FACULTY OF SCIENCES

Multiscale analysis of the Russian interbank loan network

door

Marnix Van Soom

Department of Physics and Astronomy

Promotoren: Prof. Dr. J. Ryckebusch,
Prof. Dr. K. Schoors
Scriptiebegeleiders: Msc. M. van den Heuvel

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

Academiejaar 2016–2017

GHENT UNIVERSITY

Prologue

Ik wil eerst en vooral graag Prof. Dr. J. Ryckebusch bedanken, die tijdens zijn lessen Statistische Fysica voor mij een fantastische wereld heeft geopend waar ik elke dag nog verwonderd in rondloop. Ik ben erg vereerd met het vertrouwen dat hij in mij gesteld heeft. Verder ben ik veel dank verschuldigd aan Milan voor zijn aangename begeleiding; hierbij horen ook Benjamin en Prof. Dr. K. Schoors, die mij bijzonder vriendelijk onthaalden.

Zonder mijn familie en het eindeloze geduld van mijn vrienden zou dit werk nooit gelukt zijn (en het was inderdaad bijna zo). Ik ben aan hen eindeloze dank verschuldigd.

Deze scriptie is opgedragen aan Helena.

Marnix Van Soom, juni 2017.

Multiscale analysis of the Russian interbank loan network

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

door

Marnix Van Soom

Samenvatting

Het doel van dit werk betaat erin om een uitgebreid methodologisch onderzoek van het Russische interbanknetwerk toe te voegen aan de relatief jonge literatuur over multiplexe interbanknetwerken. Het hoofdthema van deze thesis is het onderzoek naar de verschillende schalen die aanwezig zijn in complexe netwerken. We argumenteren dat interbanknetwerken complexe systemen zijn (in de formele zin van het woord) die emergente eigenschappen vertonen, en dat netwerktheorie uitermate geschikt is om dit fenomeen te beschrijven. In de netwerkliteratuur is vrij recent duidelijk geworden dat multiplexe netwerken op doeltreffende wijze een systeem dat bestaat uit spelers die op verschillende manieren met elkaar kunnen interageren kunnen modelleren. Wij hebben dit toegepast op het Russische interbanknetwerk door formeel alle leningen met eenzelfde termijn in aparte lagen te plaatsen, waarna we de topologie en de structuur van deze zogenaamde termijnlagen kunnen bestuderen. We herstellen daarna de orde door te vragen of de termijnlagen een werkelijke toegevoegde waarde bieden bij het modelleren en begrijpen van het gehele interbanknetwerk, en of bepaalde termijnlagen kunnen worden samengenomen om een meer compacte beschrijving van het Russische multiplexe interbanknetwerk te bekomen.

Als eerste resultaat hebben we een algoritme voorgesteld dat de datum van uitgifte kan schatten van transacties waarvan we wel de terugbetalingsdag kennen, alsook een ruwe schatting van de termijn van de transactie. Dit algoritme maakt een beschrijving van het netwerk op wekelijkse tijdsschalen mogelijk, vooral voor de transacties met een lange termijn. Het algoritme is beschikbaar in de vorm van een `R` en `Python` library, voorzien van documentate en veel voorbeelden, in de hoop dat het nuttig kan zijn voor verder onderzoek.

Het tweede resultaat is een karakterisering van de topologische structuur van de termijnlagen. Nadat een lijst was opgesteld met stylized facts van interbanknetwerken, is iedere termijnlaag aan een uitgebreid onderzoek onderworpen. We stelden vast dat alle termijnlagen een hubstructuur bezitten – hubs zijn gecentraliseerde banken die leningen verhandelen met een groot aantal tegenpartijen –, maar dat de financiële functie van de hubs afhangt van laag tot laag. Een eenvoudig model van liquiditeitsdissipatie toonde aan dat de hubs in de `<1d-`, `2-7d-` en `8-30d`-lagen aan financiële

intermediatie doen en zogenaamde tiering banks zijn die liquiditeit intermediëren tussen periferiebanken. Vervolgens hebben we het core-periphery-algoritme van Della Rossa et al. [1] aangepast en uitgebreid om twee doelen te bereiken. Ten eerste moet voor de originele implementatie aan de zware voorwaarde van strongly connectedness voldaan zijn. Het probleem is dat vooral voor langere termijnen de grootste strongly connected cluster maar een zeer beperkt aantal leningen en banken vertegenwoordigt, en dat de economische interpretatie van deze voorwaarde onduidelijk is, waardoor deze moeilijk te rechtvaardigen valt. Door PageRank-centralisatie te gebruiken in plaats van een standaard random walk was het mogelijk deze voorwaarde te omzeilen, zij het door het toevoegen van een nieuwe parameter in het probleem, namelijk de damping factor. Ten tweede kan het algoritme van Della Rossa et al. enkel core-peripherystructuur afleiden. Het vervangen van hun α-profiel door zogenaamde intermediatieprofielen stelde ons in staat om de intermediatie-activiteit van de hubs in eender welke laag te meten.

