Faculty of Sciences:
Department of Physics and Astronomy

Academic Year 2011–2012

Network Analysis of the Russian Interbank System

Benjamin Vandermarliere

Promotor: Prof. dr. J. Ryckebusch
Promotor: Prof. dr. K. Schoors

Scriptie voorgedragen tot het behalen van de graad van
Master of Science in Physics and Astronomy

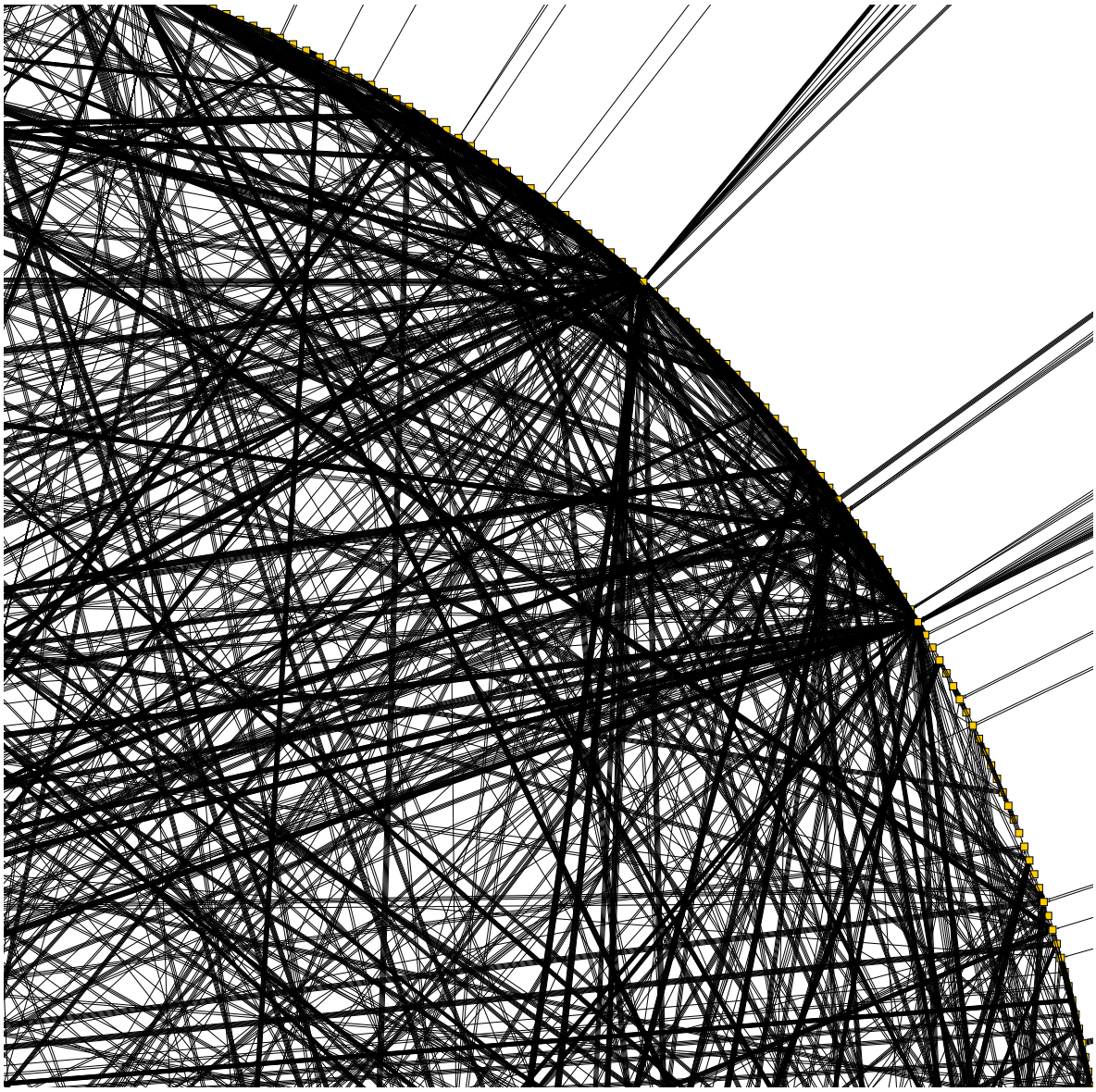


Figure 1: The Russian interbank planet. The yellow squares represent the banks and the black lines connecting them represent their interbank loans. The weight of the lines is proportional with the value of the loans. This figure shows the activity of March 2002.

Dankwoord

Het indienen van je thesis is het eindpunt van een twintigjarige opleiding. Wat ooit begon met vingerverven en in de zandbak spelen, culmineerde bij mij dit jaar met het bestuderen van de Russische interbankenmarkt. Het kan verkeren. In die twintig jaar (en de drie ervoor natuurlijk) heb ik altijd het geluk gehad om omringd te zijn door boeiende en inspirerende mensen en heb ik alle kansen gekregen om mij ten volle te ontplooien. Maar laten we niet te melig worden!

Vooreerst, bedankt Prof. Ryckebusch dat ik dit thesisonderwerp onder handen mocht nemen en daarbij de nodige vrijheid kreeg. Ik bedank je voor de vele tijd die je stak in het kritisch nalezen van mijn tekst en voor de winter school die ik op jouw kosten mocht bijwonen. Ook wil ik graag Prof. Schoors en dr. Karas bedanken voor het ter beschikking stellen van hun databank en voor hun bezoek in november. Dr. Karas maakte mij ook wegwijs in de data en bezorgde mij een opgekuiste versie die mij het leven heel wat gemakkelijker maakte. Verder stond Maarten altijd paraat om mijn computerproblemen op te lossen en was Jonathan mijn netwerkpartner in crime. Gelukkig waren Tom, Camille, Willem, Sam, Mathias, Kelly, Lesley, Sander, Vishvas, Mario (niet-exhaustieve lijst) er om de saaiheid van het INW wat te doorbreken. Ook de 'mannen van de Fysicasa' en mijn appartementsgenoot Kevin mogen niet vergeten worden. En dan zijn er nog Sarah en Pascal die zich vaak ontfermden over deze ontheemde.

Natuurlijk mogen de vrienden en familie in Poperinge (nog een West-Vlaming die afstudeert aan de UGent) niet vergeten worden. Ik bedank mijn twee zussen voor het geven van het goeie voorbeeld, Nom voor het stoom aflatende boksen, en Ban voor het entertainen met de glimlach. En ik bedank mijn ouders voor de zeer goede zorgen, ik heb een luizenleventje gehad, en de vele kansen die ik kreeg. Maar waar ik het meest geluk mee heb gehad deze laatste 23 jaar, is dat ik mijn vriendin Ellen heb mogen ontmoeten.

Benjamin

PS: Dit werk werd mede mogelijk gemaakt door: de Vlaamse Gemeenschap, S. Brin en L. Page, en de ingenieurs bij DELL.

Contents

The whole is more than the sum of the parts.

Metaphysica
Aristotle

Chapter 1

Introduction

1.1 Rethinking the Financial System

The 2007-2012 financial crisis has made it obvious that the current financial system is fragile and unstable. The problems with the subprime exposure, which were believed to be too small to lead to widespread problems, spread throughout the system like a contagious disease and made it topple. Regulators did not fully control the situation and their policies did not result in a stable and robust financial system. One could say that systemic risk[1] and financial contagion are rather poorly understood, which results in an ineffective policy. In the words of Dirk Helbing in Nature [7]:

> They [the regulators] can do a good job of tracking the economy using the statistical measures of standard econometrics, as long as the influences on the economy are independent of each other and the past is a reliable guide to the future. But the recent financial collapse was a systemic meltdown, in which intertwined breakdowns in housing, banking and many other sectors conspired to destabilize the system as a whole. And the past has been anything but a reliable guide of late.

The economy is a complex dynamical system full of nonlinear feedback mechanisms and linkages that are latent and often unrecognized. Regulators do not sufficiently acknowledge this complexity and attempt to stabilize and protect the market by imposing rules on each individual player. Basel I and Basel II, e.g., dictate what each bank has to do in order to protect itself for harsher times. The adage is that regulation of each separate player can result in control over the system as a whole. Most often, however, "the whole is more than the sum of its parts" (paraphrasing Aristotle and P.W. Anderson). In effect, by reducing the system

[1] The Bank for International Settlements defines systemic risk as the risk that failure of a participant to meet its contractual obligations, may in turn cause other participants to default with a chain reaction leading to broader financial difficulties.

to its separate players without accounting for their mutual interdependences, the regulators loose sight of the systemic level. They approach a complex system in a reductionist fashion.

But learning to understand this complex system and its risks is hard. Especially because they do not have the tools they need to predict and prevent meltdowns. Now one could envision to develop computational 'wind tunnels' that would allow regulators to test policies before putting them into practice [7]. This would lead to a beter understanding of systemic risk and would allow policies on a systemic level which could create a more stable and robust economy. If it would ever be possible, it would give the policy-makers a way to turn the right knobs and tune the system in on stability.

1.2 Complex Adaptive Networks

In two papers - "Rethinking the financial network" [19] and "Systemic risk in banking ecosystems" [20] - Haldane et al. set out to tackle the problem of systemic risk in financial systems by using science's mightiest weapon - drawing analogies. Ref. [19] starts with drawing the analogy between the epidemic SARS outbreak and the collapse of Lehman brothers. In the words of Haldane, Executive Director for Financial Stability at the Bank of England:

> The similarities between the SARS outbreak and the failure of Lehman brothers are striking. An external event strikes. Fear grips the system which, in consequence, seizes. The resulting collateral damage is wide and deep. Yet the triggering event is, with hindsight, found to have been rather modest. The flap of a butterfly's wing in New York or Guandong [the location of the SARS outbreak] generates a hurricane for the world economy. The dynamics appear chaotic, mathematically and metaphorically.
>
> These similarities are no coincidence. Both events were manifestations of the behaviour under stress of a complex, adaptive network. Complex because these networks were a cat's-cradle of interconnections, financial and non-financial. Adaptive because behaviour in these networks was driven by interactions between optimising, but confused, agents. Seizures in the electricity grid, degradation of eco-systems, the spread of epidemics and the desintegration of the financial system - each is essentially a different branch of the same network family tree. [19]

Considering the financial system as a complex adaptive network, makes it possible to export insights and models from other network disciplines - such as ecology, epidemiology, biology and engineering - to the financial sphere. One can find common ground in the systemic risk in financial networks and in ecosystems. There is especially common ground in the need to identify the conditions that dispose a system to be knocked from seeming stability into another, less optimal state. After all, the main drive of the study of complex adaptive networks is to search for 'tipping points', 'thresholds and breakpoints','phase transitions', or

'phase shifts' - all terms that describe the flip of a complex dynamical system from one state to the other [30].

1.3 Networks

Complex adaptive systems are built upon networks. Hence, studying the behaviour of these complex systems led to the development of network theory. A network is a set of items, which one calls nodes, with connections between them, called edges [35]. Systems taking the form of networks abound the world. Examples include the internet, the World Wide Web, social networks of acquintances or other connections between individuals, neural networks, metabolic networks, food webs, distribution networks, networks of citations between papers, and many others [33]. Of course not all networks can be seen as complex adaptive systems. In this thesis we focus on financial networks and in particular the interbank lending network. In such an interbank network the nodes represent banks and the edges represent their interbank loans (cfr. Fig. 2.1).

The ultimate goal of network theorists is to understand the behaviour of complex adaptive systems. Before one can understand the processes taking place on networks, one needs to fully understand the underlying network structure. This is why network theory started out as a science that studied and characterized the topological structure of networks. Statistical measures were developed which enabled one to provide a concise description of a highly complicated bunch of nodes and connections (cfr. Fig. 1.1). Later on, network theory evolved into a science that studies the processes - disease spreading, cascade failures, foodweb dynamics - taking place on networks. These processes have common ground with contagion in the financial world. Network theory possesses the potential, by using insights and models from other network branches, to asses the systemic risk in the financial network.

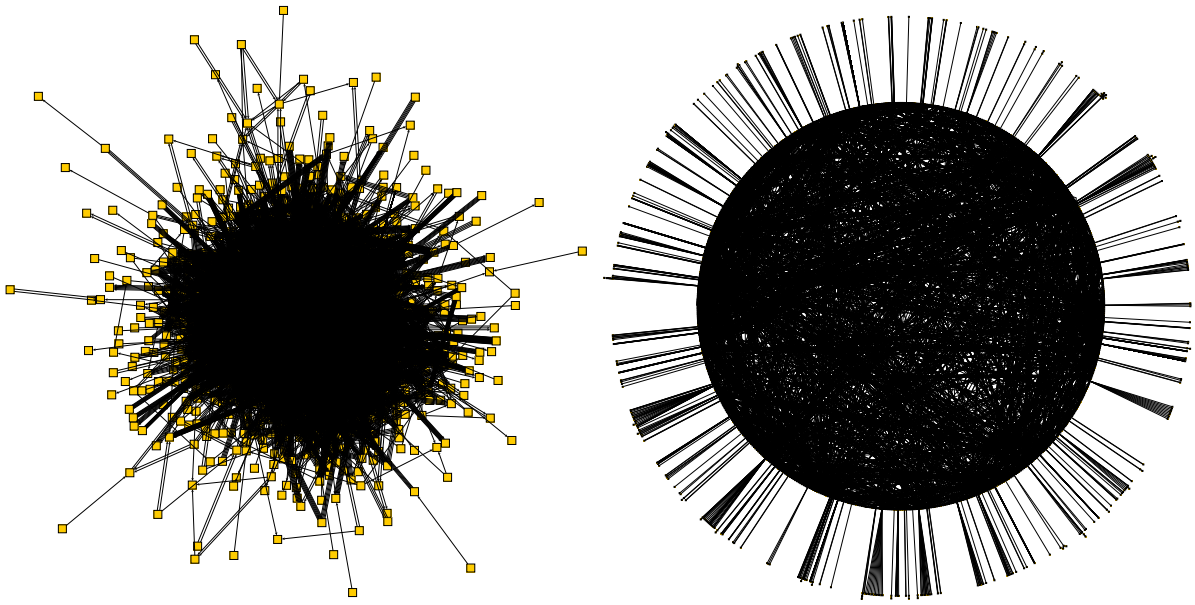
1.4 Outline of the Thesis

Before one can understand a complex adaptive network, one needs to know and understand the underlying topological network structure. It is necessary to perform an empirical characterization of the interbank network in order to obtain stylized facts - which can then be used in the development of theoretical models. These kind of analyses are increasingly used to map real world networks. Examples are food-web networks [38], social networks [41], and computer grids [9]. Empirical analyses of interbank networks have already been performed for the Austrian [6] and Brazilian [11] cases.

In this work an empirical analysis of the interbank network will be performed using real life bilateral and time-varying data on interbank exposures from Russia in the time period 1998-2004. As mentioned above, the most fascinating part of a complex adaptive system is its ability to undergo a phase transition. The fact that two major crises, which can be seen

as phase transitions, hit the Russian banking system in the time period 1998-2004, offers unparalleled opportunities. It enables us to study the network structure before, during, and after such a phase transition. We will pay particular attention to possible crisis-indicators among the network structure properties during the run-up to the crashes. Hence, the focus of this work will lie on these crises periods.

In the last chapter we will turn our attention to the concept of percolation, which is used in network theory to model spreading mechanisms. Especially its application in the modeling of cascade failures can prove to be useful for the understanding of systemic risk in the interbank network. In this work percolation theory will be used to study the resilience of the interbank network.



(a) Organic layout **(b)** Circular layout

Figure 1.1: A bunch of nodes and connections, the entire Russian interbank network in March 2002 represented using two different layouts. The yellow squares represent the banks and the black lines connecting them represent their interbank loans. The outer rim in the circular layout are banks which connect to one single other bank.

Chapter 2

The Russian Interbank Network

2.1 The Data

What kind of information does the database contain?

The data used in this master thesis were collected by Prof. K. Schoors and dr. A. Karas from a private information agency called Banksrate.ru [26]. It involves all interbank loans issued on the Russian market during a period ranging from August 1998 up to and including November 2004. These loans were reported monthly except for January 2003. Each entry or record in such a monthly report identifies the following features:

- Issuer of the loan
- Receiver
- Domestic or foreign loan
- Maturity[1] (life expectancy of a loan)
- Due Date
- Interest rate
- Begin of period balance: amount of money owed at the beginning of the month
- Debit turnover: amount of money issued during this month
- Credit turnover: amount of money payed back during this month
- End of period balance: amount of money still owed at the end of the month

[1] In this work, maturity is defined as the length of time between the commitment date and the final repayment date of a loan. One notes that maturity can also refer to the final payment date of a loan, or due date, in se.

A loan issued and fully repayed in one and the same month is recorded only once in the database. Such a record has the beginning and the end of period balances equal to zero and the debit turnover equal to the credit turnover. A loan issued in month 1 but repayed in month 2 will be recorded twice in the database. In order to avoid double counting of the loans, one only considers loans with a begin of period balance equal to zero and a positive debit turnover. The maturity of the loans is subdivided into loans with a one-day ($1d$) maturity, loans due in a week ($1 - 7d$), a month, three months, half a year, a year or even longer. In Fig. 2.2 one shows the number of loans per month with a maturity of $< 7d$. For January 2003, for which one lacks data, one uses an interpolated value. These short-term loans account for more than 80 percent of the transactions both in terms of the number and the volume. Thus, it seems reasonable to focus on these loans with a one-day and a one-week maturity. To give an idea of the amount of data: there are around 900 banks covered in the database which are responsible for on average 25000 interbank loans per month. Last but most importantly, we note that the data have already been made consistent (for example all the redundancies have been removed) by Prof. K. Schoors and collaborators. We would especially like to thank dr A. Karas here for doing this tedious task and also for explaining to us the features of this database.

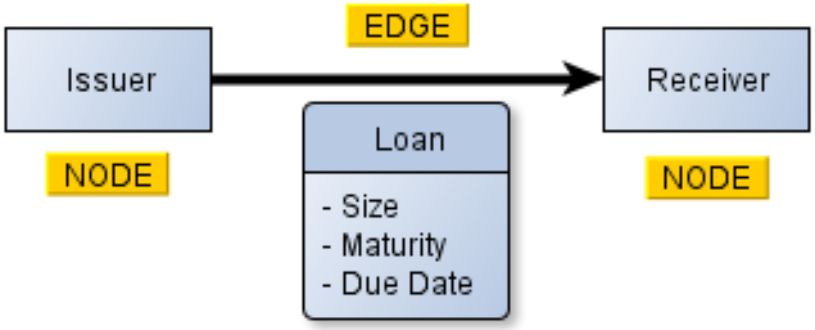


Figure 2.1: The basic components of an interbank network. A bank is represented by a node and an interbank loan by a directed edge. The three most important attributes of the loan are its size, maturity and due date.

The appeal of the data

Most current research on interbank markets, systemic risk, and cascade failures is purely theoretical [36; 5; 16; 17; 22; 15]. This research focusses on accurately replicating the contagion spreading channels but does not pay a lot of attention to the model input. Most of the assumptions made with regard to the structure of interbank networks are not based on solid empirical findings.

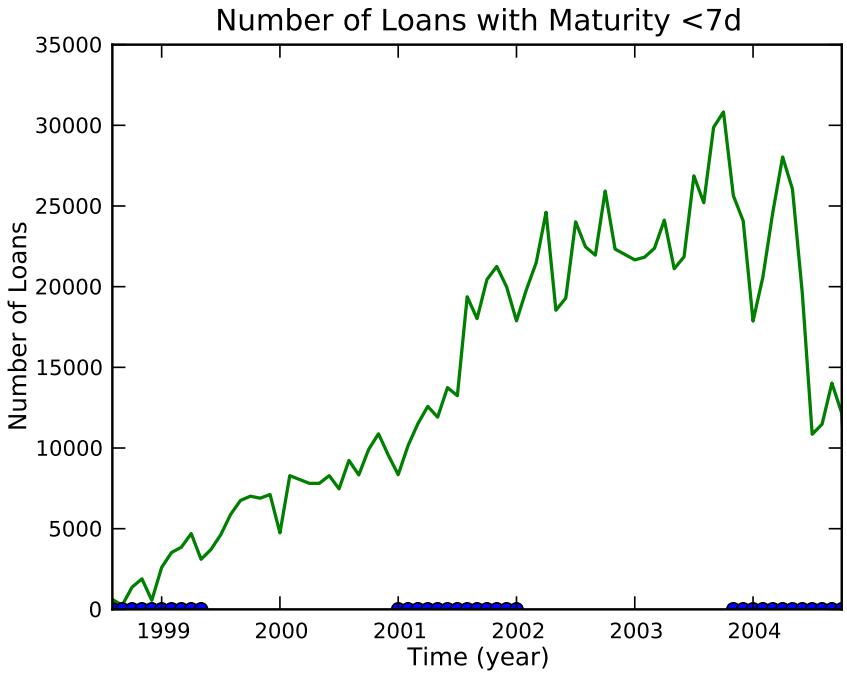


Figure 2.2: The monthly number of loans with a maturity of $< 7d$ for our entire data-period. The blue dots indicate our periods of interest. 2001 represents a normal period, whereas the other two indicated regions are the crisis periods.

The study of Prof. K. Schoors and collaborators adopts a different point of view. They developed a hybrid model which is described in a BOFIT discussion paper of 2008 [28]. It implemented a cascade or *loss, given default* model but they did start from specific empirical data as input. This was a novel approach in the research of interbank markets which was conducted along the lines of the investigations in other economic networks, like the global [8] and the European [13] international debt networks.

The theoretical research of Refs. [36; 5; 16; 17; 22; 15] can test different financial contagion scenarios, study the impact of model parameters and draw general conclusions. These studies however, do not have any real world feedback. In contrast, the study of Ref. [28] departs from real world data. But this makes it rigid for testing different interbank network structures and settings.

Very little is known about the real world structure of an interbank network. On the one hand this is because the interest for this field is fairly new and on the other hand this is because there isn't much good data available. To the best of our knowledge only two empirical analyses have already been performed.

The first one, an analysis of the Austrian network published in 2004 [6], studied the loan size distribution, the degree distribution and some global network measures which will be

explained in great detail in Chapters 3 and 4. This was supposedly the first empirical study of an interbank network. The second one, an analysis of the Brazilian network published in 2010 [11], studied exposure sizes, degree distributions, and clustering. Again these measures will be explained in Chapters 3 and 4. The second part of Ref. [11] implements a *loss, given default* model much in the same way as it was developed by Prof. K. Schoors and collaborators [28]. Both studies found a complex network structure. A complex network, as defined in Ref. [33], displays substantial non-trivial topological features, with patterns of connection between their elements that are neither purely regular nor purely random. The empirical findings of both studies are in marked contrast to interbank networks that have been studied in the theoretical economic and econo-physics literature. The networks used in the literature do not reflect reality. This alludes to the fact that theoretical models can greatly benefit from the feedback to the real world. Wherever possible, the findings from this work will be compared with the ones from the abovementioned empirical studies.

What is so interesting about this Russian data? First, the completeness of the data set is very appealing. In contrast, the Austrian study had the disadvantage of having an incomplete data set and had to resort to certain approximation techniques (like the principle of maximizing the entropy) to make the data more complete. Secondly, the availability of records for dated one-day loans allows us to study the network with a daily temporal resolution scale and to investigate the dynamics of the network. The other studies of Refs. [6] and [11] used monthly aggregated data. Finally, in the sampled period 1998-2004 one has the occurence of two crises. One during August 1998 and one in the summer of 2004. This gives us the means to look at the network structure up to and including a crisis, which offers unparalleled opportunities.

The data under consideration in this work enables one to perform an empirical interbank network analysis. This could contribute to a better understanding of interbank networks, and could help bridge the gap between the real world and theoretical models.

The focus of this work

Most of the time the 7 years of data will not be covered entirely, but we will zoom in on three specific periods. The first one represents a normal, tranquil period on the Russian interbank market. The year 2001 meets the requirements of such a normal period. The other two interesting periods involve interbank turmoil. The big crisis of August 1998 and the mini-crisis of the summer of 2004, which both resulted in the collapse of the interbank market. The events leading to and the crises themselves are explained clearly in the paper of Schoors and collaborators [28]. The first crisis got triggered on August 17, 1998 when Russia abandoned its exchange rate regime, defaulted on its domestic public debt and declared a moratorium on all private foreign liabilities. The second crisis was ignited by an investigation of banks suspected of being engaged in money laundering and sponsorship of terrorism. This led to a complete lack of trust between banks and a liquidity drought. The time ranges which we

take to investigate these periods are the following:

- **Entire Period**: 1 Aug 1998 – 30 Nov 2004
- **Crisis 1**: 1 Aug 1998 – 31 May 1999
- **Crisis 2**: 1 Oct 2003 – 30 Nov 2004
- **Normal Period**: The entire year 2001

In Fig. 2.2 these four periods are visualised. Unfortunately, the records start with those of August 1998 just as the first crisis strikes. Accordingly one does not have the chance to study the period preceding the first major crisis. On the bright side, one is lucky that the data period does indeed neatly includes this first cataclysm.

As already mentioned we focus on the short-term loans with a one-day and a one-week maturity which account for more than 80 percent of the transactions both in terms of number and volume. Another important reason to choose these loans is that they can be more or less accurately dated. When the due date is known, one knows when the loan got payed back. So one can identify the precise time period during which the link between the two banks existed. For the one-day loans one can be sure, for the one-week loan there is some uncertainty. If e.g. the loans with a maturity of three months were also included, the date of actual repayment would involve some educated guessing. The short-term loans give a unique perspective on interbank transactions with a high time resolution. This is especially interesting to study the dynamics of these transactions. It is a way to keep our finger on the pulse of the interbank market.

As a final remark, the data includes the monthly balance sheet of every Russian bank during 1998-2004. This information offers even greater perspectives when combining it with the interaction data. The analysis of this aspect is beyond the scope of this work.

2.2 The Network

Like any network, interbank networks consist of basic, fundamental entities called nodes which are connected by edges or links. In our case Russian banks are the nodes and the interbank loans are the edges (cfr. Fig. 2.1).

2.2.1 Nodes

How does one build a network from the data? One starts with a so-called edgelist which lists all selected loans in the time-period one chooses to study (cfr. table 2.1). Next one runs over every entry: the issuer gets a node, the receiver gets a node, and the loan is represented by a (directed) link between both. A node is always unique. If a node is already present in the network, the new loan will be linked to the existing node. Only the banks that participated

in a loan during the particular studied time-period will be represented in this network[2]. A bank which is not present in this specific network was not active on the interbank market during this time-period. The programs used to analyze the data are written in Python and make use of the Python module called NetworkX [18].

Table 2.1: An edgelist with simplified input-data which will be used to construct the networks in Fig. 2.4 of the following section.

Issuer	Receiver	Loan Size
A	B	1
B	C	2
C	B	3
C	B	1
C	A	1
C	D	1
D	C	3
D	A	2
D	A	2
C	E	2
D	E	3
E	D	2
D	E	1
E	D	1

When building a network time series, two issues need to be addressed. The first issue is related to the width of the time intervals used to bin the data. Does one loose information if one only looks at monthly aggregated data? Is one loosing sight of the bigger picture when consdering daily aggregated data? This is a difficult choice to be made. We will try to strike the golden mean with weekly aggregated data, but monthly blocking will also be needed to show certain effects. The other issue is related to the way one slides through time-space. One can choose not to overlap the subsequent time intervals used to bin the data. Or one can choose to let subsequent time intervals overlap and use a so-called moving average. Both of these features will always be well indicated in the caption of every figure.

[2]From now on, one will call such a participating bank an active bank.

Aggregates and Periodicity

In this part different kind of aggregates or time intervals are considered. Also the weekly and monthly periodicity of the data will be discussed.

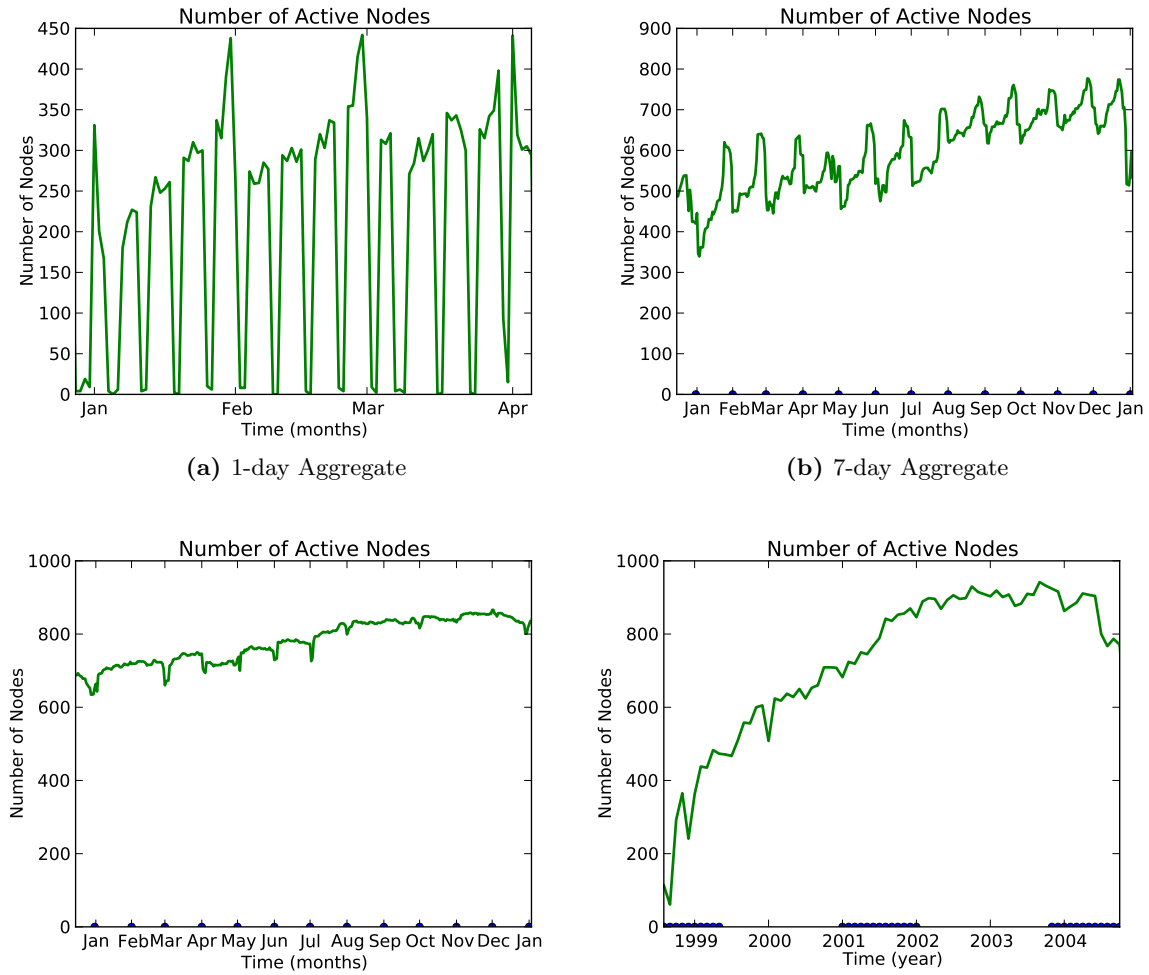
1-day Aggregate In Fig. 2.3a the number of loans with maturity $< 7d$ is shown for the first quarter of the normal period 2001, using a one-day aggregate. For example the number of loans plotted at the first of March, consist of the one-day and the $2 - 7d$ loans which have the first of March as due date. As expected one clearly sees the weekend periods with hardly any activity on the domestic interbank lending market. One also notices a spike in the activity towards the end of every month.

7-day Aggregate In Fig. 2.3b the 7-day moving average for the same $< 7d$ loans is shown. As an example the indicated number of loans at the first of March, comprise all $< 7d$ loans which have a due date in the first seven days of March. Again one notices a periodic, monthly phenomenon of less than average activity in the beginning of the month and more than average activity at the end of the month. This monthly periodicity is a consequence of regulation. Certain requirements imposed by policy makers, like the capital requirements for the individual banks, need only to be fulfilled at the end of the month. This leaves some freedom for the banks. Banks will start the month quietly, lying low on the interbank lending market. If they notice, while approaching the end of the month, that requirements will not be met, they step up their game. One also sees a slow but gradual increase in the number of active banks over a years period.

28-day Aggregate In Fig. 2.3c the same as above is done but now with 28-day aggregates. The idea behind the 28-day length is to strike a mean between the 7-day and the monthly periodicity which one discovers by the cusps at the beginning of every month[3]. This 28-day aggregate period turns out to be inconvenient for this work and will not be further used.

