x edges, and the probability to attach to every other node p_i is:

$$p_i = \frac{k_i}{\sum_j k_j} \tag{3.2}$$

The result of applying this algorithm is a scale-free network with $\gamma \simeq 3$. An example of the behaviour of this algorithm can be found in Figure 3.3.

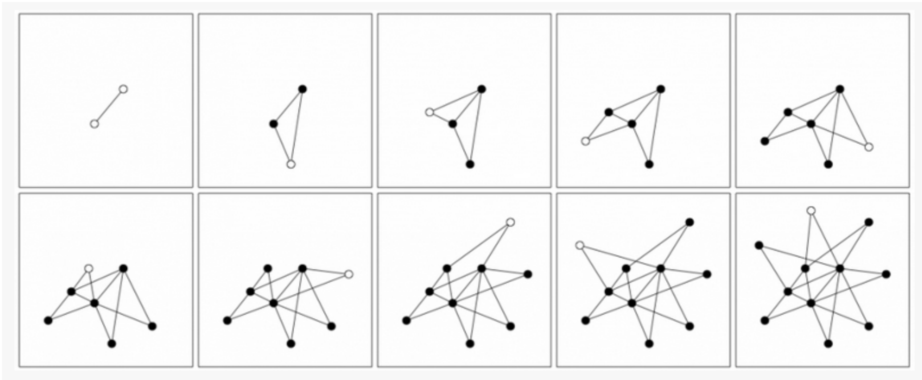


FIGURE 3.3: Illustrative example of the growth of a network following the BA algorithm as it can be found in (Barabási, 2016). The white node in each image symbolizes the new entering node which enters with two edges.

To know more about scale-free networks as well as a deeper analysis of the BA algorithm go to (Barabási, 2016).

3.3 Random Network, the Erdós–Rényi (ER) model

There are two main definitions of random network,

- $G(N, M)$ **model**: Where the network has N nodes and M links randomly distributed among the N nodes. An alternative definition would be that the graph is chosen uniformly at random from the collection of all graphs which have N nodes and M edges.
- $G(N, p)$ **model:** Where the network has once again N nodes and every pair of nodes can be connected with probability p.

The $G(N, M)$ model was introduced in (P.Erdós, 1959), while the $G(N, p)$ model was introduced in (Gilbert, 1959). Both approaches, and more generally all random graphs/networks generated in a similar fashion, receive the name of Erdós–Rényi network in honor to their contributions to the field.

The actual algorithm of the ER model follows the second approach. It follows the guidelines:

- Start with N isolated nodes (0 edges).
- Given a node pair, generate a random number in $[0, 1]$, if the number is below the probability p then an edge is introduced between the nodes.
- Repeat the second point for every pair of nodes possible. Remember that by combinatorics theory that is $\binom{N}{2} = \frac{N(N-1)}{2}$.

3.3.1 Properties of $G(N, p)$

In order to compare both models in section 7 both graphs should have an almost equal number of nodes and edges. Furthermore, one of the main differences between both the BA and the ER models is the fact that a BA network is scale-free whereas an ER network is single scale. The study of some properties of $G(N,p)$, the expected number of edges given a probability p and the degree distribution of the ER network is therefore required.

Expected number of Links

Let our final graph $g \in G(N,p)$ be such that g has exactly L edges. This had happened if and only if:

- L times an edge has been drawn between a pair of nodes.
- The rest $\binom{N}{2} - L$ of times an edge has not been drawn.

The probabilities that both this events had happened are p^L and $(1-p)^{\frac{N(N-1)}{2}-L}$ respectively. Moreover, we need to account for the total number of different combinations of L edges that can be placed in the network $\binom{\binom{N}{2}}{L}$.

Thus, the probability that our network has exactly L edges is

$$p_L = p^L \binom{\frac{N(N-1)}{2}}{L} (1-p)^{\frac{N(N-1)}{2}-L} \tag{3.3}$$

which corresponds to a binomial distribution. Therefore, a graph $g \in G(N,p)$ has on average $\binom{N}{2}p$ edges.

Degree distribution

Similar to the procedure done when analyzing the total number of edges in the whole network, if instead we look to the number of edges that one particular node has, we have that it will have precisely k links if and only if k edges had been drawn and $(N-1)-k$ edges had not been drawn. Again, we need to account for all the possible combinations of k links. This yield, just as expected, that the degree distribution follows a binomial:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \tag{3.4}$$

where N is the total number of nodes in the graph. Thus, the expected value of the degree is $\langle k \rangle = (N-1)p$.

Assuming the limit $N \gg \langle k \rangle$ we have that the binomial distribution is well approximated by a Poisson distribution, see Figure 3.4

$$P(k) \longrightarrow_{N\to\infty} \frac{\langle k \rangle e^{-\langle k \rangle}}{k!} \tag{3.5}$$

Notice how in both cases, binomial and Poisson distributions, that clearly makes the ER degree distribution single-scale in contrast with the scale-free BA. As it has been explained before most real networks follow a scale-free approach and thus a ER model fails to fit reality. Nevertheless. for the hypothetical case of a true communist financial network we will assume it would behave as a random ER graph.

To properly see the mathematical proof of 3.5 go to Advanced Topic 3.A. in (Barabási, 2016). If the reader is interested in knowing more about random graphs

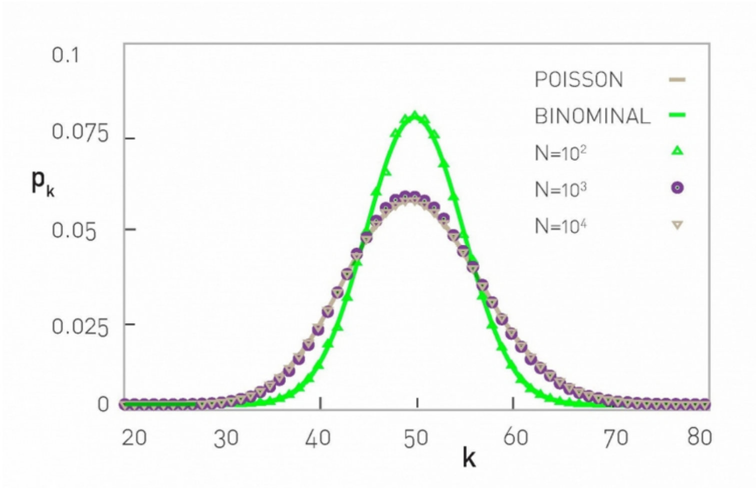


FIGURE 3.4: Figure that illustrates how both distributions converge when $N \gg \langle k \rangle$ as it is shown in (Barabási, 2016)

and their evolution we suggest (P.Erdós, 1959; P.Erdós, 1960; P.Erdós, 1961) or following section 3 in (Barabási, 2016).

Chapter 4

Study Design

In this study, we take advantage of the theory of complex networks to shed some light on the mechanisms behind default propagation. We propose a computational model, based on the dynamics of default contagion, that allows us to assess the main statistics of default diffusion at individual and network levels. We study several different topologies that resemble real financial client-supplier networks. The model enables to investigate the functioning of the default propagation and to assess the economic health of the full structure, providing some insights in which topologies are more resistant towards default propagation and which yield a higher overall risk of collapse. Furthermore, the model can give some intuitions about the systemic risk of companies of interest.

4.1 Financial client-supplier networks

Financial client-supplier networks are a combination of financial and client-supplier networks. In a financial network the objective is to simulate money flows between financial instances. Examples of these financial instances can be, but are not limited to people, companies, firms, banks. Financial networks give insight in the strength of economic relationships between instances and macro effects in financial systems. The applications extend to far ranges in the finance field. With client-supplier networks the intent is to simulate relationship between instances, in terms of supply and demand of goods or services. In this project, the network is based on transaction data between business clients of the BBVA, hence using money flows. However, it differs from a pure financial network in the sense that the money flows are not directly simulated. Instead, money flows are used to model client-supplier relationships. Using the combination of the principles of both types of networks makes it that a financial client-supplier network inherently simulates financial states as well as economic strength. Moreover, money flows are more difficult to model than solid business relationships, due to the relatively erratic behaviour of money flows. Through the use of financial client-supplier networks we obtain insights in contagion and system risk. These insights are obtained by looking at default density, but more importantly, by looking at cascades, which will be explained later in this section.

4.1.1 Network construction

The exact construction of our financial client-supplier networks depends on the implementation. In this work two methods are used to construct the networks. The first method consists of quasi-random generated weighted bidirectional graphs. These networks are generated using well-known models, the Barabási–Albert and

Erdós–Rényi models, which are explained in detail in the previous chapter. There are many more models to generate weighted bidirectional graphs, the reason of using specifically these models stems from the desire to simulate sharing economy and liberal markets.

The second method bases the network on the statistics of the BBVA network. This imitation network is sectorial, meaning that for different clusters of nodes, the characteristics change; incoming and outgoing connections. The exact process goes as follows. We set a desired number of nodes. The we partition these nodes into sectors according to the relative sizes of each sector, S_s. Next, for each of the desired number of nodes in the graph, a number of outgoing connections is drawn from a normal distribution

$$N_n \sim \mathcal{N}(\overline{k^s_{out}}, \overline{k^s_{out}}/10)$$

Here N is the number of outgoing connections for node n and $\overline{k^s_{out}}$ is the mean number of outgoing connections for sector s. Now we find N_n nodes to connect to. Since we want to satisfy the mean number of incoming nodes per sector, we have to take this into account connecting to nodes. We do this by randomly selecting a number of sectors equal to N_n, according to the sector probability P_s.

$$P_s = \frac{\overline{k^s_{in}} S_s}{\sum_s (k^s_{in} S_s)}$$

We now draw a random node from the sector for each of the drawn sectors. We end up with a network attaining the provided statistics of the BBVA network.

The imitation network is not completely representative of the BBVA network, since it is randomly generated using only the mean incoming and outgoing connections per sector. In order to get a more representative network, the complex client-supplier structures within the network are of extreme importance. Unfortunately, we were not able to remodel this. The reason for not using the full BBVA network directly is due to computational limitations and confidentiality.

4.2 Cascading failure

A cascading failure is a process in a system of interconnected parts in which the failure of one or few parts can trigger the failure of other parts and so on. Such failure may happen in many types of systems, including power transmission, computer networking, finance, human body systems, and transportation systems. Cascading failures may occur when one part of the system fails. When this happens, other parts must then compensate for the failed component. This in turn overloads these nodes, causing them to fail as well, prompting additional nodes to fail one after another.

The size of cascades is greatly affected by the topology of the network. For example, completely random networks are more resilient than networks with reinforced loops, which occur, among others, in energy distribution systems. The cascade sizes vary significantly between the three networks used in this paper. The cascade sizes between the imitation BBVA network and the actual BBVA network will vary too, since the imitation network lacks the complex structures present in the BBVA network.

Terms frequently used in financial and client-supplier networks are "too big to fail" (TBTF) and "too interconnected to fail" (TICTF). These terms refer, respectively, to instances/nodes that are so big, they have so much money flow, or have so many connections that the likelihood of them defaulting is extremely small. In financial client-supplier networks, these terms are roughly interchangeable, since many strong connections imply strong money flow. In a liberal market, there will be some TBTF nodes, while in a sharing economy these nodes are absent. The BBVA network is more similar to a liberal market system, but with the addition of partitioning. Some of these sectors are more connected than other sectors, making the system more resilient to total collapse and default of a large node. However, this system is just as prone to collapse as the liberal market system in case of a large node defaulting.
More information about this topic is on 6

4.3 Systemic risk

In finance, systemic risk is the risk of collapse of an entire financial system or entire market, as opposed to risk associated with any one individual entity, group or component of a system, that can be contained therein without harming the entire system. Whenever a company becomes systemic risk, it means that they are too big to fail. In the very unlikely case of failure of these companies, the entire financial system is endangered to collapse. Governments can use systemic risk as a justification for intervening in the economy, trying to prevent economic collapse, usually indicating a financial crisis.

In this project, systemic risk is investigated for the three networks; liberal market, sharing economy and BBVA imitation. This is done by investigating cascades after purposely defaulting nodes with many connections and comparing the results with the results from simulations without defaulting these nodes. The larger the difference in most likely and maximum cascade size, the more susceptible the network is to systemic risk. Besides the insight it gives on systemic risk in the entire system, it can also be used to analyze the systemic risk posed on the network by any company of interest.

4.3.1 Sharing economy

Sharing economy is a term for a way of distributing goods and services, a way that differs from the traditional model of corporations hiring employees and selling products to consumers. In the sharing economy, individuals are said to rent or "share" things like their cars, homes and personal time to other individuals in a peer-to-peer fashion. To analyze the part of sharing economy we will follow the model of Erdós–Rényi. In the mathematical field of graph theory, the Erdós–Rényi model is either of two closely related models for generating random graphs. They are named after mathematicians Paul Erdős and Alfréd Rényi, who first introduced one of the models in 1959, while Edgar Gilbert introduced the other model contemporaneously and independently of Erdős and Rényi. In the model of Erdős and Rényi, all graphs on a fixed vertex set with a fixed number of edges are equally likely; in the model introduced by Gilbert, each edge has a fixed probability of being present or absent, independently of the other edges. These models can be used in the probabilistic method to prove the existence of graphs satisfying various properties, or to provide a rigorous definition of what it means for a property to hold for almost all graphs.