Door dat te doen vonden we dat er drie soorten termijnen zijn, gebaseerd op de topologie van de onderliggende termijnlagen. De drie soorten termijn zijn de tiering layers, de flat layers en de source-sink star layers. De tier layers bestaan uit leningen met `<1d-`, `2-7d-`, `8-30d-` en `31-90d`-termijnen. Het interbanknetwerk in deze termijnlagen zijn georganiseerd volgens de core-peripherystructuur, waarin de hubs de grootste intermediators in het netwerk zijn. Ze gedragen zich als money centers die liquiditeit dissiperen, genereren en herdistribueren. Deze tieringstructuur veroorzaakt systemic risk die geässocieerd is een hoger risico op contaminatie wanneer een core bank problemen ondervindt of zelfs failliet gaat. De flat layers bestaan uit leningen met `91-180d-` en `.5-1y`-termijnen. Hoewel de hubs nog steeds het grootste deel van de intermediatie op zich nemen zijn de banken in de periferie rond de hubs gevoelig meer geconnecteerd vergeleken met de tiering layers. Dit zorgt voor minder steile intermediatieprofielen. De netwerken in de flat layers bestaan uit vele losse (i.e. niet verbonden) componenten, waarvan de meerderheid slechts enkele banken telt. De grootste verbonden componenten zijn dense; bovendien herbergen zij de hubs. Tenslotte zijn de hubs in de source-sink star layers (termijnen `1-3y` en `>3y`) sinks (consumenten) of sources (bronnen) van liquiditeit. Zij zijn verbonden met de periferiebanken die zelf nagenoeg niet verbonden zijn met elkaar. De source-sink star layers zijn netwerken die uit een sterachtige hubstructuur en een klein aantal losse componenten bestaan.

Het laatste belangrijke resultaat heeft te maken met de reduceerbaarheid van de termijnlagen. De bovenstaande classificatie met de drie soorten lagen houdt stand bij de vergelijking met de optimale layer bins volgens de Jensen-Shannon-afstand. We hebben de methodologie van De Domenic et al. (2015) [2] geïmplementeerd om de optimale partitie te vinden van de termijnen die aangeeft welke termijnlagen geäggregeerd kunnen worden met een minimaal verlies aan informatie. Vervolgens hebben we getoond hoe de Jensen-Shannon-afstand gebruikt kan worden voor beeldherkenning en om een epochs te vinden in temporal layers. De temporal change points die tussen de epochs voorkomen komen in grote mate overeen met onderliggende gebeurtenissen, alsook de groei– en maturiteitsperiode van het

interbanknetwerk. We vonden ook een goede kwalitatieve overeenkomst tussen de epochs voorspeld door de Jaccard afstand, wat een bemoedigend teken is voor de vooralsnog eerder obscure von Neumann-netwerkentropie. Om de waarde van de termijnlagen te testen in een robuuster formalisme hebben we Bayesiaanse modelselectie uitgevoerd om te bepalen of de termijnlaagrepresentatie voldoende informatie levert over de mesoschaalstructuur van het gehele netwerk; het antwoord was positief.

We hopen dat dit werk een waardevolle bron kan zijn voor verder onderzoek naar multiplexrepresentaties van interbanknetwerken, aangezien het zowel de tools daartoe als een uitgebreide case study van het Russische multiplexe interbanknetwerk bevat.

Academiejaar 2016–2017

Promotoren: Prof. Dr. J. Ryckebusch, Prof. Dr. K. Schoors
Scriptiebegeleiders: Msc. M. van den Heuvel

Department of Physics and Astronomy,
Faculty of Science,
Universiteit Gent.

Contents

Chapter 1

Introduction

Physicists have made outstanding contributions to diverse branches of science, such as biology and sociology. Only relatively recently, physicists and economists have started collaborating to tackle the complexity problems inherent in economics [3]. With the birth of *econophysics*, the gap is finally being closed, as this cooperation has immense potential in understanding economical aspects of the increasingly interconnected world [4].

Most economic markets are *complex systems*. A complex system may be loosely defined as a set of interconnected and interdependent agents[1] capable of producing a variety of patterns that lie somewhere between order and randomness. Complex systems are studied in *complexity science*, a branch of science that originated in the late sixties and has grown to become a broad and multi-disciplinary subject. Partly because of that, currently no universally accepted definition of a complex system exists [6]. In the context of economics, the three most important properties of a complex system are feedback, robustness and the emergence of some spontaneous order and self-organization absent (strong) central control. Somewhat ahead of business, we point out that earlier work has already studied feedback and robustness features of the interbank lending system examined in the next Chapters, respectively using control theory [7] and network percolation [8]. This thesis focuses on the third property.