Monthly Aggregate To end this discussion a monthly aggregate is considered in Fig. 2.3d. Together with Fig. 2.2 on page 7, this represents the most basic notions of our data: the monthly number of active banks and loans. There is an increase in the interbank activity, by about a factor of 5, over the 7 year period and the effect of the crises are clearly seen.

[3]Notice that February, as only 28-day month, does not have this cusp.



(a) 1-day Aggregate

(b) 7-day Aggregate

(c) 28-day Aggregate

(d) Monthly Aggregate

Figure 2.3: The effect of taking different aggregates on the number of active nodes or banks. This is done for four different aggregates.

2.2.2 Edges

Types of Edges

How do the real world loans get translated to the links in our network? Six kinds of network representations for the input-data can be distinguished. Each of them has a different level of detail, or in other words a different level of coarse graining.

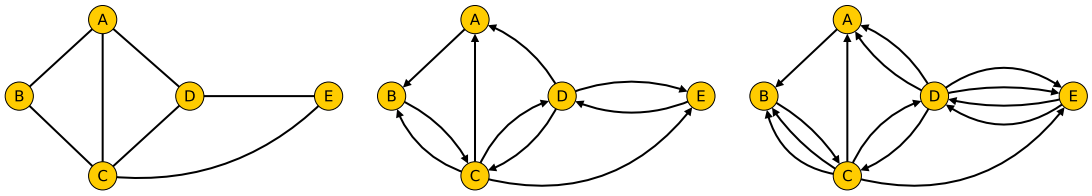
Unweighted Representations: One starts with defining the unweighted representations where one uses the simplified edgelist of table 2.1 in order to visualize the three representations in Fig. 2.4.

- **Undirected Edges:** Consider a certain aggregate period. If two banks enter into a

contract one way or another during this period, there is an edge between the two nodes. It does not matter how many times they issue a loan, or how large these loans are. There is no discrimination between issuer or lender, hence undirected, and there can only be a maximum of one link between every possible pair of nodes. As illustrated on Fig. 2.4a, in this representation an edge is either present or absent.

- **Directed Edges:** Consider a certain aggregate period. Now we discriminate between issuer or lender. If an issuer lends money to a receiver during this period, there is a directed edge starting form the issuer node and pointing towards the receiver node. Again it does not matter how many times they issue a loan, or how large these loans are. There can be a maximum of two links between two nodes. This situation is illustrated in Fig. 2.4b.

- **Multi-directed Edges:** Consider a certain aggregate period. We still discriminate between issuer and lender but now also between the individual loans. Each time an issuer lends money to a receiver during this period, the loan gets a directed edge starting from the issuer node and pointing towards the receiver node. Now it does matter how many times they issue such a loan, but still not how large these loans are. As illustrated in Fig. 2.4c, there is no limit on the number of links between two nodes.



(a) Undirected **(b)** Directed **(c)** Multi-directed

Figure 2.4: The different unweighted representations of a network using the edgelist from table 2.1 as input.

Weighted Representations If the size of the loans is taken into account, one can add an extra level of detail to the network edges. Here, one can also consider three possibilities. The visualizations would be the same as their counterparts in the unweighted cases from Fig. 2.4, but now with numbers or weights added to the links.

- **Weighted Undirected Edges:** Consider a certain aggregate period. If two banks lend money to each other during this period, there is an edge between the two nodes which is weighted with the *net exposure*[4]. There can only be a maximum of one link

[4]The net exposure is the absolute value of the difference between the sum of all loans from A to B and the sum of all loans from B to A.

between every possible pair of nodes.

- **Weighted Directed Edges:** Consider a certain aggregate period, whereby we discriminate between issuer and lender. If an issuer lends money to a receiver during this period, there is a directed edge starting from the issuer node and pointing towards the receiver node. The edge is weighted by the total of all loans in this direction. There can be a maximum of two links between two nodes.

- **Weighted Multi-directed Edges:** Consider a certain aggregate period whereby we discriminate between issuer or lender and also between individual loans. Each time an issuer lends money to a receiver during this period, the loan gets a directed edge starting form the issuer node and pointing towards the receiver node. Now it matters how many loans there are and each of these loans will be weighted by their sizes. There is no limit on the number of links between two nodes.

Edges in the Russian Interbank Network

In Fig. 2.2 on page 7 the monthly number of loans with a maturity of $< 7d$ is shown for the entire data period. There is a gradual increase in the number of loans until the beginning of 2004 where this number starts to decline. From now on we focus on the three time periods which we consider representative for 1998-2004 time period studied in this work. We will look at monthly and 7-day aggregates.

Monthly Aggregate: In Fig. 2.5 the monthly number of edges is shown for the three different types of network representations. One considers the three periods of interest separately. In both crisis periods one sees two drops in activity. For the first crisis period, the first drop in August 1998 corresponds to the default of Russia on its domestic public debt. The second drop in December 1998 happened in the aftermath of this default. For the second crisis period, the first drop in January 2004 was only a minor shock. The second drop in the number of edges during the summer of 2004 is the actual liquidity crisis. Further, one also notices that the number of edges for the undirected representation and the number for the directed representations are similar. This means that interbank lending is mainly a one-way activity, which is to be expected. When bank A is in need of money and lends from bank B, it is unlikely that bank B is also in need of money and will lend from bank A.

7-day Aggregate: Fig. 2.6 contains similar information as Fig. 2.5 but now for a 7-day aggregate instead of a monthly aggregate. The 7-day aggregate uncovers the monthly periodicity. Fig. 2.7, which shows two ratios, is interesting. First we display the ratio of the number of multi-directed edges to the number of active nodes. Or in other words the average number of loans per active bank. We also show the ratio of the number of multi-directed

edges to the number of undirected edges. In other words this is the average number of loans per pair of interacting banks.

The average number of loans per pair of interacting banks stays more or less constant at a rate of 2. Combining this observation with the conclusion of the previous paragraph that interbank lending is mainly one-way, one concludes that on average a receiver has two loans per week per issuer. The average number of loans per active bank does follow the monthly periodicity. Although the activity per active node increases at the end of the month, the average number of loans per pair of interacting banks stays more or less constant. This leads us to the following conclusion. When a bank needs to meet the requirements at the end of the month, it will not increase its activity with a counterparty it is already involved with. Instead, it will look for a new counterparty to lend money from.

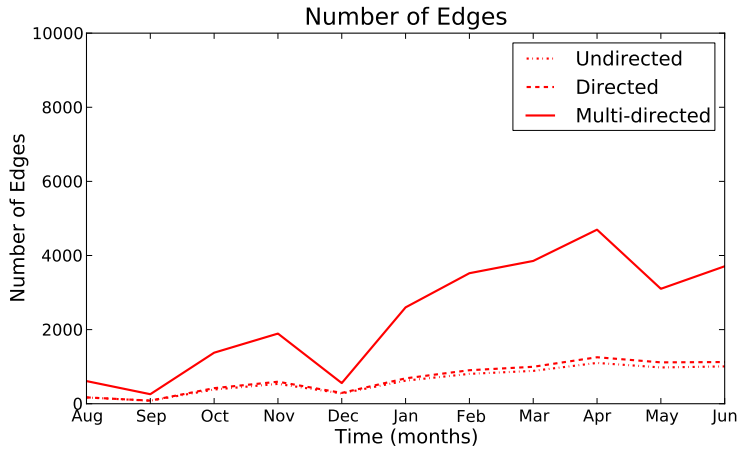
2.3 Overview

In this chapter we took a first glance at the Russian interbank network. The basic entities of a network -nodes and edges- were studied and the different possibilities of constructing a network were explained. In table 2.2 these 'construction' possibilities are listed as a summary.

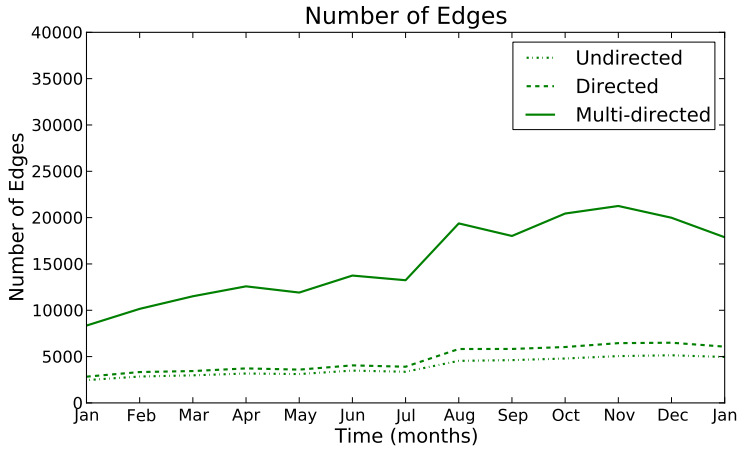
After studying the network building blocks, we are now equipped to delve into the topological structure of the interbank network. In Chapter 3, some typical network measures will be studied, while in Chapter 4 and 5 we will gather stylized facts about respectively node and edge distributions.

Table 2.2: The 'construction' possibilities explained in this Chapter.

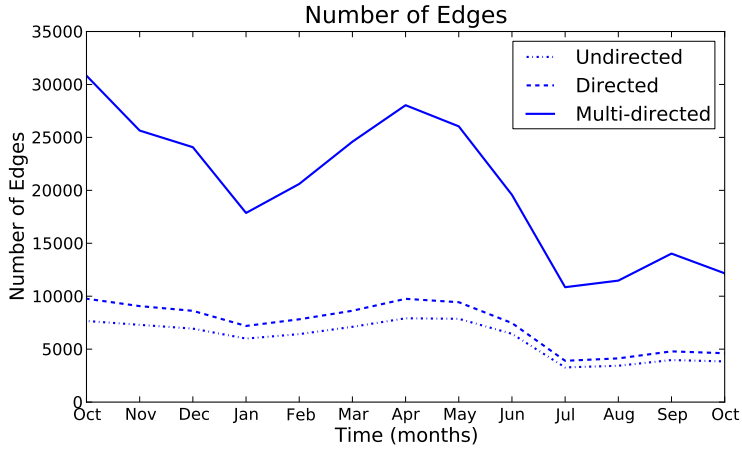
Concept	Option
Contracts	$< 1d$
	$2 - 7d$
	$< 7d$
	All Loans
Aggregate	7-day
	Monthly
	Quarterly
	Yearly
Weight	Unweighted
	Weighted
Edge Types	Undirected
	Directed
	Multi-directed
Period	Entire
	Crisis1
	Crisis2
	Normal



(a) Crisis 1 (1998-1999)

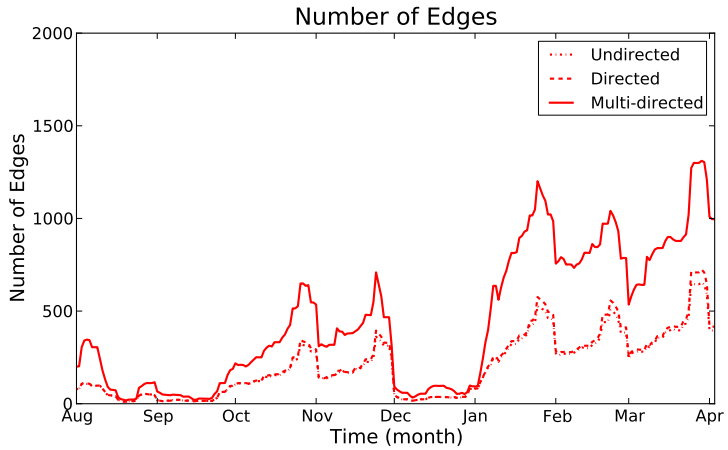


(b) Normal (2001)

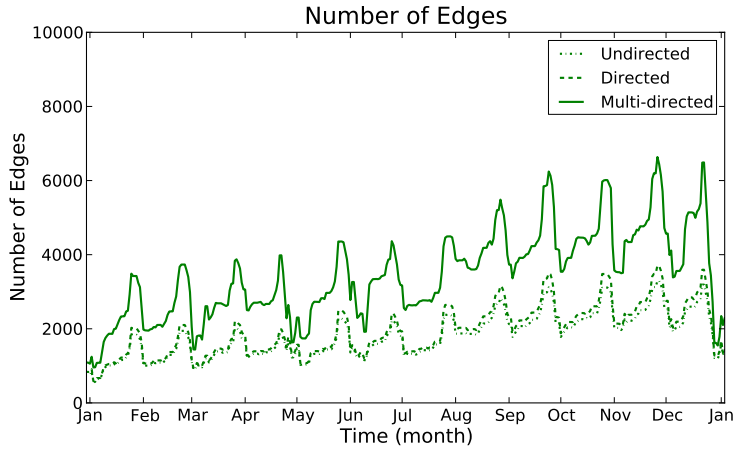


(c) Crisis 2 (2003-2004)

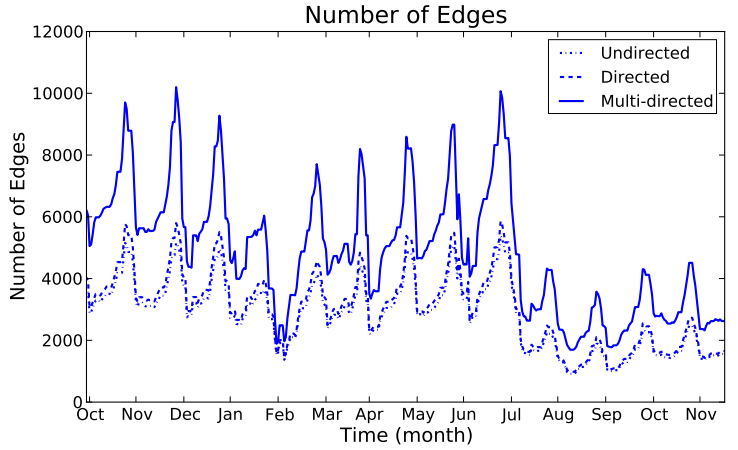
Figure 2.5: The monthly number of edges for the three different types of network representations. The three periods of interest are considered separately.



(a) Crisis 1 (1998-1999)

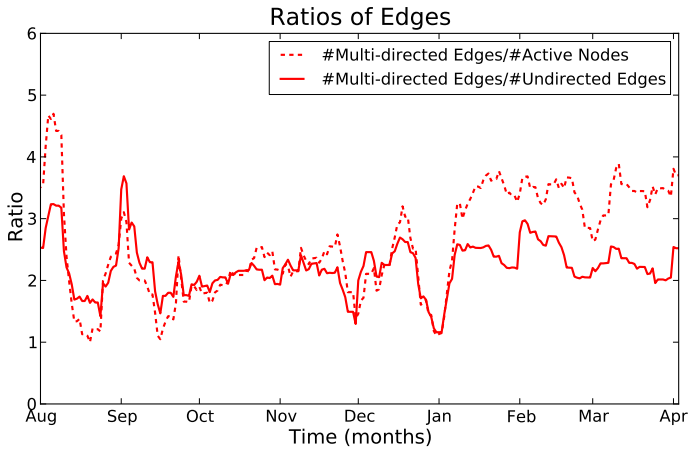


(b) Normal (2001)

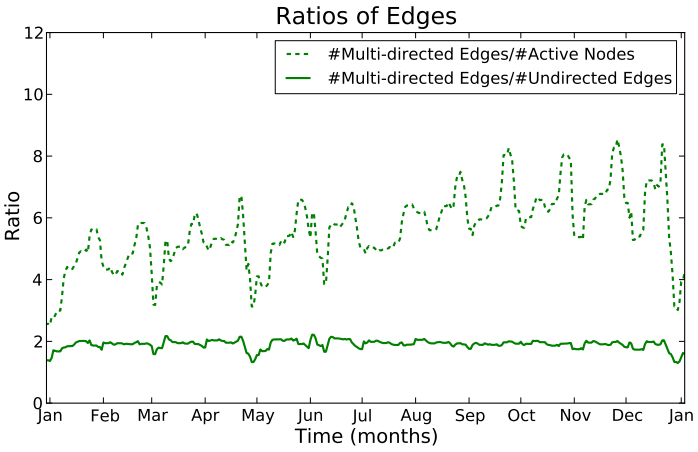


(c) Crisis 2 (2003-2004)

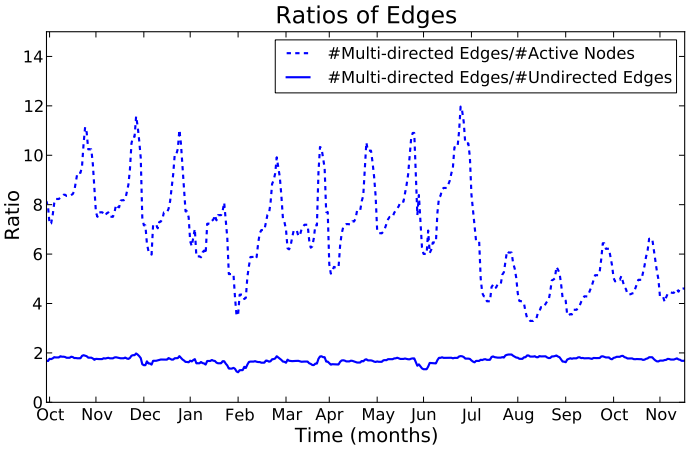
Figure 2.6: The weekly number of edges for the three different types of network using a moving weekly aggregate. The three periods of interest are considered separately.



(a) Crisis 1 (1998-1999)



(b) Normal (2001)



(c) Crisis 2 (2003-2004)

Figure 2.7: Two ratios are shown. The first is the ratio of the number of multi-directed edges and the number of active nodes. The second ratio is the number of multi-directed edges and the number of undirected edges. This is done with a moving weekly aggregate and the three periods of interest are considered separately.

Chapter 3

Network Measures

In this chapter some typical measures for quantifying network structure will be considered. These quantities capture particular features of the network topology and form an important part of the network toolbox. The density of a network, the network's distance and clustering properties and the types of components; all will pass in this concise review. Next to reporting these network measures, our goal is to search for a significant difference of these measures between crisis and non-crisis periods. Are there any network quantities which indicate a pending crisis in the period preceding a crash and could serve as early warning signals?

3.1 Density

The density d of a network is the ratio of the number of present edges and the number of possible edges. It is a measure for the completeness of a network, where for a complete network all possible edges are present. The density for undirected networks is defined as

$$d = \frac{2m}{n(n-1)}, \tag{3.1}$$

and for directed networks is

$$d = \frac{m}{n(n-1)}, \tag{3.2}$$

where m is the number of edges and n is the number of nodes in the network, with $0 \leq d \leq 1$ [18]. Fig. 3.1 illustrates an undirected network with 4 nodes and 4 edges, which has a density of $d = 2/3$. Because a multi-directed network has an infinite amount of possible edges, this definition is rendered useless.

In Fig. 3.10 on page 36, the undirected density is shown for the three periods which were considered representative. Because the number of edges in the undirected and directed representations is similar (cfr. Figs 2.5 and 2.6), we expect the undirected and directed density

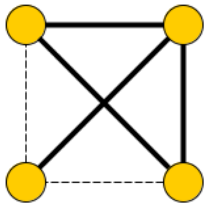


Figure 3.1: An undirected network with a density $d = 2/3$.

to be similar too, apart from a factor 2. To easily compare the three periods, the y-range is identical.

For the 'normal' period in Fig. 3.10b the density is stable over the entire year 2001. Only about one percent of all possible links between pairs of banks is actually present. The average density in 2001 is $d = 0.0095 \pm 0.0005$[1]. This density of about one percent will be referred to as the equilibrium density. Also the fluctuations about the average are very modest. This is reminiscent for a period of stability or smooth operation of the network.

For crisis 1 in Fig. 3.10a there are two spikes at the time instances when the interbank market collapsed. So during these drops in activity, i.e., less active banks, the network grew more dense with a ten percent peak. From January on, the density settles for the equilibrium density of one percent. For the data of the two weeks preceding Russia's default, the density level is on average 0.042 ± 0.005 , which is above the post-crisis equilibrium level. Also the fluctuations about the average density are ten times larger. This elevation of density could possibly be a crisis indicator or a warning signal. More pre-crisis data would be needed in order to put this observation on more solid grounds. One would need to consider a longer pre-crisis period. Was there a (gradual) increase in the density in the period preceding August 1998? Or did the pre-crisis interbank market just have a higher equilibrium density than the post-crisis equilibrium density of one percent?

For crisis 2 in Fig. 3.10c one only sees a small drop in density during the summer of 2004. The rest of the period, the density of the network stays stable at around one percent. This second mini-crisis did not have an equally distinctive effect on the density as the first crisis did.

One notes that the density tends to increase when considering larger time-aggregates. E.g. when using a one month aggregate, the average density of 2001 is $d = 0.011 \pm 0.002$[2]. Whereas the weekly average for the same year is only $d = 0.0095 \pm 0.0005$.

[1] Mean and standard deviation over the moving weekly density.
[2] Mean and standard deviation over monthly density.

3.2 Distance Measures

3.2.1 Distance Definitions

Another widely used network measure is the average shortest path length (ASPL), which is a measure for the average closeness of two nodes in a network. The length of a path between two nodes is the number of edges passed to get from one node to the other along this path. Fig. 3.2 illustrates a shortest path between two nodes[3]. The average shortest path length is defined as

$$l = \sum_{s,t \in V} \frac{d(s,t)}{n(n-1)}, \tag{3.3}$$

where V is the set of nodes in the network, $d(s,t)$ is the length of the shortest path from s to t, and n is the number of nodes in the network [18]. One cannot simply use this definition (3.3) on any given network. This is because the above definition requires that for each pair of nodes in the network, there exists a (shortest) path connecting them. But if the network e.g. consists of two disconnected parts, this way of computing the ASPL fails. In the following discussion, this problem is circumvented by computing the ASPL for the largest component of weakly connected nodes in the network. This largest weakly connected component, which will be defined in Section 3.5, accounts for on average 98 percent of the network size. The computed ASPL can thus be interpreted as representative for the total network.

The ASPL can be computed for unweighted as well as for weighted networks. The definition of the ASPL for weighted networks makes use of a weighted sum for $d(s,t)$.

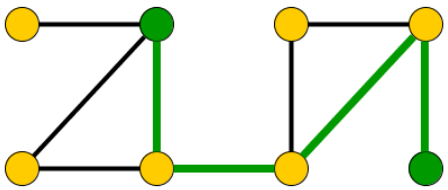


Figure 3.2: The shortest path between the two green nodes is colored green. Its length is equal to 4.

Two other commonly used distance measures are the diameter and the radius. The diameter D is defined as

$$D = \max_{s,t \in V} \{d(s,t)\}. \tag{3.4}$$

[3]We note that the direction of the edges is not taken into account so far. The direction of the edges needs not to be respected while moving along a path over the network. Including directed paths could serve as a subject for further research.

The diameter is of a network is the length (expressed in the number of edges) of the longest shortest path between any two nodes. In order to define the radius of a network, the concept of eccentricity is needed. The eccentricity of a node s, E_s, is defined[4] as

$$E_s = \max_{t \in V} \{d(s,t)\}. \tag{3.5}$$

From this, the definition of the radius of a network R follows as

$$R = \min_{s \in V} \{E_s\}. \tag{3.6}$$

3.2.2 Distance Discussion

The paper of Boss et al. [6] calculates the ASPL of the undirected as well as the directed representation of the Austrian interbank network. They found an ASPL equal to $l = 2.26 \pm 0.03$[5] for the undirected version and an ASPL equal to $l = 2.59 \pm 0.02$ for the directed version.

In Fig. 3.11 on page 37, the ASPL, diameter and radius of the largest component are shown for the three periods of interest. For the 'normal' period in Fig. 3.11b the ASPL is stable for the entire year 2001. It fluctuates slightly around the mean distance of $l = 3.67 \pm 0.20$, which will be called the equilibirum ASPL. Compared to Boss et al. [6], which found an ASPL equal to $l = 2.26 \pm 0.03$, the Russian interbank network displays a longer ASPL in 2001. The radius of the largest component on average has length 5, whereas the diameter fluctuates around length 9.

For crisis 1 in Fig. 3.11a one sees two drops in the ASPL of the largest component at the troubled months[6]. At the end of these two drops there is a clear ASPL-spike after which the ASPL settles to about the equilibrium length. The three weeks in the run-up to Russia's default, start with an ASPL of on average 3.26 ± 0.44. This differs from the so-called equilibrium ASPL. This difference might be a crisis indicator. But once again, one would of course need a bigger pre-crisis data range to investigate such claims. The radius and diameter follow the same trends as the ASPL. Most interestingly, the radius and the ASPL nearly coincide in the months of August, September, and December. This could be another crisis indicator.

For crisis 2 in Fig. 3.11c there is only a small rise in the ASPL during the summer of 2004, which persists for the following months. The radius and diameter again follow the

[4] This definition also fails for a disconnected network, as was the case with the ASPL definition. Here again only the largest component of the network will be considered.

[5] Mean and standard deviation over the 10 monthly data sets.

[6] Here, the computed distance measures are less representable than for the other two periods. This is because the largest component includes a smaller fraction of the nodes. We refer to Fig 3.13 which shows the size of the largest component.

same trends as the ASPL. As with the density, the second mini-crisis did not have an equal distinctive effect on the ASPL as the first crisis did.

Further one notices that a higher density results in a smaller ASPL and vice versa. The more crowded a network gets with edges, the less steps it takes to get from one node to the other. One notes that the distance measures tend to decrease when considering larger time-aggregates. When using, e.g., a one month aggregate, the average ASPL of 2001 equals $l = 3.13 \pm 0.15$, whereas the weekly average for the same year equals $l = 3.67 \pm 0.20$.

Instead of just looking at the average value, one can go further by determining the distribution of lengths of the shortest paths between all pair of nodes[7]. In Fig. 3.3 this is done for December 1998 and July 2001. The first is a typical example of a crisis month and the second of a normal month. For a normal period (Fig. 3.3b) the mean value is lower and the distribution is less broad compared to a crisis period (Fig. 3.3a). In times of crisis, there are stronger fluctuations in the ASPL distribution. Now suppose that one starts at a given node and wishes to reach a random other node in the network by walking along their shortest path. It is striking that for a network with 150 nodes and 300 edges (August 1998) this would take longer, than is the case for a network with 1200 nodes and 5000 edges (July 2001).

3.3 Cluster Measures

The next important general network feature to be studied in this chapter is the concept of clustering. In many networks it is found that if node A is connected to node B, and node B to node C, then there is a larger probability that node A is also connected to node C. Or in interbank network terms, the chance that two banks which are involved in money lending with a third bank, also lend money to each other. There are two commonly used measures to quantify this: transitivity and average clustering.

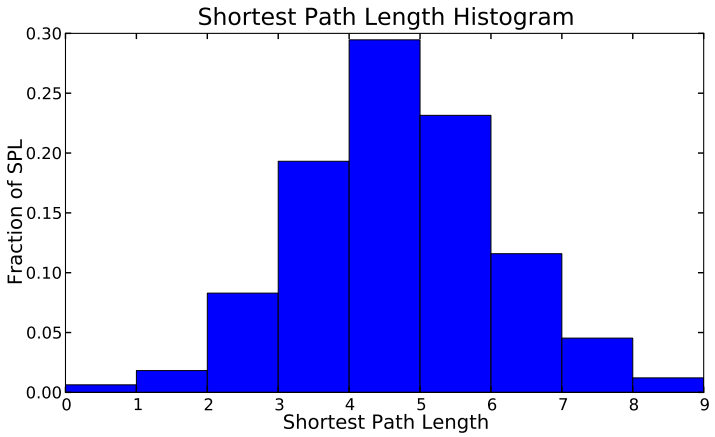
3.3.1 Cluster Definitions

In terms of network topology, transitivity means the presence of a heightened number of triangles in the network. Triangles are sets of three nodes each of which is connected to each of the others. This presence can be quantified by defining a transitivity coefficient T:

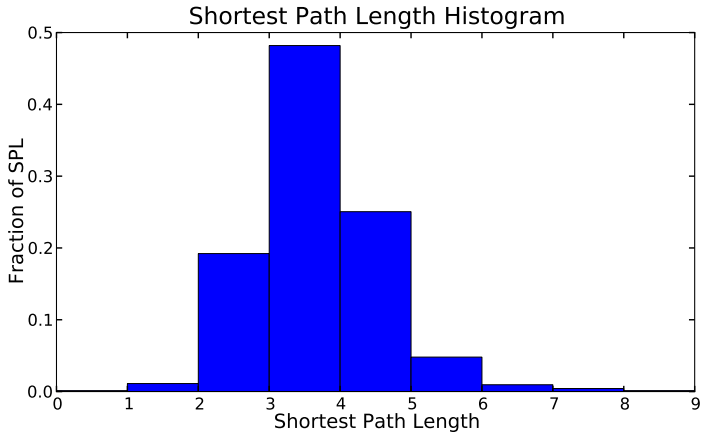
$$T = \frac{3 \times \text{number of triangles in the network}}{\text{number of connected triples of vertices}}, \tag{3.7}$$

where a connected triple means a single node with edges running to an unordered pair of others [33]. For an example we refer to Fig. 3.4.

[7] In this discussion not only the shortest paths between the nodes in the largest components are considered, but every posible shortest path in the network



(a) December 1998



(b) July 2001

Figure 3.3: A normed histogram of the shortest path distributions for August 1998 ('crisis') and July 2001 ('normal').

In essence, T measures the fraction of triples that have their third edge filled in to complete the triangle. The factor of three in the numerator accounts for the fact that each triangle contributes to three triples and ensures that T lies in the range $0 \leq T \leq 1$. In simple terms, T is the average probability that two nodes which are network neighbours of the same vertex will themselves be neighbours.

An alternative definition to quantify the degree of clustering in a network is called the average clustering coefficient. Watts and Strogatz [43] proposed to define a local value of the clustering coefficient:

$$C_i = \frac{\text{number of triangles connected to node i}}{\text{number of triples on node i}}. \tag{3.8}$$

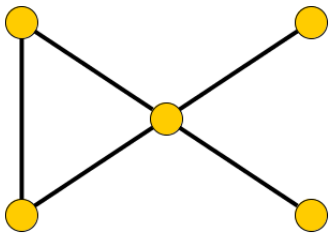


Figure 3.4: This picture illustrates the definitions of the transitivity and the average clustering coefficient. This network has one triangle and eight connected triples. Therefore its transitivity coefficient is $T = 3 \times \frac{1}{8} = \frac{3}{8}$. The individual nodes have local clustering coefficients of $1, 1, \frac{1}{6}, 0$ and 0, with an average value of $C = \frac{13}{30}$. This figure was reproduced from Ref. [33].

C_i, a ratio between 0 and 1, expresses the degree of connectedness among the neigbours of a given node [11]. The average clustering coefficient for the entire network is the mean value of these local clustering coefficients:

$$C = \frac{1}{n} \sum_i C_i. \tag{3.9}$$

This definition, in as opposed to the one for the transitivity coefficient, tends to weigh the contributions of low-degree nodes more heavily [33]. C is a global network measure with a local feature. For an example we refer to Fig. 3.4.

3.3.2 Cluster Discussion

The paper of Boss et al. [6] reports a transitivity coefficient of $T = 0.12 \pm 0.01$. They say this is small compared to other networks[8] and explain this as follows. While banks might be interested in some diversification of interbank links, the costs involved in opening a new link adds to the cost of maintaining the existing ones. In optimizing their lending costs, banks tend to mminimize the number of edges. So if for instance two small banks have a link with the same larger institution, there is no reason for them to additionally open a link among themselves.

In Fig. 3.12 on page 38, the transitivity and average clustering coefficients are shown for the three periods of interest. For the 'normal' period in Fig. 3.12b the two clustering coefficients are stable around the value of $T = 0.088 \pm 0.013$ and $C = 0.125 \pm 0.032$. Both have a monthly cycle which is more pronounced for the the average clustering coefficient. Recapitulating that the average clustering coefficient is influenced more by the lower degree

[8] The T of the interbank network is of the same size as the WWW. Technological networks mostly have even smaller values for T of about 5 percent [33].

nodes, this implies a higher clustering among these lower degree nodes at the end of the month.

For crisis 1 in Fig. 3.12a the clustering coefficients are very volatile. They even hit zero in August and December. In contrast to the normal period, the average clustering coefficient is smaller than the transitivity coefficient, which implies that fewer low degree nodes are clustered than was the case for the normal period.

For crisis 2 in Fig. 3.12c the observed trend in the cluster measure coefficients move along the same lines. The periodicity effect is less outspoken compared to the normal period. During the two dips in activity, shown in Fig. 2.5c on page 17, both coefficients tend to become equal.