4.3.2 Liberal market

Liberal market is an economic system based on the private ownership of the means of production and their operation for profit. Characteristics central to liberal market include private property, capital accumulation, wage labor, voluntary exchange, a price system, and competitive markets. In a liberal market economy, decision-making and investment are determined by every owner of wealth, property or production ability in financial and capital markets, whereas prices and the distribution of goods and services are mainly determined by competition in goods and services markets. To analyze the part of liberal markets we will follow the model of the Barabási–Albert (BA) model. This is an algorithm for generating random scale-free networks using a preferential attachment mechanism. Several natural and human-made systems, including the Internet, the world wide web, citation networks, and some social networks are thought to be approximately scale-free and certainly contain few nodes (called hubs) with unusually high degree as compared to the other nodes of the network. The BA model tries to explain the existence of such nodes in real networks. The algorithm is named for its inventors Albert-László Barabási and Réka Albert and is a special case of a more general model called Price's model.
More information about this topic is on 3.3, 3.2

Chapter 5

SIS - Agent-based models

In this section the SIS epidemiology model is introduced as well as different agent-based models and their policies following the structure of the SIS model.

5.1 SIS model

The SIS model is a basic epidemiology model where all individuals are susceptible of becoming infected and can recover from such infection. When an individual recovers from the infection it is immediately susceptible to becoming infected again. This is a key difference from other epidemiology models such as SIR/SIRS where the recovered individual is resistant to becoming infected for a while (SIRS) or flat out immune (SIR).

Given a constant population of N individuals, and denoting I as the number of infected individuals and $S = N - I$ as the number of individuals susceptible of becoming infected, the first order differential equations (ODE's) governing the system are:

$$\begin{aligned} \frac{dS}{dt} &= -\frac{\beta S I}{N} + \mu I \\ \frac{dI}{dt} &= \frac{\beta S I}{N} - \mu I = -\frac{dS}{dt} \end{aligned} \tag{5.1}$$

where β represents the infectious rate and μ the recovery rate, $\beta, \mu \geq 0$.

One might be interested in knowing when the system reaches an equilibrium (i.e. $\frac{dI}{dt} = 0$). A quick look into 5.1 gives us:

$$\frac{dI}{dt} = 0 \Longleftrightarrow I = 0, \text{ or } I = \left(1 - \frac{\mu}{\beta}\right) N. \tag{5.2}$$

Thus, it makes sense to define the reproductive number as $H_0 = \frac{\beta}{\mu}$.

- At the $\lim_{H_0 \longrightarrow 0}$ we have $I \longrightarrow -\infty$, but since the number of infected individuals is positive ($I \geq 0$) we recover $I = 0$, meaning the infection disappears from the population. This limits corresponds to the case $\mu \gg \beta$, i.e. the recovery rate is much higher than the infectious rate.

- At the $\lim_{H_0 \longrightarrow \infty}$ we have $I \longrightarrow N$, so all the population becomes infected. This limits corresponds to the case $\beta \gg \mu$ i.e. the infectious rate is much higher than the recovery rate. Observe how both limits behave exactly as one would expect they should.

- Moreover, for the infection to spread it is required that $H_0 > 1 \iff \beta > \mu$. Since the amount of infected individuals at the equilibrium state is found by 5.2, notice how, for $H < 1 (\iff \beta < \mu)$, the total number of infected individuals at equilibrium would be negative and, since that is not possible, would essentially be 0. Meaning, the infection would die. Also, note that if $\mu = 0, \beta > 0$ the infection spreads to all the population. Thus, both extreme cases, fully recovered and fully infected, can be obtained without the need to go to the limits $\mu \gg \beta, \beta \gg \mu$, see Figure 5.1.
- Notice that whenever $\mu = \beta$, equation 5.2 would give us $I = 0$ and $dI/dt = 0$ which is incompatible with a set of initial conditions where for instance $I_0 \neq 0$.

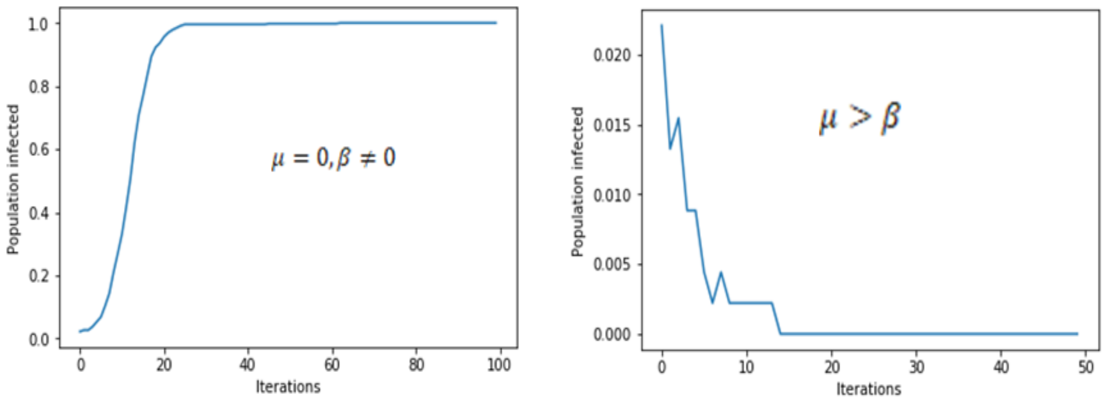


FIGURE 5.1: Simulations that show the two extreme SIS behaviours, all infected (left) and all cured (right).

Let's discretize the first order differential equations. Given two consecutive measures of time t_i, t_{i+1} equations 5.1 become:

$$
\begin{aligned}
S(t_{i+1}) &= S(t_i)\left(1 - \beta \Delta t \frac{I(t)}{N}\right) + \mu I(t) \\
I(t_{i+1}) &= I(t_i)\left(1 + \beta \Delta t \frac{S(t)}{N}\right) - \mu I(t) = -S(t_{i+1})
\end{aligned}
\tag{5.3}
$$

where $\Delta t = t_{i+1} - t_i$. Redefining the rates as $\beta' = \beta \Delta t, \mu' = \mu \Delta t$, one can write

$$
\rho(t_{i+1}) = \rho(t_i) - \mu \rho(t) + \beta \rho(t)(1 - \rho(t)) \tag{5.4}
$$

where the primes ($'$) have been omitted and $\rho(t_i) = \frac{I(t_i)}{N}$ is the fraction of infected individuals at instant t_i.

In a microscopic approximation the fraction of infected individuals becomes a good estimator of the probability of infection. Of course, in our interpretation we need to account also for the weights of the edges between the nodes, making the probability of the infection being spread from your neighbours to you proportional to the weight. To keep the probability sense we need to normalize the out-degree weights to 1 for every node.

The resultant equation that governs the system is:

$$p_i(t+1) = \underbrace{(1-q_i(t))(1-p_i(t))}_{(1)} + \underbrace{(1-\mu)p_i(t)}_{(2)} + \underbrace{\mu(1-q_i(t))p_i(t)}_{(3)} \tag{5.5}$$

where

$$q_i(t) = \Pi_{j=1}^{N}(1-\beta r_{ji} p_j(t)) \tag{5.6}$$

Then, $(1-q_i)$ accounts for the probability that your neighbours infect you. More precisely, each term in 5.5 can be interpreted as follows,

- **(1):** Is the probability that your neighbours infect you given that you were not infected in the previous iteration.
- **(2):** Is the probability that you remain infected given that you were infected in the previous iteration.
- **(3):** Is the probability that your neighbours infect you given that you were infected in the previous iteration and you healed yourself in this one.

Finally, note that r is the matrix of weights between nodes, thus r_{ji} is the weight of the edge connecting nodes j and i. If those nodes are not connected the weight between them is 0.

5.1.1 Stochastic approach

In our approach we will follow a stochastic $(0,1)$ rule for the probability of a company being defaulted at each iteration i. We will still use equation 5.5 but at the end of each iteration we transform the continuous probabilities p into either 0 or 1. This is done by generating a random number between 0 and 1. If the generated number lies below the continuous probability p_i then the company becomes defaulted $p_i = 1$, otherwise $p_i = 0$.

The main difference between this case and the continuous probabilistic one is on the edge rewiring. Thus, the policies our agents followed may not work properly anymore. Remember the policies are the criteria that a business (node) follows to look for a new supplier when the previous provider has entered default. Let's look at how each policy used in the previous work 5.2.1 behaves in the stochastic scenario.

- The hard or strict policy will force all companies to change their defaulted suppliers to non-defaulted ones. Thus, it will become stagnant after the first iteration. All companies will become non-defaulted if the recovery rate (μ) is non-zero and constant otherwise.
- The random policy is always an option and we will still consider it viable. Nevertheless, since at the beginning there are much more non-defaulted companies than defaulted ones and the edge-rewiring only takes place when the supplier is defaulted, it is much more likely that the new supplier is not defaulted.The end result is that this policy will also become stagnant after a short while.
- The soft policy simply differs from the strict one in the sense that two defaulted companies would still do business with one another. Since an infected company can not get infected again this policy is as irrelevant as the strong or

random one and will become stagnant in a short time. See for instance Figure 5.2.

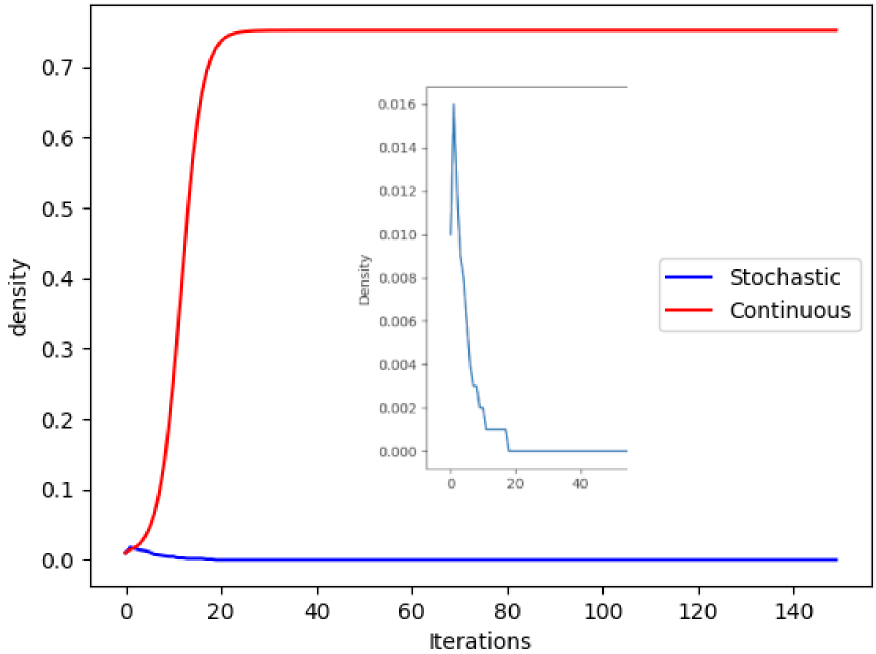


FIGURE 5.2: Comparison of the evolution of the default in a financial network for the stochastic and continuous cases applying the SOFT policy given $\mu > 0, \frac{\mu}{\beta} = 0.25$. Includes a Zoom into the stochastic curve to see the behaviour.

Thus, new policies regarding edge rewiring need to be defined.

5.2 Agents and policies

The agents move through the financial network following different policies. This policies are guidelines to change suppliers. It is important to note that a business will only attempt to change suppliers when one of them becomes defaulted.

5.2.1 Previous work policies

In the continuous case, a node considers that one of his suppliers is defaulted whenever the defaulted probability of that supplier is higher than its own defaulted probability. The previously used policies were:

- **Random**: The business chooses it's new supplier randomly. That means a healthy node can be reconnected to either healthy or infected nodes.
- **Soft**: The business only chooses suppliers whose probability of infection is equal to or lower than that of itself. That means, a healthy node will be reconnected to healthy nodes and an infected node will be connected to infected or healthy nodes.

- **Strong**: A business only chooses suppliers whose probability of infection is strictly lower than that of itself. In essence, the literal interpretation would make, in a stochastic case, healthy nodes unable to reconnect. Nevertheless, the interpretation should be that all nodes will only reconnect to healthy nodes.

where it has been included a brief discussion of what happens in the stochastic approach as it has been previously discussed in 5.1.1. Since there are only 2 states $(0, 1)$, defaulted and non defaulted, the companies that enter default are quickly isolated for all the previous policies. Then, if the system has a recovery rate higher than 0, the system will be fully recovered and, if the recovery rate is 0, there will be for the entire process a constant number of infected companies. See Figure 5.2.

5.2.2 New Policies

For defining the new policies what it is proposed is to create some added modifications to complement the previous policies. This modifications are the following ones:

- **Delay**: The delay represents the number of steps a company waits to change suppliers by following one of the above policies. In essence, delay measures the patience of a company with its suppliers. Clearly, when the delay is 0 one encounters the stagnation that was described before and when the delay parameter systematically increases it becomes more similar to not doing edge rewiring at all.

- **Weight transfer**: If a company has two or more suppliers of the same sector it transfers weight from the defaulted connection to the non-defaulted connections in the same sector. If the company has no more suppliers in that sector then, it creates a new edge into another company in the sector (using one of the previous criteria) but keeps the previous connection and, then, transfers weight. The amount of weight transferred can be $1/4, 1/6, ...1/\eta$ of the total weight such that after η weight transactions the edge disappears.

Then, the resultant policies our agents will follow will be of the form: SOFT + delay 3, RANDOM + Weight-transfer,...

One interesting question is how much patient, loyal, a company can be to its suppliers without compromising itself and/or the complete financial network? This issue will be discussed for some particular financial networks in 7.