An interesting paragraph from Jaynes (2003) illustrates these concepts in a macroeconomic setting.

> It seems likely [...] that the 'turbulence' of individual variations in economic behavior is the engine that drives macroeconomic change in the direction of the equilibrium envisaged by Adam Smith. The existence of this turbulence was recognized by John Maynard Keynes (1936), who called it 'animal spirits' which cause people to behave erratically; but he did not see in this the actual cause that prevents stagnation and keeps the economy on the move.
>
> In the next level of understanding we see that Adam Smith's equilibrium is never actually attained in the real world because of what a physicist would call 'external perturbations', and what an economist would call 'exogenous variables'

[1] An agent is the basic unit in the system under consideration. Four typical examples from the literature are molecules, banks, transportation stops and people [5].

> which vary on the same time scale. That is, wars, droughts, taxes, tariffs, bank reserve requirements, discount rates and other disturbances come and go on about the same time scale as would the approach to equilibrium in a perfectly 'calm' society.
>
> The effect of small disturbances may be far greater than one might expect [...]. If small individual decisions (like whether to buy a new car or open a savings account instead) take place independently, their effects on the macroeconomy should average out [...], to show only small ripples with no discernible periodicity. But seemingly slight influences (like a month of bad weather or a 1% change in the interest rate) might persuade many to do this a little sooner or later than they would otherwise. That is, a very slight influence may be able to pull many seemingly independent agents into phase with each other so they generate large organized waves instead of small ripples.
>
> Such a phase-locked wave, once started, can itself become a major influence on other individual decisions (of buyers, retailers, and manufacturers), and if these secondary influences are in the proper phase with the original ones, we could have a positive feedback situation; the wave may grow and perpetuate itself by mutual reinforcement [...]. Thus, one can see why a macroeconomy may be inherently unstable for reasons that have nothing to do with capitalism or socialism. Classical equilibrium theory may fail not just because there is no 'restoring force' to bring the system back to equilibrium; relatively small fortuitous events may set up a big wave that goes instead into an oscillating limit cycle – perhaps we are seeing this in business cycles. To stop the oscillations and move back toward the equilibrium predicted by classical theory, the macroeconomy would be dependent on the erratic behavior of individual people, spreading the phases out again. Contrarians may be necessary for a stable economy! [3, p. 233]

Looking at Jaynes' macroeconomy as a complex economic system, the agents are the buyers, retailers and manufacturers that interact by buying and selling services and goods. The collective outcome of individual intelligent activities depends on the partly competing, partly cooperating interactions of the multitude of agents, which causes the market's self-regulation capacity [9]. We can identify this with the emergence of order and self-organization in complex systems. The self-regulation implies that the system is quite robust, although every so often local feedback may provoke nonlinear effects on larger scales [10, 11].

The use of the terminology of complex systems in economics has spread considerably in the last decade. For example, the following words are from Bernanke, then appointed Chairman of the Federal Reserve, about the 'Flash Crash' of May 6, 2010:

> The brief market plunge was just a small indicator of how complex and chaotic, in the formal sense, these systems have become. Our financial system is so complicated and so interactive — so many different markets in different countries and so many sets of rules.
>
> What happened in the stock market is just a little example of how things can cascade or how technology can interact with market panic. [12]

Let us now turn our attention not to stock markets or macroeconomic systems, but to the interbank lending market.

1.1 Interbank lending markets

If we recognize that complex systems constitute promising models for various economic markets [10], then evidently *interbank lending (money) markets* may be described by them as well [13]. This thesis studies the Russian interbank market from August 1998 until November 2004 derived from a data set containing the interbank exposures during this period. How this is done will be discussed at length in Chapter 2.

As complex systems

The interbank money market reallocates the liquidity originally supplied by the central bank, which we will approximate as an external influence in this thesis. It is crucial for the transmission of monetary policy, mainly via the central bank's lending and deposit rates in normal times. One reason for the aforementioned reallocation is the offset of anticipated and non-anticipated daily liquidity imbalances. Furthermore, banks are motivated to participate in the interbank market for speculative purposes [14]. In short, banks lend money to each other if for some reason this is more beneficial than turning to the central bank.

Interbank lending markets are considered to be extremely liquid, but this can change under the influence of mutual distrust induced by asymmetrical information, possibly leading to liquidity hoarding and a complete dry-up of funding. This is reminiscent of the feedback effects mentioned above. We can rephrase this by saying that banks are susceptible to *market sentiment* [10].

Furthermore, we should expect that banks organize their lending and borrowing activity in function of risk minimization. One possible way of doing this is by establishing long-term *trading relationships* [10, 13, 15]. As long as their counterparties meet their expectations, they should have little reason to seek new, unfamiliar partners.