One notes that the clustering tends to increase when considering larger time-aggregates. When using, e.g., a one month aggregate, $T = 0.14 \pm 0.01$ and $C = 0.23 \pm 0.03$. Whereas the weekly averages for the same year were only $T = 0.088 \pm 0.013$ and $C = 0.125 \pm 0.032$.

3.4 Small World Network

Equipped with the concepts of the ASPL and the local clustering coefficient, one now can have a look at the so-called *small world effect*. This effect refers to networks where, although the network size is large and each node has a small number of direct neighbours, the distance between any two nodes is very small compared to the network size [11]. This effect was first described by Stanley Milgram in his famous experiment in the 1960s. He examined the average path length for social networks of people in the United States. It suggested that human society is a *small world* type network characterized by short path lengths. Although the results of Milgram are considered very controversial, this experiment became well-known in popular culture and it is often associated with the phrase "six degrees of separation" [42].

The Austrian study of Boss et al. [6] found an ASPL equal to $l = 2.59 \pm 0.02$ for the undirected network. From this result they concluded that the interbank network looks like a *very small world* network with about three degrees of separation. The paper of Cont et al. [11] criticises this. Cont et al. raise the point that a small ASPL itself does not characterize the *small world* property[9]. They look at another signature[10] of the *small world* property: while the ASPL is bounded or slowly increasing with the number of nodes, the local clustering coefficient of nodes remain bounded away from zero [12]. In other terms, as the size of the network increases, the clustering cannot decrease to zero.

Fig. 3.5 [11] shows the local clustering coefficient versus the degree of a node for the Brazilian interbank network. The degree of a node v is the number of edges connected to

[9] E.g. complete graphs, where every node is connected to every other node, do not have the *small world* property. This is because they do not have a small number of direct neighbours.

[10] Another definition for a network to have the *small world effect* is that the ASPL scales logarithmically or slower with network size for fixed mean degree [33]. This definition is hard to use for empirical data sets, because it is impossible to study the scale dependence.

v. One observes nodes with an arbitrary small clustering coefficient. Cont et al. argue that the absence of uniform clustering is a clear indication of the fact that the Brazilian interbank system cannot be considered as a *small world* network.

In Fig. 3.6 the same is done for the Russian interbank network. Because one might expect a difference in the *small world* properties of crisis and non-crisis periods, both are investigated. Here again, December 1998 is used as a prototypical example of a crisis month, whereas July 2001 is used as a typical non-crisis month.

Fig. 3.6b for July 2001, has the same structure as Fig. 3.5. Following the definitions of Cont. et al., one can conclude that the Russian interbank network is not a *small world* network in normal periods of operation.

The question whether or not the interbank network of December 1998 is *small world*, is considerably harder to adress. Fig. 3.6b considers December 1998. One notes that although there are 150 active nodes, most of them have zero clustering and a low degree. Therefore, most markers are plotted on top of each other and one has fewer distinctive data points. We will not draw any conclusions about the *small world* property of this network, due to the lack of data points. Obviously, the difference between Figs. 3.6a and 3.6b implies that the distribution of the degree of a node versus its local clustering coefficient is a distinctive feature in order to discriminate between a 'normal' and 'abnormal' operation of the interbank network.

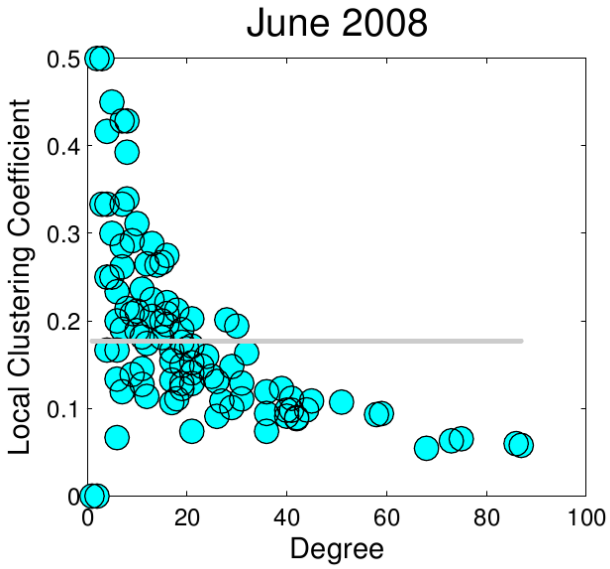
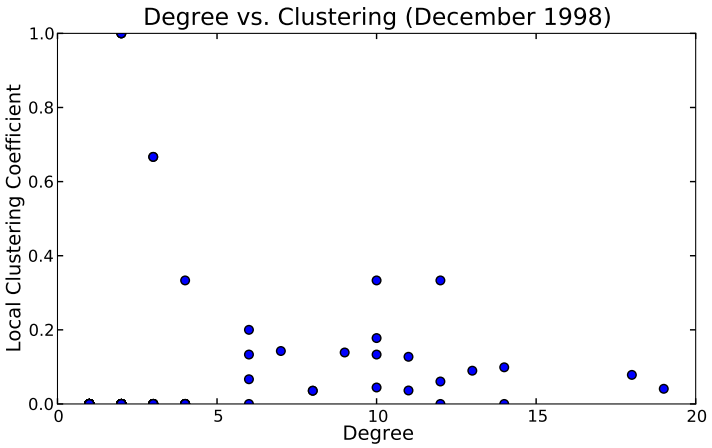
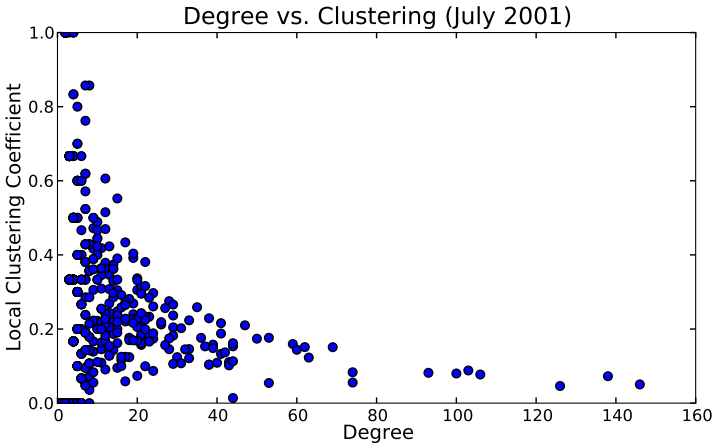


Figure 3.5: The degree of a node versus its local clustering coefficient for the Brazilian interbank network for June 2008. The grey line indicates the average clustering coefficient. This figure was reproduced from Ref. [11].



(a) December 1998



(b) July 2001

Figure 3.6: The degree of a node versus its clustering coefficient for the Russian interbank network for December 1998 ('crisis') and July 2001 ('normal').

3.5 Components

3.5.1 Components Definitions

Whereas the undirected network representation was considered in the previous sections of this Chapter, we will now turn our attention to the directed representation. Each edge has a direction and while moving along a path over the network, the direction of the edges needs to be respected. The component to which a node belongs is that set of nodes that can be reached from it by paths along the directed edges of the network. One can define several different components.

- **Weakly Connected Component:** For each pair of nodes u, v in a weakly connected component (WCC), there needs to be an *undirected* path from u to v. On such an

undirected path, the direction of the edges are neglected so an edge can be passed in both directions (Fig 3.7a).

- **Strongly Connected Component:** For each pair of nodes u, v in a strongly connected component (SCC), there needs to be a *directed* path from u to v and from v to u (Fig. 3.7b).

- **Out-component of a Particular Node:** The out-component of a node v is that set of nodes that can be reached by moving along all possible directed paths starting from v (Fig. 3.7c). Mapping this on the interbank network, the banks of the out-component of a bank v are all banks which v lends money to explicitly, its debtors, and implicitly, the debtors of its debtors and so on. This collection of banks can potentially infect bank v when they default.

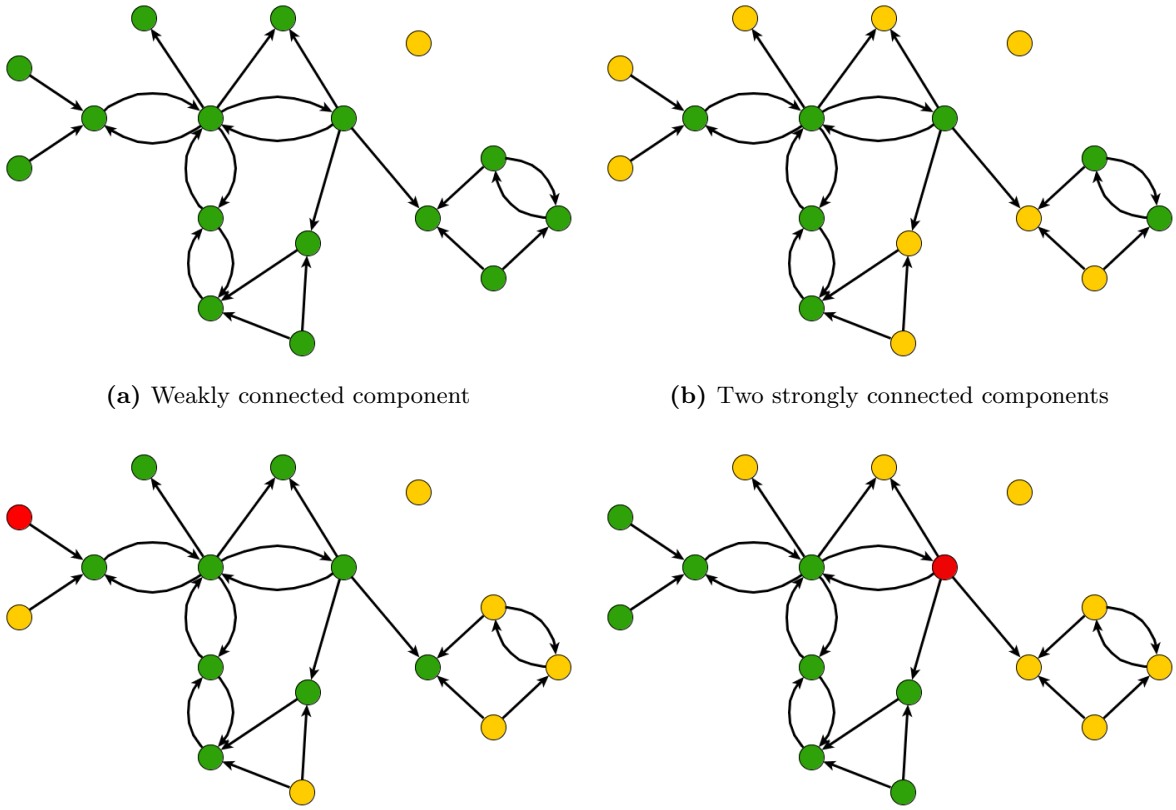
- **In-component of a Particular Node:** The in-component of a node v is that set of nodes from which v can be reached by moving along directed paths (Fig. 3.7d). The banks of the in-component of a bank v are all banks which v lends money from explicitly, its creditors, and implicitly, the creditors of its creditors and so on. This collection of banks can potentially be infected by bank v when it defaults.

One notes that members of an out- and in-component depend on the choice of the starting node. Such a component is said to have a seed. Choose a different starting node for an out-component and the set of reachable nodes may change. Thus an out-component is a property of the network structure and the starting node and not (as with strongly and weakly connected components) of the global network structure. This means that a node can belong to more than one different out-component for example.

A few other points are worth noticing. First, it is self-evident that all the members of the SCC to which a node u belongs are also members of u's out-component. Furthermore, all vertices that are reachable from u are necessarily also reachable from all the other nodes in the SCC. Thus it follows that the out-components of all members of a SCC are identical. It would be reasonable to say for this case that out-components really belong not to individual nodes, but to SCC. The same can be said of in-components [35].

3.5.2 Components Discussion

In Fig. 3.13 on page 39, the size of the largest SCC and the size of the largest WCC is shown. For the 'normal' period in Fig. 3.13b the largest WCC covers on average 96 ± 2 percent of the nodes. The banks not present in this WCC are isolated banks which form smaller interbank networks among each other. The largest SCC includes on average 27 ± 9 percent of the nodes. There is a 10 percent rise in August for the largest SCC. This rise cannot be explained by any assignable event.



(a) Weakly connected component

(b) Two strongly connected components

(c) Out-component of red node

(d) In-component of red node

Figure 3.7: Four types of components.

For crisis 1 in Fig. 3.13a the largest WCC starts at a 97 percent coverage just before the crisis and then, as the crisis strikes, drops to only 20 percent coverage. In Fig. 3.8 the number of WCC is shown for August 1998. As the crisis strikes, the network clearly shatters into many smaller disconnected parts. The SCC, the core of the interbank network, stays at a level of on average 5 ± 2 percent.

For crisis 2 in Fig. 3.13c the largest WCC covers, with a 97 ± 2 percent rate, nearly the entire network. The largest SCC declines from covering 45 percent in October 2003 to only 30 percent in November 2004. In the summer of 2004 when the second crisis hit the system, only a decrease in the largest WCC can be seen. Again the effects of the first and the second crisis on the interbank network differ greatly.

To investigate the in- and out-components in the interbank network, one starts with calculating the size of the in- and out-component of every single node. In Figs. 3.9a and 3.9b one shows the histogram of respectively the in- and out-component sizes for March 2002. In Fig. 3.9a there are only six different in-component sizes present in the network and these six sizes can be subdivided into two groups. Either a node has a very small or no in-component or

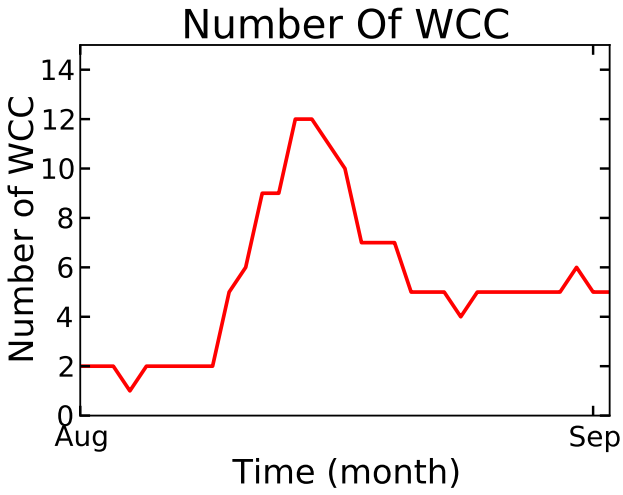


Figure 3.8: The weekly moving average of the number of WCC in August 1998.

a node has an in-component that spans about 90 percent of the network. One observes the same grouping for the out-component histogram in Fig. 3.9b. To determine the percentage of nodes in both groups, a larger bin width was used. From Fig. 3.9c one can conclude that 35 percent of the nodes has a very small or no in-component, whereas 65 percent has a large in-component of 85 percent. Fig. 3.9d shows that 15 percent of the nodes have a very small out-component and 85 percent has a large out-component that reaches 65 percent of the nodes. One observes that the percentage of nodes which have the large out-component, is the size of the large in-component. Vice versa, the percentage of nodes which have the large in-component, is the size of the large out-component. Referring to the interbank network, this becomes obvious. Saying that 65 percent of the banks can infect 85 percent of the banks, is the same as saying that 85 percent of the banks can be infected by 65 of the banks. The large out-component can be seen as the possibly infecting banks, whereas the large in-component can be seen as the possibly infected banks. Of course a bank can belong to both the large in- and out-component.

Is this dichotomy between nodes with the same large out-component size and isloated nodes with just a small out-component, a constant throughout the data period or just a happenstance? To adress this question, similar figures as Fig. 3.9 were plotted for the entire data-period 1998-2004. The dichotomy appears the first time for March 1999 for the in- as well as out-component and then persists for the rest of the data-period.

In 3.14 on page 40 the size of the large in- and out-component is shown for the normal and the second crisis period. The first crisis period is not included in this discussion because there was no clear dichotomy between a small and a large in/out component. If we consider e.g. September 2001, then the large in-component covers 70 percent of the network and the large

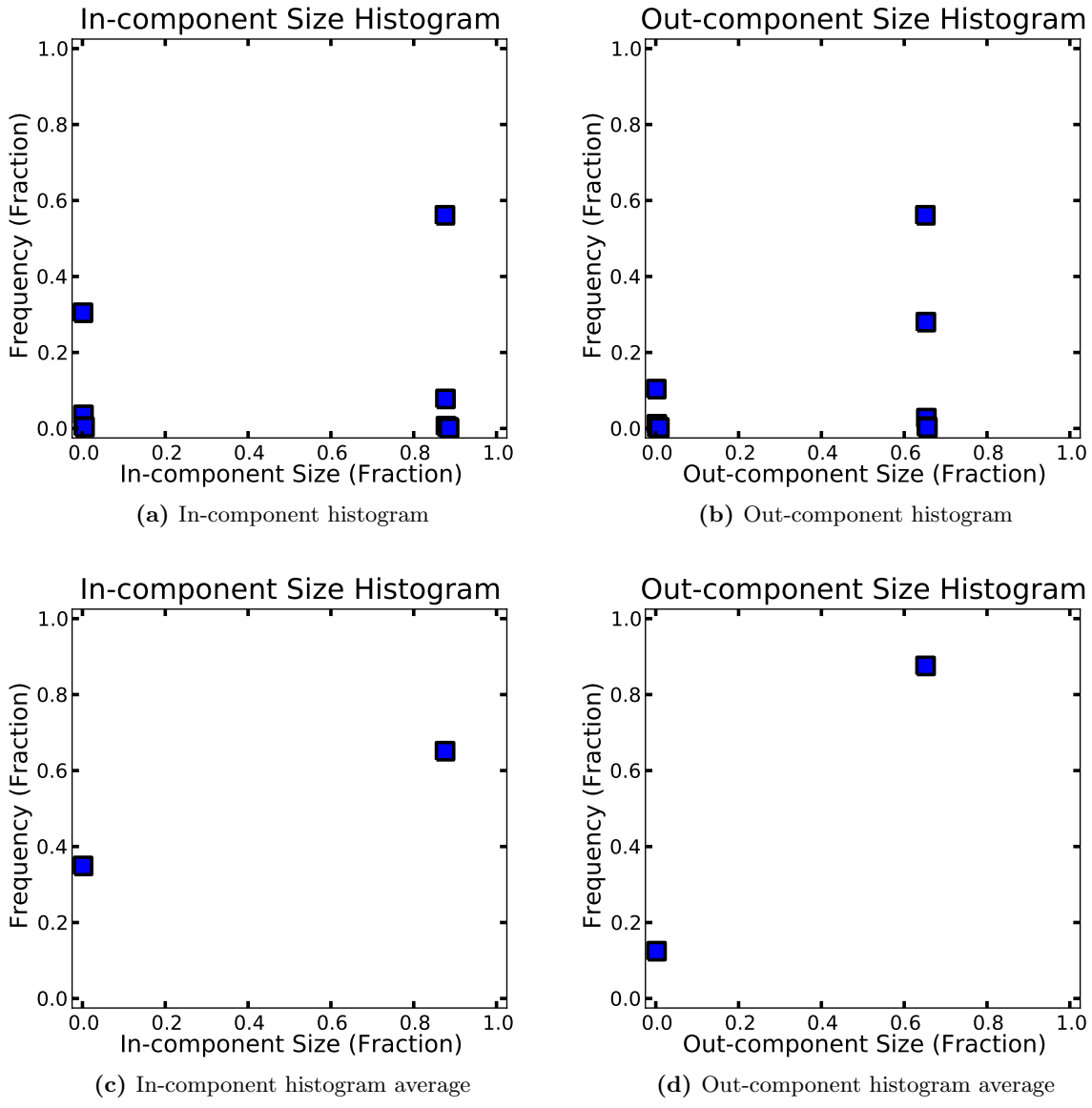
out-component covers 50 percent. This means that 50 percent of the banks have the ability to infect a total of 70 percent of the banks. This 50 percent can be seen as systemic risk carrying banks. Whereas the other 50 percent of the banks can only infect a small negligible percentage of the network and do not carry systemic risk.

The size of the large in- and out-components of the 'normal' period are shown in Fig. 3.14a. The out-component stays constant including around 50 percent of the network. The in-component spikes up at the end of each month. This implies there are a lot of new issuers at the end of the month which enter the large in-component and expose themselves to the systemic risk. In Fig. 3.14b the size of the large in- and out components are shown for the second crisis period. In the three months preceding the crisis, the large out-component declines. When the crisis strikes both the large in- and out-component contract to include 40 percent of the nodes.

3.6 Overview

In this Chapter, we took a deeper look at the network structure and paid particular attention to the crisis periods. Our first conclusion was that the effect of the second crisis on the network structure is certainly less drastic than the effect of the first crisis. In every network measure the first crisis was very outspoken whereas the second only caused a ripple. Next, we identified some typical differences between crisis, referring only to the first, and non-crisis periods. As a crisis strikes there is: an increase in density, a drop in the distance measures, and the clustering coefficients become very volatile. Further, the network shatters into many smaller and disconnected components. The local degree versus the local clustering coefficient and the shortest path length distribution can both distinguish between normal and dysfunctional operating periods. To identify pre-crisis early warning signals, we need additional data covering a larger pre-crisis time period. Last, we concluded that the interbank network, following the reasoning of Cont et al., is not a small-world network and we investigated the in- and out-components of the interbank network.

After studying these global network measures, we now shift our attention to the distribution of the local node and edge attributes in the following two chapters.



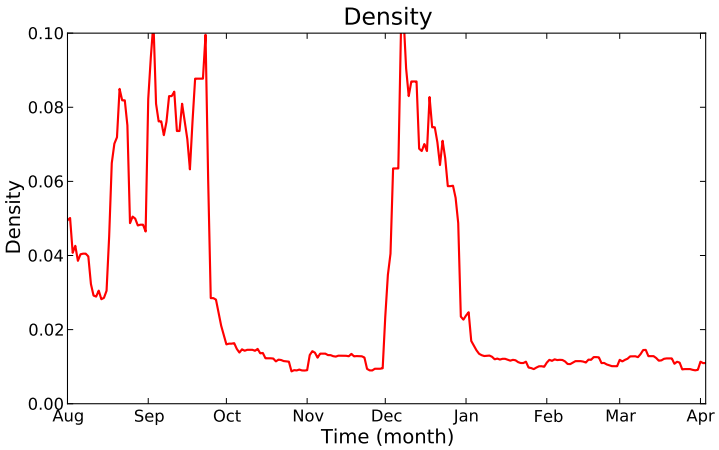
(a) In-component histogram

(b) Out-component histogram

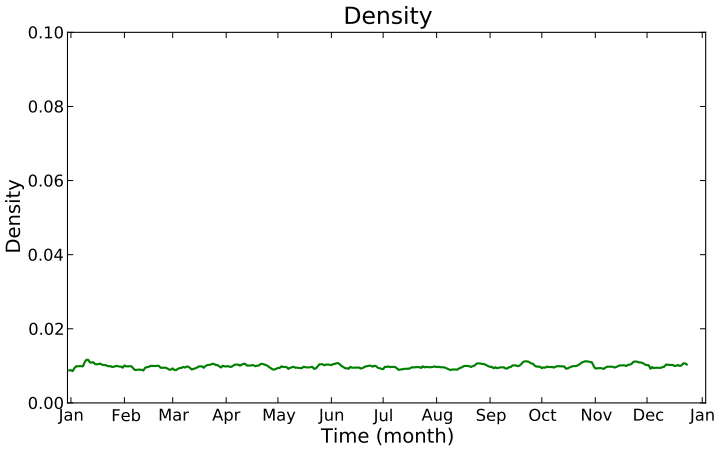
(c) In-component histogram average

(d) Out-component histogram average

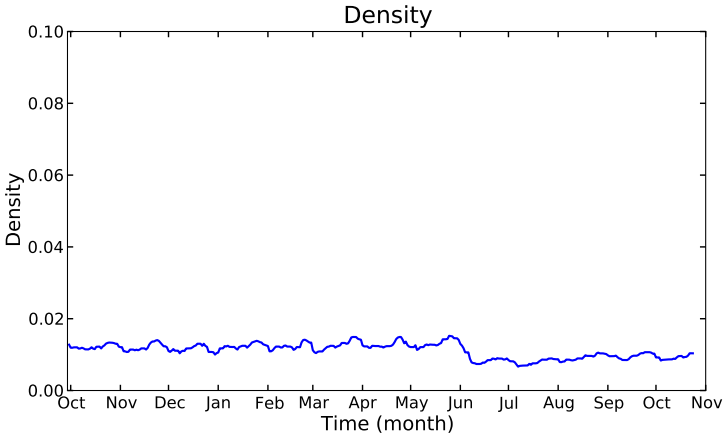
Figure 3.9: The in- and out-component size histograms. The data used is from March 2002.



(a) Crisis 1 (1998-1999)

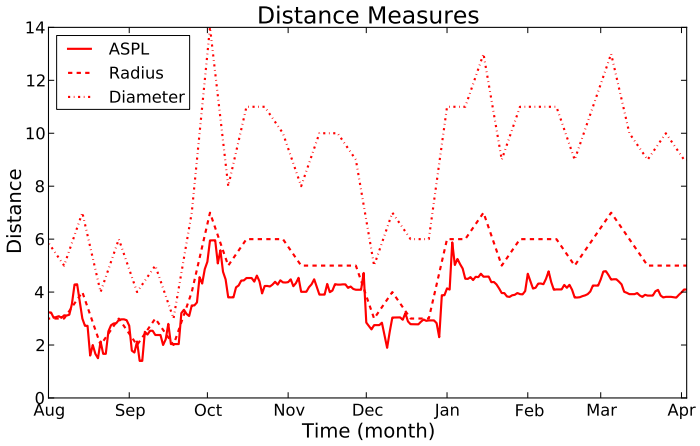


(b) Normal (2001)

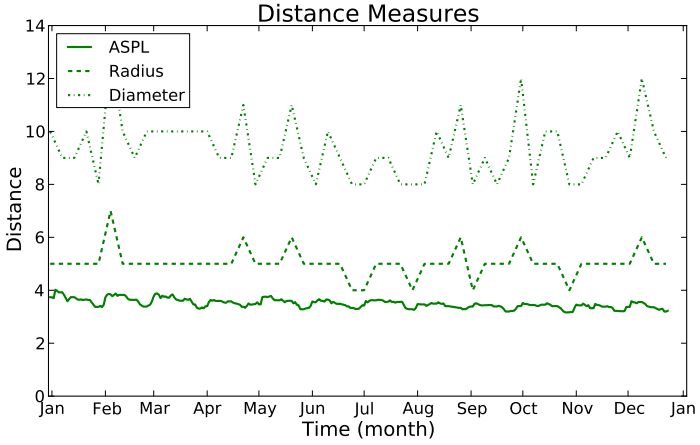


(c) Crisis2 (2003-2004)

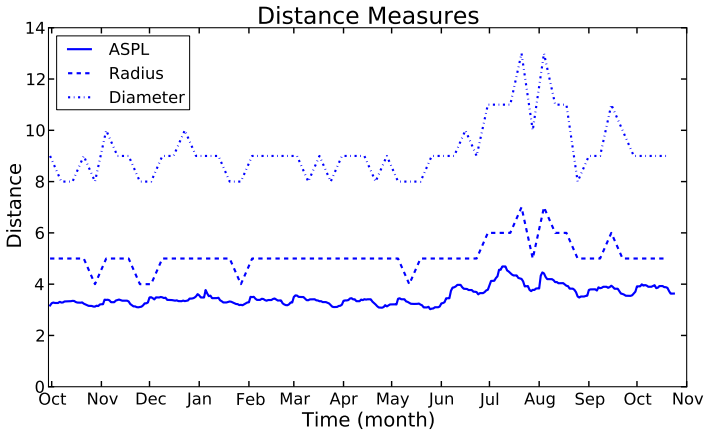
Figure 3.10: The moving weekly average of the density coefficient, shown for the three periods of interest.



(a) Crisis 1 (1998-1999)

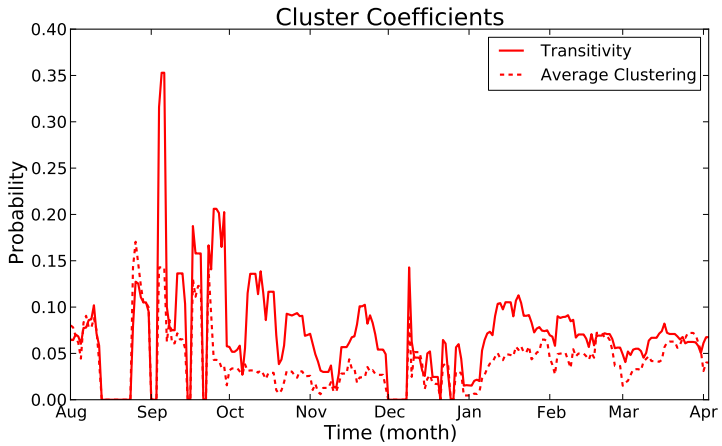


(b) Normal (2001)

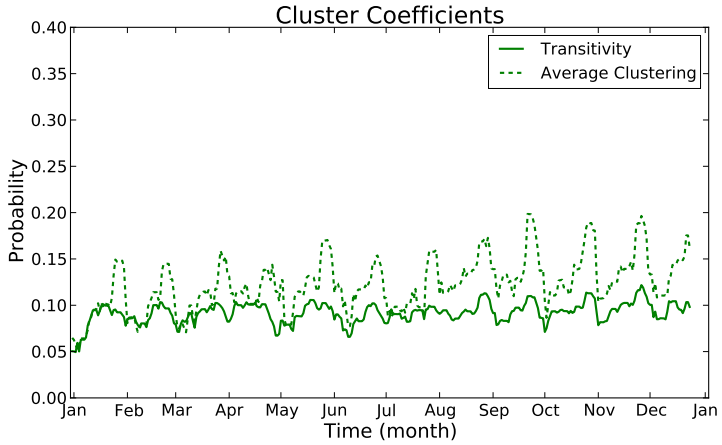


(c) Crisis 2 (2003-2004)

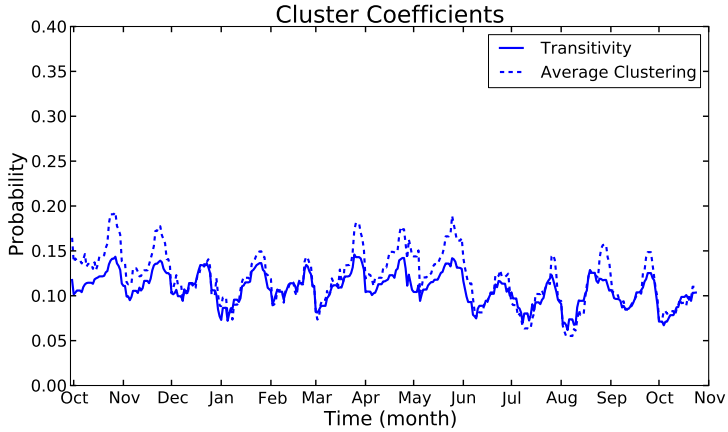
Figure 3.11: The ASPL, radius and diameter, shown for the three periods of interest. For the ASPL the moving weekly average is displayed, while for the radius and diameter skip-varying was used.