Observation 5.1 The weight transfer modification is too expensive in computational time and will not be implemented extensively. This happens because as the infection spreads through the financial network more and more suppliers are defaulted, making a huge amount of companies look for new suppliers. However, they keep the connection to the previous ones, at least for a while. Then, the number of edges grows fast making the algorithm increasingly slower in the following iterations.

Chapter 6

Failure Cascades

When looking at the dynamics of the system one of the most interesting points is to analyse the existence and, if so, the behaviour of failure cascades. That is, the analysis on how a node failing (entering default, becoming infected) can trigger a succession of failures in neighboring nodes that can ultimately collapse the whole network, or a great portion of it. This analysis may show which nodes have a deeper influence in the network and which connections can then be a potential risk for the company as well as for the whole network.

6.1 Properties/Attributes

Let us define:

- **Cascade Depth (Cd)**: How many levels the default propagates down to.
- **Cascade Size (Cs)**: How many nodes become defaulted. Note we will not consider the original node as part of the final cascade.

For instance, the failure cascade for a simple network consisting on 3 companies the first pointing to the other two where all three companies become defaulted when the first company defaults, has size 2 and depth 1.

When looking at the failure cascades in a directed complex network one can follow two basic approaches:

- **Back-propagation**: Find a node in default that points to no other node in default and backtrack the origin of the failure cascade obtaining the cascade's size and depth.
- **Forward propagation**: Find a defaulted node not pointed by any defaulted node and go forward in the direction of the edges to find the cascade size and depth.

Observation 6.1: An interesting point about the default propagation in a financial network is that it follows the inverse of the edge directions. A business can become defaulted if one of his suppliers is defaulted. Remember edges go in the direction of cash flow so, they point towards the suppliers. As a result, for a network where default propagates in the inverse direction of the edges, the notion of back-propagation becomes forward propagation and vice versa.

A slightly different way to look at the failure cascades to make sure the whole network is properly analysed would be the following:

- **Controlled propagation**: The initial companies that enter default are not selected at random but instead they are fixed and in successive simulations different starting defaulted companies are selected until all of them have been chosen at least once.

  Note that one has to be very careful with which companies sets into default because one of them could be in the path of another and thus contribute to the failure cascade becoming larger when, for instance, it wouldn't have grown so much. Then, the only sure way for a general complex network to actually perform this analysis is to set them initially one by one to default in successive simulations or to select companies that are not connected to each other following any possible path.

6.2 Cascades with Edge Rewiring

What happens when we allow edge rewiring in our financial network?

The general hypothesis is that one expects the cascade sizes and depths to be significantly reduced. More precisely, one expects the probability of the cascade size to be smaller than a certain value Cs, $P(cs < Cs)$ (i.e. the cumulative distribution) to increase for lower values of Cs. Meaning, edge rewiring stops failure cascades from reaching higher values than they would.

Looking at the policies that will be used in the study, 5.2, this effect should be more significant for the Soft policy than the Random one. Of course, as it was shown in 5.1.1, for the stochastic approach the difference between Random or Soft is minimal and this difference may not be observed. Moreover, edge rewiring without delay stops default contagion almost immediately and therefore failure cascades do not occur. As delay increases, cascades should become more prominent, since we get closer to the no edge rewiring case.

Observation 6.2 Looking at the details of the algorithm, clearly one needs to keep track of all the nodes (business) that had been connected to every other node in the past in order to be able to deconstruct how the cascades had originated. Furthermore, the first iteration for which the node becomes defaulted should also be kept. This extra information stored indirectly provides us a nice piece of knowledge, which are the origin nodes that can potentially form a cascade.

Therefore, edge rewiring clearly will help to reduce the cascades size and the higher the delay parameter is, the larger the expected cascades should be. Also, the higher the reproductive number $H_0 = \frac{\beta}{\mu}$ the larger this cascades will be. However, there are still answer to be obtained. For example, which model, ER or BA, works better in this front? Which financial sector is more susceptible to generate a failure cascade? Which less? We will try to answer all this questions in 7.

6.3 Risk in a financial network

Failure cascades are one of, if not the main, option to evaluate the risk in a financial network. In this section we propose different approaches to assess the risk in a general financial network. Note that the computational cost of risk assessment tends to by considerably huge since multiple repetitions need to be performed for the results to be statistically significant.

6.3.1 Basic risks

Let's start by defining some basic risks that do not require failure cascade computation.

- **Personal risk** (R_p)**:** Accounts for the risk for a certain node (business) to become defaulted given a set of initial conditions. This can be assessed in a strict way, looking only to the end result, have I became defaulted at some iteration? Or in a softer way, taking into consideration at which iteration the company entered default and reducing the risk as the iterations increase. The proposed equations are:

$$\begin{aligned} R_p &= \frac{N_{defaulted}}{N_{total}} && \text{(Strict)} \\ R_p &= \frac{1}{N_{total}} \sum_j \left(1 - \frac{i_j}{i_{total}}\right) && \text{(Soft)} \end{aligned} \tag{6.1}$$

  where $N_{defaulted}$ is the number of repetitions for which the node has being defaulted at some point, N_{total} are the total number of repetitions and i_j, i_{total}, are the first defaulted iteration and the total number of iterations of repetition j.

- **Network risk** R_n**:** Accounts for the risk for the overall network of default propagation given a set of initial conditions. Again, two approaches (Strict,Soft) can be implemented. The proposed equations are:

$$\begin{aligned} R_n &= \frac{N_{eq}}{N_{total}} && \text{(Strict)} \\ R_p &= \frac{1}{N_{total}} \sum_j \left(1 - \frac{i_j}{i_{total}}\right) && \text{(Soft)} \end{aligned} \tag{6.2}$$

  where in this case N_{eq} represents the amount of simulations that have achieved the defaulted equilibrium $\rho = (1 - \frac{\mu}{\beta})$, and i_j the iteration for which the system has first arrived to that equilibrium.

6.3.2 Cascade related risks

Moving on from the basic risks, lets define some new risks that take into consideration the failure cascades generated in the financial network. The risks in this approach will be represented by a pair of values (x, y) where x will be the expected size of the cascade and y the maximum size.

- **Individual risk,** R_i**:** Accounts for the risk for the overall network of one node becoming defaulted.

- **Sectorial risk,** R_s**:** Accounts for the risk for the overall network of one sector entering partially in default. It assesses the potential risk of one sector,for instance transportation, reacting to an outrageous increase of the petroleum cost, which results in a small but significant percentage of the sector entering default. May show some light into the importance of each sector in the network.

Knowing the risk of a failure cascade originating and the expected size of it may provide critical knowledge for financial institutions and governments to act accordingly before they occur in order to prevent them. Also, from a business standpoint, one company might be interested in knowing the risk he is taking by changing or acquiring a certain supplier or by not doing so. Some risk assessment will be done in section 7 but it will be kept mostly general.

Chapter 7

Results

In this section we show the obtained results throughout the study. First, we will focus on the analysis of a Sharing Economy (Cooperative) Market, here represented by an Erdós-Rényi network topology, vs a Liberal Market, here represented by a Barabási-Albert network topology (see sections 3.2, 3.3, for more information about these models). Afterwards, we will show the results obtained with the simulated BBVA client-supplier network (see Table 2.2).

All studies for both datasets were carried out with the two re-connection policies, namely random + delay and soft + delay, for different values of the Delay parameter, as they were defined in Section 5.2.

Data management and statistical analysis were performed using Python 3.6.5 with the following libraries:

- Numpy
- Pickle
- Networkx
- Matplotlib
- Sys
- Joblib

For all statistical tests a nominal significance level 5% was fixed.

7.1 ER model

The first analysis attempts to create a random sample to verify how a financial network would work without having a hierarchical structure among entities with different status and without predefined sectors having different default distribution. This network resembles an ER topology.

sector	size(%)	k_{in}	k_{out}	default (%)
1	20.700	6.188	5.778	0.966
2	21.300	5.986	6.000	0.000
3	19.100	5.927	5.990	0.000
4	20.100	5.831	6.045	0.498
5	18.800	6.032	6.176	1.064

TABLE 7.1: ER network information.

Figure 7.1 shows $\rho(\beta,\mu)$ for each economical sector. As this client-supplier network is created randomly it can be observed that all sectors behave in the same way regarding default dynamics. Notice that the fact some sectors are initially defaulted, and some are not, does not affect at all the behaviour of the end result for each sector for any ratio.[1] Also, notice that up to ratio 3 propagation does not kick off. This is

[1]Ratio = Reproductive number = $\frac{\beta}{\mu}$

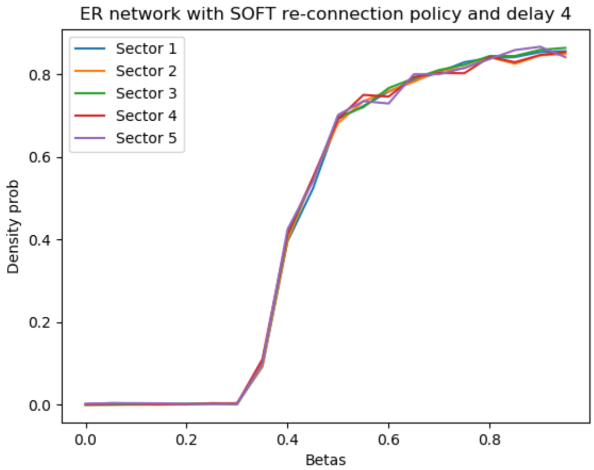


FIGURE 7.1: Average sectorial default density at equilibrium for the random Erdos-Renyi graph with a recovery rate $\mu = 0.1$.

due to the stochastic effect as this would not happen in a continuous default probabilities setting. This effect was shown for delay 0 in Figure 5.2 and will be studied in further detail for higher delays in the next section.

7.1.1 Study of critical delay

Here, given a value of $\mu = 0.2$ we will analyze for some values of $\beta = 0.4, 0.6$ (i.e. reproductive ratio $H_0 = 2, 3$) which is the critical value of the delay, for several different delays. That is, for which delay the network shifts from full recovery, $\rho = 0$, to the upper defaulted equilibrium, $\rho = \left(1 - \frac{\mu}{\beta}\right)$. An extensive analysis of the probability of reaching each one of the equilibrium states is also included.

First scenario. $\beta = 0.4$

Figure 7.2 shows for which delay the upper equilibrium is reached. Since we are assuming that the only two possible states of equilibrium are $\rho = 0$ and $\rho = 1 - \frac{\mu}{\beta}$, then, whenever the average default density of the 30 simulations becomes stable for a non zero value this automatically implies that the upper equilibrium has been reached for some of the simulations. [2] The delay that first behaves as just described is delay 4. Thus, 4 is the so called critical delay.

It is important to observe that, as expected, every delay increases the average default density until delays $9 - 11$, which are around the expected equilibrium, are reached. Both underlying policies (soft & random) do not seem to matter in an stochastic setting. For the same delay both default density curves are pretty close one with the other. The random ones seem slightly higher for some curves but lower for others. More repetitions should be performed to obtain a more robust result. To mitigate this issue, an extensive study for each delay is performed below.

[2] The mean between a simulation reaching the upper equilibrium and another fully recovered is a curve that mimics the stable behaviour of the first one but only reaches half the value.

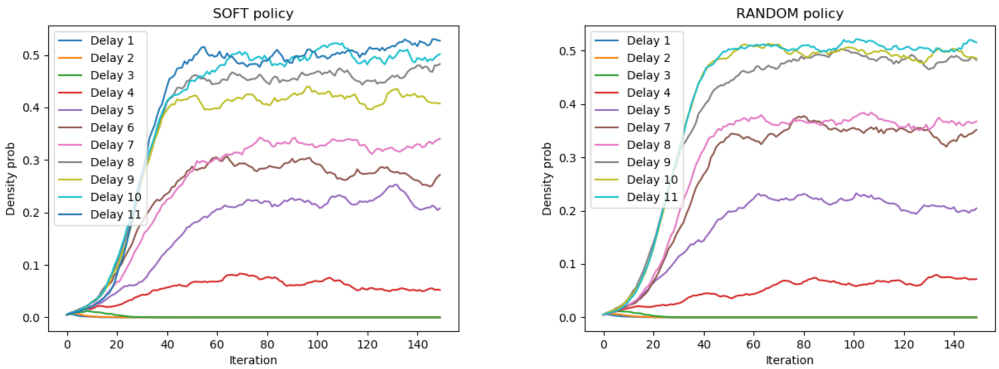


FIGURE 7.2: Average default density evolution for 30 simulations given different values of the delay for random and soft policies

It was done a Monte-Carlo procedure to have a better idea about how the default density is really distributed after replicating the experiment $n = 100$ times for the previous values of the parameters. After this procedure we want to analyze how many of them really go to the upper equilibrium point (after 200 iterations) and how many remain at the end by an intermediate point (meaning that they have not yet converged). These results are shown with some plots, see Figure 7.3, and within a table, see Table 7.2. The table will contain the percentages of how many times at the end of the simulation the density ends at the point of balance or in the 0. Therefore, this percentage will not add up to 1 in some cases, because on the intermediate delays some simulations will not converge after 200 iterations. The plots will represent the average values of the default density plus the confidence interval of 95% of the average in order to have a better idea of how it differs during the whole process.

soft

Delay = 3	Delay = 4	Delay = 5	Delay = 6
(1.00,0.00)	(0.90,0.00)	(0.27,0.17)	(0.11,0.71)
Delay = 7	Delay = 8	Delay = 9	Delay = 10
(0.06,0.92)	(0.09,0.91)	(0.07,0.93)	(0.03,0.97)

random

Delay = 3	Delay = 4	Delay = 5	Delay = 6
(1.00,0.00)	(1.00,0.00)	(0.20,0.12)	(0.10,0.70)
Delay = 7	Delay = 8	Delay = 9	Delay = 10
(0.08,0.90)	(0.07,0.93)	(0.09,0.91)	(0.03,0.97)

TABLE 7.2: Percentages of the density (lower, upper) that reaches each respective equilibrium point for different values of beta and delays taking into account the policies applied.