Susceptibility to market sentiment and trading relationships both relate to *behavioral finance* [16]. The interaction of many adaptive agents leads to coordination at the global level that would have been hard to infer from the individual level. The complex system as a whole generates emergent phenomena [10]. In the case of interbank lending markets, it can be conveniently implemented by a *complex network*[2].

As complex networks

Traditionally, the interbank market has been considered without taking the network structure into account [10]. Early work modeling the interbank lending market as a network includes Allen and Gale (2000), who consider contagion effects of defaulting banks in simple stylized network structures [17]. Since then, increasingly sophisticated network models have been proposed and fitted to empirical interbank exposures [13].

[2]Perhaps gratuitously but for completeness, we note that a complex network is simply defined as a network that represents a complex system.

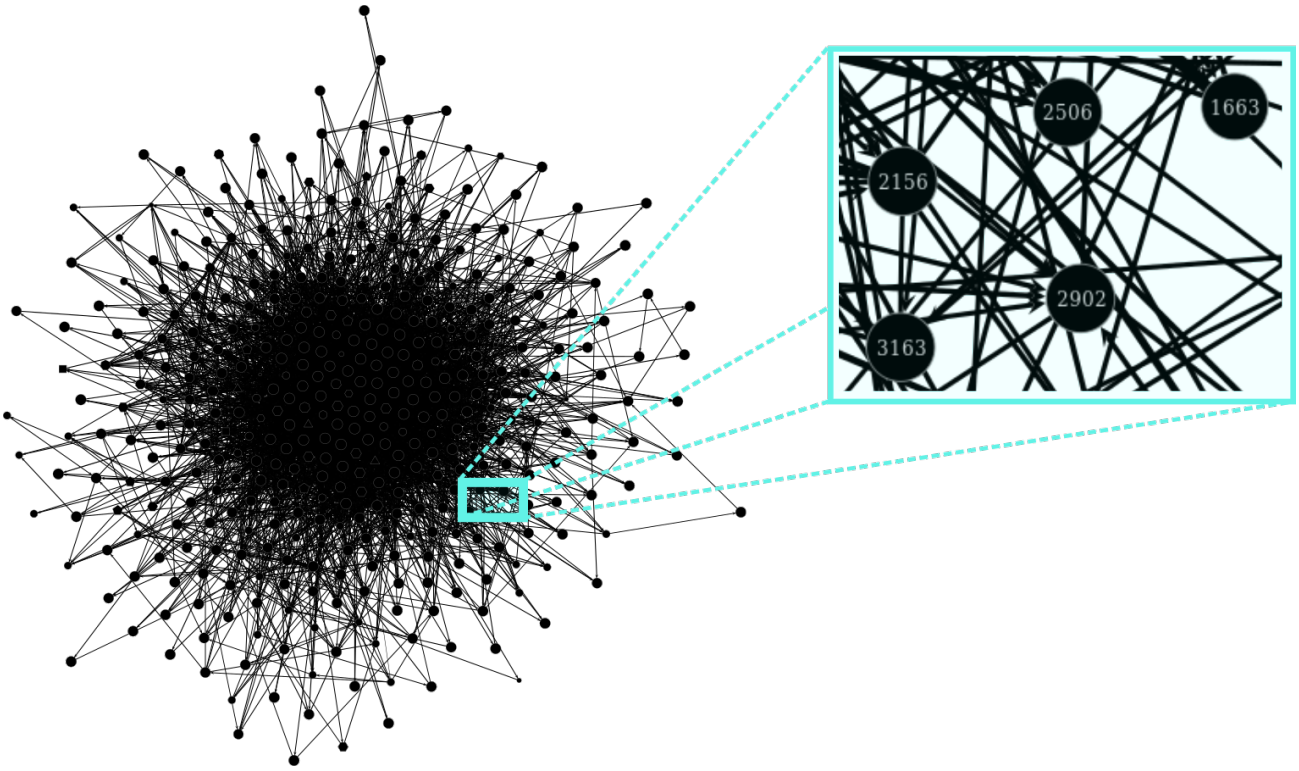


Figure 1.1: A snapshot of the largest connected component (see Chapter 3) of the Russian interbank lending network during the week of April 19, 2004. The network contains 416 banks and 3948 exposure links. The width of the edges is proportional to the order of magnitude of the total exposure between pairs of lenders and borrowers during that week. We observe that the cluster, which accounts for 74% of the total lending activity, has a core of densely connected banks. The magnified picture shows the banks with their registration numbers and the directed edges.

But why model the interbank market as a network? According to Hüser (2015), banks are highly interdependent, being connected both via the asset and liability sides of their balance sheets. Credit exposures create links between banks. These connections can be conveniently represented by a network where banks – nodes – are connected by loans – edges – pointing from lender to borrower [13]. One may also generalize this by using aggregated exposures during some period of time instead of loans. Figure 1.1 may help picturing this; the detailed procedure for the Russian interbank network will be set forth in Chapter 3. We also mention that banks are indirectly linked through holding similar portfolios or sharing the same mass of depositors [18]; models have been proposed that take this into account as well [19, 20]. This will not be done in this thesis.