(a) Crisis 1 (1998-1999)

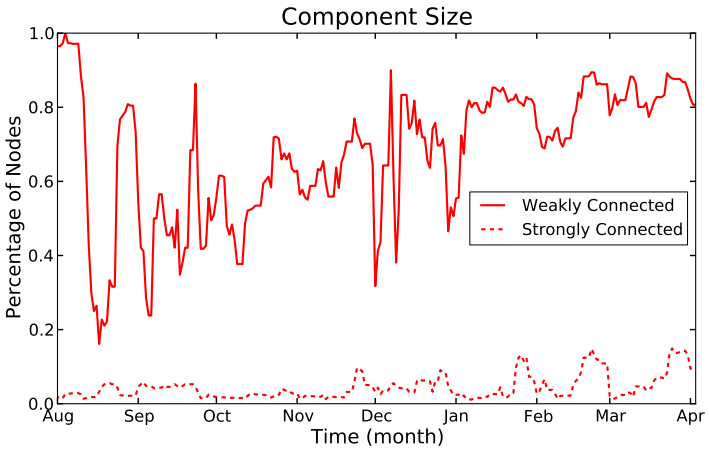


(b) Normal (2001)

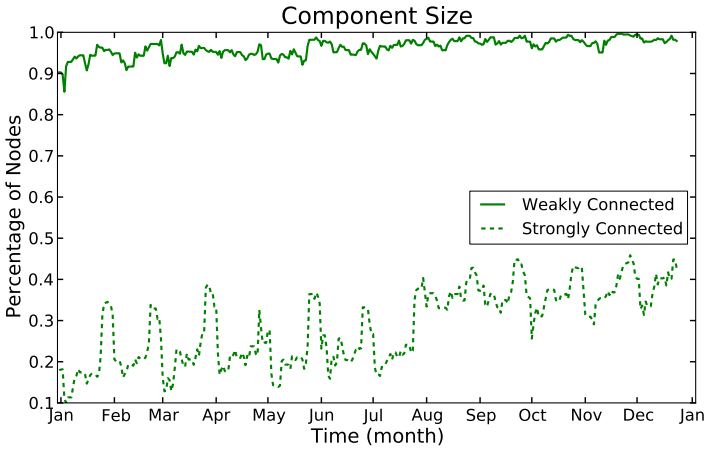


(c) Crisis 2 (2003-2004)

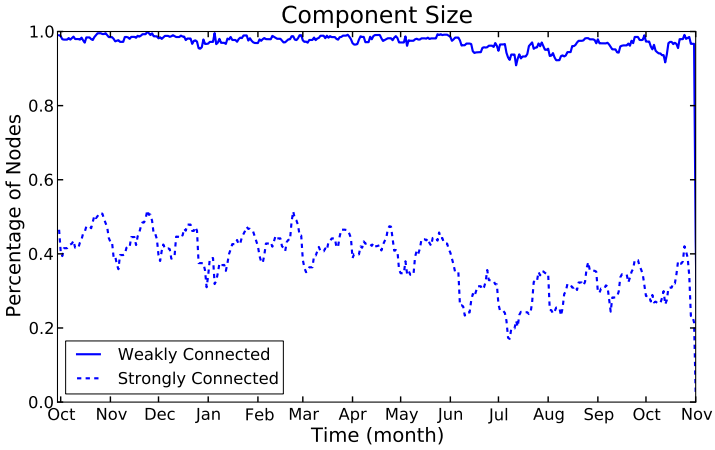
Figure 3.12: The moving weekly averages of the transitivity and the average clustering coefficient, shown for the three periods of interest.



(a) Crisis 1 (1998-1999)

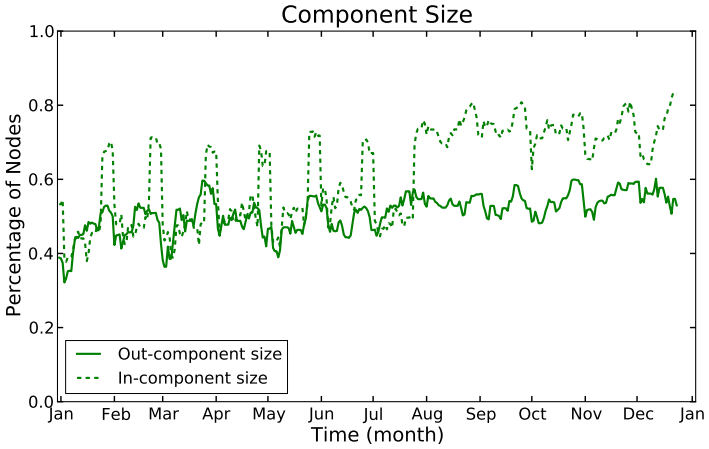


(b) Normal (2001)

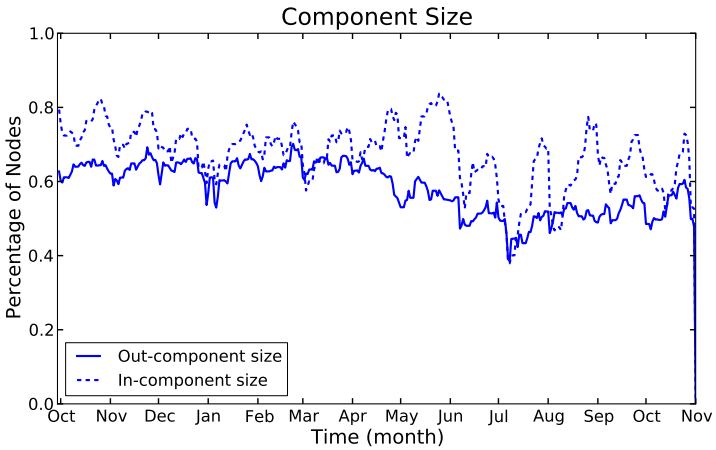


(c) Crisis 2 (2003-2004)

Figure 3.13: The moving weekly average of the size of the strongly and the weakly connected clusters, shown for the three periods of interest.



(a) Normal (2001)



(b) Crisis 2 (2003-2004)

Figure 3.14: The moving weekly average of the size of the large in- and out-component. The first crisis period is not included because there was no clear dichotomy between a small and a large in/out component.

Chapter 4

Node Distributions

In the next two chapters we focus on distributions. We discriminate between measures that can be attributed to nodes and measures concerning edges[1]. The edges will be discussed in Chapter 5 and the nodes in this Chapter.

4.1 Degree Distributions

To begin with we consider degree distributions, one of the most important and intuitive measures of a network. We analyze the undirected, directed and multi-directed versions and find that all of them can be discribed by the same distribution.

As mentioned before the degree of a node in a network is the number of edges connected to that node. p_k is defined as the fraction of nodes in the network that have degree k. Equivalently, p_k is the probability that a node chosen uniformly at random has degree k. For any given network a plot can be made by making a histogram of the degrees of the nodes. This histogram is the degree distribution for the network. For directed graphs each node has both an in-degree and an out-degree. Therefore the degree distribution becomes a function p_{jk} of two variables, representing the fraction of nodes that simultaneously have in-degree j and out-degree k [33].

To give an example of a degree distribution, lets consider the simple network model of uniform random graphs[2]. To construct a uniform random graph, one takes a number n of nodes and connects each pair (or not) with probability p (or $1-p$)[3]. Then the probability p_k for large n of a node having degree k is

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \backsimeq \frac{z^k \exp^{-z}}{k!}, \tag{4.1}$$

[1] An edge, e.g., does not have a degree and a node does not have a contract size.
[2] In mathematics a network is called a graph.
[3] In the light of Chapter 6 this will be called bond percolation on a complete network.

with the mean degree $z = p(n-1)$. From Eq. (4.1) we conclude that a uniform random graph has a Poisson degree distribution [33].

Another important type of network is the scale-free network which is defined as a network having a power-law degree distribution

$$p_k \sim Ck^{-\alpha}, \tag{4.2}$$

where C is a normalization constant and α is a parameter whose value is typically in the range $2 < \alpha < 3$. These scale-free networks seem to abound the real world. They appear in social networks with e.g. the famous film actor network [3]; in information networks with e.g. the well-investigated citation network [39]; in technological networks with e.g. the ubiquitous internet [9]; and in biological networks with e.g. the vital metabolic network [23].

4.1.1 Previous Research

What did the abovementioned studies of Refs. [6] and [11] and have to say about the degree distributions of the interbank network which was studied?

The paper of Boss et al. [6] studied the undirected as well as the directed versions of the Austrian interbank network. The data used to make the plots in Figs. 4.1 and 4.2 is aggregated from 10 quarterly single month periods between 2000 and 2003. We also note that an estimation routine based on local entropy maximization has been used to complete the data set. Figs. 4.1 and 4.2 show the histograms of the degree distributions. Fig. 4.1 considers the undirected network and Figs. 4.2a and 4.2b respectively the out-degree and in-degree of the directed network. In all three cases Boss et al. find two regions which can be fitted with a power-law. The tail region has a steeper slope than the region with the small degrees.

The other paper of Cont et al. [11] only studied the directed interbank network. The data used to make the plots in Fig. 4.3 are all interbank transactions of June 2008. Figs. 4.3a and 4.3b show the cumulative distribution function (cdf) of respectively the out- and in-degree data. Again one chooses to use a power-law to fit the data, but now only the tail was considered.

For a power-law distribution $P(x)$ (exponent α) the cdf is also a power-law with exponent $\alpha + 1$. Using this, we can compare the findings of Refs. [6] and [11] but only for the directed networks and in the tail part. Note that other methods and data intervals were used by these two papers while fitting, so this comparison is only qualitatively. The results are displayed in Table 4.1.

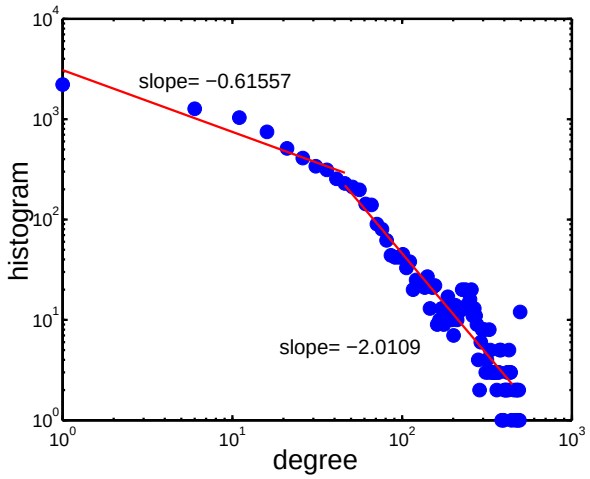
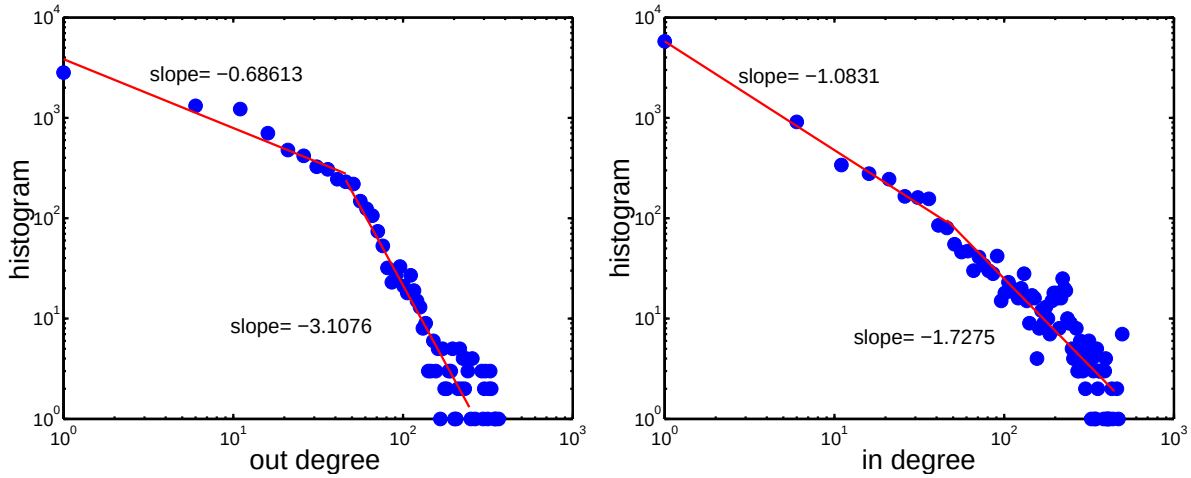


Figure 4.1: The histogram of the degree distribution for the undirected network. This figure was reproduced from Ref. [6].

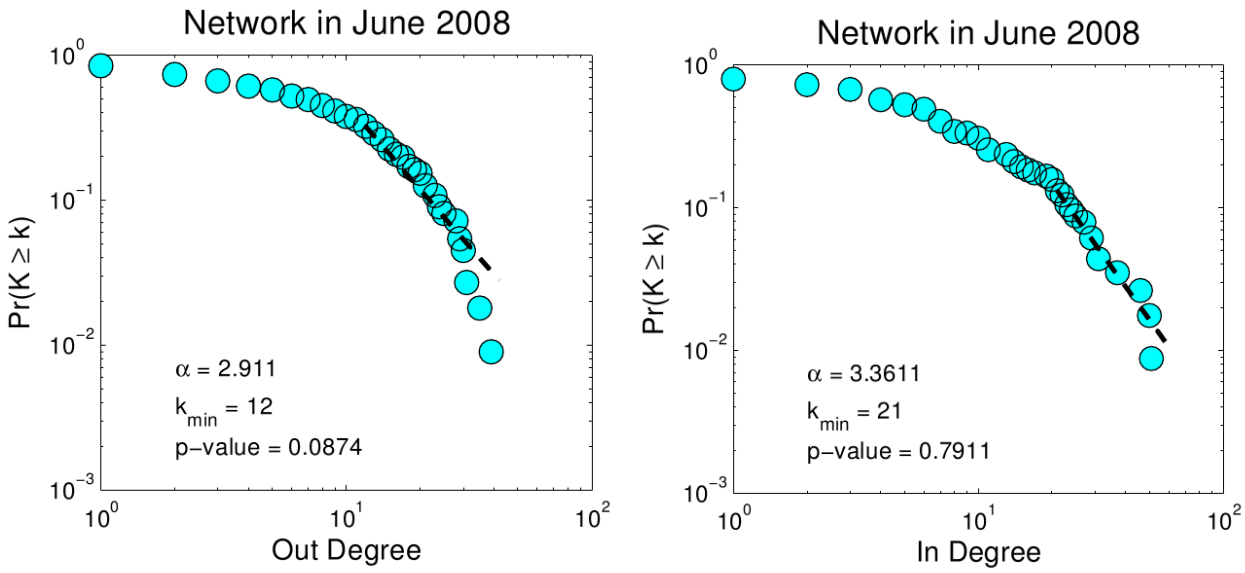


(a) Out-degree **(b)** In-degree

Figure 4.2: The histgrams with the in- and out-degree distribution for the directed network. This figure was reproduced from Ref. [6].

Table 4.1: A comparison of the exponents of the power-laws, which were fitted to the directeed degree distributions of the tails in Refs. [6] and [11].

	Ref. [6]	Ref. [11]
out-degree tail	-3.1067	-1.911
in-degree tail	-1.7275	-2.3611



(a) Out-degree **(b)** In-degree

Figure 4.3: The first plot (a) shows the cumulative distribution for the out-degree and the second plot (b) shows the in-degree cumulative distribution. The data in these plots is aggregated from the entire month of June 2008. This figure was reproduced from Ref. [11].

4.1.2 Undirected Edges

In this section we will discuss in detail the adopted fitting procedure.

Fitting data via the cumulative distribution function

After extracting the degree of every node from the undirected network, we wish to quantify this information in a convenient way and to find out if there is some underlying theoretical distribution. In order to achieve this, one can use a (normalized) histogram of the probability p_k. Next, one can try to fit an appropriate pdf to this data. Alternatively, one can plot the cumulative distribution of the data, defined as

$$P_k = \sum_{k'=k}^{\infty} p_{k'}, \tag{4.3}$$

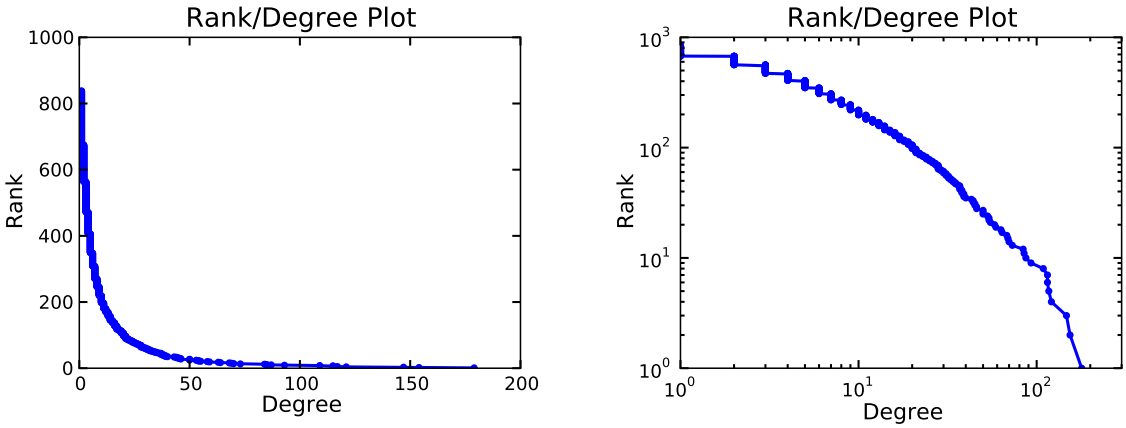
which gives one the probability that a randomly chosen node has a degree larger than or equal to k. This representation of the data enables one to fit the cdf of some theoretical distribution to this data. This is a small sidestep but a very convenient one in our case. Indeed, we can identify a distribution through its pdf as well as through its cdf.

When we make a conventional histogram by binning, any differences between the values of data points that fall in the same bin are lost. In contrast, when we use the cdf, one does

not need to bin and therefore no information is lost. The cdf also reduces the noise in the tail. On the downside, the plot does not give a direct visualization of the degree distribution itself, and adjacent points on the plot are not statistically independent, making correct fits to data tricky [33].

Constructing the cumulative distribution

We wish to construct the cumulative distribution P_k of the degree distribution in the interbank network. To this end we will use a so-called rank/frequency plot, or for this case better coined a rank/degree plot. Here, all we need to do is to sort the nodes in decreasing order of degree, number them giving the node with the highest degree number one, and then plot their ranks as a function of their degree [34]. Instead of plotting the cumulative distribution of the degrees, defined in other words as the *fraction* of nodes with degree more than or equal to k, we plot the *number* of nodes with degree greater than or equal to k. The latter differs from the former only in its normalization: the cumulative distribution P_k is proportional to these rank/degree plots. In Fig. 4.4 we see a rank/degree plot for the undirected network which uses as aggregate the first quarter of the 2001. A linear scale (a) as well as a log-log scale (b) was used.



(a) linear plot **(b)** log-log plot

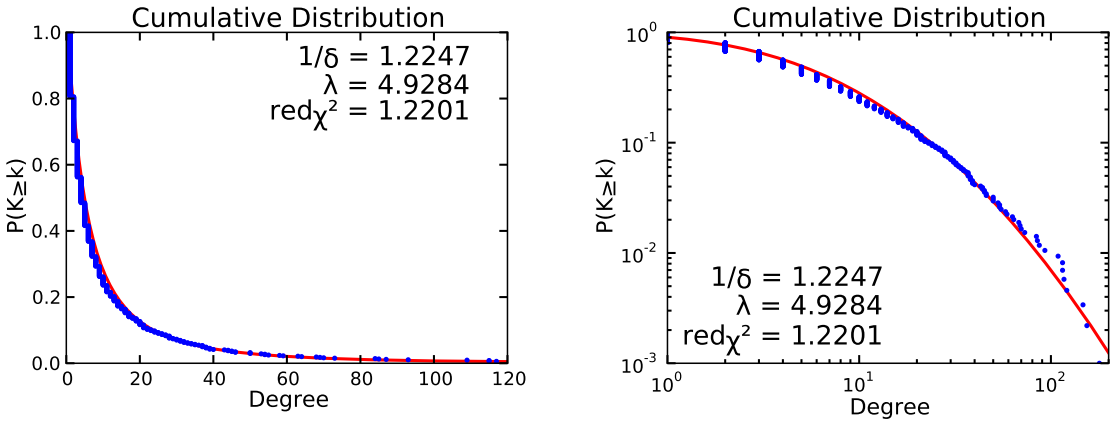
Figure 4.4: A rank-degree plot for the undirected network using the data of the first quarter of 2001.

Fitting the data with a log-normal

After normalizing the rank/degree plot from Fig. 4.4, one can plot the cumulative distribution of the degree data. In Fig. 4.5 we did this and compared to Fig. 4.3[4] we find a similar shape.

[4]Although it does not plot exactly the same measure, it is our only lead on an empirical cumulative degree distribution.

Cont et al. [11] used a power-law to fit the distribution, probably inspired by the interest of the network community for scale-free networks. We believe that a log-normal distribution suits better. If a quantity X is log-normally distributed, then $log(X)$ is normally distributed. For more information on log-normal distributions we refer the reader to Appendix 8.1. We do find a neat fit in Fig. 4.5 with a reduced χ^2-value of 1.2201. The best fit parameters are also included in the figure. One concludes that the degree distribution for the undirected representation of the domestic Russian interbank network, aggregated from the first quarter of the year 2001, is log-normal distributed. We note that we fit a continuous distribution to discrete data. This is appropriate in view of the fact that the data covers three orders of magnitude.



(a) linear plot **(b)** log-log plot

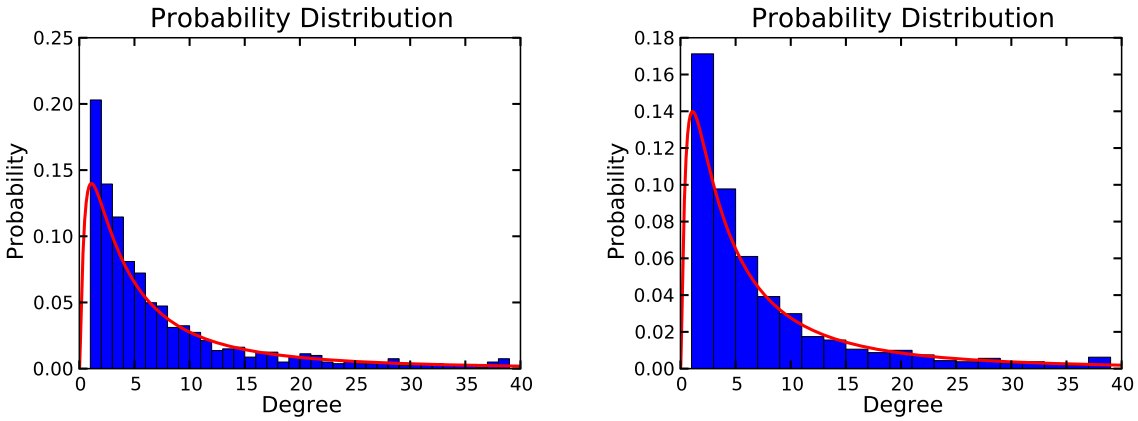
Figure 4.5: A cumulative distribution for the undirected network fitted with a log-normal and using the data of the first quarter of 2001.

Looping back to the degree distribution

Now we return to the pdf by plotting the degree data in a different way. We try to visualize the probability density by plotting the normalized histogram of this data. The red line in Fig. 4.6 is the log-normal pdf, which is drawn with the parameters obtained in the fit of Fig. 4.5[5].

If we take a closer look at the distribution in Figs. 4.6 and 4.7 the fit is not that good. In the range of degrees from 1 to 5 the real probability exceeds the theoretical log-normal and vice versa for the range of degrees from 5 to 25. This could indicate that we chose the wrong distribution to use as a fit. A less elegant explanation could be that the region for

[5]In general it would be more useful to use logarithmic binning and a logarithmic scale, but to state the discussion in this section a linear scale was necessary. When we bin the pdf of the loan size distribution, e.g., the logarithmic case will be used.



(a) bin width 1 **(b)** bin width 2

Figure 4.6: A degree histogram for the undirected network fitted with a log-normal pdf and using the data of the first quarter of 2001.

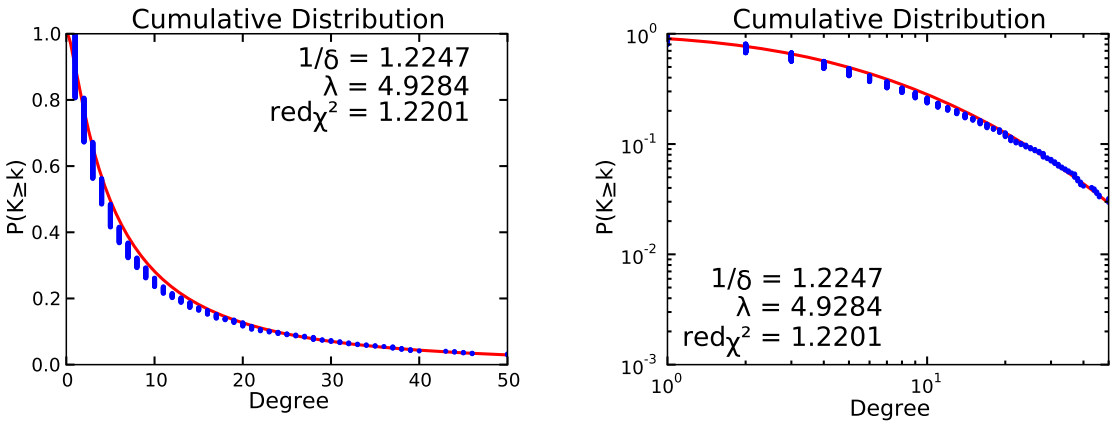
small degrees does not follow a log-normal and only the tail region can be considered to be log-normally distributed. This would mean we have to split up the degree-axis into separate regions. In the following section we will look for a better suited distribution, if there is any.

We note that there is an important condition which a distribution should fulfill to get elected as a degree distribution in an undirected network. The chance to find a node with degree equal to zero should be equal to zero. A node with a degree equal to zero is not connected to any node in the network, and is not part of the network at all. A log-normal distribution satisfies this condition. A power-law degree distribution predicts an infinite amount of nodes with zero degree, which makes no sense. Therefore power-law degree distributions cannot cover the entire degree range for undirected networks.

Fitting the data with a Johnson S_B

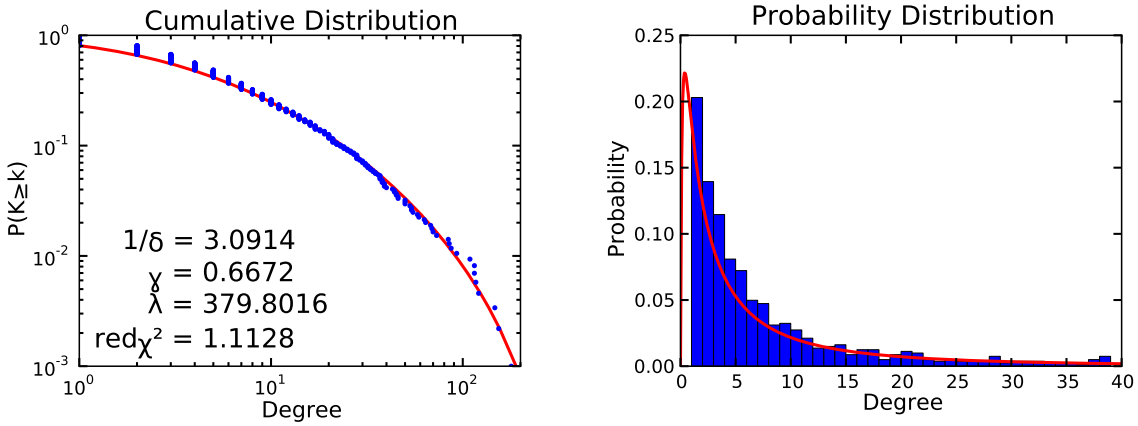
Our goal here is to find a better fit for the degree distribution of the undirected network. In the SciPy library [25] a more appropriate candidate was found, the so-called Johnson S_B distribution. This distribution was developed in biometrics [24] and describes e.g. the distribution of tree heights and diameters, and of leaf area in forest stands[37]. We refer to Appendix 8.2 for more information and a complete definition of this distribution.

In Figs. 4.8a and 4.8b the same data is shown as respectively in Figs. 4.5b and 4.6a. If we fit a Johnson S_B to this data, it has a reduced χ^2-value of 1.1128. The Johnson S_B also describes the bins with smaller k-values better. The difference both fits may not be that big for the currently investigated undirected network, it certainly is for the directed and multi-directed network degree distributions. We will return to this comparison in the Section 4.1.4 where the discrepancy between the two fits will become clearer. So we state a new conclusion:



(a) linear plot **(b)** log-log plot

Figure 4.7: A zoom in on Fig. 4.5.



(a) cdf **(b)** pdf

Figure 4.8: A cumulative distribution and a degree histogram for an undirected network, both fitted with a Johnson S_B distribution and using the data of the first quarter of 2001.

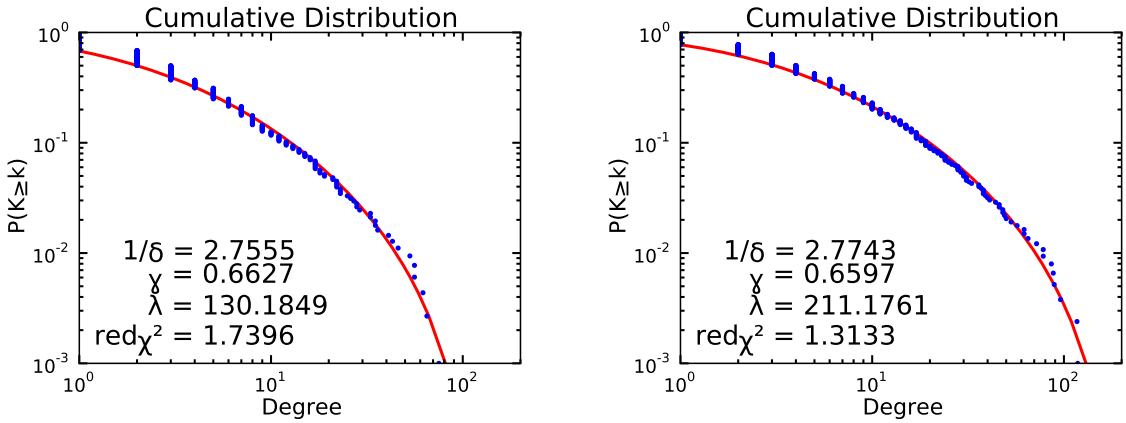
the degree distribution for the undirected representation of the domestic Russian interbank network, aggregated from the first quarter of the year 2001, is Johnson S_B distributed.

This fit is better than the log-normal because there is an additional term in the exponential of the pdf which is linear in $logX$. This term provides the ability to tilt the distribution in order to better align with the data points. To see the effect of this extra term we refer to Appendix 8.1.

Different aggregates

We found that that the degree distribution for the undirected representation of the domestic Russian interbank network, aggregated from the first quarter of the year 2001, is Johnson S_B distributed. We now adress the question whether this conclusion can be generalized to different aggregates and other time periods. To this end we considered the year 2001 and we split it up into weeks, months and quarters and plotted for each of these time periods a figure similar as Fig. 4.8a.

Because it is impossible to show all these fits (more than 60), we selected two examples. In Fig. 4.9a the fit to the degree distribution of the first week of March 2001 is shown and in Fig. 4.9b the fit to the degree distribution of the whole of March 2001. As expected, one can see an increase in the reduced χ^2-value of the fits: the larger the aggregate, the more data one has, the better the fit.



(a) First week March 2001 **(b)** March 2001

Figure 4.9: Two examples of a Johnson S_B fit for different aggregates. (a) uses the data of the first week of March 2001 and (b) uses the data of the whole of March.

Upon inspection, all of the abovementioned weekly, monthly and quarterly degree distributions were neat fits. So we may conclude that this particular distribution is the universal one for the degree distribution of the undirected network representation in normal operation periods.

In Table 4.2 the average of the abovementioned fitted parameters is shown for the different aggregates. Fig. 4.10 visualizes these distributions accordingly. We find that larger aggregates have a more spread out distribution than smaller ones, as can be concluded from the scale parameter λ. We also added the distribution fitted to the data entire of the entire year 2001.

Table 4.2: The averaged Johnson S_B parameters for four different aggregates in the year 2001.

	shape γ	shape δ	scale λ
Week	2.65015402	0.69657197	91.43388955
Month	2.90981391	0.65001354	241.55709894
Quarter	3.13677831	0.67067210	404.05282522
Year	3.28687111	0.69574938	684.61439099

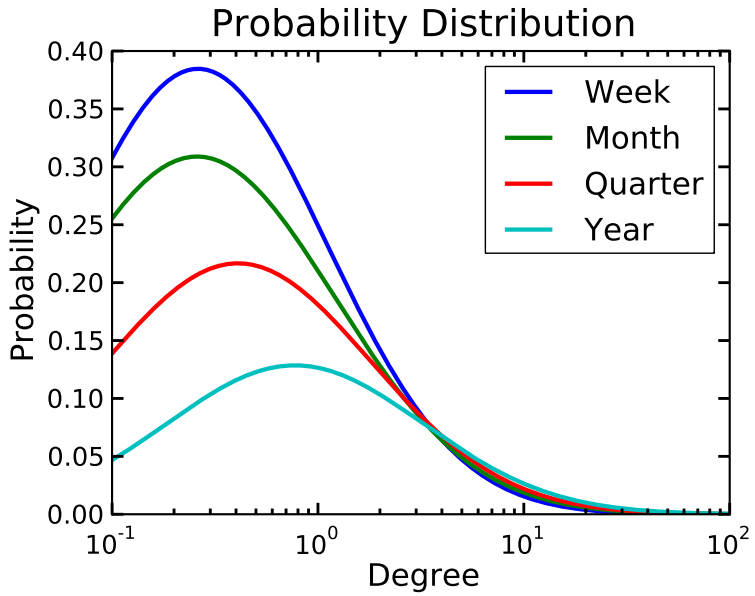


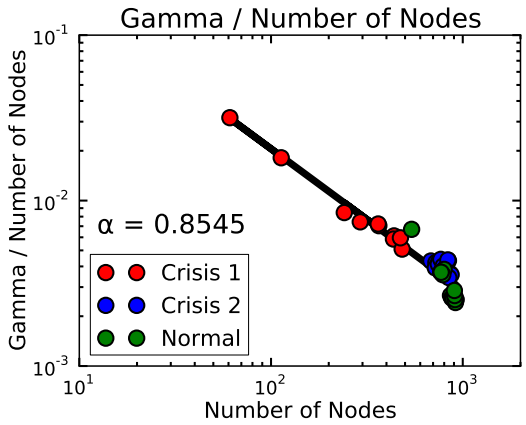
Figure 4.10: The average Johnsons S_B distributions for the different aggregates in the year 2001. We used a log scale for the x-axis where only the region for $k \geq 1$ is relevant for a degree distribution.