From Table 7.2 it is observed that when the delay increases the percentage of values that goes to the upper equilibrium also increases as it was expected. This trend continues until the delay is high enough such that it resembles the non edge rewiring case. At first, one could think that without edge rewiring (i.e. ∞ delay), the expected result is that all of the simulations will reach the upper equilibrium. Actually, this is not entirely the case always as we will see in 7.3.

Looking more closely at the values of the policies, it can be observed that the random policy does not work particularly worse than the soft. In fact, for many results

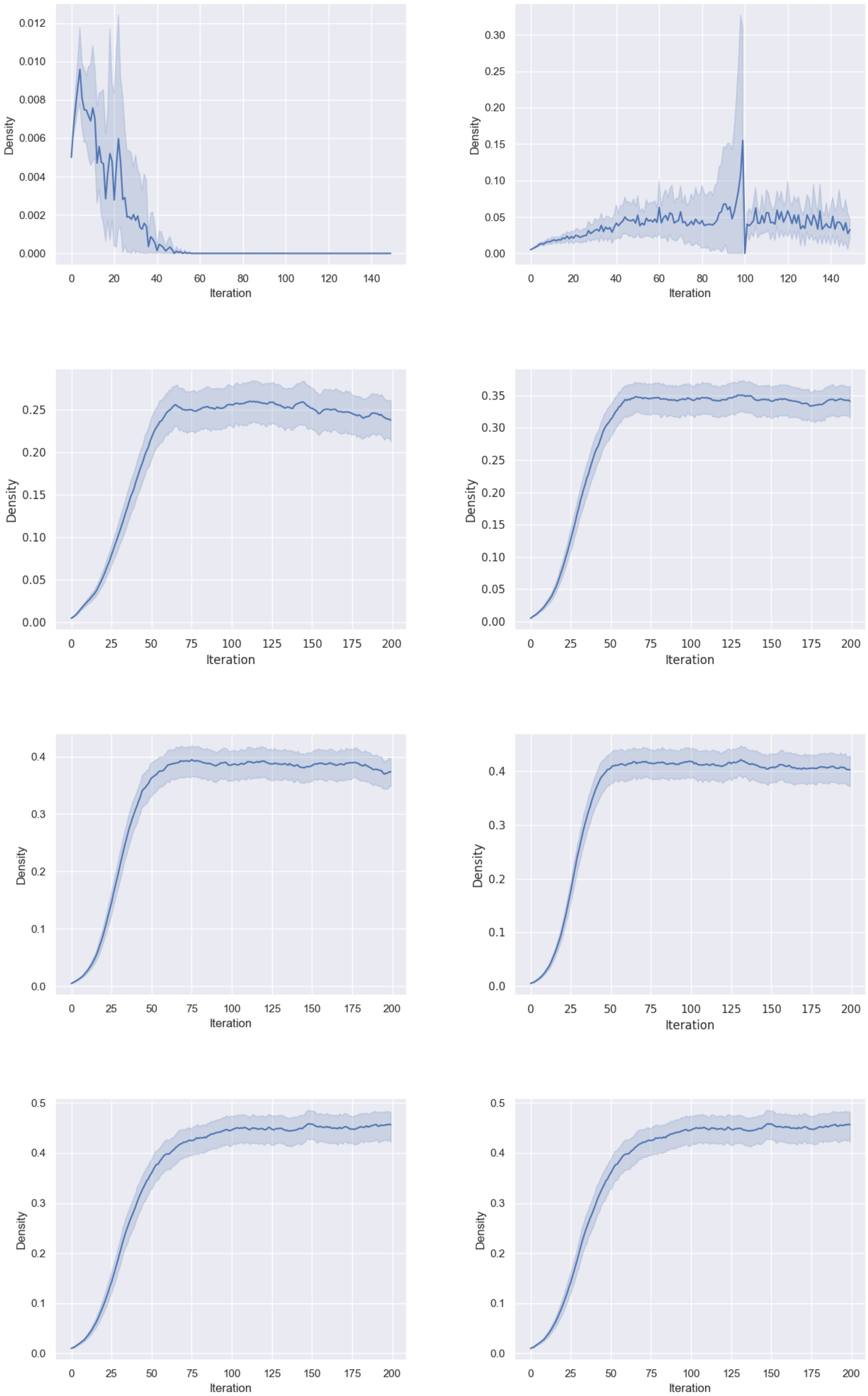


FIGURE 7.3: 100 simulation the mean of the density with the IC of the mean for different values of the delay(3 to 10) from right to left and for the policy soft

the random policy actually behaved slightly better, which is a clear indicator that both underlying policies seem to produce the same effect on the network dynamics.

Moreover, we can detect how for delay 5 we have almost all the simulations still in a middle ground and for delay 4 almost all of them are down. Therefore, the most interesting delay is by far 5. A comment on this strange balance obtained for delay 5 is done in the Second scenario, $\beta = 0.6$, for comparison purposes.
Similar conclusions are elucidated from figures 7.3.

Second Scenario. $\beta = 0.6$

Again we start by studying for which delay a simulation reaches the upper equilibrium first, i.e. the critical delay.

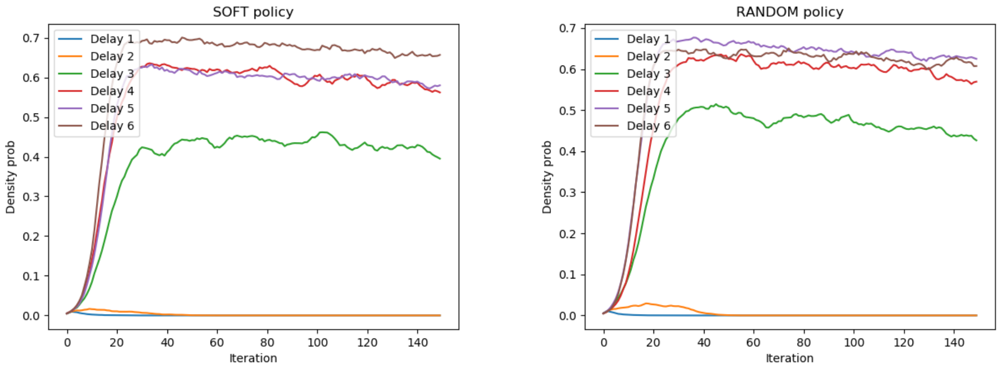


FIGURE 7.4: Average default density evolution for 30 simulations given different values of the delay for random and soft policies

In this case the critical delay is equal to 3. Moreover, now for delay 4 the average default density is already significantly close to the upper equilibrium. Meaning there are not that many intermediate delays which are the more interesting to study.

The difference between following a random or a soft policy seems greater in this case, although may not be significant. To study if in fact the random policy yields greater default densities as average than the soft let us move to the same study we have already did for $\beta = 0.4$.

soft

Delay = 2	Delay = 3	Delay = 4	Delay = 5
(1.00,0.00)	(0.17,0.83)	(0.03-0.97)	(0.02,0.98)

random

Delay = 2	Delay = 3	Delay = 4	Delay = 5
(1.00,0.00)	(0.05,0.94)	(0.04,0.96)	(0.00,1.00)

TABLE 7.3: Percentages of the density (lower, upper) that reaches each respective equilibrium state for different values of the delay.

From table 7.3, we see that the upper equilibrium point is reached much faster than in the case of $\beta = 0.4$. Meaning, the upper equilibrium is reached for a lower delay. Moreover, there are no real intermediate delays, as delay 5 was before, where the default density has not reached any of the two possible equilibrium states after

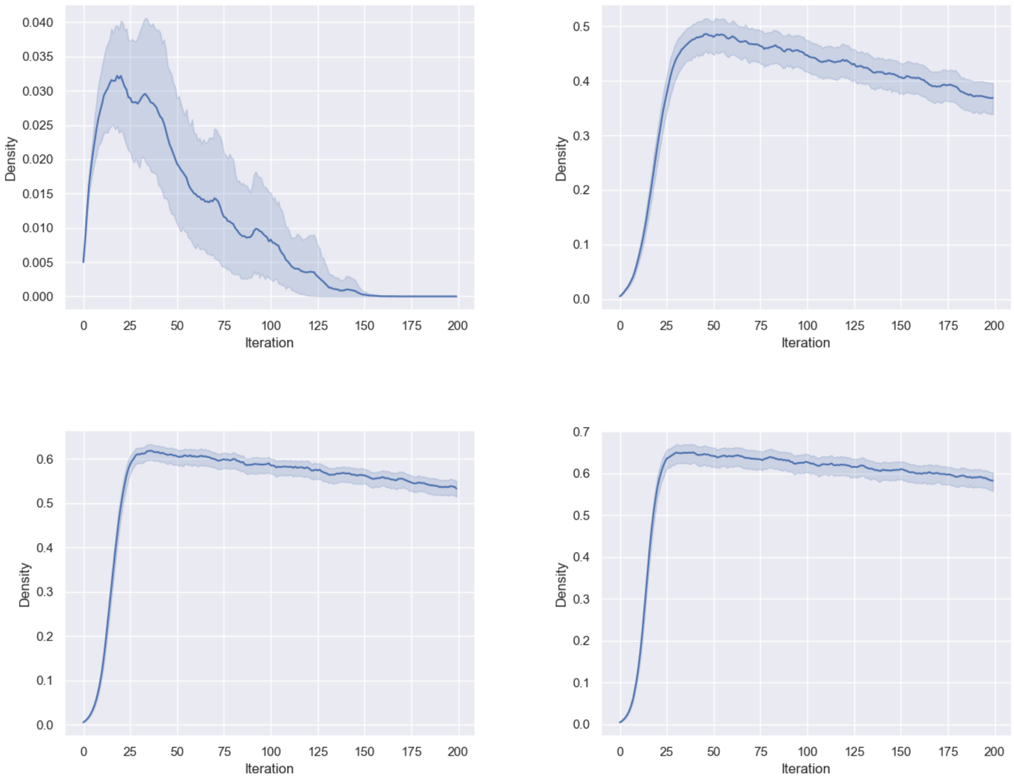


FIGURE 7.5: 100 simulation the mean of the density with the IC of the mean for different values of the delay(2 to 5) from right to left and for the policy soft

the 200 iterations. This is due to the higher ratio/reproductive number $H_0 = \frac{\beta}{\mu}$. Let's explore why.

The effect the re-connection policies have in our stochastic setting is to reduce the speed of the default spreading. Alternatively, it could be viewed as an increase of the overall recovery . This effect depends on the amount of companies re-connecting and the ratio. Thus, depends on the parameters μ, β, and delay.
For delay 5 it existed sort of a balance between the speed at which companies actually became defaulted and the amount of companies that could heal themselves. On paper, this is not an isolated result, as other possible combinations of delays and ratios might yield the same intermediate behaviour. A change in delay implies an automatic change on the amount of companies re-connecting. Therefore, given a ratio, this balance can only exist for one delay. However, as we see in Table 7.3, for ratio 3 no value of the delay is able to achieve the balance. It should be pointed out that this false balance is highly unstable, and the higher are the forces involved (recovery and default spreading) the harder is to achieve that balance. Thus, the higher the reproductive ratio, the harder this balance exists.

For this higher ratio, the soft policy works better than the random one, especially for the only intermediate delay, 3. Remember that the soft policy will only reconnect healthy nodes with healthy nodes, but in the random policy they can be reconnected to anybody. That behaviour,given a higher ratio, produces a bigger difference between the policies than a lower ratio would as it penalizes each wrong re-connection harshly.

From Figure 7.5 we see that even for a very small delay (delay 2) it takes much longer to reach $\rho = 0$ than it did in Figure 7.3. Moreover, we can see that when the default spreads through the network, the amount of infected companies quickly exceeds the expected upper equilibrium and then slowly converges to it.

The same plots have also been made for the random policy, for more information check the Appendix A.

7.1.2 Default cascades study

In this section, the random policy is omitted because, as seen in the previous section, results were very similar to the ones obtained with the soft policy. Moreover, the soft policy may represent a more realistic scenario.

Therefore, for the same values of the delay and the reproductive number H_0 as in the previous study, we will examine how the failure cascades behave in the Erdós Rényi model network with soft policy.

$\beta = 0.4$		
Delay	$\mathbb{E}(Cs)$	$\max(Cs)$
2	1.22	14
3	4.07	50
4	87.39	845
5	145.04	947
6	184.71	968
7	184.19	956
8	172.72	953
9	173.94	955
∞	158.83	889
$\beta = 0.6$		
Delay	$\mathbb{E}(Cs)$	$\max(Cs)$
1	1.18	8
2	16.66	659
3	189.39	989
4	197.62	984
5	197.00	982
6	190.22	977
∞	156.72	848

TABLE 7.4: Expected and maximum cascade size, $\mathbb{E}(Cs), \max(Cs)$, for different values of the delay for the ER network with $\mu = 0.2$.