The interbank network representation is capable of grasping several styl-

ized facts of interbank markets [13]. First, the *systemic risk*[3] inherent in interbank markets can be assessed by network measures such as DebtRank [21]. The literature has seen a vast increase in research concerning systemic risk and debt contagion after the global financial crisis of 2007, using network theory in the majority of cases [10].

Second, it is convenient to weigh the edges by the loan size, as is done in Figure 1.1, and different types of loans can be associated with each edge. In this thesis, we will discriminate between the loan terms (maturities) – particularly we will investigate in Chapter 3 which network patterns emerge when we look only at loan of a certain term.

Third, and most important, a network representation is the only viable option to study the *topology* of the interbank exposures, i.e. the way the banks connect themselves. In particular, we can assess core-periphery patterns [22] or even arbitrary *community structure*[4] [23]. Classical or modular community structure implies that the network is roughly partitioned into groups of nodes that are densely connected within them, with a lower density of edges between the groups – we will return to this in Chapter 4. It has been shown that community structure tends to increase contagion risk [10,17]. Communities of banks are expected to play a role in the Russian interbank network because it is a large sector with considerable geographical spread[5]. To conclude, we note that most existing studies on the structure of real interbank markets have been conducted by physicists trying to get an idea of the topology of directed interbank networks [24].

1.2 Scales in complex networks

The self-organization present in complex systems can be related to[6] their tendency to display structure on one or more scales. Paraphrasing an influential article by Ladyman and Lambert (2012), a necessary property of a complex system is robust order (which we have called self-organization in the context of economics) which enables a macroscopic level to arise out of microscopic interaction. Furthermore, this macroscopic level is stable on its time or length scale [6, p. 27].

'Length scales'

These rather abstract propositions about complex systems readily translate to complex *networks*. If we ignore the time scale for now, we can identify the length scale in physical systems with *mesoscale structure* in complex networks. The conventional approach of representing mesoscale structures is to

[3]The risk that a failure of a single institution poses to the system as a whole [18].

[4]In the literature, communities in networks are also called blocks or groups. We will use the three terms interchangeably.

[5]Research on the Dutch interbank network that is relatively small compared to ours indicates a negligible community effect [10].

[6]We are reluctant to use 'causes' here, since causality in complex systems is a subtle topic.

separate the nodes into groups that have a similar role in the network topology [25]. Thus *one can obtain the mesoscale structure of a complex network by finding its community structure.* This is an important insight. Indeed, as most real-world networks are simply vast, one can only hope to understand them by statistical analysis [26], which essentially boils down to finding a meaningful lower-dimensional representation of a large amount of data. Important clues for this difficult task lie in the different scales exhibited by the system. Returning to the topology of large complex networks, we now argue that community structure plays the role of summarizing the network topology. Moreover, the mesoscale structure appears if the inferred community structure is clear. We will do this in Chapter 4 using the *stochastic block model* (SBM) [23].

Community structure is not the only 'length scale' present in complex networks, as other methods exist. In particular, smaller scales may be studied using *motif analysis* [5, 7]. Motifs are small sub-graphs recurring within a network with a frequency higher than expected in random graphs. Their importance resides in the fact that they can be understood as basic building blocks, each associated with specific functions within the global system. Consequently they indicates scales defined slightly above the single node level. In light of this, we can think of communities as mesoscale structures mainly defined at a global level [5]. See Figure 1.2 for an application of both concepts on the Russian interbank network.

For completeness, we note that community structure can also be generalized in the following way: one aggregates each community of nodes into so-called block nodes, obtaining a simple directed block graph that has no parallel edges. The block nodes are then connected by directed edges weighted by the amount of links between the original communities. Finally we again look for communities, but now in in the block graph. Iterating this procedure yields a description of the network topology at multiple scales. These *hierarchical block structures* [28] can be inferred in a non-parametrical way using nested SBMs [23]. What is particularly convenient about using them is that one need not not impose a scale *a priori*, for example by constraining the maximal number of communities. Rather, the number of appropriate mesoscales is inferred by the nested SBM itself. In this thesis, we will use the non-nested SBM mainly for simplicity[7].

Time scales

One way of identifying time scales in time series of complex networks is by finding different phases in the underlying dynamics, which is equivalent to the detection of 'change points' in networks [30]. For example, in the case of the Russian interbank network we expect to be able to discern the growth and maturity phase and the crises in 1998 and 2004. A time series is defined by a set of snapshots of the network, possibly aggregated during time periods such as weeks or months.

Many strategies exist for finding change points [25, 30–32]. In Chapter 4 we approach the problem heuristically by thinking of the monthly snap-

[7] Another practical reason is that the nested SBM algorithm in [29] was unstable when applied to our large multi-directed interbank networks.