Parameter behaviour

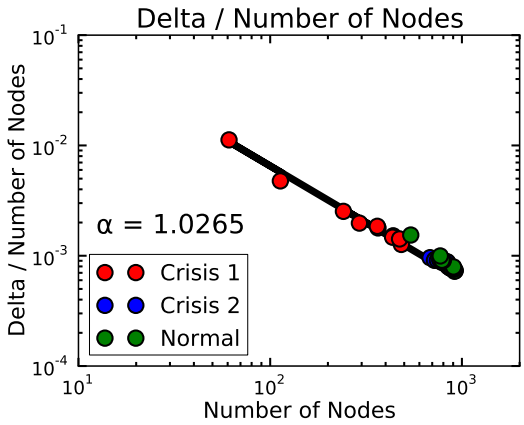
To end the discussion of the undirected degree distributions, we have a look at the variation and the dependence of the fitting parameters. We also broaden our spectrum and now include the two crisis periods next to the normal operating period. The values of the parameters used in this Subsection were obtained by fitting the undirected degree distribution to every monthly aggregate in the three periods of interest.

The first question we adress is if these parameters of all the monthly samples would happen to have an underlying dependence on the number of nodes and the number of edges. Figs 4.11 and 4.12 try to answer this question. In 4.11a we plot the ratio of the first shape parameter γ to the number of nodes in function of the number of nodes. Fig 4.11b does the same thing for the second shape parameter δ. Fig 4.11c represents the ratio of γ to the number of edges

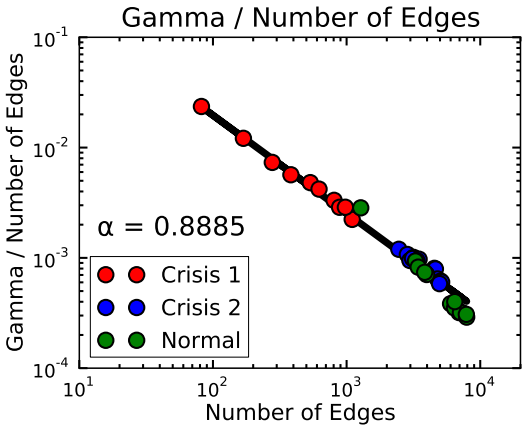
in function of the number of edges. Again the following Fig. 4.11d does the same thing for the second shape parameter δ. Each figure discriminates between the first crisis period, the second and the normal period (cfr. Section 2.1). We clearly get a power-law dependence in each of the four figures. The exponent α is included in every figure. What is also very interesting to note is that the three periods lie on the same power-law, this could mean that for the monthly aggregates there is no fundamental difference between stressed and normal periods when considering degree distributions. This power-law dependence can most likely be derived using theoretical statistics and using the underlying connection between the number of nodes/links in a network and the degree distribution of a network. But this is beyond the scope of this work.



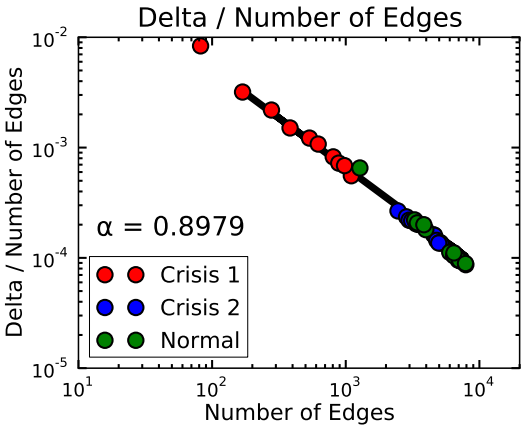
(a) Gamma and Number of Nodes



(b) Delta and Number of Nodes



(c) Gamma and Number of Edges

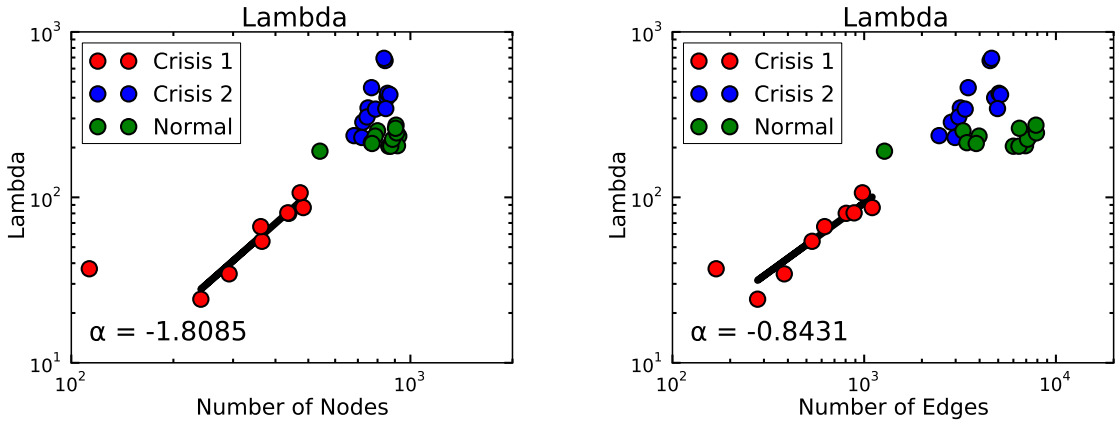


(d) Delta and Number of Edges

Figure 4.11: The dependence of the shape parameters on the number of nodes and edges for the monthly aggregates of the three periods of interest.

In Fig. 4.12 the dependence of the scale parameter λ is shown. Figs. 4.12a and 4.12b

show λ respectively in function of the number of nodes and edges. The data for the first crisis period could be said to follow a power-law[6], but for the other two periods we fail to find a dependence. This is probably because the scale parameter is very sensitive to the outliers of a sample. This means that the scale parameter will fluctuate strongly according to these outliers. If we would discard the outliers before fitting, we expect to find a relation between λ and the number of nodes or edges. This is because λ can be seen as a measure that quantifies the size of the network (cfr. Appendix 8.1).



(a) Lambda and Number of Nodes

(b) Lambda and Number of Edges

Figure 4.12: The dependence of the scale parameter on the number of nodes (a) and edges (b) for the monthly aggregates of the three periods of interest.

The next question that pops into mind is if there is any dependence between the number of nodes and the number of links[7] and also if there is any dependence between the first γ and second δ shape parameter. Fig 4.13a shows the number of edges (y) in function of the number of nodes (x) again for the monthly aggregates covering the three periods of interest. We find an exponential dependence with the parameters C and ζ defined as

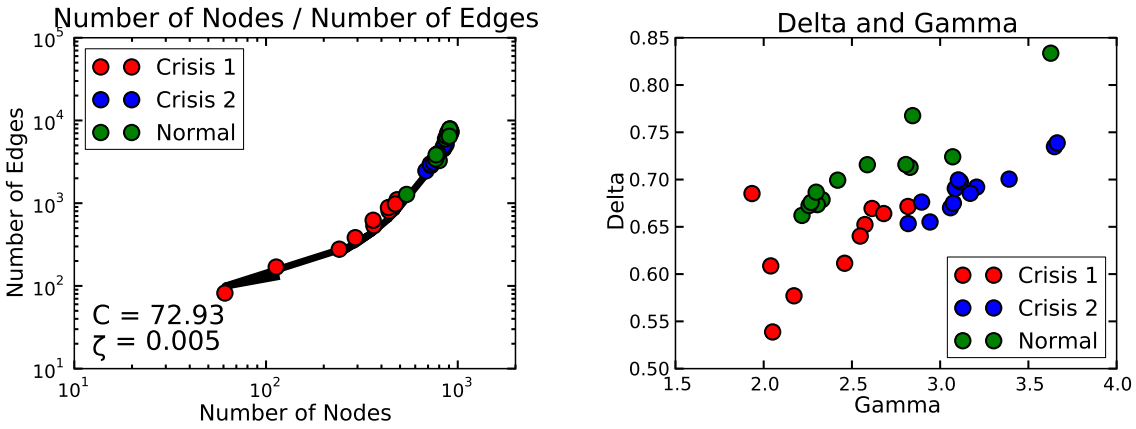
$$y = C \exp(\zeta x). \tag{4.4}$$

The other Fig. 4.13b shows the first shape parameter δ as a function of the second γ. We note that we used a linear scale for this figure. We find a strong positive correlation between γ and δ.

The discussion in this subsection, although interesting, was concise and will not be repeated for the other types of node or edge distribution due to a lack of time.

[6]Note that we excluded August 1998 for the fit.

[7]This discussion might have been more appropriate in the first chapter.



(a) Number of Nodes and Number of Edges

(b) Gamma and Delta (linear scale)

Figure 4.13: The dependence of the number of edges on the number of nodes (a) and the dependence of shape parameter δ on γ (b) for the monthly aggregates of the three periods of interest.

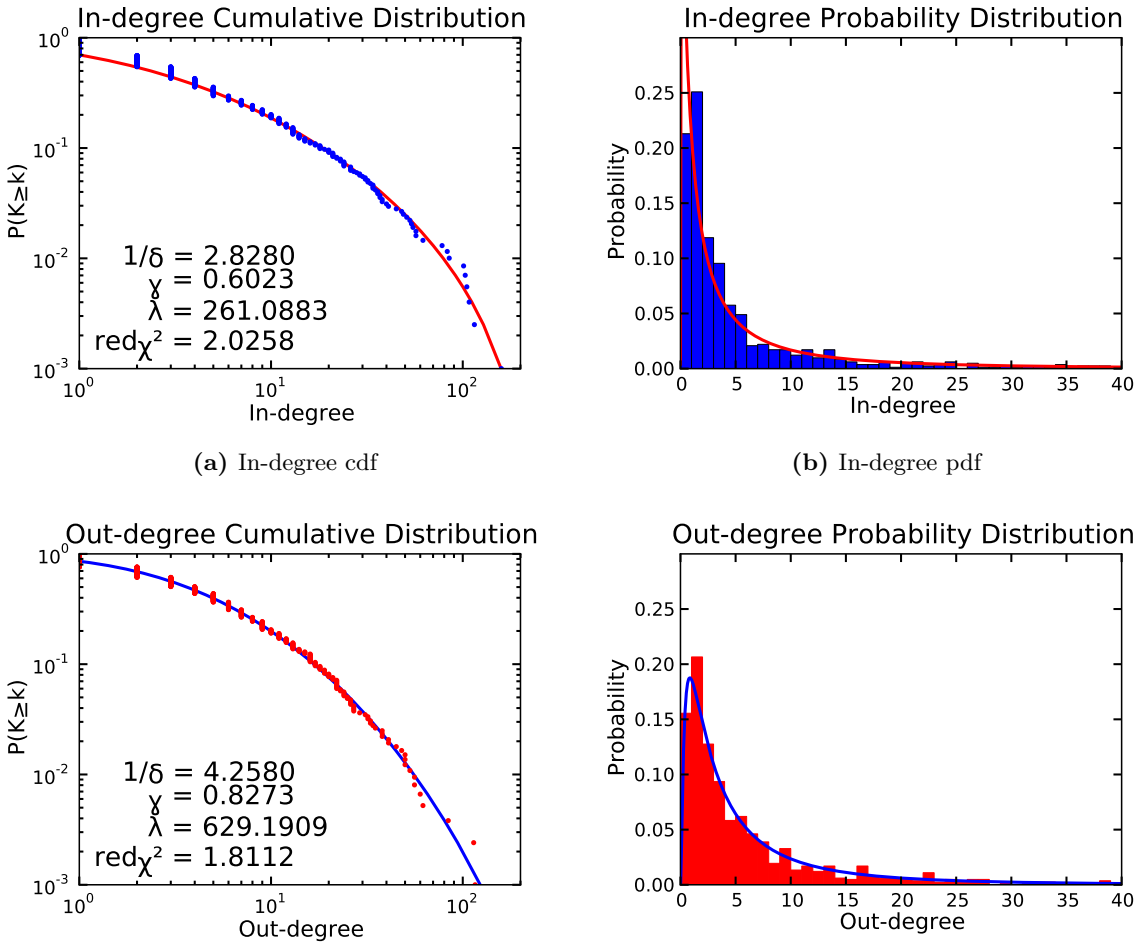
4.1.3 Directed Edges

In a directed network each node now has an in-degree and an out-degree, next to the regular degree which is the sum of both. In Fig. 2.5b one observes that the ratio of the number of directed links to the number of undirected links is close to one. So we expect no big difference between the total degree distributions in both network representations. Much more interesting is to investigate the in- and out-degree distribution and their mutual interdependence.

In Fig. 4.14 the in- and out-degree distributions represented as cdf and as pdf are shown for the data of the first quarter of 2001[8]. The Johnson S_B distribution was used to fit the in- as well as the out-degree distribution and again provided an accurate description. A big difference with the undirected case is that a node can have, e.g., an in-degree but no out-degree. This explains the fact that the bins for the in- and out-degree equal to zero of Figs. 4.14b and 4.14d are not empty. One sees that the Johnson S_B distribution actually fails to describe these nodes. The in- and out-degree distributions for the directed network representation for the first quarter of 2001 are Johnson S_B distributed. To generalize this conclusion to all aggregates and to the entire normal period, we use the same method as in the previous section and we find that the Johnson S_B distribution is also the universal one for the in- and out-degree distribution of the directed network representation in normal operation periods.

What is the difference between the in- and out-degree distributions? To investigate this we determine the in- and out-degree distribution fit of each month of 2001 and then plot the Johnson S_B distribution using the averaged monthly parameters which are tabulated in

[8]In this and the next chapter, we will always choose the first quarter of 2001 to visualize the pdf and cdf of the studied distribution. We do this to maintain some uniformity in the discussion.



(a) In-degree cdf

(b) In-degree pdf

(c) Out-degree cdf

(d) Out-degree pdf

Figure 4.14: The in- and out-degree distributions represented as cdf and as pdf for the directed version of our network using the first quarter of 2001 as aggregate. We used the Johnson S_B distribution to fit the data.

Table 4.3. Fig. 4.15 shows these average distributions together with the average total degree distribution for an undirected and a directed network. One clearly observes that the in-degree distribution is more concentrated on the smaller degrees, whereas the out-degree distribution is more spread out. This means that mostly an issuing bank lends to more counterparties than a receiving bank will have counterparties from which it lends. So we can conclude that there is a difference between the in- and out-degree distributions. The two normal degree distributions depicted in Fig. 4.15 further confirm that there is no real difference between the undirected and directed representations of our network concerning the total degree distribution.

The next step is to look if there is any correlation between the in- and out-degree of a node. We choose to only visualize this interdependence, but network theory has already developed

Table 4.3: The averaged Johnson S_B parameters for the monthly aggregates in the year 2001. We give the parameters for the undirected degree, directed degree, directed in-degree and directed out-degree distribution.

	shape γ	shape δ	scale λ
Undirected Degree	2.90981391	0.65001354	241.55709894
Directed Degree	3.05627462	0.65263928	342.10790048
Directed In-degree	2.77984460	0.55666054	172.34692396
Directed Out-degree	4.51177612	0.78325136	676.90433668

some measures to study this correlation quantitatively (cfr. Ref. [33]). Figs. 4.18, 4.19 and 4.20 show the aggregated data of the first quarter of 2001.

Fig. 4.18 is a two-dimensional histogram of the in/out-degree distribution. We used a bin width of two to keep the figure as readable as possible. Fig. 4.19 consists of three figures which are mutually connected. In the left bottom corner we have a scatterplot showing the in- and out-degree of the nodes. On top and on the right we have 1-dimensional histograms respectively showing the in- and out-degree distribution of our directed network. The colour red represents the top 20 percent of the nodes with the highest total degree. The scatterplot can be seen as the projection along the z-axis of Fig. 4.18. The upper histogram can then be seen as the projection along y-axis and the histogram on the right as the projection along the x-axis, both also of Fig. 4.18. We use Fig. 4.20 to investigate the high-degree nodes. This figure plots the same three elements as Fig. 4.18 but now with only the top 20 percent nodes. Considering these three figures, we find a positive correlation among the in- and out-degree of a node.

4.1.4 Multi-directed Edges

The last degree distribution we consider is the one from the multi-directed version of the interbank network. We plot the cumulative in- and out-degree distributions in Fig. 4.16 and meanwhile try to convince the reader again that the log-normal distribution is an inferior choice. Fig. 4.16a and Fig. 4.16b have a Johnson S_B as fit whereas Fig. 4.16c and Fig. 4.16d have a log-normal as fit. One clearly sees that the reduced χ^2-values are closer to one for the Jonhson S_B than for the log-normal fit. This implies that the former is a better fit-candidate than the latter. Again we find, using the method of section 4.1.2, that the Johnson S_B is a good fit for different aggregates and for the entire normal period. We also looked at the correlation between the in- and out-degree of a node and again found a positive correlation.

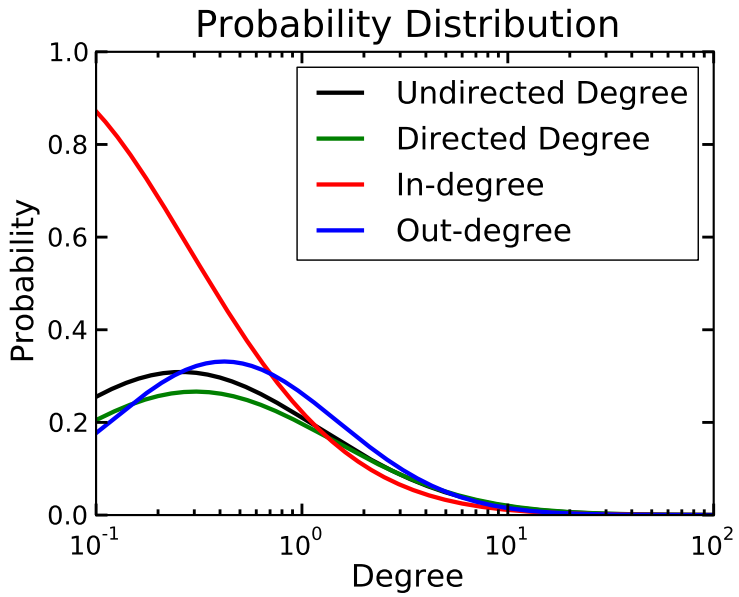


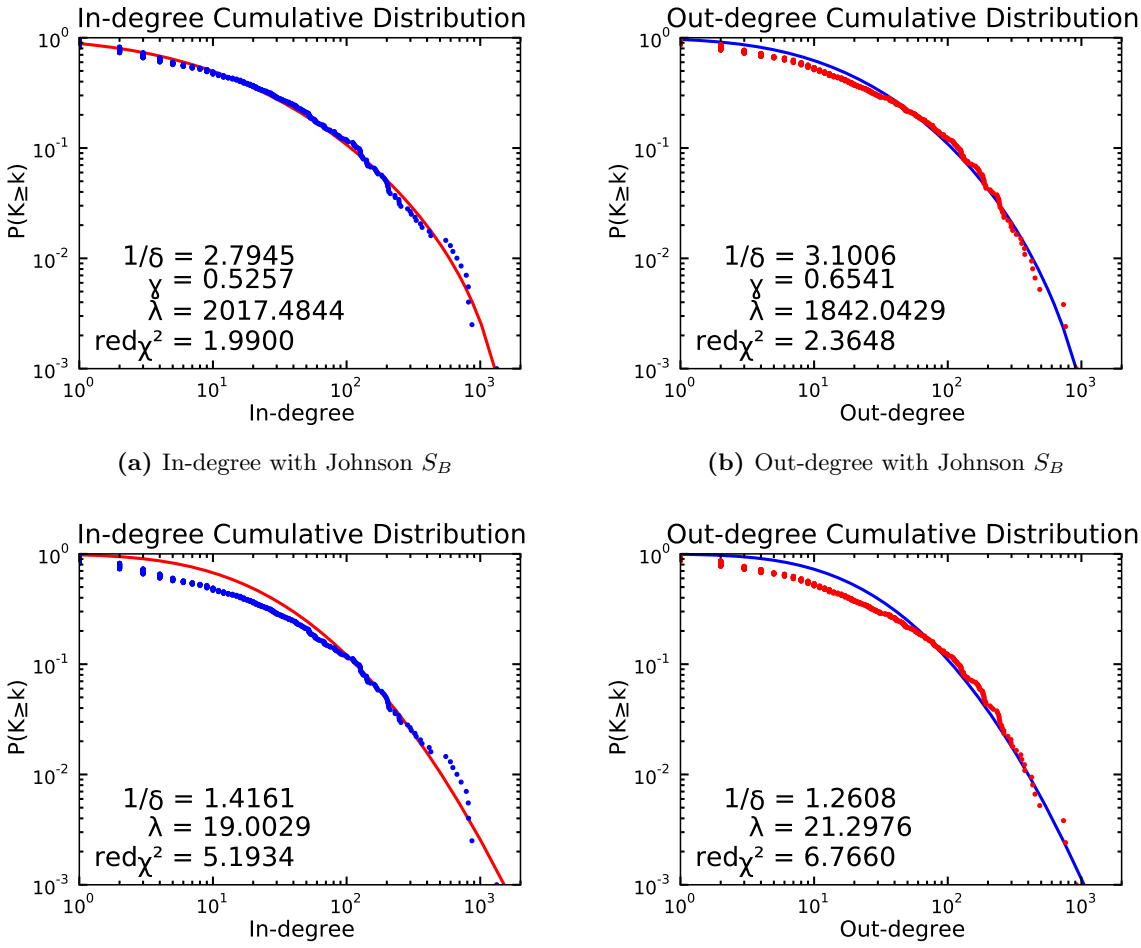
Figure 4.15: The averaged Johnson S_B distributions for the monthly aggregates in the year 2001. We show the undirected total degree, directed total degree, directed in-degree and directed out-degree distribution. We used a log scale for the x-axis where only the region for $k \geq 1$ is relevant for a degree distribution.

4.2 Strength Distribution

In Section 4.1 we studied unweighted representations of the interbank network. We now introduce weighted edges. The strength of a node is nothing more than a kind of weighted degree. The in-strength represents the total of received money in the aggregate period for a node and the out-strength represents the total of issued money in the aggregate period for a node.

In Fig. 4.17 the cumulative distribution of the in- and out-strength is shown for the aggregated data of the first quarter of 2001. We see that the Johnson S_B distribution is a good candidate to describe this data. Using the method from section 4.1.2 we are able to generalize our findings. We note that in contrast, e.g., to the degree distributions, of Figs. 4.16a and 4.16b the strength distributions span seven orders of magnitude. Another important difference with the degree distributions is that we now have a quasi-continuous distribution, due to the non-discreteness of the loan sizes.

We also visualize the interdepedence of the in- and out-strength of a node. Fig. 4.21 shows a two-dimensional histogram of the strength distribution using two as bin width. When comparing this to Fig. 4.18, the two-dimensional histogram of the degree distribution, we now have a negative correlation between the in- and out-strength of a node in the area shown. A bank, e.g., which is an issuer in the area shown in the figure, will not likely be a receiver



(a) In-degree with Johnson S_B

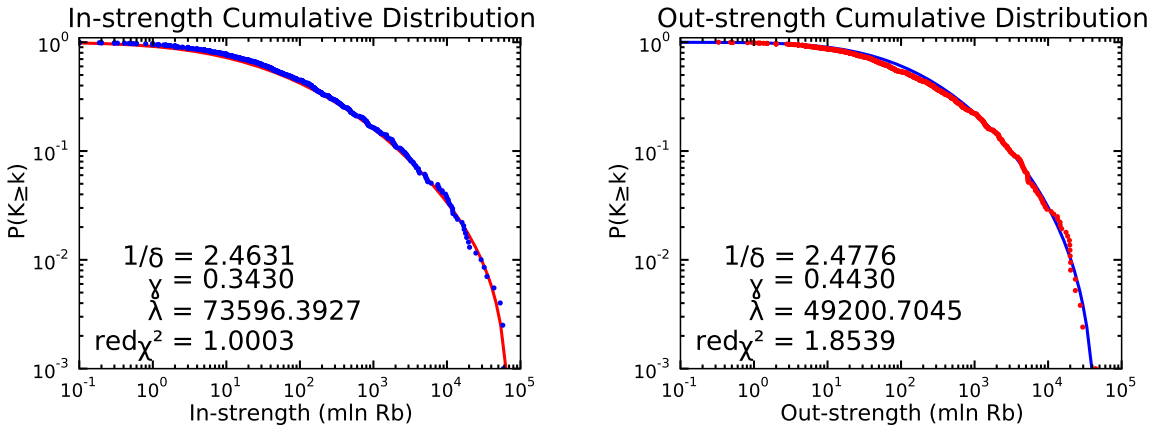
(b) Out-degree with Johnson S_B

(c) In-degree with log-normal

(d) Out-degree with log-normal

Figure 4.16: The in- and out-degree distributions for the multi-directed representation of our network. We use a Johnson S_B as well as a log-normal distribution as fit for comparison. The data used is from the first quarter of 2001.

and vice versa. Fig. 4.22 corroborates this observation. Whereas Fig. 4.19 showed a positive correlation between the in- and out-degree of a node, Fig. 4.22 leads to the conclusion that there is a rather negative correlation between the in- and out-strength of a node.



(a) In-strength

(b) Out-strength

Figure 4.17: The in- and out-strength cumulative distributions acquired from the weighted directed network and fitted with a Johnson S_B distribution using the data of the first quarter of 2001.

4.3 Overview

In this Chapter, the distributions of two important node attributes were studied. We investigated the degree distribution of the three unweighted representations and the strength distribution. We found they all can be described quite well with a Johnson S_B distribution. Alongside, we saw a positive correlation between the in- and out-degree of a node and a rather negative correlation between the in- and out-strength of a node.

There are many other node attributes of which the distributions could be studied, like, e.g., the k-core, or the local clustering coefficient of the nodes (cfr. Chapter 3). We choose not to treat them but the methods used here could easily be extended to these other measures.

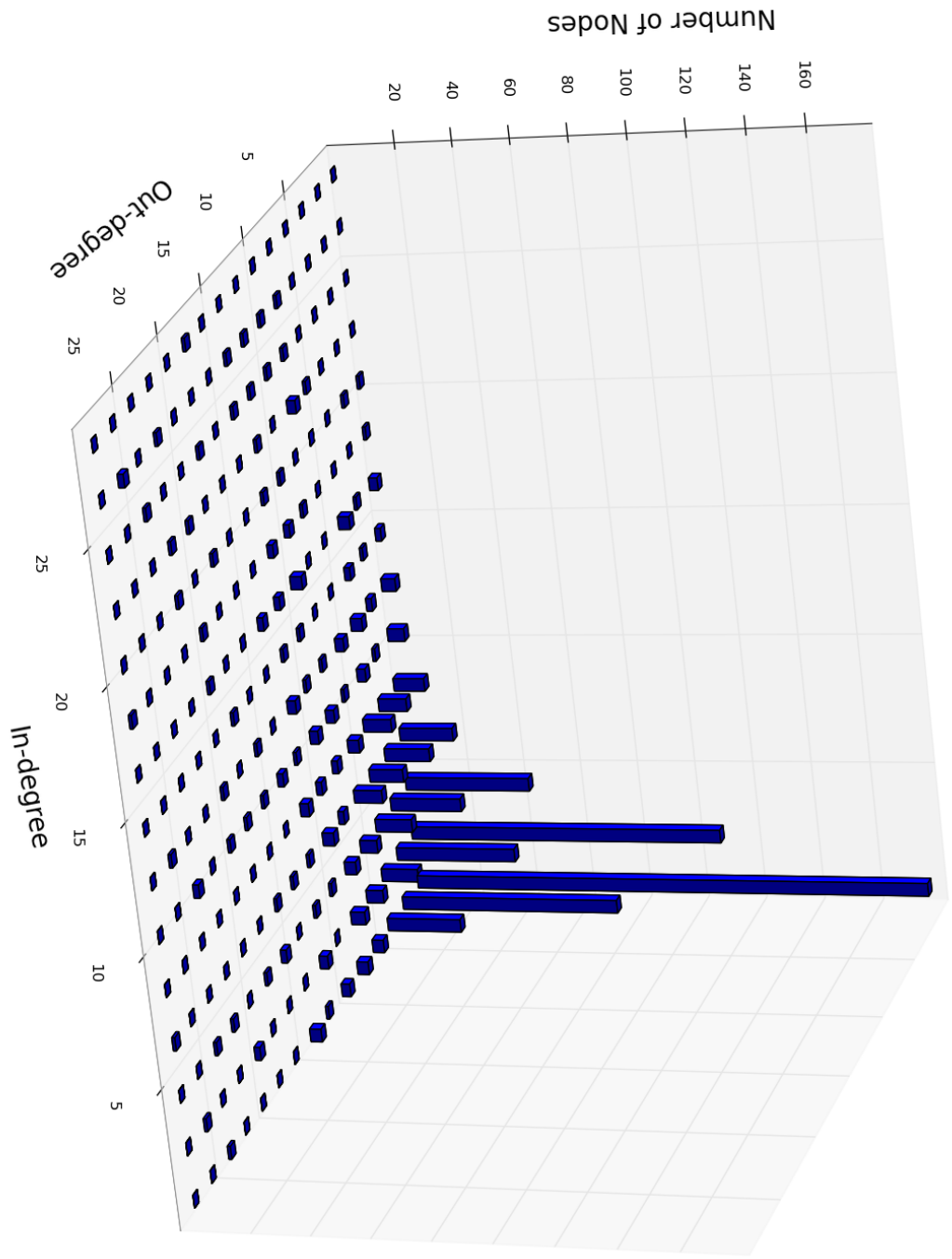


Figure 4.18: A two-dimensional histogram of the degree distribution of a directed network for the first quarter of 2001 using two as bin width.