From Table 7.4 we can see that even for low values of delay the maximum size of the cascades is huge. For a delay of 4 and 2 for the ratios 2, 3 respectively the maximum cascade size already approaches the maximum possible, even if the expected number of cascades, $\mathbb{E}(Cs)$, is not too big. This means that we have a long tail of the distribution and a large concentration of values close to 0.

The most interesting result of the table however, is that if no edge rewiring is performed the maximum size of the cascades and their expected value is lower than for the higher considered delays. Meaning, waiting too much to do edge-rewiring is worse than not doing it at all. This is a really surprising result as not doing edge rewiring is the same as delay $\to \infty$. The root of this discrepancy is in the fact that re-wiring connections allows two parts of the network, on paper separate, to become connected. This might stops cascades that should reach a size of ~ 100 in a size of ~ 10 but allows others to reach a size of ~ 900.

Figure 7.6 shows the inverse cumulative distribution of cascade sizes.

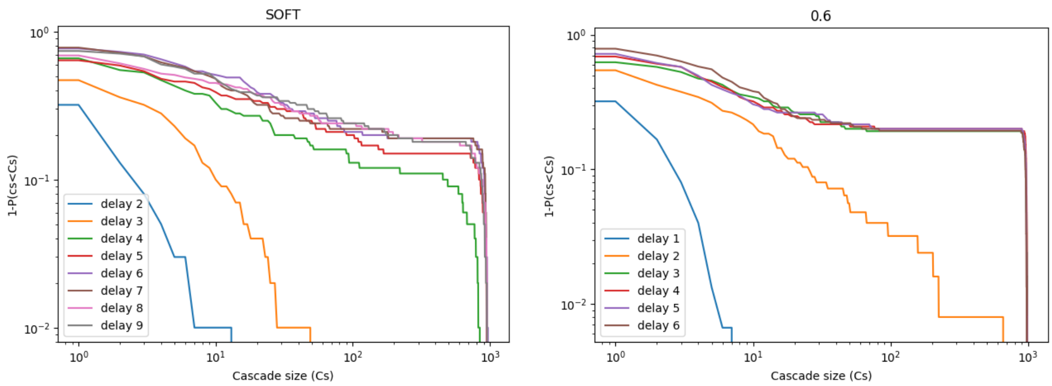


FIGURE 7.6: Inverse cumulative distribution of the cascade size (Cs) for different delays, soft policy, $\mu = 0.2$ and $\beta = 0.4$ (left) , $\beta = 0.6$ (right).

7.2 BA Model

Now we will follow the same procedure as in Section 7.1 for the structured, hierarchical network created following the Barabási-Albert model.

Let us start by showing the sectorial behaviour given Table 7.5. Notice that the in and out-degree as well as the size(%) have been kept almost the same between sectors. We have assigned the Hubs to the different sectors in order. Our goal is to figure out if just the Hub structure by itself already results in a sectorial differentiation.

sector	size(%)	k_{in}	k_{out}	default (%)	Hubs rank
1	18.900	6.265	6.265	1.587	1
4	20.500	5.615	5.615	0.488	4
5	20.200	5.401	5.401	0.495	5
3	21.700	6.217	6.217	1.843	3
2	18.700	6.455	6.455	0.535	2

TABLE 7.5: BA network information.

Figure 7.7 shows that the Hub structure just by itself does not prompt any sectorial differentiation. Of course, having one node blocking default propagation does not influence the whole network/sector if the rest of the nodes have random connections. In real client-supplier networks Hubs shield uneven parts of the network, meaning, one Hub protects some sectors more than others and this, more than the sector of the Hub itself, is what prompts the sectorial differentiation. A deeper discussion on the matter will take place in Section 8.

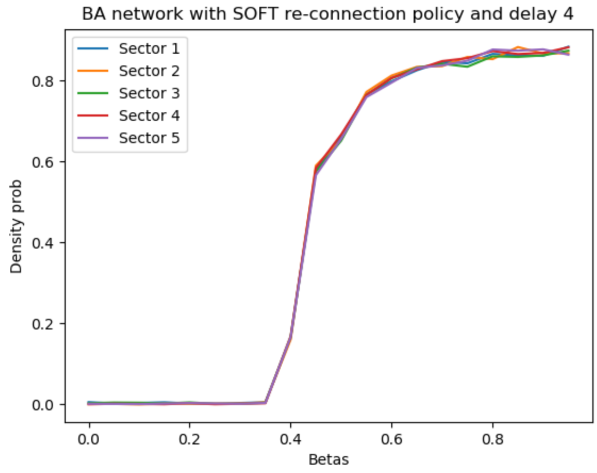


FIGURE 7.7: Average sectorial evolution for the BA network given a recovery rate $\mu = 0.1$.

7.2.1 Study of the critical delay

Again, given a value of $\mu = 0.2$ we will see for some values of β which is the critical value of the delay, for a battery of different delays. That is, for which delay the network shifts from full recovery, $\rho = 0$, to some simulation achieving the upper equilibrium of $\rho = \left(1 - \frac{\mu}{\beta}\right)$.

First scenario. $\beta = 0.4$

Figure 7.8 shows that for the BA network the critical delay is around 5. A deeper study will be performed as we did before for the ER network.

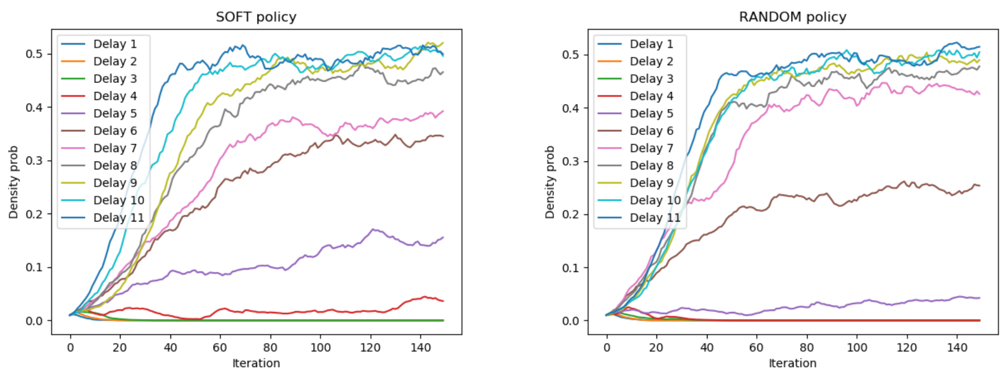


FIGURE 7.8: Average default density evolution for 30 simulations given different values of the delay for random and soft policies

From Table 7.6 it can be observed that when the delay increases the percentage of values that goes to the equilibrium also increases, as expected. Looking closely at the values of the random and soft policies, we see a major difference with the ER network. The BA network seems to have a significant difference between both policies, with the random policy yielding worst results than the soft one for all delays, soft

Delay = 3	Delay = 4	Delay = 5	Delay = 6
(1.00,0.00)	(1.00,0.00)	(0.70,0.02)	(0.58,0.27)
Delay = 7	Delay = 8	Delay = 9	Delay = 10
(0.24, 0.63)	(0.11,0.87)	(0.09,0.90)	(0.04,0.96)

random

Delay = 3	Delay = 4	Delay = 5	Delay = 6
(1.00,0.00)	(1.00,0.00)	(0.48,0.06)	(0.28,0.62)
Delay = 7	Delay = 8	Delay = 9	Delay = 10
(0.11,0.85)	(0.1,0.9)	(0.04,0.96)	(0.00,1.00)

TABLE 7.6: Percentages of the density (lower, upper) that reaches the equilibrium for different values of beta and delays taking into account the policies applied in BA model

especially the intermediate ones. Moreover, in this case we do not seem to have a delay for which the values remain mostly between the two points of equilibrium after 200 iterations. A deeper analysis of the differences between both model networks is performed in section 7.3. Look at the plots in Figure 7.9.

Second scenario. $\beta = 0.6$

Figure 7.10 shows that the critical delay is clearly 3 as it was for the ER network.

soft

Delay = 2	Delay = 3	Delay = 4	Delay = 5
(1.00,0.00)	(0.23,0.77)	(0.03,0.97)	(0.00-1.00)

random

Delay = 2	Delay = 3	Delay = 4	Delay = 5
(1.00,0.00)	(0.27,0.72)	(0.05,0.95)	(0.02,0.98)

TABLE 7.7: Percentages of the density (lower, upper) that reaches the equilibrium for different values of beta and delays taking into account the policies applied in BA model

From Table 7.7 we observe mainly two things. First, that once again there is no intermediate delay for which the simulations stay between both equilibrium states after 200 iterations. Second, that the soft policy does not work better than the random one in this case.

Looking at Figure 7.11 we can see some interesting behaviours when comparing with the ER network. For delay 2 we observe that there is a small peak but then goes straight to 0 with far less discrepancy. For the next delays the infection does not spread above the upper-equilibrium in the first few iterations and simple converges to it as usual.
The same plots have also been made for the random policy, for more information you can check the appendix A.

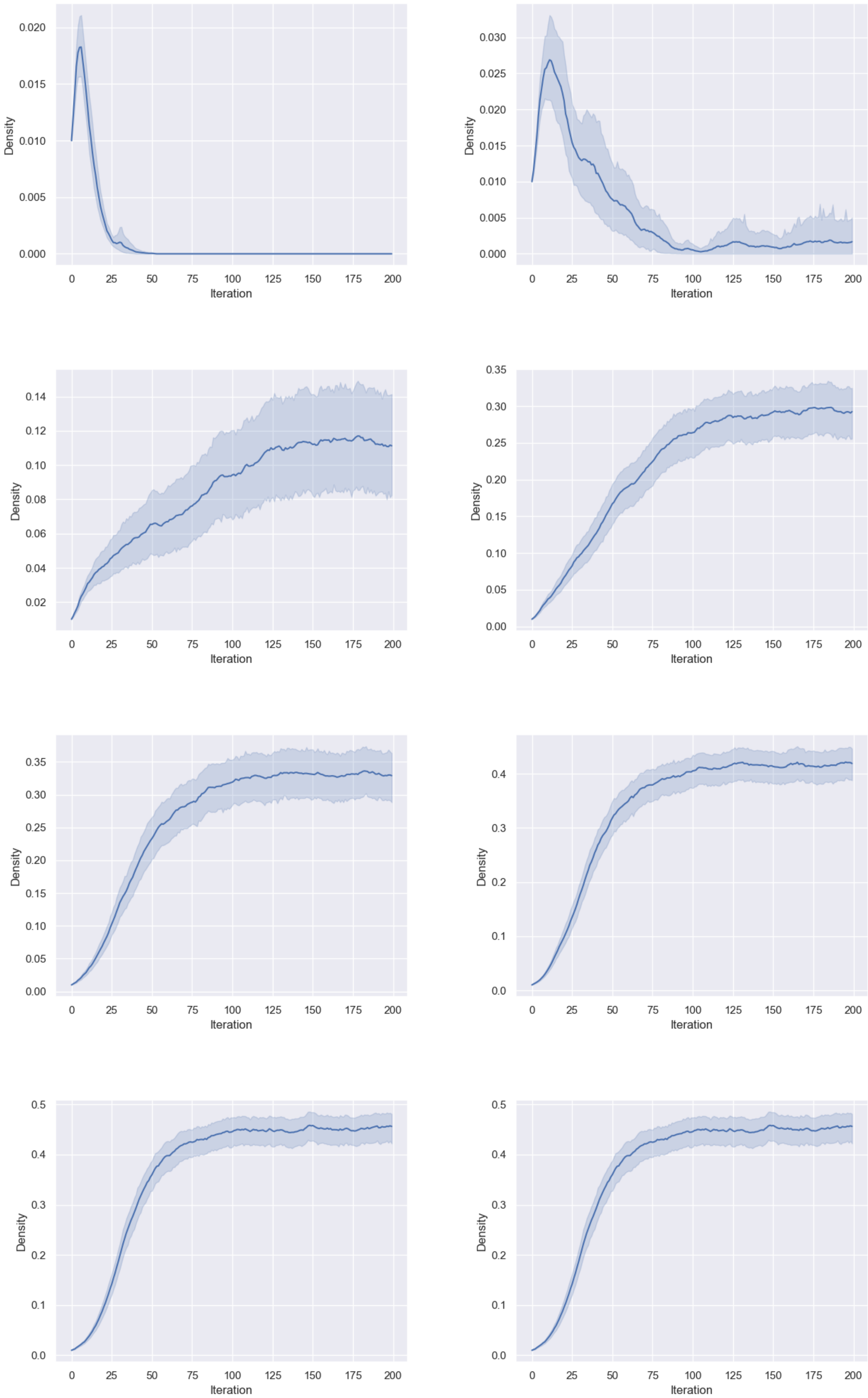


FIGURE 7.9: 100 simulation the mean of the density with the IC of the mean for different values of the delay(3 to 10) from right to left and for the policy soft.