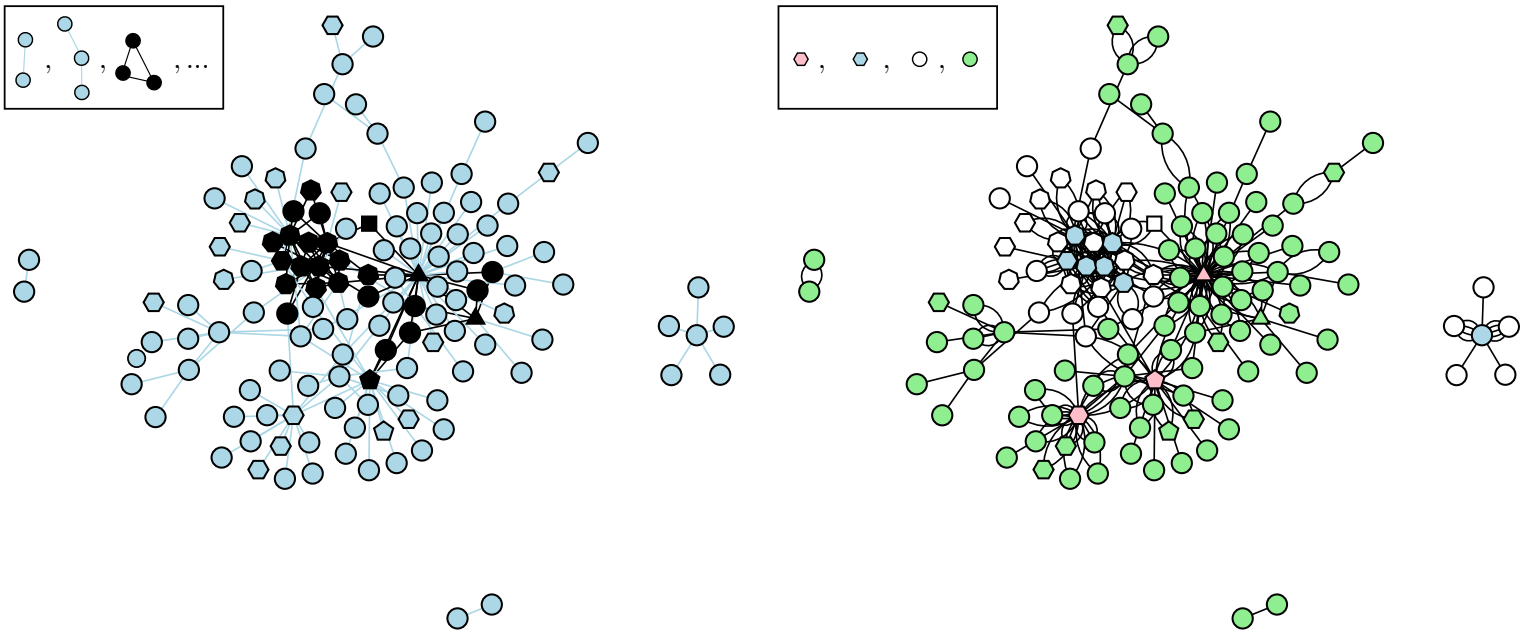


Figure 1.2: Illustration of two different scales in a snapshot of the undirected interbank network during the month of September 1998. (left) Two banks are connected if they had a loan contract at least once during this period. All triangle motifs are indicated in black. We see that they are near the network core. One can measure the network *transitivity* using triangles, which are second-order motifs. In the language of social networks, transitivity measures how likely the friend of your friend also is your friend [26]. We come back to this in Chapter 3. (right) Edges represent undirected loan contracts. The community structure is inferred using the non-degree corrected SBM (which we'll discuss in Chapter 4) with a maximum of $B = 5$ blocks. We clearly see a core-periphery structure for white-blue and green-pink. In the language of the Basel III accord, the pink nodes might be characterized as tier banks, or as 'money centers' [27]. Three disconnected components are also present, which implies reduced systemic risk for the banks belonging to these components.

shots of the undirected Russian interbank network as different *layers* of the network, as is done for three consecutive months in Figure 1.3. Then we define a *distance metric* to measure the 'distance' between a pair of layers. The idea is then to aggregate (or 'bin') layers that are close by, i.e. similar, according to some *quality function* that controls the aggregation process. It does this by reaching a maximum for the optimal partition of layers in layer bins, i.e. *time epochs*. In principle, the change points then lie at the boundaries of the optimal partition.