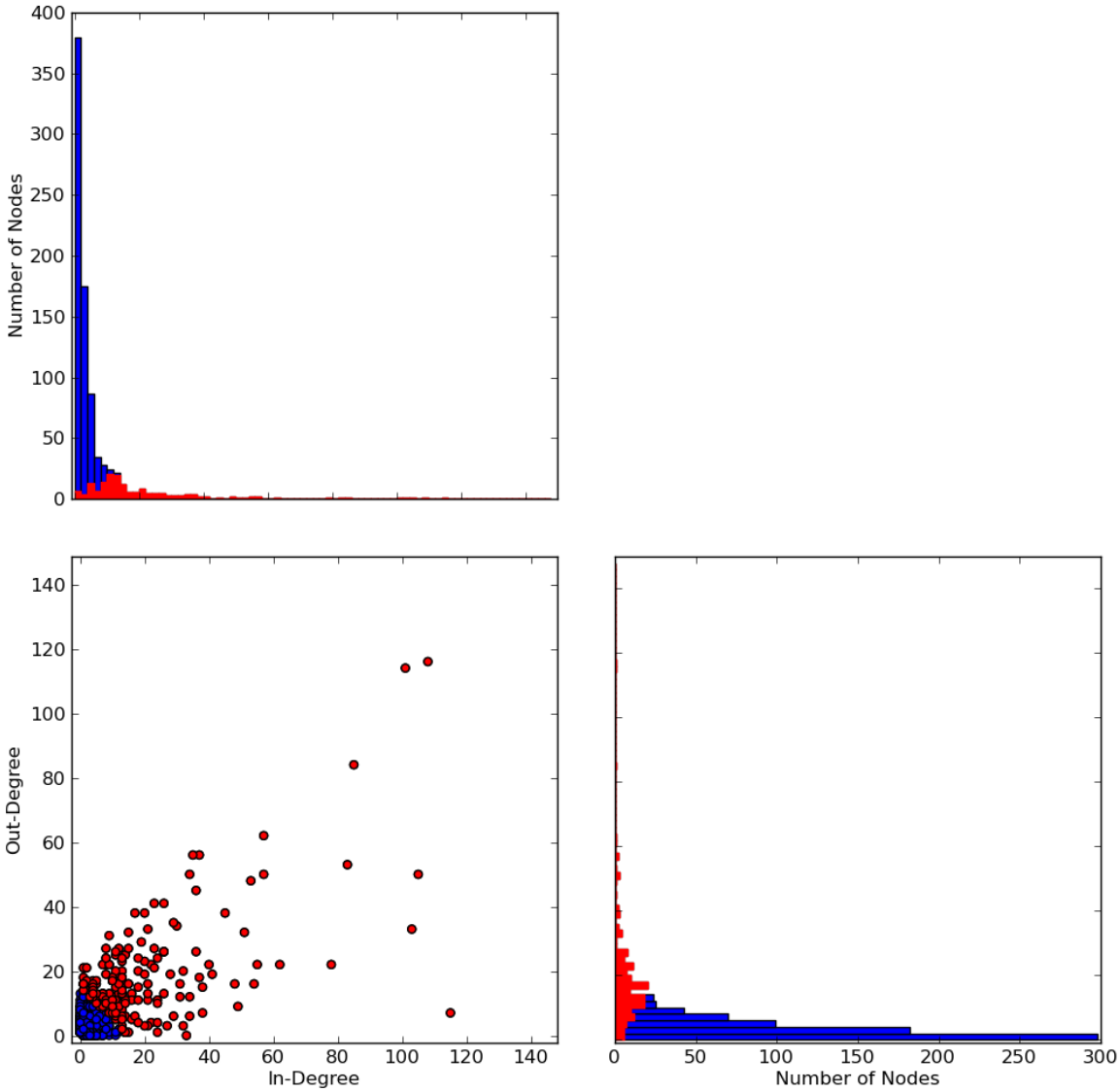


Figure 4.19: In the left bottom corner we have a scatterplot showing the in- and out-degree of the nodes. On top and on the right we have 1-dimensional histograms respectively showing the in- and out-degree distribution of the unweighted directed network. The colour red represents the top 20 percent of the nodes with the highest total degree. The data used is the aggregated data of the first quarter of 2001.

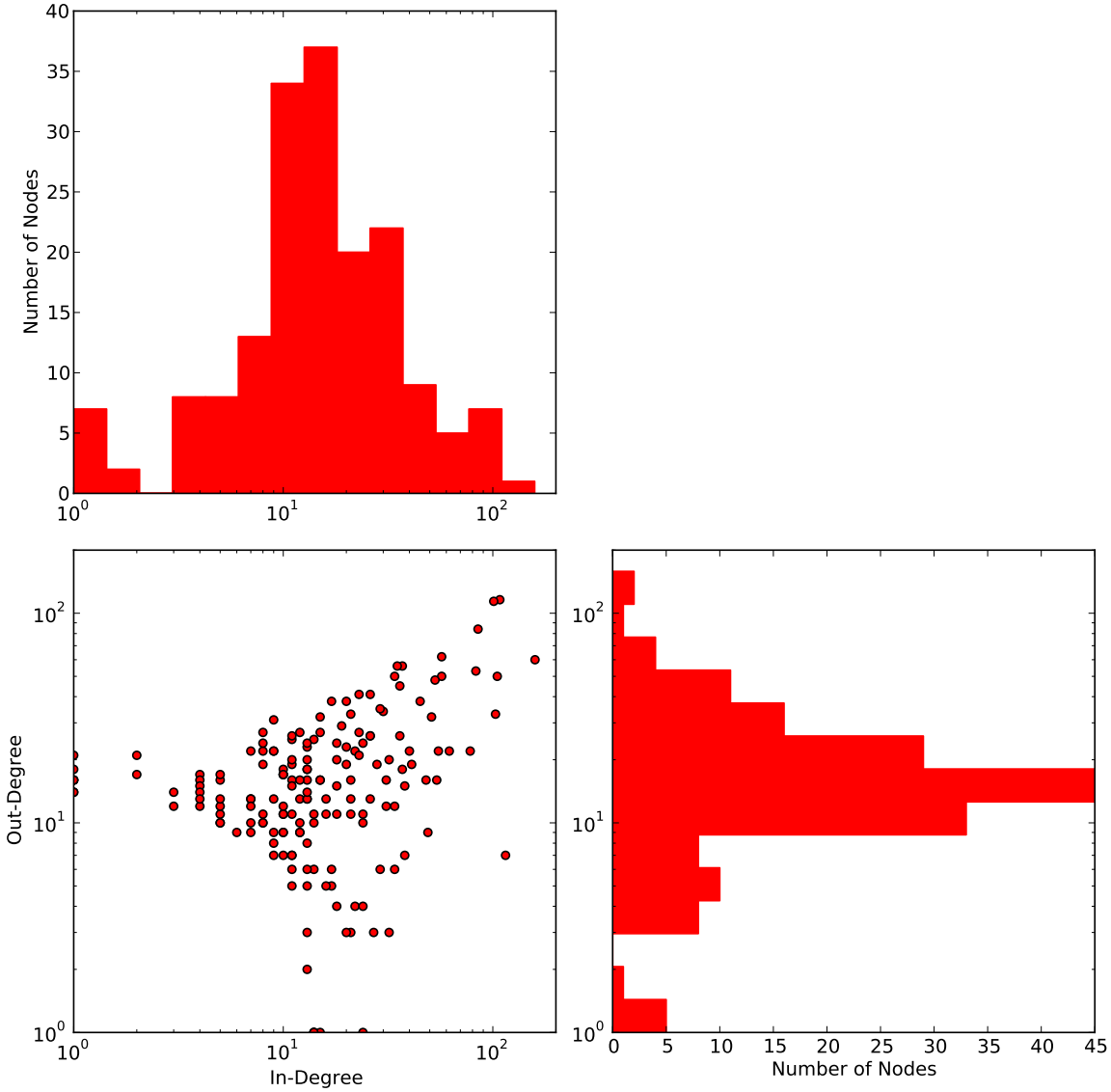


Figure 4.20: This figure has the same three elements as Fig. 4.19 but it now only shows the top 20 percent nodes with the highest degree. We also note that we use logarithmic binning for the histograms and logarithmic scaling for the scatterplot.

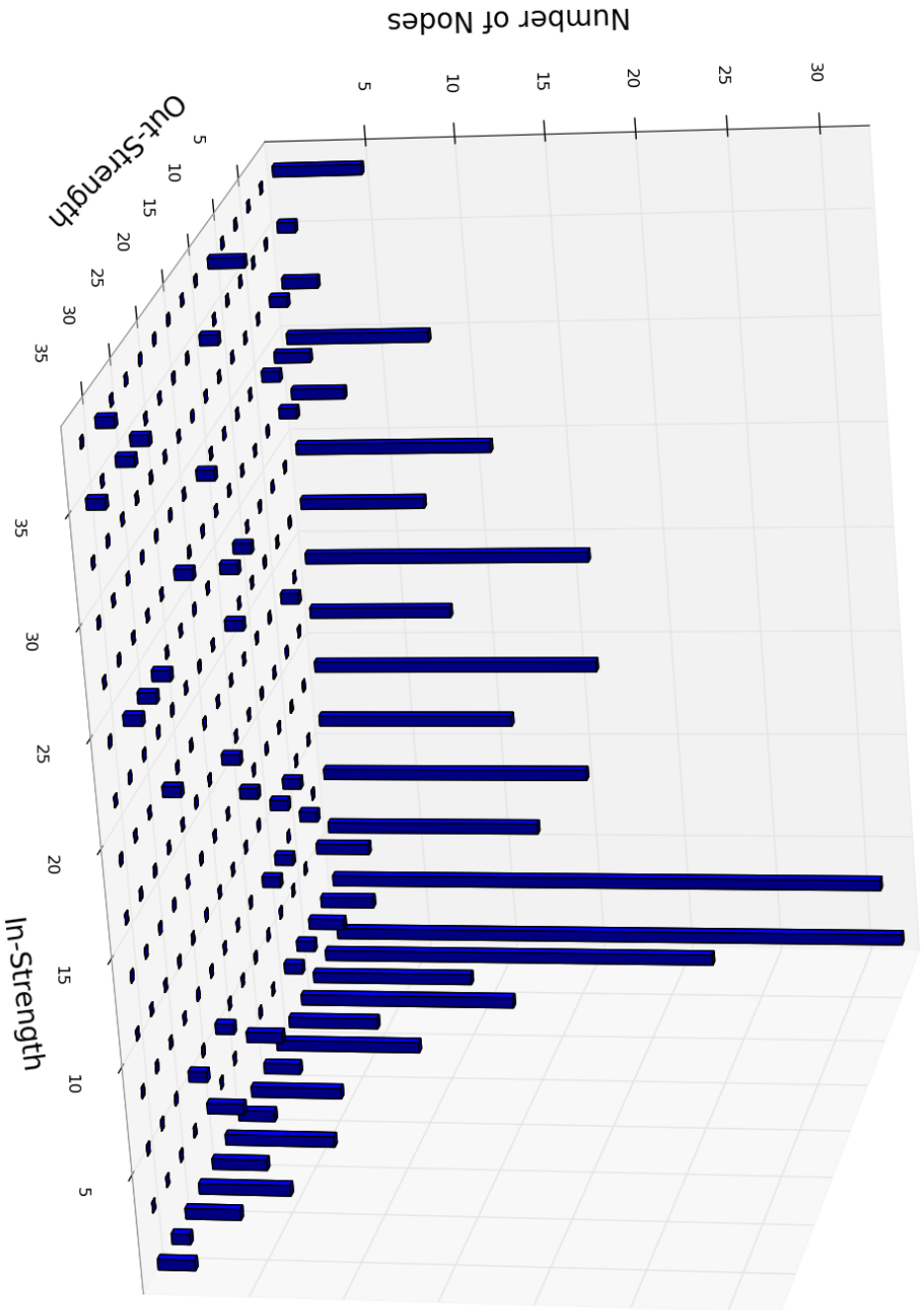


Figure 4.21: A two-dimensional histogram of the strength distribution for the first quarter of 2001 using two as bin width.

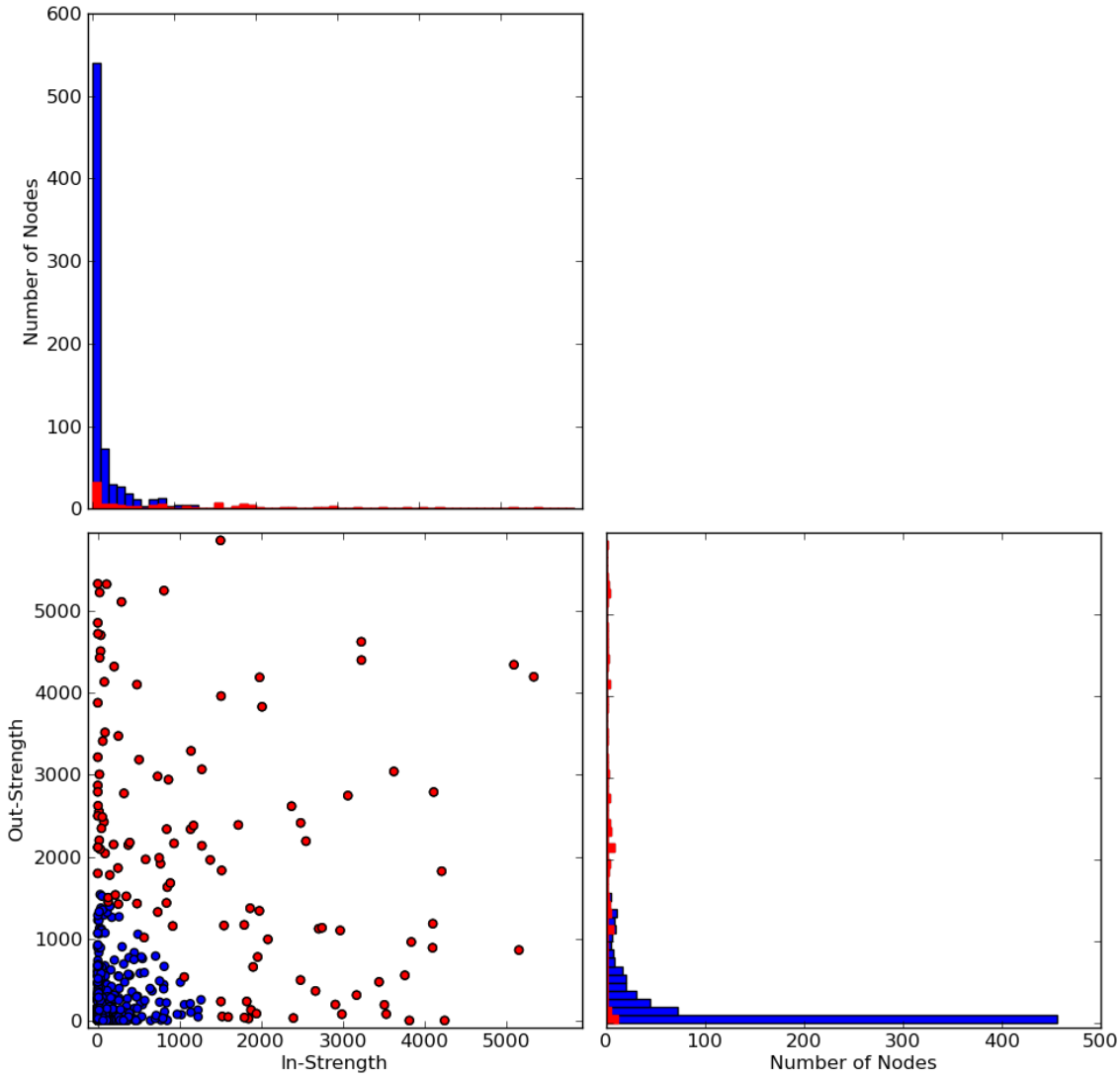


Figure 4.22: In the left bottom corner we have a scatterplot showing the in- and out-strength of the nodes. On top and on the right we have one-dimensional histograms respectively showing the in- and out-strength distribution of our weighted directed network. The colour red represents the top 20 percent of the nodes with the highest total strength. The data used is the aggregated data of the first quarter of 2001.

Chapter 5

Edge Distributions

In this Chapter the attention is turned to the edge distributions. We will again study the normal, non-crisis period of the year 2001. The adopted techniques are along the lines of those used in the previous Chapter. This chapter will be concluded with the question of how we can use all of these distribution from this and the previous chapter to construct a model interbank network from scratch.

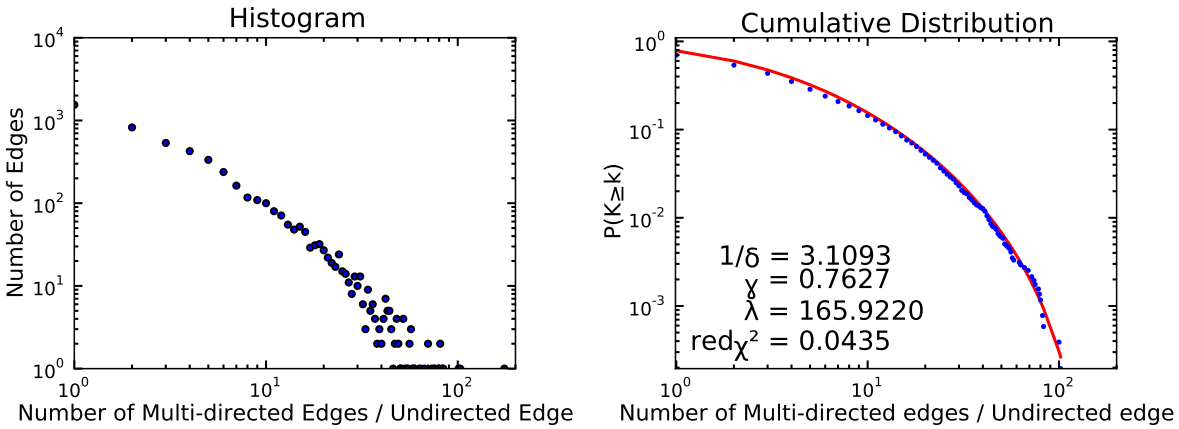
5.1 Number of Multi-directed Edges per Undirected Edge

We start this chapter with asking ourselves how the loans get distributed among the edges when we change from the undirected representation to the multi-directed representation of an unweighted network. We will first study the distribution of the number of loans in the multi-directed representation with respect to an edge in the undirected representation. Second, we will look at this problem from a two-dimensional perspective.

To investigate the first, we loop over every undirected edge and keep track of the number of underlying loans in the multi-directed case. This means that we discard any sense of direction for the multi-directed edges. Fig 5.1a shows a histogram for this distribution for the aggregated first quarter of 2001. We did not use logarithmic binning, in order to show the typical spread at the higher x-values. To fit a theoretical distribution we will again use the cumulative distribution which is shown in fig 5.1b. As expected the Johnson S_B is once again a valuable candidate. Using the method of section 4.1.2 we can generalize the Johnson S_B distribution to other aggregates and to the entire normal operation period.

When studying nodes, we could define incoming and outgoing. When studying edges separately, there is nothing to pin down a preferred direction. Therefore the configurations of Figs. 5.2a and 5.2b are considered identical when considering edges, just as the configurations of Figs. 5.2c and 5.2d. Obviously, this method leads to a loss of information.

If we wish to study the two-dimensional distribution of the number of multi-directed edges per undirected edge, we need to classify these multi-directed edges per undirected edge while



(a) Histogram **(b)** Cumulative Distribution

Figure 5.1: The histogram and cumulative distribution of the number of loans per undirected edge for the first quarter of 2001. We used a Johnson S_B distribution to fit the data.

accounting for direction. We do the following. We count the number of loans in one direction, the number of loans in the other direction and make an ordered pair putting the maximum number first. Fig. 5.2c and 5.2d, e.g., will both have as ordered pair (2,1). Fig. 5.10 at the back of the chapter shows the histogram of this two-dimensional edge distribution using this classification. The x-axis represents the number of loans in the maximum direction and the y-axis represents the number of loans in the minimum direction.

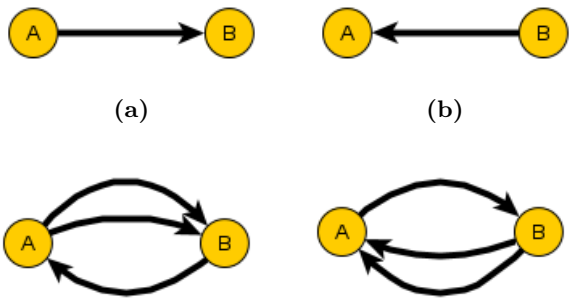
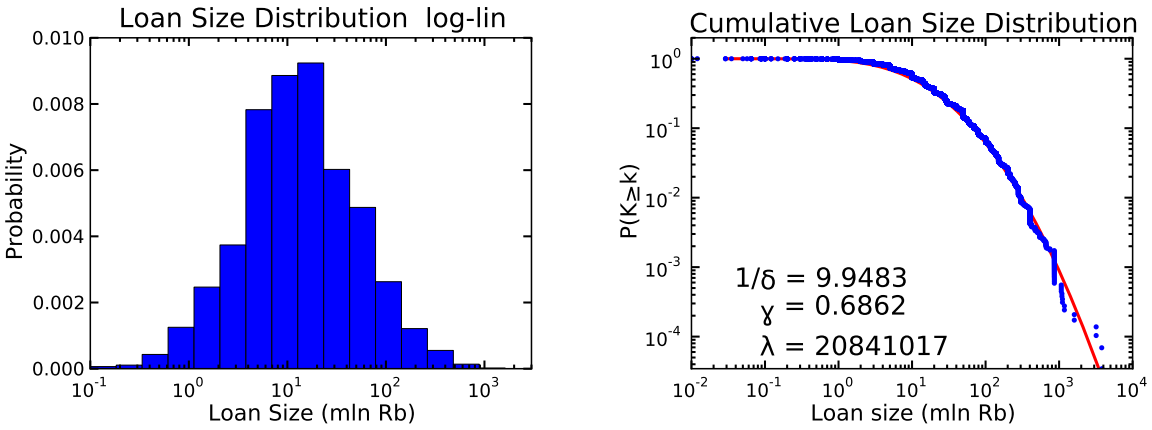


Figure 5.2: Four possible different underlying multi-directed configurations for one and the same undirected link between node A and node B.

5.2 Loan Size Distribution

Next, we consider the loan size distribution. In other words we want to know the distribution of the weight of the individual links in the multi-directed representation, which is also an important measure in network theory. We fit the distribution via the cdf. In Fig. 5.3b we performed the fit succesfully with the Johnson S_B distribution, although we failed to compute a χ^2-value. Again this good fit is also the case for other aggregates and for the entire normal period. Fig. 5.3a shows the normalized loan size histogram with logarithmic binning.



(a) Histogram **(b)** Cumulative Distribution

Figure 5.3: The loan size distribution for the first quarter of 2001. (a) shows the normed histogram with logarithmic binning and with logarithmic x-scale and (b) shows the cumulative distribution.

5.3 Exposure Size Distribution

As a last measure, we study the distribution of the size of the exposure which one node has to another single node. We adopt the following definition: in a given period, the exposure from bank A to bank B is the total amount of money bank A has issued to bank B during that period. In other terms, it is the size of the weighted directed edge running from node A to node B (cfr. Section 2.2.2). In Section 4.2 we studied the strength of a node, which can be seen as the total exposure a node has to the outside world. The exposure size distribution can be seen as an additional level of detail.

The paper of Boss et al. [6] mentions this quantity which they call the contract size instead of the exposure size. In Fig. 5.4 the histogram the exposure sizes is shown. They proposed a power-law to fit the data.

Cont et al. [11] also studied the exposure size distribution of the Brazilian interbank network. In Fig. 5.5 the cdf for the exposure size distribution is shown for June 2008. Here

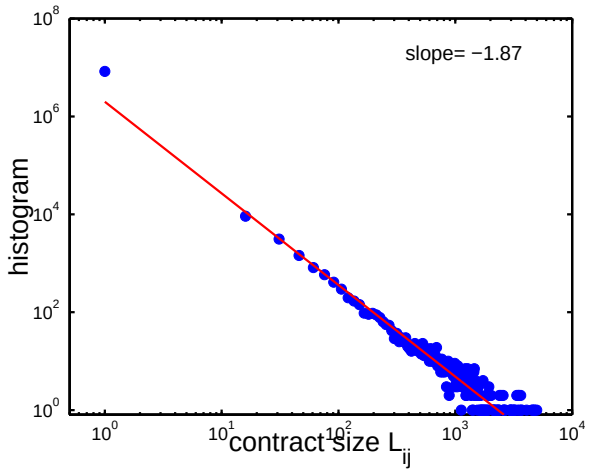


Figure 5.4: The exposure size distribution for the Austrian interbank network. We note that the label of the x-axis reads contract size instead of loan size. This figure was reproduced from Ref. [6].

again, they chose to fit the tail of the distribution with a power-law.

Fig. 5.6 shows the cumulative exposure size distribution for the Russian network. The Johnson S_B provides a good fit for the depicted aggregate with a reduced χ^2-value of 1.0128. Again this fit can be used for different aggregates. If we link the Brazilian Real (BRL) to the Russian Rouble (Rb), Figs. 5.5 and 5.6 can be compared qualitatively[1]. The average exchange rate over the period 2001-2008 is about 15 Rb/BRL[2]. Hence, one mln BRL (10^{-4} on the x-axis) equals 15 mln Rb. When we compare the cumulative distributions of both figures at these exposure sizes we find very similar behaviour. Considering the maximum exposure sizes of both cases, there is a difference in order of magnitude. The maximum for the Brazilian interbank network is 10 bln BRL (=150 bln Rb) whereas the maximum for the Russian network is only 10 bln Rb. In Ref. [11] a power-law with $\alpha = 2.3778$ was fitted to the tail part which only considered the exposures larger than 1 bln BRL. We do the same in Fig. 5.6b and find a power-law with $\alpha = 1.4739$, which differs a lot of the one for the Brazilian case. We conclude that both distributions are similar but certainly not the same.

[1] We are comparing monthly aggregated exposures to quarterly aggregated.
[2] www.oanda.com

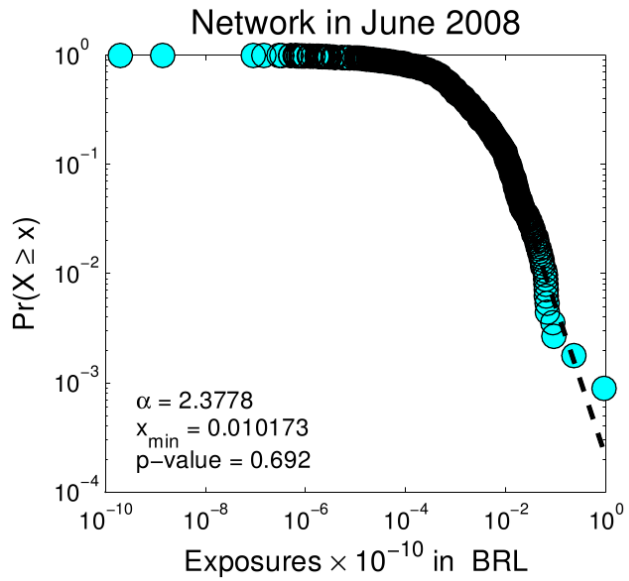
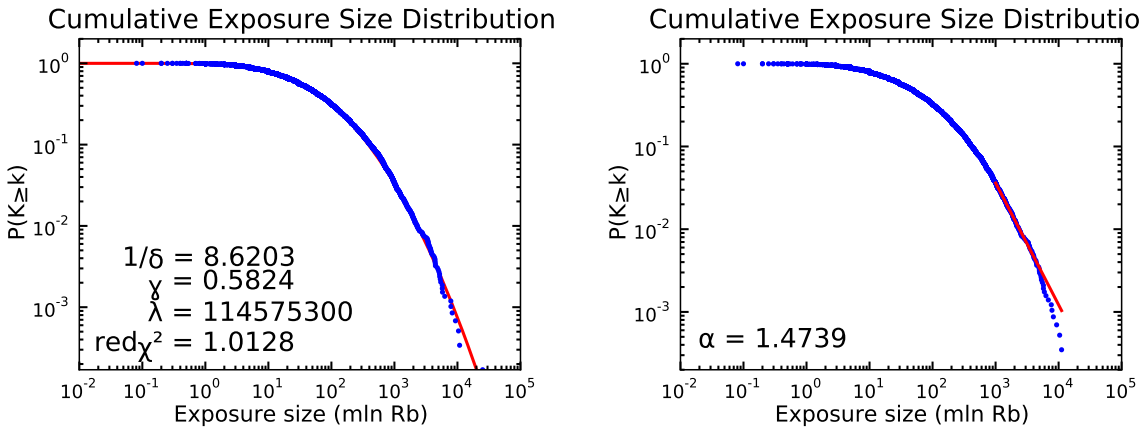


Figure 5.5: The cumulative exposure size distribution for the Brazilian network. This figure was reproduced from Ref. [11].



(a) Johnson S_B fit

(b) Power-law tail fit

Figure 5.6: The cumulative exposure size distribution for the Russian network. Notice that the second quarter of 2001 was used as aggregate period due to fitting problems.

5.4 Constructing a Model Interbank Network

We now have a plethora of probability distributions. But how do all these concepts we studied in this and the previous section entwine? To show their interdependence we try to build a model interbank network, Russian style, from scratch. We suppose this is the best way to see the mutual interconnections.

We think there are three possible ways to build such model networks. In this section we develop these three very similar algorithms but we do this just in a basic and conceptual way. We surely do not pretend that these algorithms, as described here, would produce an interbank network resembling the real-world data. This is mainly because we do not yet understand the correlation between the different distributions. Do nodes with a high degree, e.g., also have on average higher exposure sizes or do nodes with a high degree have a different multi-directed distribution (cfr. Fig. 5.10) as nodes with a lower degree? This is a fundamental problem which still needs to be resolved.

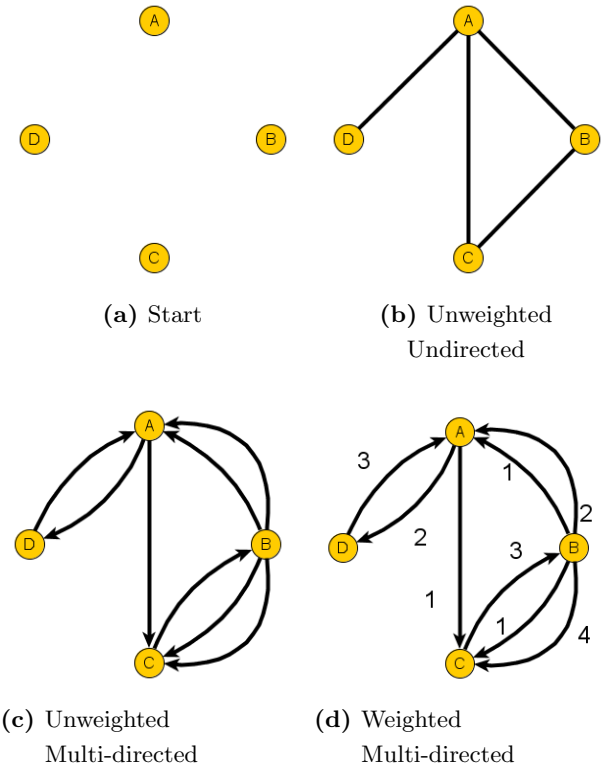
The present the following algorithms are presented in their imperfection. We note that the steps in the algorithms where we assign a node or edge with a certain probability are certainly not straightforward. These parts are very tricky and this has to do with conditional probability. The papers of Hurd et al. [17; 22] provide a framework to tackle this problem.

Algorithm 1: Undirected Degree (Fig. 5.7)

1. Choose number of nodes.
2. Determine the number of undirected edges needed for this network (cfr. Fig. 4.13a and Eq. 4.4 on page 53).
3. Assign each undirected edge to two nodes respecting the undirected degree distribution.
4. Assign to each undirected edge a multi-directed edge combination respecting the multi-directed per undirected distribution.
5. Assign to each multi-directed edge a weight according to the loan size distribution.
6. End with a weighted multi-directed network

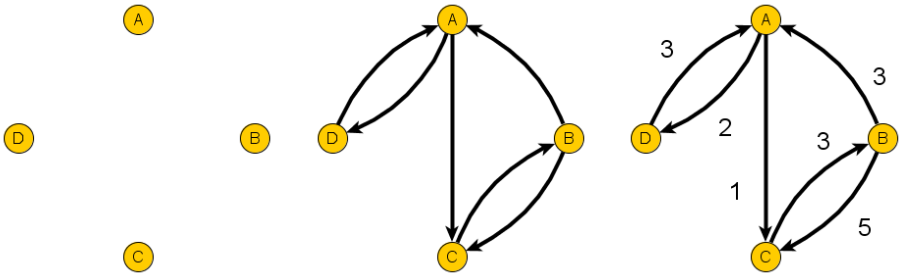
Algorithm 2: Directed Degree (Fig. 5.8)

1. Choose number of nodes.
2. Determine the number of directed edges needed for this network.
3. Assign each directed edge to two nodes respecting the directed in-degree distribution as well as the directed out-degree distribution. (A difficulty here, e.g., would be to incorporate the correlation between both distributions.)
4. Assign to each directed edge a weight according to the exposure size distribution.
5. End with a weighted directed network



(a) Start **(b)** Unweighted Undirected **(c)** Unweighted Multi-directed **(d)** Weighted Multi-directed

Figure 5.7: Schematic representation of algorithm 1



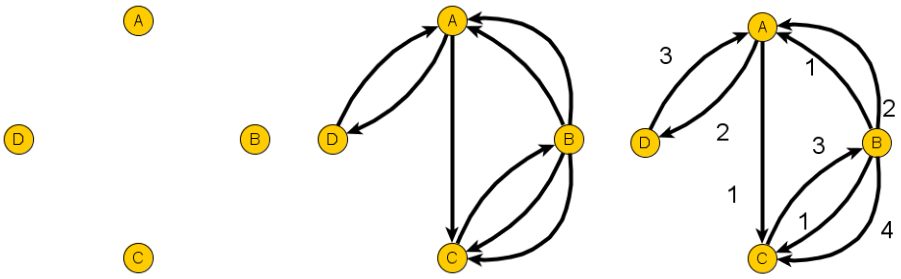
(a) Start **(b)** Unweighted Directed **(c)** Weighted Directed

Figure 5.8: Schematic representation of algorithm 2

Algorithm 3: Multi-directed Degree (Fig. 5.9)

1. Choose number of nodes.
2. Determine the number of multi-directed edges needed for this network.
3. Assign each multi-directed edge to two nodes respecting the multi-directed in- and out-degree distribution.