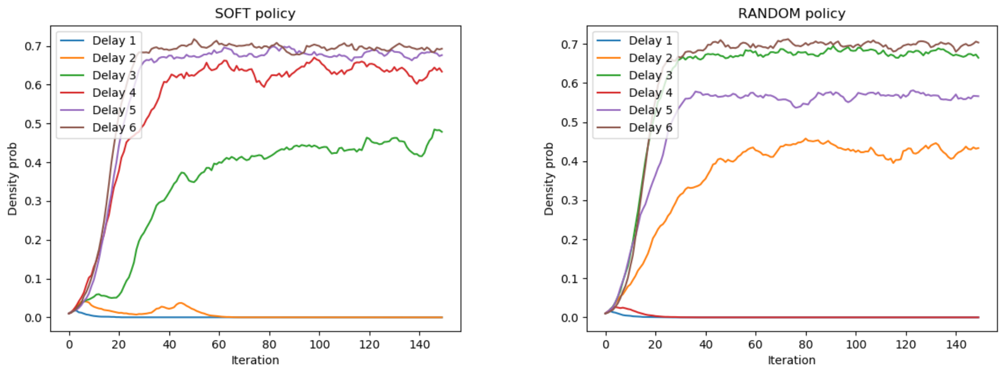


FIGURE 7.10: Average default density evolution for 30 simulations given different values of the delay for random and soft policies

7.2.2 Default cascades study

In this section, the random policy is omitted as we did for the ER random network. Note that for the BA network we have seen a significant difference between both underlying policies for $\beta = 0.4$ but not for $\beta = 0.6$. For this reason, as well as because the soft policy is more realistic we decided to exclude the random policy from the default cascades study.

Lets examine how the failure cascades behave in this structured hierarchical network. For random starting defaulted companies it is expected that the presence of big companies (Hubs) will help reduce the cascade sizes significantly.

$\beta = 0.4$		
Delay	$\mathbb{E}(Cs)$	$\max(Cs)$
2	1.11	9
3	2.35	47
4	12.89	627
5	36.57	902
6	52.79	939
7	80.97	942
8	94.97	925
9	96.68	961
∞	73.16	821
$\beta = 0.6$		
Delay	$\mathbb{E}(Cs)$	$\max(Cs)$
1	0.76	9
2	4.03	131
3	66.50	970
4	91.53	964
5	98.10	977
6	94.75	971
∞	79.06	746

TABLE 7.8: Expected and maximum cascade size, $\mathbb{E}(Cs), \max(Cs)$, for different values of the delay for the BA network with $\mu = 0.2$.

As expected, see Table 7.8, the expected cascade size is significantly smaller than

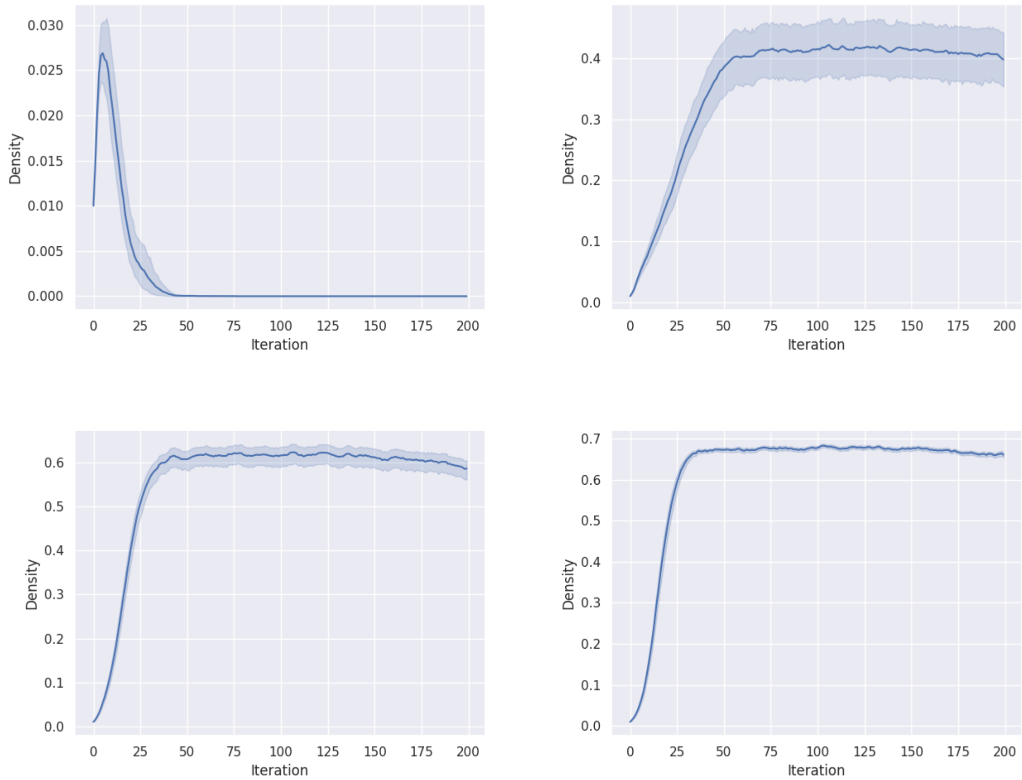


FIGURE 7.11: 100 simulation the mean of the density with the IC of the mean for different values of the delay(2 to 5) from right to left and for the policy soft

for the ER network. As it should, the expected cascade size, $\mathbb{E}(Cs)$, and the maximum cascade size, $\max(Cs)$, increase with the delay until they stabilize. However, the same behaviour for delay ∞ (i.e. no re-connection) is obtained. Meaning, there must exist a certain value of the delay for which both values start going down converging towards the no re-connection case.
A deeper comparison between both model networks, ER and BA, will be done in 7.3. To see the actual inverse cumulative distributions of the cascade sizes for each value of β see Figure 7.12.

7.3 Comparison between ER and BA networks.

First, let us compare the results obtained in Tables 7.2, 7.3, 7.6, and 7.7. Since we have seen that even for a certain high delay there are some repetitions that still differ,they end up in a different equilibrium state, we will compute the amount of repetitions that achieve each equilibrium point, without edge rewiring, for a recovery rate of $\mu = 0.2$. That would be like having an infinite delay.

From Table 7.9 we can see that the upper equilibrium is not always reached for a ratio of 2 ($H_0 = \frac{\beta}{\mu} = 2$). This is a behaviour intrinsic to the stochastic approach, as in the continuous default probability case this would not happen. Remember the result for delay 0 seen when describing the stochastic approach. (Figure 5.2).

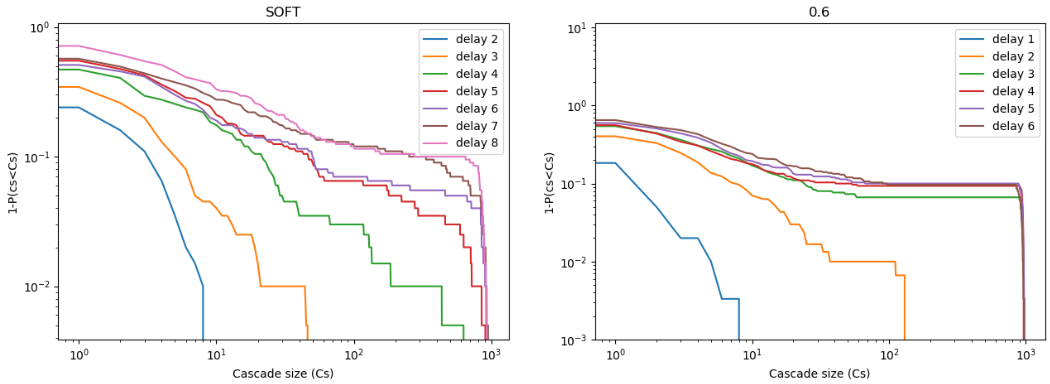


FIGURE 7.12: Inverse cumulative distribution of the cascade size(Cs) for several different delays, soft policy and for $H_0 = 2$ (left) and $H_0 = 3 =$ (right).

β	0.4	0.6
ER	(0.04,0.96)	(0.00,1.00)
BA	(0.03,0.97)	(0.00,1.00)

TABLE 7.9: Amount of repetitions that achieve the lower and upper equilibrium respectively without edge rewiring for some values of β for each model network.

The main differences we observe between the ER and the BA model results are on the intermediate delays. For $\beta = 0.4$, delays $5, 6, 7$ show a significant difference, with the ER network achieving the upper equilibrium in a much higher number of occasions, or what is the same, a BA network is much more resistant to default propagation for intermediate delays. This is hint number one that a Liberal Market network, with a clear hierarchical structure and companies with huge importance (i.e. the Hubs), is much more resistant to default propagation than a Sharing Economy Market network.

Now, let us shift focus to the failure cascades observed in both networks. Look back at Figures 7.6,7.12, as well as Tables 7.4,7.8. The first thing that is clearly observed is that the expected cascade size, $\mathbb{E}(Cs)$, is much larger for the random ER network than the BA one. This happens not just for the intermediate delays but for all of them. The maximum cascade size is similar for both networks, yet for small delays, meaning critical delay downwards, the random ER network has failure cascades that reach a significant larger size. This is the second hint/proof that a network following a Liberal Market approach is more resistant towards default propagation than a network following a Sharing Economy Market approach. The presence of Hub structures (larger companies with high degree) act as barriers to default propagation saving the system from large cascade failures in many cases.

Of course, if one of this Hubs is defaulted, then, the network will most certainly face a huge failure cascade. Meaning, Hubs are two way agents, they shield the network from default propagation when healthy but they act as their doom when defaulted. To see this behaviour more clearly let us perform a risk analysis using some of the risks defined in 6.3. Let us focus in the Individual risk, R_i, for both networks. The analysed nodes are:

- The node with the highest degree and the node with the highest betweenness centrality.[3] **Betweenness centrality** measures how much nodes need to pass through you to reach the rest. It is a measure of the importance of your location inside the network.The chosen nodes where 0, 7 for BA and 871, 899 for ER.

- A node in BA with the same degree (or similar) to the node with highest degree in ER, that is node 70.

See Table 7.10 for the results.

$\beta = 0.4$				
node_id	Network	delay	$\mathbb{E}(Cs)$	$\max(Cs)$
0	BA	5	137.66	678
7	BA	5	106.23	644
70	BA	5	19.47	468
0	BA	6	251.98	725
7	BA	6	269.83	645
70	BA	6	105.55	600
871	ER	5	383.98	982
899	ER	5	671.15	981
871	ER	6	386.40	987
899	ERs	6	514.87	987
$\beta = 0.6$				
node_id	Network	delay	$\mathbb{E}(Cs)$	$\max(Cs)$
0	BA	3	637.06	935
7	BA	3	658.21	923
70	BA	3	268.67	932
871	ER	3	608.06	994
899	ER	3	789.31	996

TABLE 7.10: Expected and maximum cascade size, $\mathbb{E}(Cs), \max(Cs)$, for different values of the delay for the BA,ER networks with $\mu = 0.2$ when the only node that starts at default is node_id.

In the end, it turns out that the structured hierarchical network, BA, out-performs the random ER model network even when the highest degree or the highest betweenness node defaults for each network.

From Table 7.10 we can immediately deduce that the biggest difference is on the maximum size that the cascades reach. For a random network, once default propagates enough, it reaches almost the whole network. On the other hand, for the hierarchical network with Hubs, it seems like the other big companies are able to contain the default from reaching the rest of the network. Moreover, this effect only happens when the reproductive number/ratio is small enough. As we can see from the $\beta = 0.6$ table, no difference between the networks when this aforementioned nodes default is easily spotted. Meaning, when default spreads fast enough not even the Hubs are able to contain it.

A different approach to look at this results that does show that the hierarchical network is more dependent on the biggest company than the random one is to compare this result with the ones obtained in Tables 7.4,7.8. For delays 5, 6, $\beta = 0.4$, the ER model doubles it's expected cascade size when defaulting the highest degree

[3]When this node is the same, as happens for the BA network, we simply select the second node with highest betweenness

node, while the BA network more than triples the expected size for delay 5 and becomes five times more for delay 6, all of that while keeping the maximum cascade size to $\frac{2}{3}$ of the previous maximum. Thus, the effect of defaulting the highest degree node for each network is relatively much higher for the BA network which goes more in the line of the expected.

There where mainly two more results that were truly expected. First, when comparing the biggest company in the random ER network to a similar size company in the hierarchical structured BA network, the structured network always behaves much better in containing the default propagation. This difference is also observed for the higher ratio. Second, in a network were all the companies are roughly equivalent, such as ER, the company with the most central position in the network, node 899, has a stronger importance in the default propagation than the biggest company, node 871.

7.4 BBVA Simulated Client-Supplier Network

In this section we will repeat the analysis we did for the random ER network and the hierarchical structured BA network with the BBVA simulated client-supplier network. We will only consider the underlying soft policy since it is the more realistic scenario of the two. Moreover, we have seen that the random policy behaves pretty similar to the soft policy in most of the cases.

Let us begin by checking the behaviour of the overall default density for different values of the delay and β, as it is seen in Figure 7.13.
In this section we will not perform a deeper study of the critical delay and the probability of reaching each equilibrium state given a ratio and a delay. Therefore, we opted to increase the number of simulations to produce Figure 7.13 up to 100.

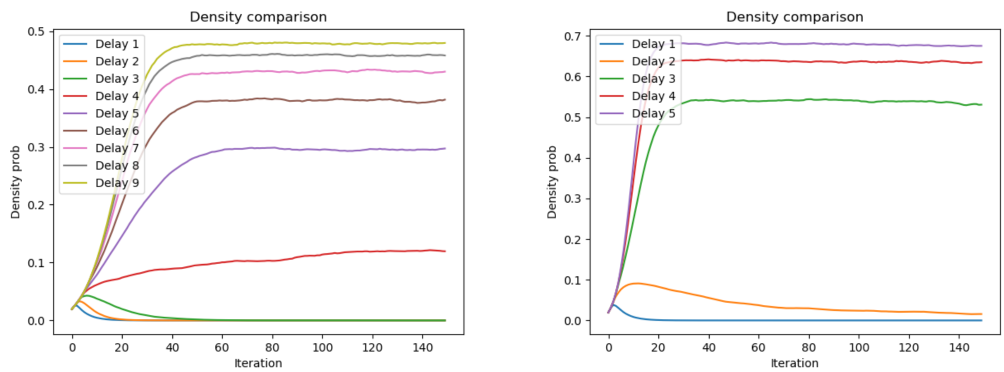


FIGURE 7.13: Average default density comparison for different values of the delay for the soft policy with $\mu = 0.2$ and $\beta = 0.4$ (left) , $\beta = 0.6$ (right).