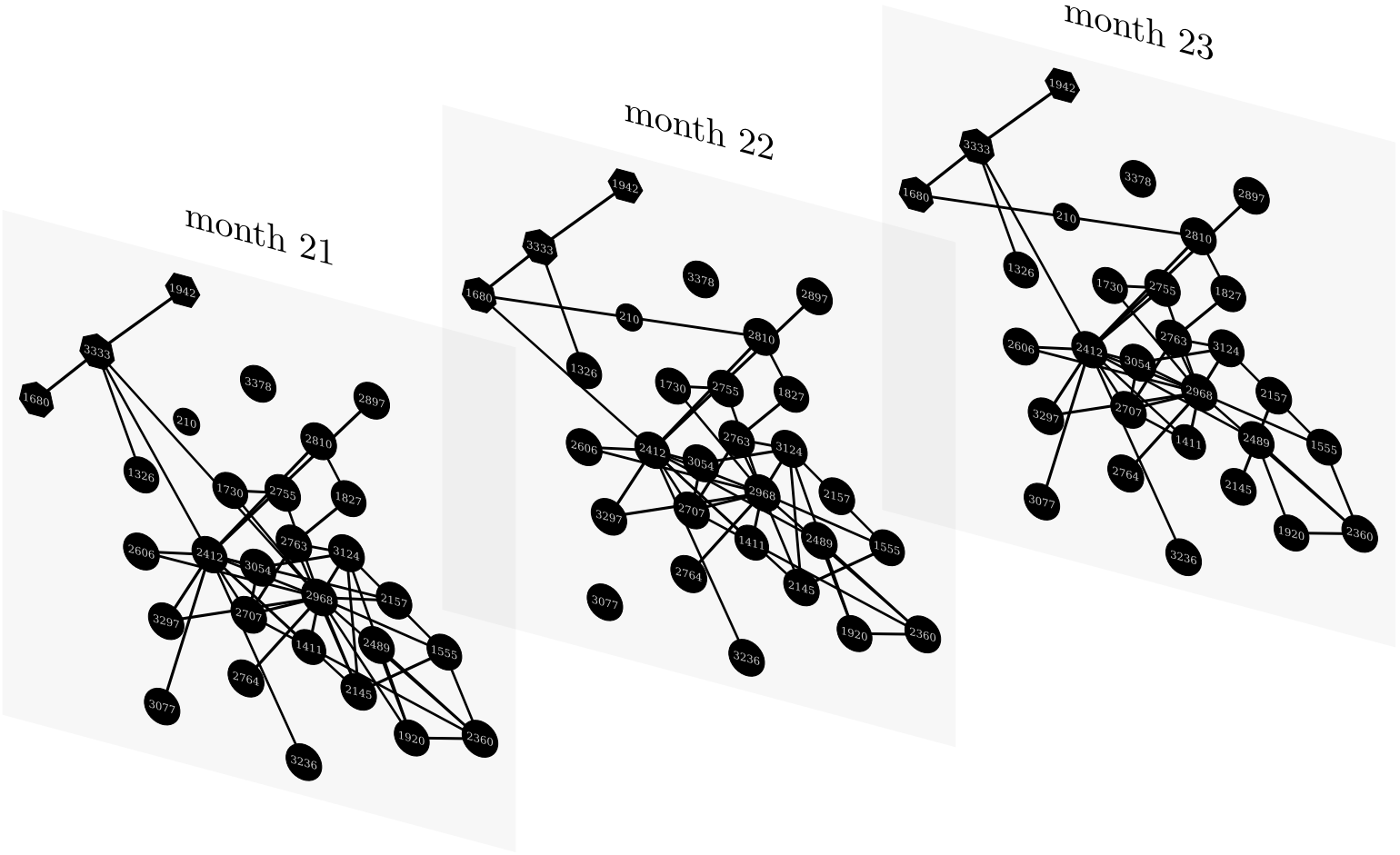


Figure 1.3: Three snapshots of the neighborhood of bank `2968` during April, May and June 2000, laid out as layers. Banks are connected if they had a loan contract at least once during that month. The banks presented here may be thought of as being in the 'address book' of bank `2968`. Note how the majority of edges persist throughout the layers [33].

Layers

The above strategy is based on the idea of layer aggregation from De Domenico et al. (2015) [2], with the layers being the monthly network snapshots in our case. The procedure is fully generic, however, and can be applied to any concept of 'layers' in networks.

Networks with inherent or synthetic layers are called *multilayered networks*. The Russian interbank network has a natural multilayered representation because each loan has a distinct *term* (maturity). The different term categories are listed in Table 2.2 on page 11.

One expects that the lending and borrowing strategies of banks depend

primarily on the loan terms, *ceteris paribus*, since various lending risks[8] are associated with the term's length. This implies immediately that the topology of the term layers could differ considerably [20, 34–36]. Therefore in the Russian interbank network we may consider the term layers to be a third scale. Of course, we do not know *a priori* whether this scale is meaningful; for example, if the topologies are quite similar separating the loans into the term layers may be redundant. We investigate the term topologies in Chapters 3 and 4.