4. Assign to each multi-directed edge a weight according to the loan size distribution.
5. End with a weighted multi-directed network



(a) Start **(b)** Unweighted Multi-directed **(c)** Weighted Multi-directed

Figure 5.9: Schematic representation of algorithm 3

Note 1 A first level of checking our results is easy and quite powerful. If we determine the strength distribution (cfr section 4.2) of the endnetworks, it should have the same distribution as the real-world stregth distribution.

Note 2 Algorithm 1 and 3 resemble each other. The steps 2, 3 and 4 from the algorithm 1 and the steps 2, 3 from algorithm 3 have the same result.

Note 3 Algorithm 1 and 3 end with a weighted multi-directed network, whereas algorithm 2 ends with a weighted directed network. So algorithm 1 and 3 provide a higher level of detail.

5.5 Overview

In this chapter we studied the edge distributions. We examined the distribution of the number of loans per undirected edge. We looked at the distribution of the size of the loans. And we investigated the distribution of the size of the exposure of a node to another single node. The latter concept can be seen as the synthesis of the first two. The distribution of all of these can be described well with the Johnson S_B distribution. Further, we tried to demonstrate the interconnectedness of the distributions studied in this and the preceding chapter by thinking about how to build a model network.

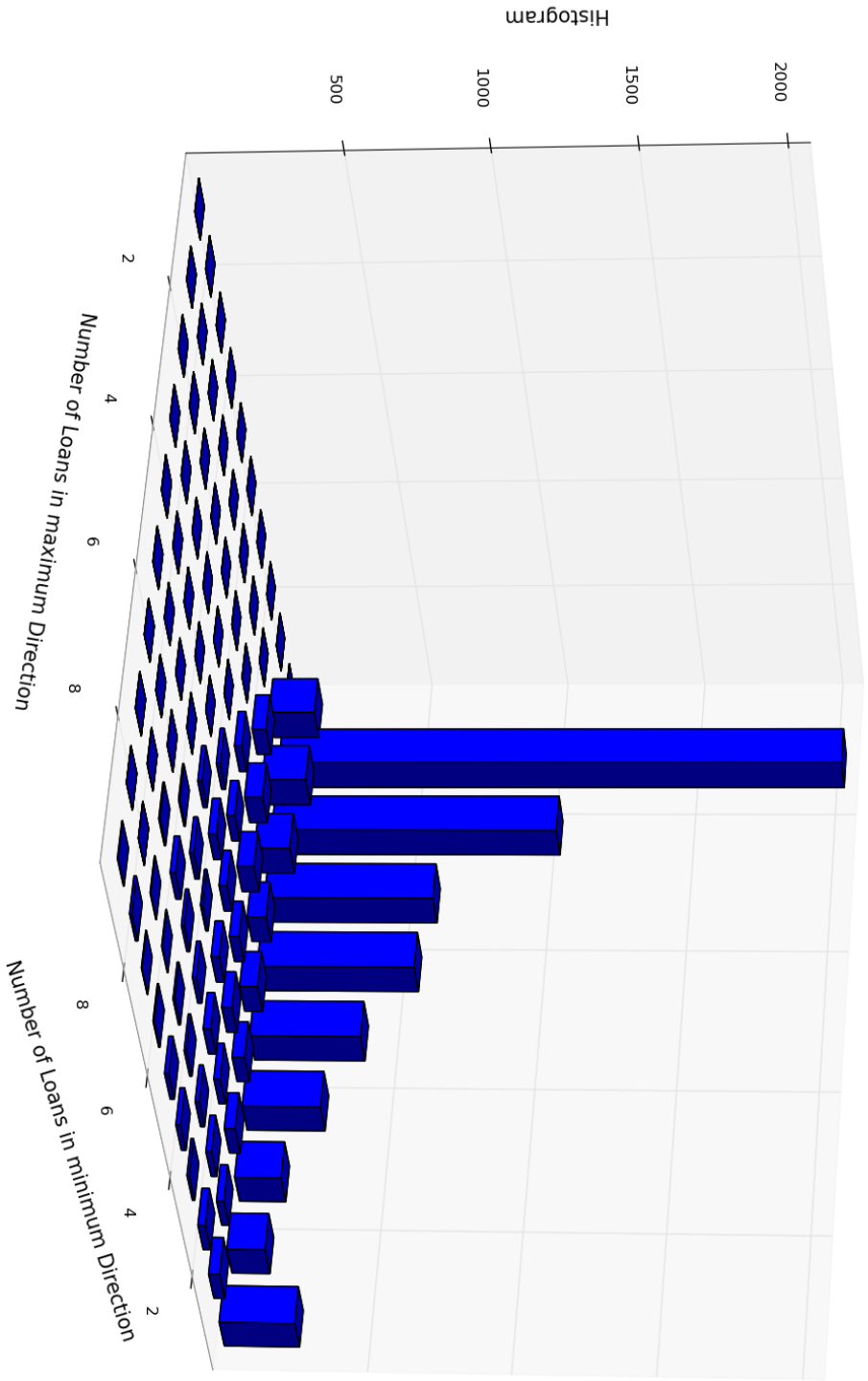


Figure 5.10: A two-dimensional histogram of the distribution of the number of multi-directed edges per undirected edge for the first quarter of 2001 using one as bin width.

Chapter 6

Percolation and Network Resilience

In previous Chapters the focus was on the characterization of the interbank network topology. In this chapter, we switch from statistics to dynamics and turn our attention to the processes taking place on this network topology. In Sect. 6.1 percolation theory will be discussed together with the theory's possible application for assessing systemic risk in interbank systems. In Sect. 6.2 percolation theory will be used to study the resilience of the Russian interbank network by 'attacking' it. We stress the fact that we attack a real-world interbank network and not fictuous networks as in other works. Sect. 6.2 will illustrate the theoretical ideas and measures developed in Sect. 6.1.

6.1 Percolation

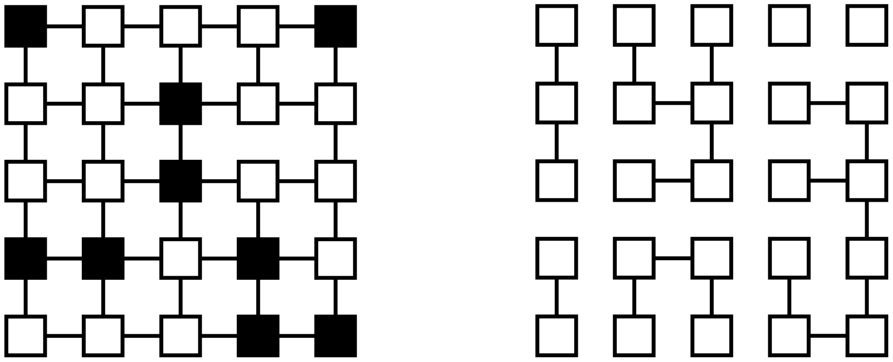
6.1.1 Introduction

[1]Percolation relates to a large number of problems in many branches of science and is an universal concept. Consider, e.g., a fluid flowing through a porous material. The answer to the question whether the fluid can pass through the material, depends on the concentration and nature of the porous channels. For another example, consider a social network. The answer to the question of whether or not a contagious disease can spread throughout the network and turn into an epidemic, depends amongst other things on the structure of the underlying network. Percolation theory adresses these questions and is used to predict the conditions for which a system percolates. For these two examples, the system is said to percolate when the fluid can pass through the material or the disease infects all nodes of the entire network. Percolation theory has been used as a model to study oil recovery from porous media; bush-fire spreading; metal-insulator transitions; epidemic modeling; and many more.

[1]The text in the first two Subsections is based on the slides of the course *Statistical Physics 2* given by Prof. Ryckebusch.

6.1.2 Percolation on a Lattice

The first models of percolation were introduced on lattices. In Fig. 6.2 two possible percolation models are shown. In site percolation, each site in the lattice is stochastically occupied (or unoccupied) with a probability p (or $(1-p)$). In bond percolation, a bond between the lattice sites is stochastically present with probability p.



(a) Site percolation **(b)** Bond percolation

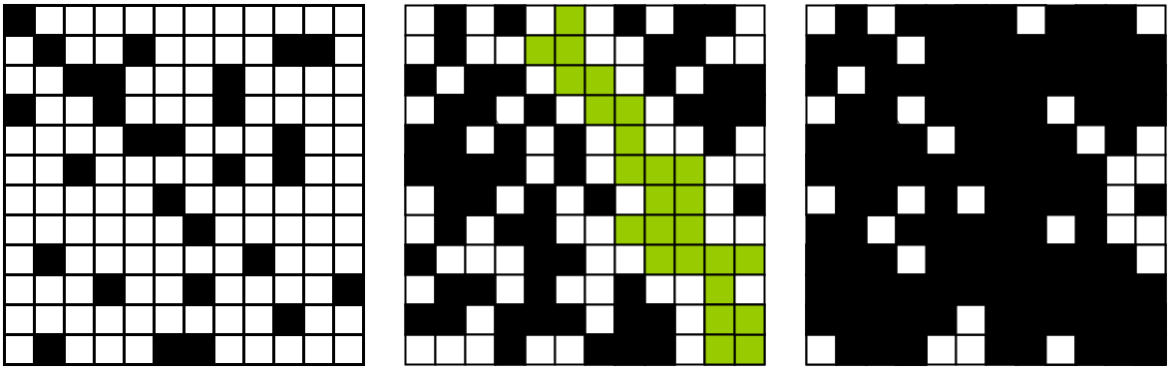
Figure 6.1: Site (a) and bond (b) percolation on a lattice. For site percolation there are 2 clusters of size 1; 2 of size 2; and 1 of size 3. For bond percolation there are 2 clusters of size 1; 1 of size 2; 1 of size 3; 1 of size 4; 1 of size 6; and 1 of size 8.

Percolation theory studies the change in the connectivity properties of the lattice, the connectedness, while varying the occupation probability p. The essential variables in percolation theory are the properties of the clusters. In site percolation, a cluster is a group of connected occupied sites. In bond percolation, a cluster is a group of sites connected by bonds.

Getting back to the example of a liquid trying to pass through a material. Considering bond percolation as a model, one can state that unoccupied sites are impermeable and occupied sites permeable. For a specimen to be permeable, there needs to be a cluster of permeable sites spanning the specimen from top to bottom. Hence, the question whether or not this system percolates, is the same as asking whether or not there is a spanning cluster of permeable or occupied sites. In Fig. 6.2 a two-dimensional site percolation lattice model is shown for different values of the occupation probability p. One notes that the probability that a certain site colours black equals the fraction of sites which colour black. For $p = 0.20$ in Fig. 6.2a there is no spanning cluster and for $p = 0.80$ in Fig. 6.2c there clearly is a spanning cluster. The most interesting question is what is the smallest value of p $p = p_c$ for which a spanning cluster appears. For two-dimensional site percolation on a square lattice it is found that $p_c = 0.59$. If a little more than half of the sites are occupied, the system percolates. The p_c is said to be the critical probability for which the system changes from subcritical ($p < p_c$) to supercritical ($p > p_c$) behaviour. At $p = p_c$ the system crosses the percolation threshold

and goes through a (geometrical) phase transition.

This critical regime is fascinating because it has remarkable scale-invariant properties. It is one of the best studied subjects in physics because these tipping points are important for understanding real-world systems.



(a) $p = 0.2$ **(b)** $p = 0.59$ **(c)** $p = 0.8$

Figure 6.2: Site percolation on a lattice. p is the probability that a certain site colours black (or is occupied). For $p = 0.2$ there is no spanning cluster, for $p = 0.59$ the spanning cluster appears (sites coloured green), and for $p = 0.80$ the spanning clusters nearly incorporates the entire lattice.

Cluster measures

We will now list some measures which are used to quantify the connectedness of the system. As will be shown in the following section, these measures can be used to characterize the critical regime due to their highly non-linear behaviour for $p \approx p_c$. We first define the order parameter P_∞ as

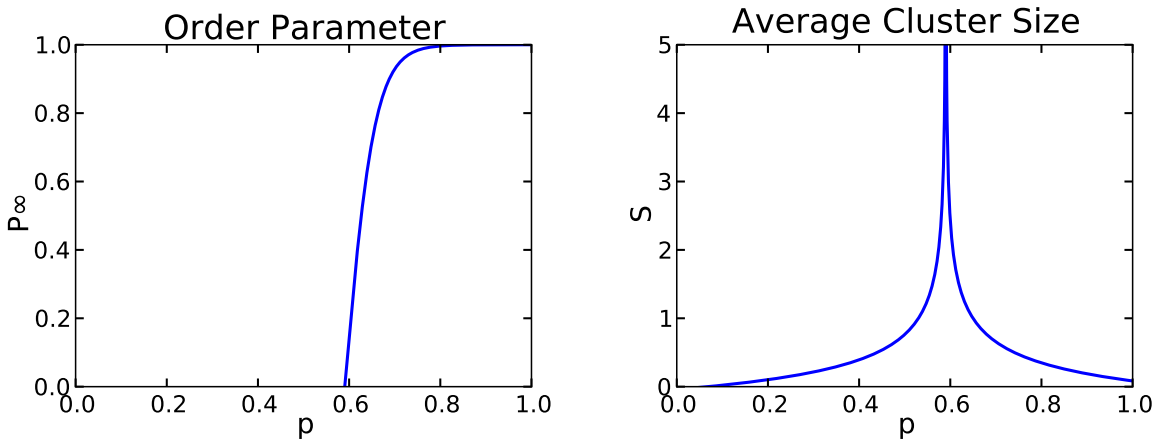
$$P_\infty = \frac{\text{number of sites in the spanning cluster}}{\text{total number of sites}}. \tag{6.1}$$

P_∞ is the probability that an arbitrarily selected site is part of the spanning cluster. In Fig. 6.3a the variation of the order parameter is shown. In the subcritical regime $P_\infty = 0$, whereas in the supercritical regime $0 < P_\infty \leq 1$. At the percolation transition P_∞ grows in a non-linear fashion[2].

The distribution of the cluster sizes $n_s(p)$ for any given p is defined as

$$n_s(p) = \frac{\text{mean number of clusters of size s}}{\text{total number of lattice sites}}. \tag{6.2}$$

[2]Theoretically speaking, the abrupt threshold of Fig. 6.3a appears only in systems with infinite size. Systems with a finite size have a smoother transition, as can be seen in Fig. 6.6



(a) Order Parameter **(b)** Average Cluster Size

Figure 6.3: The stylized order parameter $P_\infty(p)$ (a) and the average cluster size $S(p)$ (b) for the two-dimensional site percolation model of Fig. 6.2. The transition happens for $p_c = 0.59$.

We note that for $p > p_c$ the spanning cluster is excluded from $n_s(p)$.

Next, using $n_s(p)$, one also can define $w_s(p)$ as

$$w_s(p) = \frac{s n_s(p)}{\sum_s s n_s(p)}, \tag{6.3}$$

which is the probability that an arbitrarily selected occupied site belongs to a cluster of size s.

The mean number of sites in a cluster $S(p)$ can be defined as

$$S(p) = \sum_s s w_s(p) = \frac{\sum_s s^2 n_s(p)}{\sum_s s n_s(p)}, \tag{6.4}$$

where the sum $\sum_s$ spans over all cluster sizes except for the spanning cluster ($s = \infty$). In Fig. 6.3b the variation of this average cluster size is shown. For $p \approx p_c$ there is a drastic non-linear increase in $S(p)$, whereas for $p \to 0$ and $p \to 1$ the average $S(p)$ is equal to one. In these regimes an average cluster is only of size one, excluding the percolation cluster.

To end with, one can define the characteristic length of the clusters ξ or in other terms the connectedness length:

$$\xi^2 = \frac{\sum_s w_s s R_s^2}{\sum_s w_s s}, \tag{6.5}$$

with

$$R_s^2 = \frac{1}{s}\sum_{i=1}^{i=s}(\vec{r}_i - \bar{\vec{r}})^2, \qquad \text{and with} \qquad \bar{\vec{r}} = \frac{1}{s}\sum_{i=1}^{i=s}\vec{r}_i. \tag{6.6}$$

The measure is able to capture the shape and extend of the clusters. In Fig. 6.4 two clusters of size 4 are shown, one compact and one extended. A compact cluster has a smaller value for R_s^2 than an extended cluster.

(a) Compact Cluster

(b) Extended Cluster

Figure 6.4: A compact and extended cluster of size 4. Cluster (a) has an $R^2_{s=4}$ equal to 0.5, whereas cluster (b) has an $R^2_{s=4}$ equal to 1.25.

Cluster measures at the critical point

The behaviour of the order parameter P_∞ and average cluster size S clearly indicate the percolation threshold (see Fig. 6.3). This is also the case for the connectedness length ξ. For $p \approx p_c$ all three respond in a highly non-linear way. In the vicinity of p_c they can be approximated as:

$$P_\infty(p \approx p_c) \sim (p - p_c)^\beta, \tag{6.7}$$

$$S(p \approx p_c) \sim \mid p - p_c \mid^{-\gamma}, \tag{6.8}$$

$$\xi(p \approx p_c) \sim \mid p - p_c \mid^{-\nu}, \tag{6.9}$$

where β, γ and ν are called critical exponents, which depend only on the dimension used in the percolation model[3]. P_∞, S and ξ follow a power-law in the neighbourhood of the critical point $p = p_c$. In Subsection 6.2.4 P_∞ and S will be consideed for the Russian interbank network.

For the cluster size distribution n_s, its most important property is that for $p = p_c$:

$$n_s(p = p_c) \sim s^{-\tau}, \tag{6.10}$$

[3]Underlying these critical exponents, are the incredibly powerful physical concepts of universality and renormalization theory. These are beyond the scope of this work. An introduction can be found in Ref. [10].

where τ is a critical exponent. The cluster size distribution is a power-law at the critical point $p = p_c$ which means that it is a scale-free distribution. In Subsection 6.2.4 an example of a cluster distribution at the percolation threshold will be shown.

Cluster measures as indicators of criticality

The highly non-linear behaviour of these cluster measures near the critical regime can be used to figure out whether one is operating in the neighbourhood of the percolation threshold. To illustrate this, consider a liquid trying to pass through a porous material. Suppose one is in the subcritical regime (not permeable) and the location of the percolation threshold is unknown. Suppose one could identify and track a certain average cavity size. If it happens that the average cavity size does not change much, while varying the control parameter p, the system does not operate near the critical threshold. On the other hand if the average cavity size does change in a highly non-linear fashion, while varying the control parameter p, the system is operating near the critical regime.

In order to determine whether the system is in the neighbourhood of a percolation transition, one could try to identify a measure similar to the average cluster size S or the connectedness length ξ. A non-linear increase of such a measure could indicate the vicinity of the critical regime. Returning back to the financial system, such a measure could be an indicator of the stability of the system or serve as an early-warning indicator that one is approaching the critical regime. A critical regime which might bring the system down as a whole.

6.1.3 Percolation on a Network

It is possible to transpose percolation from a lattice to the topological structure of a network (Fig. 6.5a). For site percolation (Fig. 6.5b), the nodes get designated 'occupied' or 'unoccupied' whereas for bond percolation (Fig. 6.5c) the edges are turned on and off. The definition of the order parameter P_∞ changes to:

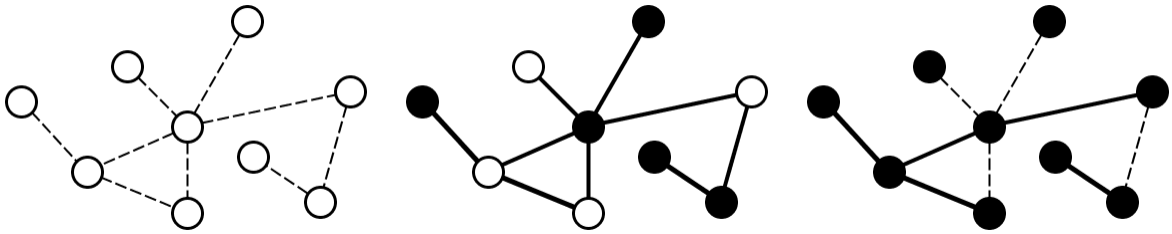
$$P_\infty(p) = \frac{\text{number of nodes in the largest cluster of the network}}{\text{total number of nodes}}, \tag{6.11}$$

and a system on a network is said to percolate if the largest cluster[4] includes a large non-zero fraction of the nodes. Because the largest cluster is used in the definition, there is no clearly defined percolating threshold as in Fig. 6.3a. The order parameter varies smoothly during the transition as can be seen in Fig. 6.6. The definition of $n_s(p)$ changes to

$$n_s(p) = \frac{\text{mean number of clusters of size s}}{\text{total number of nodes}}. \tag{6.12}$$

[4]The term cluster is a synonym for the term component defined in Section 3.5.

Bond percolation on a network can be used, e.g., to study the spreading of diseases. Consider a social network of interacting people for which one wishes to predict the size of a disease outbreak. When an ill person interacts with a healthy one, the probability of passing the disease is p, and the probability of not passing the disease is $1 - p$. This problem can be mapped onto bond percolation. If one determines a priori for every edge if it will be a succesful disease transmission channel (bond present) or not (bond not present), one obtains a set of vulnerable clusters on the network. Then, if a disease strikes one of the nodes in such a vulnerable cluster, the infection will be succesfully transmitted throughout the entire cluster.



(a) Underlying network topology **(b)** Site percolation **(c)** Bond percolation

Figure 6.5: Site and bond percolation on a network. In site percolation, the edges are present but the nodes are either occupied (solid circles) or unoccupied (open circles). Studies focus on the shape and size of the clusters of occupied nodes. In bond percolation, the edges are occupied or not (black or dashed lines) and the nodes that are connected together by occupied edges which form the clusters or components of interest. This figure was reproduced from Ref. [33].

The extraction of predictions about epidemics from the bond percolation model is rather straightforward: the distribution of clusters corresponds to the distribution of the sizes of the disease outbreaks that are initiated bby a randomly selected initial carrier. The percolation transition corresponds to the 'epidemic threshold' above which an epidemic outbreak is possible, i.e., a vulnerable spanning cluster appears [33].

We would like to stress the word *'possible'* in the previous sentence. It does not necessarily imply that when there is systemic risk inherent to the system, the system will actually collapse. Also it certainly is not true that a system which passes the percolation threshold will collapse immediately. The moment that vast components of vulnerable clusters connect to form a systemic wide vulnerable percolation cluster, is called the percolation transition. This transition should not be confused with the phase transition that the system undergoes when a trigger event actually hits the system and starts propagating.

Referring to our interbank network, the actual crashes are phase transitions and can be identified by the time series of well-selected variables. The first occurence of systemic risk in

the system is way harder to identify, if even possible.

6.1.4 Cascading Failures

Next to disease spreading, percolation can also be used to model cascading failures[5]. In some networks, such as electrical power networks, that carry a load or distribute a resource, the operation of the network is such that the failure of one node results in the redistribution of the load on that node to other nearby nodes. If nodes fail when their load exceeds some maximum capacity, this mechanism can result in a cascading failure or avalanche in which the redistribution of load pushes other nodes over their threshold, leading to further redistribution [33].

The modeling of cascade failures is well-developed in network theory and some of these models make use of site percolation. The difference between models of disease spreading and models of cascading failures is the following. For disease spreading, an edge of the underlying topological network is turned on or off to model whether or not the edge is a succesful disease transmission channel. For a cascading failure, all edges are considered succesful transmission channels. The focus lies on the nodes which are turned on or off whether or not they are vulnerable to failure.

The cascading models consist of three steps. In the first step, one states the thresholds at which a node is said to fail. Next, one determines for each node, by making assumptions about its neighbours' influence, whether it is vulnerable to failure or not. In the last step, one looks if there is a percolating cluster of vulnerable nodes to assess the systemic risk of the system as a whole.

These models are comparable to the *loss, given default* simulations in economic literature [36; 5; 16; 17; 22; 15]. In these simulations the stress of a single defaulting bank, or a set of banks, is propagated throughout the system by all kind of feedback channels (cfr. Muller [32] for an overview). When a counterparty of an initial failing bank is hit too hard, it also defaults which affects its counterparties, and so on.

The combination of the well-developed percolation theory and the current economic research offers possibilities. The study of systemic risk and financial contagion in the financial system could benefit from the concepts developed in network and percolation theory. This combination would not revolutionalize the current economic research, but it could bring a different perspective which may lead to a better understanding of the operation of the interbank network. As explained in the previous subsection, one could look, e.g., for measures in the network structure which indicate whether or not the system is near its critical regime.

[5]Financial contagion, despite its name, must be seen as a cascade process on a network, rather than an epidemiological process.

6.2 Network Resilience

Another application of percolation theory is the assesment of a network's resilience [**?**]. To test the resilience of a network, one attacks it by removing nodes one by one. Mapped to site percolation a removed node is labeled 'unoccupied' and the surviving nodes are labeled 'occupied'. While gradually removing the nodes, one monitors typical percolation measures. A good measure to evaluate a network's resilience is the order parameter P_∞, which is defined here as

$$P_\infty(p) = \frac{\text{number of nodes in the largest weakly connected component (LWCC)}}{\text{total number of initial nodes}}, \quad (6.13)$$

with p the fraction of the original amount of nodes which are left in the network. Because the LWCC (see Section 3.5) is used in the definition, there is no clearly defined percolating threshold as in Fig. 6.3a. In Fig. 6.6 typical behaviour for the order parameter for a network under attack is shown. The percolation threshold is not abrupt. P_∞ is able to capture the moment when a network starts to desintegrate into small, unconnected clusters. It is important to mention that one only monitors the underlying network structure and not some kind of vulnerable cluster. One does not discriminate between vulnerable and invulnerable nodes.

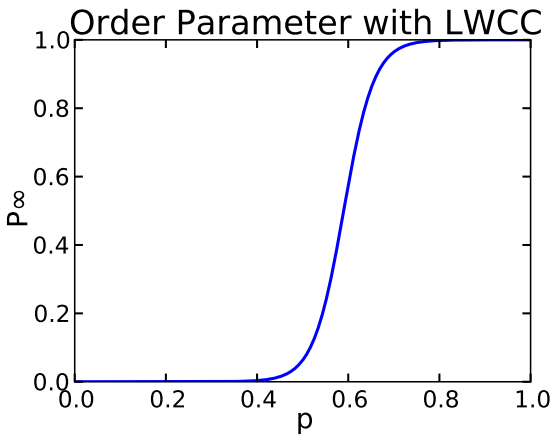


Figure 6.6: The order parameter for a network under attack.

Albert et al. [**?**] tested the resilience of the internet and the WWW. They concluded that these real-world complex networks are highly resilient to a random attack, which is the random deletion of nodes. On the other hand, if one performs a targeted attack, on the high degree nodes, these networks are found to be very fragile. They desintegrate as soon as a small percentage of the nodes is removed. Albert et al. coined these complex networks

robust-yet-fragile. Resilience studies of the American power grid and road network, exposed the same robust-yet-fragile features [35].

In this section, one random and two targeted attacks on the topological structure of the undirected representation of the Russian interbank network will be performed. This Section can be seen as an illustration of percolation.

6.2.1 Random Attack

To perform a random attack on a network, one removes randomly selected nodes sequentially. When a node is removed, all edges connected to that node are also removed. After a node is deleted, not only the relative size of the largest weakly connected cluster $P_\infty(p)$ is determined, but also other network parameters are remeasured; such as the cluster size distribution $n_s(p)$ or the average cluster size $S(p)$.

6.2.2 Targeted Degree Attack

In a targeted degree attack, the nodes with the highest degree are removed. At each step, one redetermines the node with the highest degree, which is subsequentky removed. If there are several nodes with the same degree, one of them is randomly chosen. After each node deletion, the relative size of the largest weakly connected cluster $P_\infty(p)$, the mean cluster size $n_s(p)$, and the mean number of nodes in clusters $S(p)$ are determined.

6.2.3 Targeted Core Number Attack

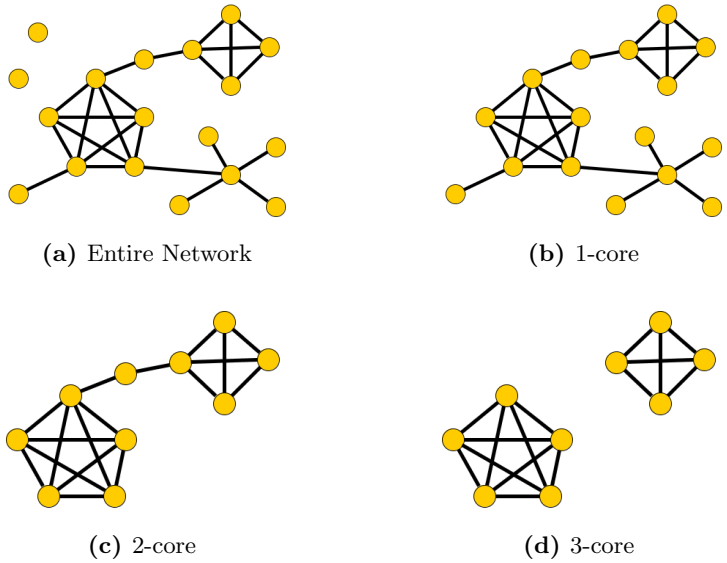
The k-core of a network is the sub-network wherein all nodes have at least a degree k. The pruning rules to determine, e.g., the k-core for $k = 2$ are as follows:

1. Remove from the network all nodes with degree $k = 1$ together with all edges connected to them;
2. Some remaining nodes may now have degree $k = 1$;
3. Repeat the previous two steps until there are no nodes with degree $k = 1$.

The remaining network is called a 2-core. And this process can be repeated for higher k to extract other cores. In Fig. 6.7 a k-core decomposition of a network is shown.

In a targeted core number attack, one removes the nodes according to their k-core number in the same way as with a targeted degree attack. The core number of a node is the largest value k of a k-core containing that node.

The choice for performing a targeted core number attack stems from the success the core number has with identifying influentual spreaders. Kitsak et al. [29] showed that, in contrast to common belief, there are plausible circumstances where the best spreaders do not



(a) Entire Network **(b)** 1-core

(c) 2-core **(d)** 3-core

Figure 6.7: Illustration of the k-core decomposition of a network.

correspond to the nodes with the highest degree. Instead, they find that the most efficient spreaders are those located within the highest k-core of the network as identified by the abovementioned k-shell decomposition analysis.

In the Ref. [27], Karas et al. studied the correlation between several node attributes, such as degree and core number, and the capacity of a node to spread financial contagion. They concluded that the coreness of a bank, as measured by its core number, is the only robust and reliable predictor of a bank's potential to spread contagion.

6.2.4 Testing the Resilience of the Russian Interbank Market

In this Subsection, the unweighted undirected representation of the interbank network of 2001 is used as a target. It consists of 1023 nodes and 8420 edges.

Order parameter

In Fig. 6.8 the order parameter, or in other terms the fraction of nodes in the LWCC, is shown for a random, a targeted degree, and a targeted core number attack. Each type of attack was performed 50 times and the figures show the average values. The random attack shows a gradual decrease of the order parameter which implies that the network is robust to random attacks [35]. The random attack will not be discussed any further because it shows no interesting percolation behaviour. The targeted attacks have a typical percolation threshold. Under a targeted core number attack the network percolates until more than 40 percent of the nodes is removed. Under a targeted degree attack the network desintegrates faster, already when 30 percent of the nodes is removed. One concludes that the network is fragile

under a targeted attack and that a targeted degree attack is more succesful in destroying the connectedness than a targeted core number attack. The interbank network can be categorized as robust-yet-fragile.