- First scenario, $\beta = 0.4$. For the first three delays all simulations end at the fully recovered state. The critical delay is 4, and as expected, for all the successive delays we see an increase in the average default density, meaning, more simulations reach the upper equilibrium. One can also see how the curves tend towards the upper-equilibrium but they still do not reach it. The overall behaviour is reminiscent of the ER and BA network with some differences.

For instance, default propagates faster as it reaches the upper-equilibrium for lower delays.

- Second scenario, $\beta = 0.6$. For the first two delays the re-connection is able to stop the default contagion and both curves tend to the fully recovered state. Notice how for delay 2 the curve is still trending towards that fully recovered state, meaning, the re-connections are just barely able to overcome the default speed. Therefore, for the next delay we see a huge jump towards the upper-equilibrium since now most of the simulations will end there. Notice that for delay 5 we seem to be slightly above the expected upper-equilibrium.

Once we know how the overall default density behaves for each value of the delay, it is time to see if each financial sector behaves differently in this simulated client-supplier network. Figure 7.14 shows that there is no clear difference between the sectors. The only sector that behaves differently is Financial Institutions, however, this has more to do with the small size it has in the simulated network than anything else. The main reason is, as we discussed in the hierarchical structured BA network, that just knowing the degrees and/or the sectors of the Hubs provides no clear sectorial differentiation. The root of this differentiation runs deeper in the network structure with clustering being a major part of the equation. That is, knowing how each company is tightly connected to others forming communities (or smaller networks) inside the real big network that are only connected to the rest by few nodes, most probably by a big Hub node that shields/protects the entire structure from collapsing when default occurs outside it but that quickly collapses if default occurs inside it.

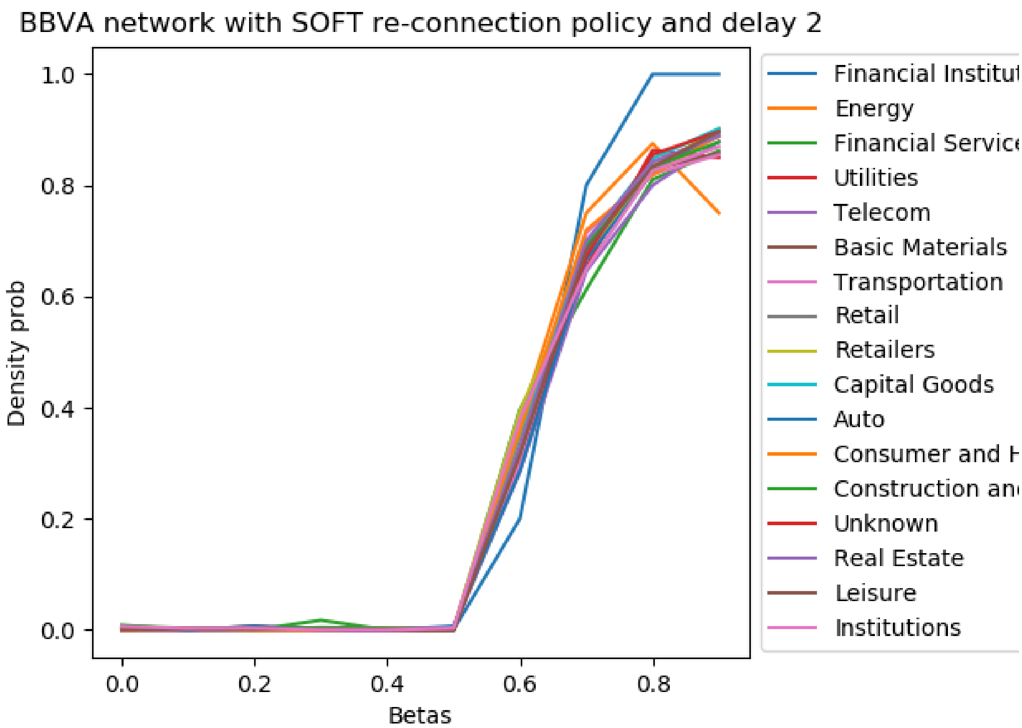


FIGURE 7.14: Average sectorial default density at equilibrium given a recovery rate $\mu = 0.1$.

7.4.1 Default Cascades

Let us analyse the failure cascades in our simulated BBVA network. Figure 7.15, as well as table 7.11, show the obtained results. The obvious results, of the expected cascade size, $\mathbb{E}(Cs)$, and the maximum cascade size, $\max(Cs)$, increasing as delay increases are obtained. Moreover, the trend that failure cascades behave better for the non edge rewiring case than for high delays continues.

Notice that, as well as it happened for the overall average default density, this behaviour is more reminiscent of the previously simulated model networks than it is of a real client-supplier network, where it is pretty difficult to obtain cascades of such a large size. In fact, this is one of several proofs that our simulated network is far from reality, since the clustering of the businesses and their intrinsic relation with one another is far more systematic and non-random.

Another curious result is the one obtained for $\beta = 0.6$, where we see that for delay 2 the probability of having a smaller cascade is actually less than for the bigger delays.

The most amazing result of all of them is observed when comparing this failure cascades with the ones obtained in the ER and BA model networks. The expected size is closely the same as for the structured BA network but the overall size of the network is 10 times larger. Thus, relatively speaking, the resistance to default contagion is 10 times higher in this network than in the BA one. Of course, it is also better than the ER network.

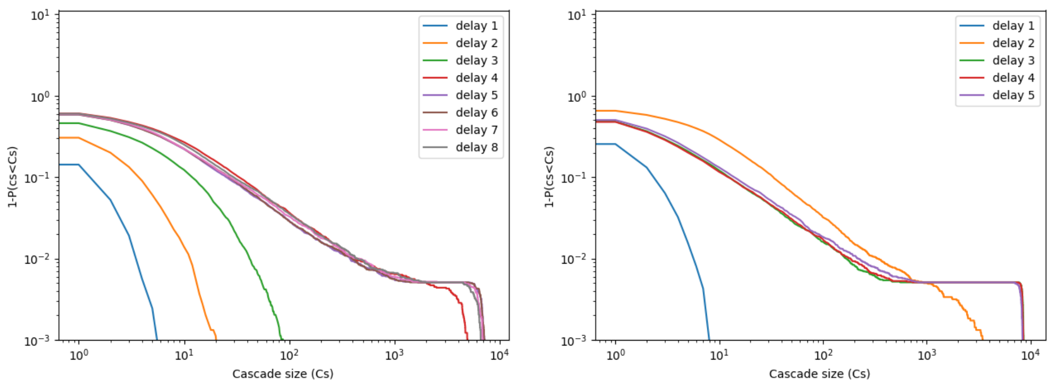


FIGURE 7.15: Inverse cumulative distribution of the cascade sizes(Cs) for different values of the delay for $\beta = 0.4$(left) and $\beta = 0.6$ (right).

7.4.2 Risk assessment

The original purpose of this section was to analyze which financial sector posted a higher risk for our network. However, as seen in Figure 7.14 , our simulated network presents almost no sectorial differentiation. Therefore, instead of analyzing the sectorial risk (R_s) we will simply look at the individual risk of some nodes as we've done in section 7.3. The considered node will be:

- Node 1: Is the node with the highest degree and highest betweenness centrality on the network.

$\beta = 0.4$		
Delay	$\mathbb{E}(Cs)$	$\max(Cs)$
1	0.60	9
2	1.45	41
3	4.45	194
4	38.62.	6290
5	47.74	7601
6	48.13	7706
7	48.01	7528
8	47.71	7401
∞	39.50	2378
$\beta = 0.6$		
Delay	$\mathbb{E}(Cs)$	$\max(Cs)$
1	1.02	13
2	27.80	5178
3	49.51	8723
4	49.55	8799
5	49.46	8550
∞	34.63	1318

TABLE 7.11: Expected and maximum cascade size, $\mathbb{E}(Cs), \max(Cs)$, for different values of the delay for the BBVA simulated network with $\mu = 0.2$.

node_id	β	delay	$\mathbb{E}(Cs)$	$\max(Cs)$
1	0.4	5	1846.55	8847
1	0.4	6	3877.82	9243
1	0.6	3	6769.40	9777

TABLE 7.12: Expected and maximum cascade size, $\mathbb{E}(Cs), \max(Cs)$, for different values of the delay, β for the BBVA simulated network with $\mu = 0.2$.

From 7.12 one can see some very interesting results. When the biggest node defaults, the expected cascade size, $\mathbb{E}(Cs)$, becomes around 30 times larger for delay 5 and 60 times larger for delay 6 with $\beta = 0.4$. For $\beta = 0.6$, it becomes around 150 times larger for delay 3. Moreover, in all cases the maximum cascade size also becomes larger. The relative increase is huge and is what we expected from a hierarchical structured network such as BA.

This massive difference has his root in the failure cascades behaviour in the original state. If we look at the behaviour just when the highest degree node defaults, we can see that actually the BBVA simulated network behaves poorly, with a higher percentage of nodes defaulting and a higher cascade size. The difference, as we stated, is in the original state. For our simulated network the nodes with highest degree are all non defaulted. In fact, the two sectors with highest average degree, Financial Institutions and Services, are all fully healthy. The first sector with one defaulted node is Institutions which is the third in importance. This might mean that there are too many bigger healthy nodes that act as barriers for the default to expand through the network containing most of the failure cascades at lower values.

Chapter 8

Discussion

The results line up with our hypothesis that the liberal market network will be much more resilient to large cascades or even entire collapse of the system than the shared economy network. The reason for this resilience is mainly due to the presence of Hubs, large companies with high out-degree (many suppliers), which act as barriers to the default propagation. This Hubs can shield the default from expanding from one part of the network to another reducing the expected and maximum cascade size. Of course, a Hub also tends to have a high in-degree (many clients) which means that if the Hub becomes defaulted, the default may expand quickly through the network, creating large cascades and thus, posting a high risk for the overall integrity of the network.

However, the simulated BBVA network does not follow our expectations. Figure 7.7 shows that the Hub structure just by itself does not prompt any sectorial differentiation. We expect that this is due to the extreme simplification we applied in the generation of the imitation network. Most likely, in the BBVA network there are complex structures present like reinforced loops and clustering, which will not occur (frequently) in the simulated network. The combination of this complex structures with the presence of Hubs makes the expansion of the default less homogeneous and yields the expected sectorial differentiation. In future research, regarding remodelling transaction data it is recommended to also gather the following statistics.

- Standard deviation of incoming and outgoing connections per sector
- Number of connections between each of the sectors
- Average weight of connections to each sector
- Clustering coefficient per sector

At the moment, there are no methods to incorporate all the statics in a fully satisfiable way, since some statistics interfere with other statistics. Nonetheless, more statistics can be incorporated, leading to more representative results. Of course, using real networks based on actual transaction data would deliver more interesting results.

We discovered that for a stochastic probability a new feature had to be introduced to make the algorithm work, called default delay. In essence the edge rewiring needs to be retained for some iterations, otherwise the whole network heals itself almost instantly. Meaning, the time a company waits to change suppliers when one of them enters default can not be instantaneous. For this reason the delay parameter has been introduced. Adding an extra parameter is not the only way, nor necessarily the best way, of circumventing the problem, since this parameter makes the computation more complex. For future research, we suggest to experiment with an apparent stochastic approach. This approach does not set the default probabilities to 0 or 1 on the node itself, but instead keeps the continuous probabilities. However, when nodes request the probabilities from other nodes, they get a 0 or 1 default probability, based on the default probability of that node and the applied policy. In this way, the system is still stochastic, but avoids the instant healing of the network without the necessity of adding another parameter.

Nevertheless, the addition of the delay parameter has yielded mainly one surprising result. Failure cascades behave better when there is no edge rewiring at all than for medium-high delays. This means that hesitation is the worst possible policy when dealing with a supplier entering default. The explanation for this behaviour is that re-connection to new suppliers allows two separate parts of the network to become connected, in some cases, avoiding the protection of Hubs. If these re-connections do not stop the default from propagating then they are worse than not having them at all.

As for our objective of risk assessment, the results are quite promising. We have seen that the behaviour of financial client-supplier networks is easily explainable and, to some extent, verifiable. From the results in Section 7.3 we can see how the simulations can be used to inspect systemic risk of individual companies. The results show that in liberal market networks, if we default one large node, the expected and maximum cascade sizes increase significantly, posing enormous risk for the system. This result compares well with reality. Since we could not test the sectorial risk algorithm on the actual BBVA data, further research will need to be done in order to perfect it for risk assessment of actual companies. In order to test the network on real life situations, known cases of a defaulted large company clearly affecting the economic stability could be used. Simulations could be run initially defaulting that large company and experimenting with setting different parameters. In this way the functioning of the algorithm can be compared with reality, concluding in more insights for usage of the algorithm in risk assessment in real scenarios.

The algorithm is very computationally heavy, even using a parallelized approach on a machine with 64 2.3GHz single-core processors. Therefore, a different way of calculating default probabilities could be created, so that is possible to calculate all the updated default probabilities in parallel. This system can then be ran on a GPU, making it run much faster. Another option, would be to rewrite the entire system so that it is modelled just using probability distributions. This would enable the use of Markov Chain Monte Carlo simulations, making the simulations much faster. However, this would most likely require an even further simplification of reality. Faster algorithms are desired if analysis has to be done on many companies in the network.