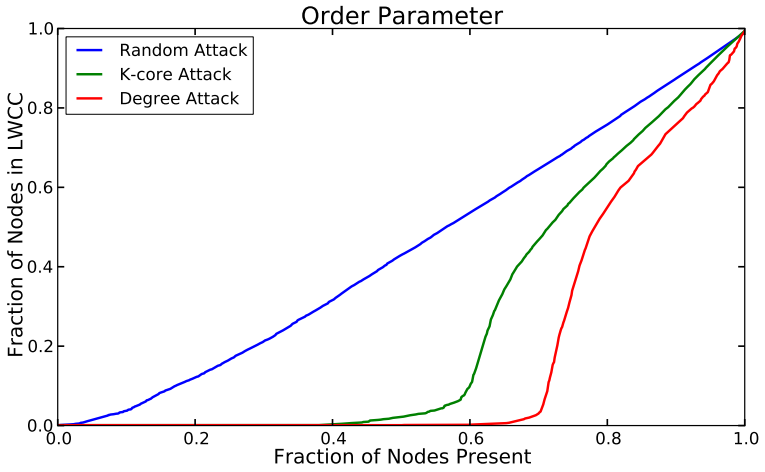


Figure 6.8: The order parameter for one random and two targeted attacks. The interbank network of 2001 is used as a target.

The location of the percolation threshold under a targeted degree attack varies widely for real world networks. The percolation thresholds of the abovementioned networks are listed in Tab. 6.1. One sees that the Russian interbank network for the year 2001 is not that fragile under attack as compared to these other real world networks. Threshold values for a targeted core number attack were not found in the literature.

One can also represent the variation of the order parameter as a function of the node with the highest degree (or highest core number) remaining in the network. In Fig. 6.9

Table 6.1: Percolation thresholds for real world networks under a targeted degree attack. Where the threshold value represents the percentage of deleted nodes of the network. The values for the internet and WWW were found in Ref. [**?**], the values of the power grid and road network in Ref. [35].

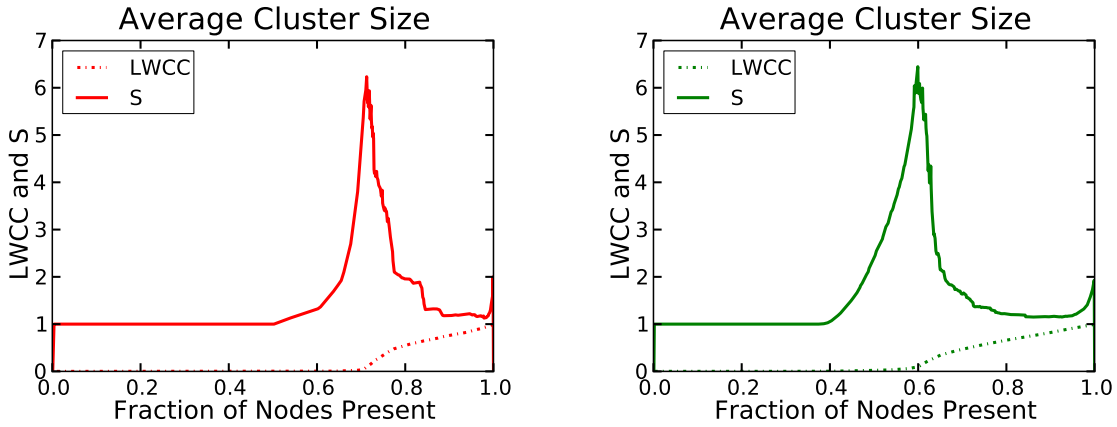
Network	Threshold
internet	3 pct.
WWW	7 pct.
power grid	10 pct.
road network	20 pct.
Russian interbank network	30 pct.

this is shown for the same simulations as used in Fig. 6.8. For Fig. 6.9a the percolation transition takes place while removing nodes with degree equal to three. While in Fig. 6.9b the percolation transition takes place while removing nodes with core number equal to two.

Average cluster size

The varation of the average cluster size S during the attack simulations is shown in Fig. 6.10. The order parameter is also shown in these two figures to indicate the percolation transition. For the targeted degree attack (Fig. 6.10a) as well as for the targeted core number attack (Fig. 6.10b), there is indeed a spike in the average cluster size at the percolation threshold.

We recall the discussion of Subsection 6.1.2 about the percolation transition indicators. Consider the targeted degree attack of Fig. 6.10a. After removing 20 percent of the highest degree nodes, the average cluster size will starts rising dramatically. From this rise, one can derive that one is in the vicinity of the critical regime. Hence, the average cluster size is a good indicator for the percolation transition.

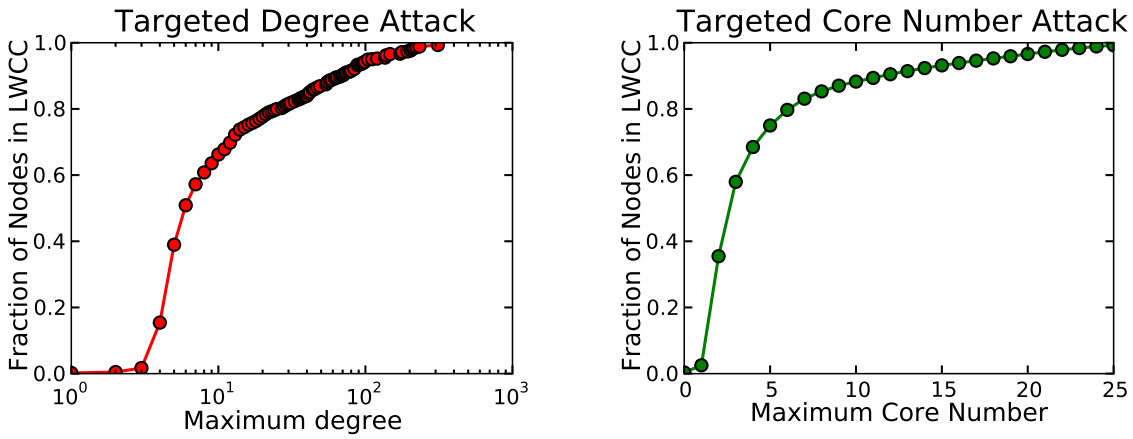


(a) Degree Attack **(b)** Core Number Attack

Figure 6.10: The average cluster size S and the fraction of nodes in the LWCC as the highest degree (a) or highest core number (b) nodes are being removed.

Cluster size distribution

In Fig. 6.11 the normalized cumulative cluster size distribution is shown for the targeted degree attack. The cumulative cluster size distribution for the core number attack is similar and will not be further discussed in this work. This cumulative distribution was determined at three points during the attack; when 10 percent of the highest degree nodes were removed (red curve); when 25 percent were removed (green curve); and at the percolation transition when 30 percent of the nodes are removed (blue curve). After removing 5 percent of the nodes in the network, the largest cluster (excluding the LWCC) had size four. Nearing the



(a) Degree Attack **(b)** Core Number Attack

Figure 6.9: The fraction of nodes in the LWCC as the highest degree (a) or highest core number (b) nodes are being removed. The curve is shown as a function of the node with the highest degree (or highest core number) remaining in the network. Fig. (a) uses a lin-log scale.

percolation transition, the observed cluster sizes increase to a maximum of 55. In Section 6.1.2, one discussed that the cluster size distribution at the critical point is a power law. In Fig. 6.11 this indeed is approximately the case for the real world interbank network.

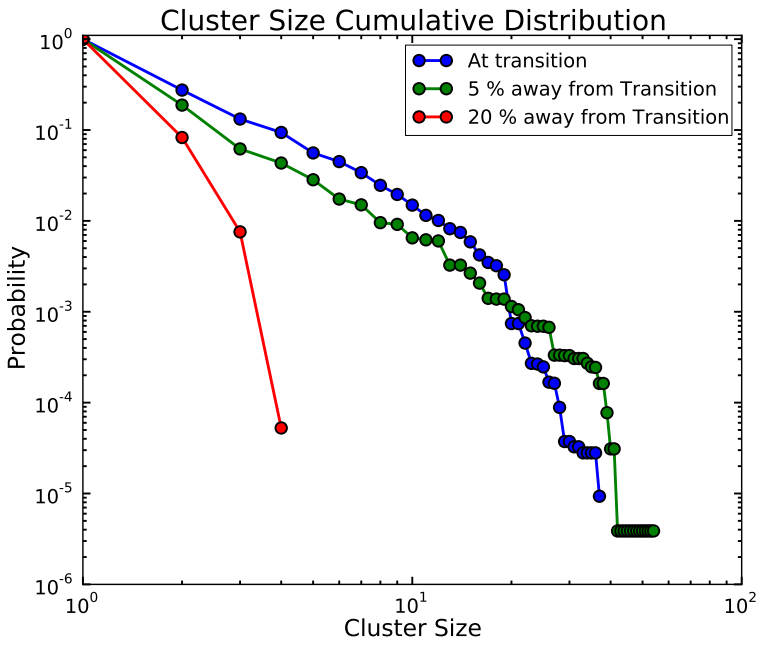
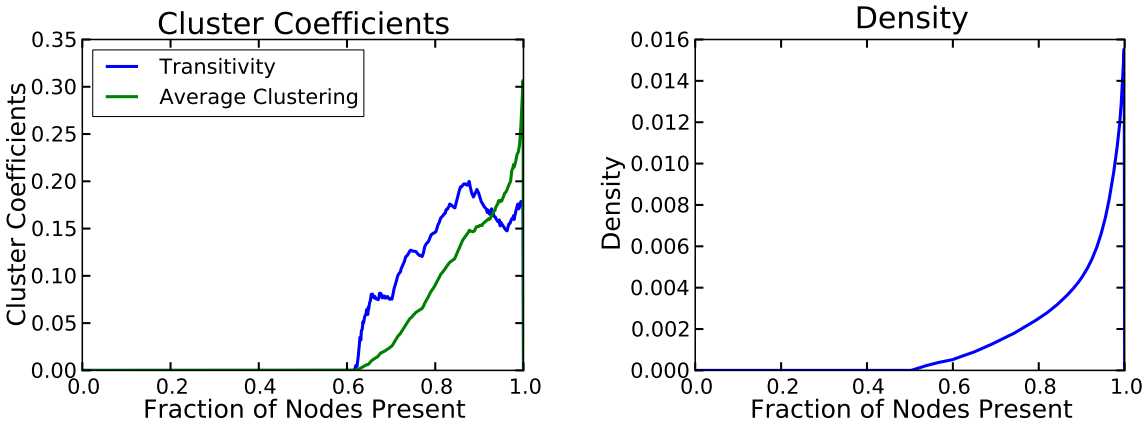


Figure 6.11: The normalized cumulative cluster size distribution for the targeted degree attack at three different moments during the attack.

Cluster coefficient and density

During the targeted attacks, the two cluster coefficients (see Section 3.3) and the density (see Section 3.1) were also monitored[6]. In Fig. 6.12a the transitivity and the average clustering coefficient is shown and in Fig. 6.12b the density is shown, both for the targeted degree attack. The variation of these measures for the core number attack is similar and will not be discussed further in this work.

The density clearly shows that after removing 50 percent of the highest degree nodes, along with the edges, all edges are removed from the network. This results, as can be seen in Fig. 6.10, in an average cluster size equal to one. As 50 percent of the highest degree nodes is removed, there are only isolated nodes left in the network. Knowing this, it is to be expected that the cluster coefficients, which quantify triangles of edges, will hit zero before 50 percent of the nodes is removed, which is the case in Fig. 6.12a.



(a) Clustering Coefficients

(b) Density

Figure 6.12: The transitivity and average clustering coefficient (a) and the density (b) both from the targeted degree attack.

6.2.5 The Variation of the Percolation Threshold in Time

Does the location of the percolation threshold change over time? In other terms, is the network more robust to attack at one time or the other?

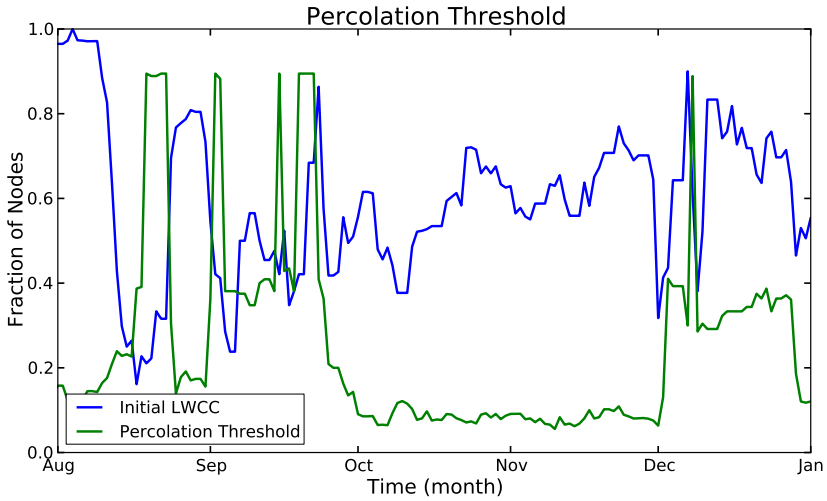
In this part, the location threshold for a targeted degree attack will be monitored while attacking a moving weekly aggregate. This weekly aggregate implies that the expected location of the percolation threshold will not be the 30 percent threshold of the yearly aggregate. Due to computational concerns, these attacks will only be performed for the two crisis pe-

[6]One points out that in the definition of these three measures used here, one does include only the nodes that are still occupied. In the definition of the density ($d = m/n(n-1)$), e.g., n is the number of nodes present at that stage of the attack, and not the number of initial nodes.

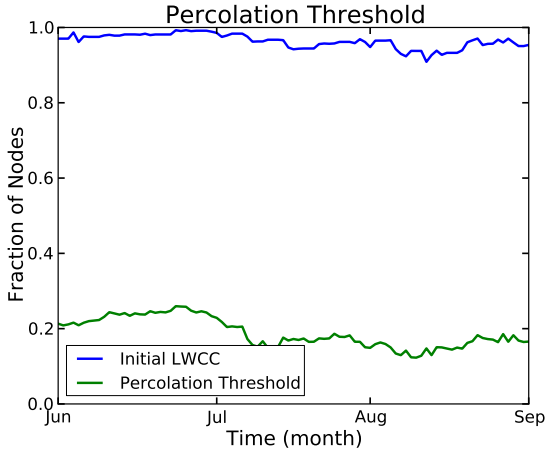
riods, which are likely to be the most interesting. The location of the percolation threshold is defined in this work as the instance during the attack at which less than 5 percent of all nodes remain in the LWCC. Or in other terms, the moment at which the order parameter drops beneath 5 percent. This threshold of 5 percent is somewhat arbitrary and could be tightened or loosened somewhat.

In Figs. 6.13a and 6.13b respectively crisis 1 and 2 are considered. Again, as could be seen in Chapter 3, there is a big difference between the first and the second crisis. The second crisis does not seem severe in comparison to the first. Every data point in these figures is the average result of 25 targeted degree attacks. The blue curves in Fig. 6.13 indicate the initial size of the LWCC, which is a good indicator of the initial clustering of the network. The green curves in Fig. 6.13 indicate the fraction of already removed nodes at the instance the percolation threshold is reached.

Fig. 6.13b shows a stable regime for the summer months of 2004. The network starts with nearly all nodes in the LWCC, and when 20 percent of the nodes are removed, the network passes the percolation threshold. In contrast, Fig. 6.13a does not show a stable regime at all. The fraction of nodes in the LWCC was already discussed in Subsection 3.5.2. The LWCC starts in August with a nearly full coverage of the network, shatters at August 22 as Russia defaults, and then fluctuates around a 50 percent coverage. The percolation threshold behaves as expected from October to the beginning of December. When 10 percent of the nodes is deleted, which is less than for crisis 2, the network looses its percolating cluster and desintegrates. If one considers the third week of August, one would need to remove 90 percent of the nodes in order for the percolating cluster to hit the threshold. This seems absurd. To understand this, one refers to Fig. 3.8 on page 33 which shows the number of WCC during this third week. It takes 90 percent of the nodes to be removed, because there are several WCC of similar size at that particular moment. Instead of attacking one and the same LWCC until it breaks, the attack is spread over the WCCs of the already desintegrated network. The attack looses its efficiency, because there is not one distinct LWCC. We conclude that in the third week of August the topological structure of the network is already in the subcritical regime.



(a) Crisis 1 (1998)



(b) Crisis 2 (2004)

Figure 6.13: The initial size of the LWCC and the location of the percolation threshold for the crisis period 1 (a) and crisis period 2 (b).

6.3 Overview

In the first part of this Chapter, percolation theory and its possible role in the study of systemic risk was explained. We concluded that the identification of vulnerable clusters, needed to assess this systemic risk, poses a serious problem. In the second part, the topological structure of the real-world Russian interbank network was attacked in order to illustrate percolation theory. We concluded that the network, as for many real world networks, is robust to random attacks, yet fragile to targeted attacks.

Chapter 7

Summary

We analyzed data containing real life bilateral and time-varying Russian interbank exposures from the time period 1998-2004. The fact that two major crises hit the Russian banking system during this period offers unparalleled opportunities. The first crisis occurs in august 1998, the second in the summer of 2004.

The first goal of this thesis was to perform an empirical characterization of the Russian interbank network in order to obtain stylized facts which can then be used in the development of theoretical models. Particular attention was payed to the difference between the network structure of the crisis and non-crisis periods. We also searched for possible crisis-indicators among the network structure properties during the run-up to the crashes. After exploring the basic entities of the network, we looked at global network measures and at the node and edge distributions.

Our first conclusion is that the effect of the second crisis on the network structure is certainly less drastic than the effect of the first crisis. In every network measure the first crisis is very outspoken whereas the second only causes a ripple. The typical differences between crisis, referring only to the first, and non-crisis periods are the following. As a crisis strikes there is: an increase in density, a drop in the distance measures, and the clustering coefficients become very volatile. Further, the interbank network falls into many smaller and disconnected pieces. The local degree versus the local clustering coefficient and the shortest path length distribution can both distinguish between normal and dysfunctional operating periods. To identify pre-crisis early warning signals, we need additional data covering a larger pre-crisis time period. We also conclude that the interbank network, following the reasoning of Cont et al. [11], is not a small-world network. All node and edge distributions are well described by the Johnson S_B distribution. Finally, we argue that the Brazilian interbank network (cfr. Ref. [11]) and the Russian to have universal features like, e.g., their exposure size distribution.

The second goal of this thesis was to search for network percolation transitions in the run-up to a crisis. The concept of percolation was developed and the link between *cascade failures*

in network theory and *loss, given default* simulations in economic literature was drawn. We argue that defining a 'vulnerable' node is model-dependent and that a percolating 'vulnerable' cluster, the cause of systemic risk, might be inherent to an interbank system. Therefore, it may be hard to use the concepts of percolation theory to assess systemic risk in interbank networks.

As a last item, the resilience of the topological structure of the interbank network was tested. We conclude from several percolation measures that there is a geometrical phase transition for the network under a targeted degree and core number attack. The interbank network is found to be more resilient than other real-world networks, but it still can be considered robust-yet-fragile.

Chapter 8

Distributions

8.1 Log-normal Distribution

A random variable x is log-normally distributed with parameters δ, λ and ξ if it obeys a probability density function (pdf) of the type

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi z}} \exp\left[-\frac{1}{2}(\delta \ln z)^2\right], \tag{8.1}$$

with

$$z = \frac{x-\xi}{\lambda} \quad \text{and} \quad \xi < x; \delta, \xi > 0. \tag{8.2}$$

We call γ en δ the shape parameters, ξ the location parameter, and λ the scale parameter [24]. Because there is no need to translate the distribution, the location parameter ξ is set to 0.

In this work the cumulative distribution function (cdf) is defined as the function $F_X(x) = P(X \geq x)$ which for a real-valued random variable X gives for every real number x, the probability that the random variable X takes on a value **larger** than or equal to x. We note that usually the cdf is defined as the function $F_X(x) = P(X \leq x)$ which represents the probability that the random variable X takes on a value **less** than or equal to x. Because, we fit rank/frequency plots in this work, it is more convenient to use the first definition. The two definitions relate to each other as $P(X \geq x) = 1 - P(X \leq x)$.

The formula for the cumulative distribution function of a log-normal distribution, with $\xi = 0$, is

$$F(X \geq x) = \Phi(\delta \ln \frac{x}{\lambda}) \qquad x \geq 0; \delta, \lambda > 0, \tag{8.3}$$

with Φ the cumulative distribution function of the normal or Gaussian distribution.

8.2 Johnson S_B Distribution

A random variable x is said to be Johnson S_B distributed with parameters γ, δ, λ and ξ if it obeys the pdf

$$f(x) = \frac{\delta}{\lambda\sqrt{2\pi}(1-z)^2} \exp\left[-\frac{1}{2}(\gamma + \delta \ln \frac{z}{1-z})^2\right], \tag{8.4}$$

with

$$z = \frac{x-\xi}{\lambda} \quad \text{and} \quad \xi < x < \xi + \lambda; \lambda, \delta, \xi > 0. \tag{8.5}$$

γ en δ are called the shape parameters, ξ the location parameter, and λ the scale parameter [24]. This expression is an extension of Eq. (8.1) with an additional shape paremeter γ. The effect of this γ will be discussed later on. For values $z \geq 1$ and $0 \geq z$ the probability equals zero which results in a bounded distribution[1]. Because there is no need to translate the distribution, the location parameter ξ is set to 0, and we get the following new definition for the variable z:

$$z = \frac{x}{\lambda} \quad \text{with} \quad 0 < x < \lambda \tag{8.6}$$

For more information on the properties of this distribution we refer to the appendix of Ref. [24].

The formula for the cumulative distribution function of the Johnson S_B distribution with $\xi = 0$ is

$$F(X \geq x) = \Phi(\gamma + \delta ln \frac{x}{\lambda - x}) \qquad x \geq 0; \gamma, \delta, \lambda > 0 \tag{8.7}$$

To give the reader an idea of the parameter dependence of this not so well-known distribution, the cdf of Eq. (8.7) is displayed for different parameter ranges in Figs. 8.1 to 8.4. In Fig. 8.1 the dependence of the scale parameter λ is shown. In this plot λ is varied with the two shape parameters $\gamma = \delta = 1$. The value of λ clearly reflects the upper boundary of the distribution. The scale parameter does not have any influence on the shape of the distribution.

In Fig. 8.2 the shape parameter γ is varied with $\delta = \lambda = 1$. In Figs. 8.3 and 8.4 the shape parameter δ is varied with $\gamma = \lambda = 1$. When we compare the shape of, e.g., the empirical cumulative distributions of Fig. 4.8a on page 48 to the curves in Fig. 8.3 and Fig. 8.4, we find the same shape for $\delta \approx 0.5$.

[1] The B in Johnson S_B is short for Bounded.

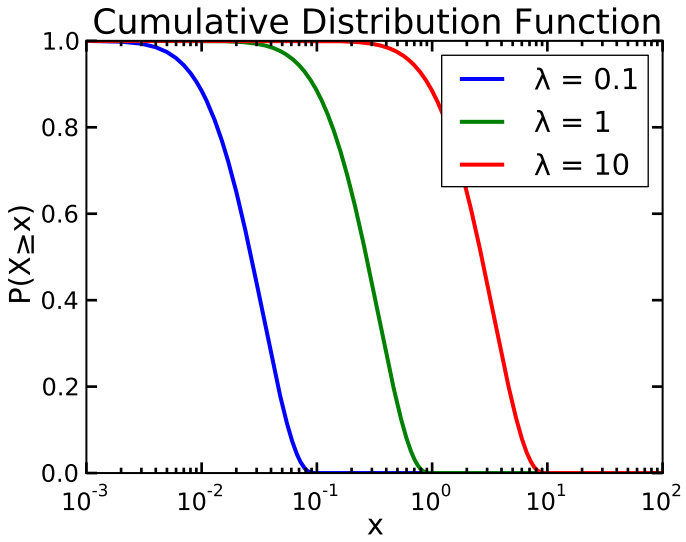


Figure 8.1: The scale parameter λ is varied the two shape parameters $\gamma = \delta = 1$.

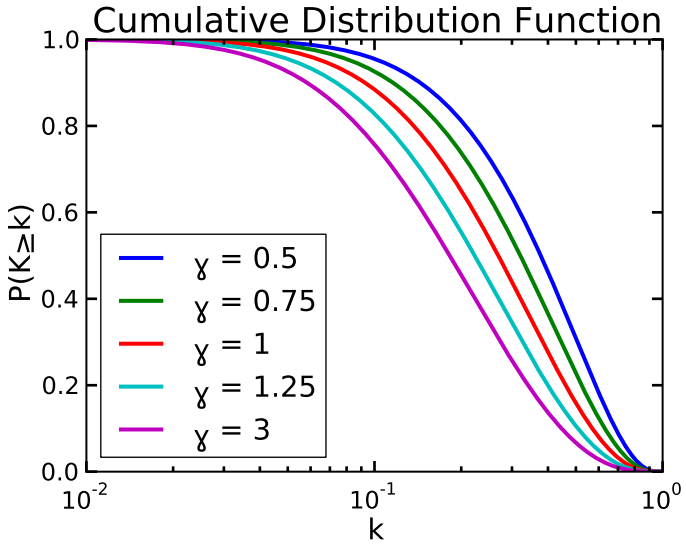


Figure 8.2: The shape parameter γ is varied with $\lambda = \delta = 1$.

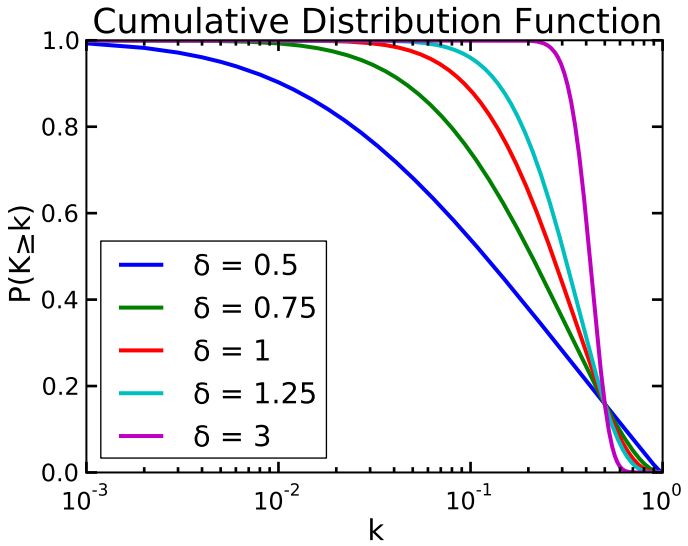


Figure 8.3: The shape parameter δ is varied with $\lambda = \gamma = 1$.

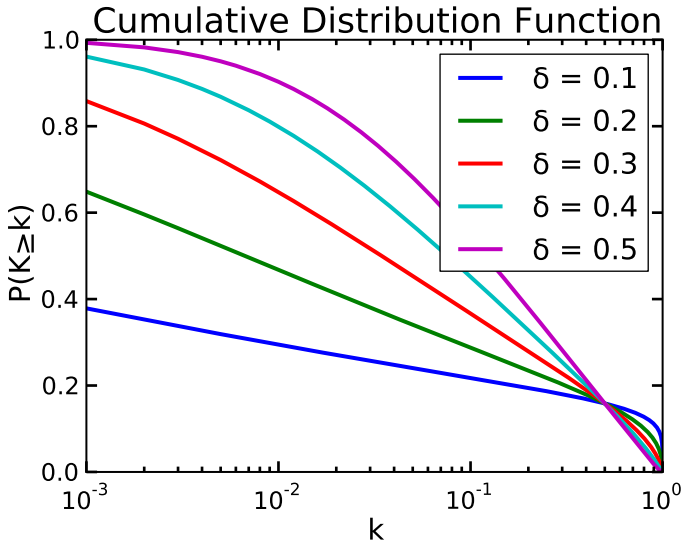


Figure 8.4: The same as in fig 8.3, i.e. vary the shape parameter δ with $\lambda = \gamma = 1$, but now for a different parameter range.

Bibliography

[1] M. Buchanan. Meltdown modelling. *Nature*, 460(7256):680–682, Aug 2009.

[2] A.G. Haldane. Rethinking the financial network, Apr 2009. Speech delivered at the Financial Student Association, Amsterdam.

[3] A.G. Haldane and R.M. May. Systemic risk in banking ecosystems. *Nature*, 469(7330):351–355, Jan 2011.

[4] R.M. May, S.A. Levin, and G. Sugihara. Complex systems: Ecology for bankers. *Nature*, 2008.

[5] M.E.J. Newman. *Networks: An Introduction*. Oxford University Press, Inc., New York, NY, USA, 2010.

[6] M.E.J. Newman. The structure and function of complex networks. *SIAM Review*, 45(2):167–256, 2003.

[7] S.L. Pimm. *Food Webs, 2nd ed.* University of Chicago Press, Chicago, 2002.

[8] J. Scott. *Social Network Analysis: A Handbook, 2nd ed.* Sage Publications, London, 2000.

[9] Q. Chen, H. Chang, R. Govindan, and S. Jamin. The origin of power laws in internet topologies revisited. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 608–617, 2002.

[10] M. Boss, H. Elsinger, M. Summer, and S. Thurner. The network topology of the interbank market. *Quantitative Finance*, 4(6):677–684, Dec 2004.

[11] R. Cont, A. Moussa, and E. Bastos-Santos. Network structure and systemic risk in the banking system. Available at SSRN: http://ssrn.com/abstract=1733528, Dec 2010.

[12] A. Karas and K.J.L. Schoors. Russian interbank market, Oct 2008. Private communication.

[13] E. Nier, J. Yang, T. Yorulmazer, and A. Alentorn. Network models and financial stability. Bank of England Working Paper No. 346, Apr 2008.

[14] S. Battiston, D. D. Gatti, M. Gallegati, B. C. Greenwald, and J. E. Stiglitz. Liaisons dangereuses: Increasing connectivity, risk sharing and systemic risk. NBER Working Paper 15611, Dec 2009.

[15] P. Gai and S. Kapadia. Contagion in financial networks. Bank of England Working Paper No. 383, Mar 2010.

[16] J. Gleeson, T. Hurd, S. Melnik, and A. Hackett. Systemic risk in banking networks without Monte-Carlo simulations. Working Paper, 2011.

[17] T. Hurd and J. Gleeson. A framework for analyzing contagion in banking networks. arXiv:1110.4312v1, Oct 2011.

[18] P. Gai, A. Haldane, and S. Kapadia. Complexity, concentration and contagion. *Journal of Monetary Economics*, 58(5), 2011.

[19] A. Karas, K.J.L. Schoors, and G. Lanine. Liquidity matters: Evidence from the Russian interbank market. BOFIT Discussion Papers No. 19/2008, Nov 2008.

[20] J.A. Chan-Lau. Balance sheet network analysis of too-connected-to-fail risk in global and domestic banking systems. IMF Working Paper 10/107, Apr 2010.

[21] T. Dette, S. Pauls, and D.N. Rockmore. Robustness and contagion in the international financial network. arXiv:1104.4249v2, Jul 2011.

[22] A.A. Hagberg, D.A. Schult, and P.J. Swart. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Aug 2008.

[23] D.J. Watts and S.H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, June 1998.

[24] D.J. Watts. *Six Degrees: The Science of a Connected Age*. W. W. Norton, New York, 2003.

[25] R. Cont and E. Tanimura. Small World Graphs: Characterization and Alternative Constructions. *Advances in Applied Probability*, 40(4), 2008.

[26] L.A.N. Amaral, A. Scala, M. Barthélémy, and H.E. Stanley. Classes of small-world networks. *PNAS*, 97(21), 2000.

[27] S. Redner. How popular is your paper? an empirical study of the citation distribution. *European Physical Journal B*, 4(2):131–134, August 1998.

[28] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A. L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.

[29] M.E.J. Newman. Power laws, pareto distributions and zipf's law. *CONTEMPORARY PHYSICS*, 46:323, 2005.

[30] E. Jones, T. Oliphant, P. Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

[31] N.L. Johnson. Systems of frequency curves generated by methods of translation. *Biometrika*, 36(1/2):pp. 149–176, 1949.

[32] M. Palahi, T. Pukkala, E. Blasco, and A. Trasobares. Comparison of beta, johnson sb, weibull and truncated weibull functions for modeling the diameter distribution of forest stands in catalonia (north-east of spain). *European Journal of Forest Research*, 126:563–571, 2007. 10.1007/s10342-007-0177-3.

[33] K. Christensen and N.R. Moloney. *Complexity And Criticality (Imperial College Press Advanced Physics Texts)*. Imperial College Press, illustrated edition edition, December 2005.

[34] J. Mueller. Interbank credit lines as a channel of contagion. *Journal of Financial Services Research*, 29(1):37–60, 2006.

[35] R. Albert, H. Jeong, and A.-L. Barabasi. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, Jul 2000.

[36] M. Kitsak, L. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. Stanley, and H. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6(11):888–893, Aug 2010.

[37] A. Karas and K.J.L. Schoors. Interbank liquidity runs, coreness and contagion on interbank markets, Apr 2012. Private communication.