Chapter 9

Conclusion

The problem of assessing systemic risk in a sectorial interconnected client-supplier network is an interesting and complex task. This risk is commonly associated with the size a failure cascades can and may reach in the network. In this project we have used the bases of an epidemiology model such as SIS (Susceptible-Infected-Susceptible) to approach default dynamics in two model networks, BA and ER, that represent different economical markets, Liberal and Sharing Economy respectively, and in a simulated network that tries to approach the real Spanish economical market.

Let us do an overview of the main conclusions that have been extracted during this project.

- In a stochastic setting the agent-based models need to be redefined to allow the infection to spread. In essence the edge rewiring needs to be retained for some iterations, otherwise the whole network heals itself almost instantly. Meaning, the time a company waits to change suppliers when one of them enters default can not be instantaneous. For this reason the delay parameter has been introduced. Alternatively, an agent that based itself on weight transfer could also exist.

- The hypothesis that the Liberal Markets perform better than the Sharing Economy Markets in terms of default dynamics is supported by the models. This is highlighted by the fact that failure cascades achieve greater expected values and larger maximum sizes in the ER than in the BA model network for the same conditions. Moreover, despite being tightly dependent on their biggest company, the BA network shows that even when this lead company enters default the overall failure cascades remains larger in the ER network given the same circumstance.

  In other words, the power that large companies (Hubs) have over smaller companies is highlighted, showing that companies with great power present a higher resistance towards entering default. Moreover, this big companies shield a certain part of the network from defaulting. On the other hand, in a random model with no clear lead company, such as in the ER model network, once default starts spreading there are no intrinsic barriers that protect parts of the network from defaulting giving room for the development of huge cascades.

- The hypothesis that the financial sectors within a client-supplier network behave differently in terms of default dynamics was left unproved. Using the metadata from the BBVA bank a network was created that approached the given data. However, the sectors in the simulated network were still far from

approaching reality. This prompted another conclusion an that is how difficult it is to simulate a real complex networks given all the variables that are in place, number of connections, clustering, degree distribution,...
In fact, even if we had all that information of the original network generating a simulated copy that matched all those variables would be almost unfeasible as it is still an open problem in Complex Networks.

In the end our hope is that the acquired knowledge in systemic risk analysis will be useful for the BBVA bank, associated companies and for future research on the subject.

Appendix A

Random policy plots

A.1 ER plots

Reproduction of the plots made for the BA for soft policy ,but for random policy and $\beta = 0.4$ and $\beta = 0.6$: As it can be observed from the plots A.1 we obtain more or

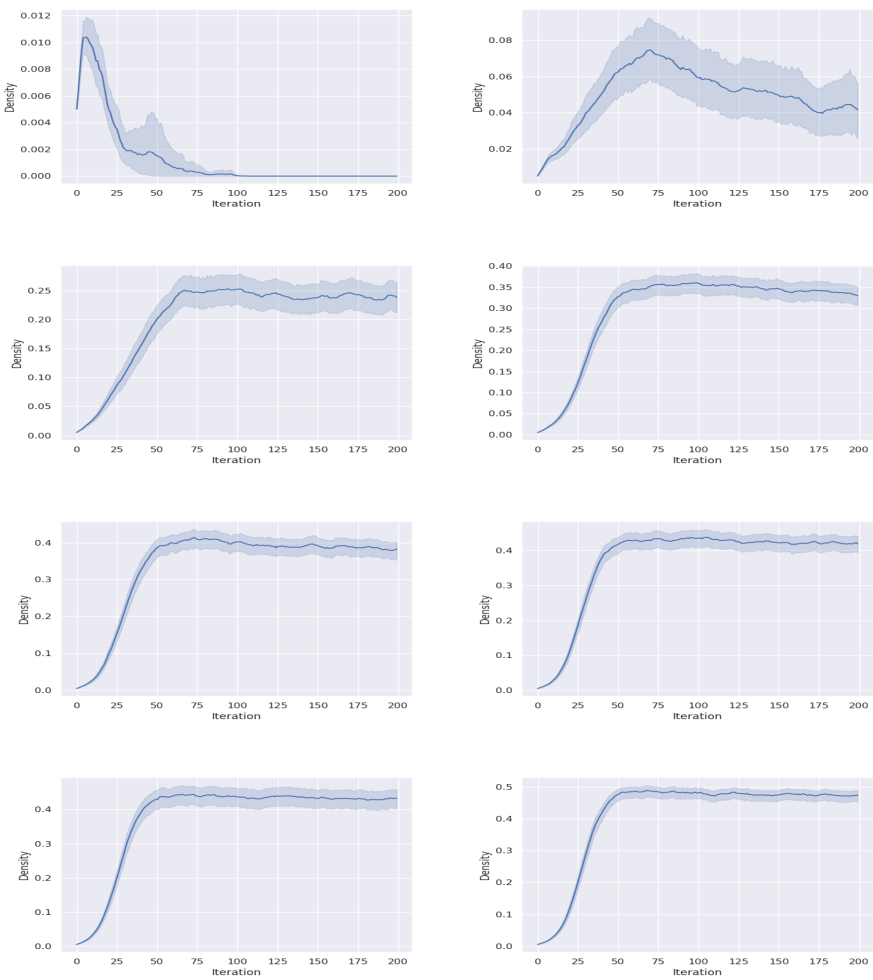


FIGURE A.1: 100 simulation the mean of the density with the IC of the mean for different values of the delay(3 to 10) from right to left and for the policy RANDOM

less same results as we obtain for soft policy. As it can be observed from the plots A.1 we obtain more or less same results as we obtain for soft policy.

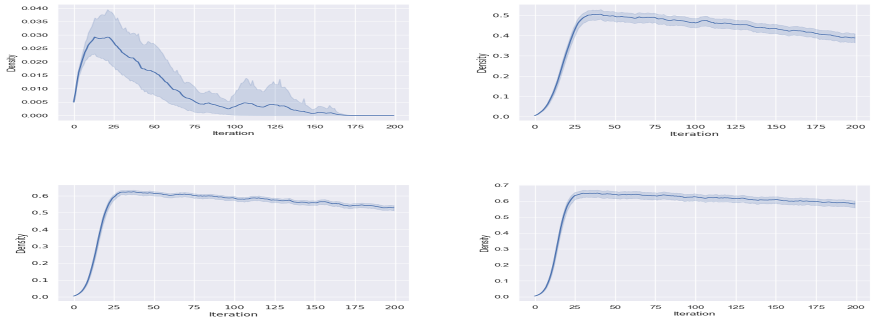


FIGURE A.2: 100 simulation the mean of the density with the IC of the mean for different values of the delay(2 to 5) from left to right and for the policy RANDOM

A.2 BA plots.

Reproduction of the plots made for the BA for soft policy ,but for random policy:

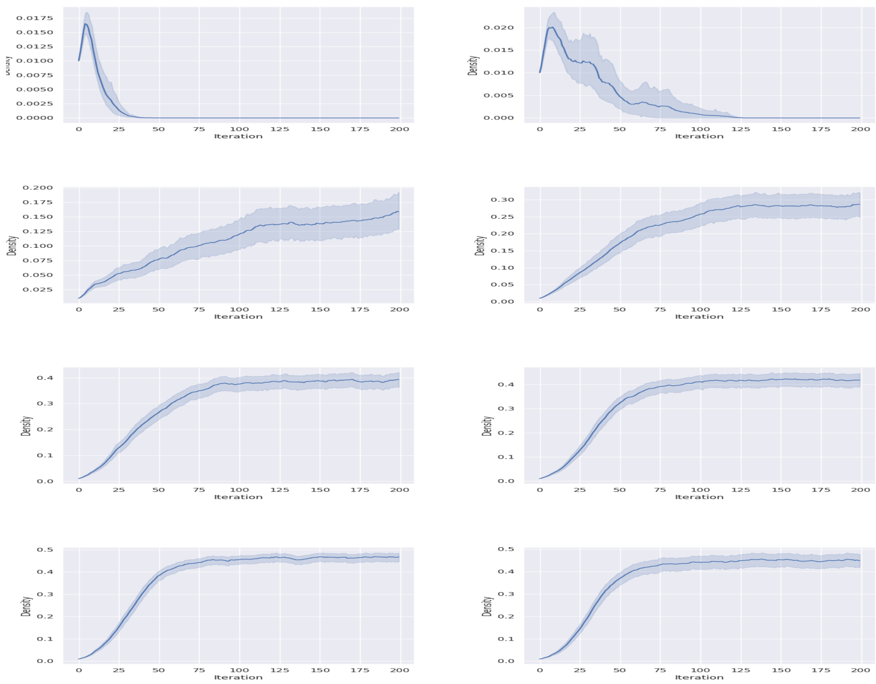


FIGURE A.3: 100 simulation the mean of the density with the IC of the mean for different values of the delay(3 to 10) from right to left and for the policy RANDOM

As it can be observed from the plots A.1 we obtain more or less same results as we obtain for soft policy.

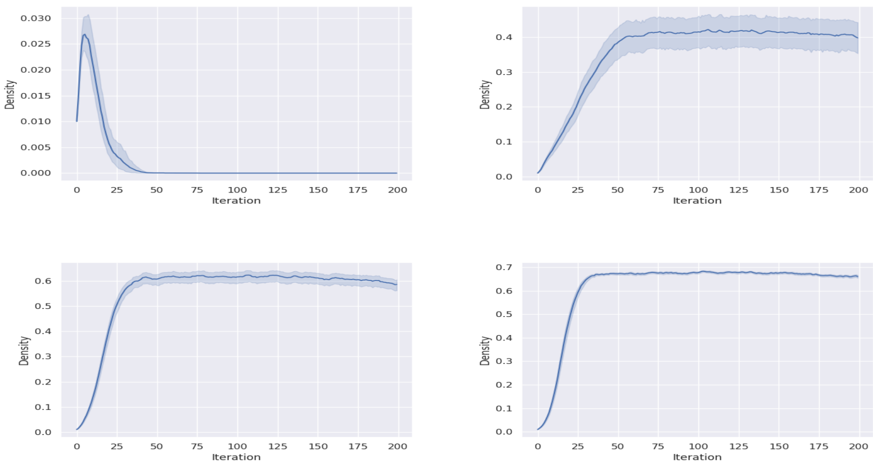


FIGURE A.4: 100 simulation the mean of the density with the IC of the mean for different values of the delay(2 to 5) from left to right and for the policy RANDOM

Appendix B

Team member contribution

In this section we describe the contribution of each member to the overall project. We will divide it in two groups, code and report.

B.1 Code

The code produced during the development of the project can be found in `https://github.com/bbva-rfn/master-project`. Also, a high level overview of the whole project can be found there. We will only include the file name of the main contributions.
The contributions for the code have been:

- Philippe Van Ameronggen
  - Main algorithm that simulates the SIS behaviour with the different agents and policies described. (SecNet.py)
  - BBVA simulated network generation. (BBVA/gen.py)
  - Parallelization of all the functions to use the multiple cores of a machine.
  - High level overview
  - Agent-based model implementations
- Ramon Mir Mora
  - Functions that use the main algorithm to compute some the results in 7. Mainly, all the failure cascades/risk code (cascades.py), the average default density evolution compared over multiple delays (sis_delay_comparison.py) and the sectorial behaviour (sectorial_density_functions.py and the sectorial part in plots_sis.py).
  - ER and BA model networks generation. (gen.py in each folder)
  - Agent-based model implementations
- Sergi Sánchez de la Blanca Contreras
  - Code to perform the Monte-Carlo simulations (replicate_density.py and the rest of plots_sis.py)
  - Simulation execution
  - Agent-based model implementations
  - Refining and cleaning code

B.2 Report

Here we stipulate the contributions of each member to the present report.

- Philippe Van Ameronggen
  - Narrative parts (Abstract, Introduction, Study Design, Conclusions, Discussions)
- Ramon Mir Mora
  - Theoretical parts (Complex Networks, SIS model,Failure Cascades)
  - Results analysis (Cascades, risks and ER-BA comparison)
  - Conclusions, Discussions and Appendix B
- Sergi Sánchez de la Blanca Contreras
  - Database/previous work explanation/ Study Design/ Conclusions/ Discussions
  - Results, execution & analysis (Study of critical delay ER/BA)
  - Appendix A

Bibliography

Adrià Barja Alejandro Martínez, Alex Arenas Pablo Fleurquin Jordi Nin José J. Ramasco Elena Tomás (2019). "Assessing the Propagation of Default in Interconnected Sectorial Financial Networks." In: *IEEE COMPUTATIONAL INTELLIGENCE MAG.*

Barabási, Albert-László (2016). "Network Science". In: URL: http://networksciencebook.com/.

Gilbert, E.N. (1959). "Random Graphs". In: *Ann. Math. Statist.* 30.4, pp. 1141–1144. URL: https://projecteuclid.org/euclid.aoms/1177706098.

P.Erdós, A.Rényi (1959). "On random graphs, I." In: *Publicationes Mathematicae (Debrecen)* 6, pp. 290–297.

— (1960). "On the evolution of random graphs". In: *Publ. Math. Inst. Hung. Acad. Sci.* 5, pp. 17–61.

— (1961). "On the evolution of random graphs". In: *Bull. Inst. Internat. Statist.* 38, pp. 343–347.