1.3 Outline of the thesis

The remainder of this thesis is structured as follows.

In Chapter 2 we scrutinize the data set and devise a complete framework that allows one to study the interbank network at weekly time scales efficiently. Furthermore the yield curve and several other economical features are extracted from the data and reviewed.

Chapter 4 constructs the multiplex interbank network from the data by putting edges representing loan contracts with the same term in separate layers. We perform a small literature study to enumerate the stylized facts of interbank lending networks and discuss the basic terminology and utility of multiplex networks, including multiplex interbank networks. Finally, we perform a thorough network analysis of the term layers to characterize their nature. We discuss and find evidence for financial intermediation. The most important result is the discovery that the terms exhibit a natural classification, which emerges from the fact that groups of layers exhibit similar topological structure and intermediary functions. We consolidate this finding by naming the groups of layers; these are the tiering layers, the flat layers and the source-sink star layers.

Building upon these findings, Chapter 4 investigates whether the layers can be aggregated into layer bins using the rather experimental and recent concept of the von Neumann entropy of graphs. We introduce temporal layers and organize the months into seven epochs with the same technique. As an additional check, we perform model selection using generative models of layered networks to infer whether the term layers really are informative with respect to the mesoscale of the interbank network.

Finally, we summarize our findings and suggest possible future research in Chapter 5.

[8] We list a few in Section 2.6.

Chapter 2

Overview of the interbank data

The interbank data was provided by Schoors and Karas [37] and offers a rich snapshot of Russian bank activities over a six year span. Although carefully assembled from different sources, both public and private, it contains many entries that demand specific interpretation. In this chapter, the data is presented and prepared for network analysis.

The data contains two main panels: monthly bank balance sheets and domestic interbank transactions. Both are reported to the legislator on the first of each month throughout August 1998 up until November 2004, with the exception of January 2003. The second panel is of interest to this thesis, since these ≈ 3.3 million transactions can be mapped to a network of banks linked by unsecured loans and deposits. This feature allows one to study the topology of interbank transactions, making this data set particularly useful and unique in the literature [38, 39]. On average, about half of the Russian banks are active on the interbank market [39].

Table 2.1 lists the most important columns from a selection of the transaction data. Each transaction consists of a volume lent out to the borrower at a certain price, the interest rate. On the due date, when the transaction term has elapsed, i.e. at maturity, the lender is repaid. The different terms, together with some comment, are listed in Table 2.2. The `reported` column indicates that the interbank loan or deposit happened in the month *before* that date[1]. As an example, consider the first row in Table 2.1. Since the loan's term is unambiguous, it is possible to deduce the issuance date of the loan: November 12, 1998. Indeed, the reported date is next December. Estimating when loans and deposits were issued is crucial for any time series analysis, and this subject will be discussed extensively.

Table 2.1: A small sample of bank 2285's transactions, with several inconveniences. Observe that r is denoted in percent, not in basis points.

lender	borrower	r (in %)	volume (millions of rubles)	term	type	reported	due
2003	2285	35	1	<1d	loan	1998-12-01	1998-11-13
339	2285	0	0	31-90d	loan	1998-11-01	
574	2285	21	0	8-30d	deposit	1998-08-01	
574	2285	17	2.1	2-7d	loan	1998-12-01	1998-12-10

Table 2.2: Enumeration of the interbank loan and deposit terms. Terms marked in grey are not considered in the next chapters.

term	note	term	note
overdraft	Deficit in a bank account caused by drawing more money than the account holds [40]. Overdraft loans are charged high interests.	31-90d	
		91-180d	
		0.5-1y	Six months to one year.
<1d	Overnight loans, to be repaid the next day.	1-3y	
2-7d	Lowest interest on average.	>3y	
8-30d		demand	Variable term and interest.

2.1 Brief timeline

The following history of the Russian interbank network between August 1998 and November 2004 is based upon previous work by van den Heuvel [7] and Vandermarliere et al. [8, 39].

The data started recording a few weeks before the August 1998 ruble crisis. On August 17, 1998, Russian officials decided to default on the domestic debt, abandon the exchange rate regime and declare a moratorium on all private foreign liabilities. The interbank market during the months preceding this crisis may be assumed to have partly anticipated the default, which results in structural change in the topology of the underlying network of interbank loans and deposits. Unfortunately, there is no record of this period. The interbank market collapses a second and third time in September and December 1998, respectively. The crisis completely paralyzed the interbank market. The recovery took more than a year [38].

A second and less severe crisis was one of trust, inducing a liquidity drought during the summer of 2004. It was triggered by the launch of an investigation into the suspected money laundering and terrorism sponsoring of certain banks. Almost a year before, the announcement of the investigation already caused a smaller trust crisis. In this thesis, 'trust crises' denote both the announcement crisis in the end of 2003 and the summer of 2004 crisis. Because there is a full recording of the network before these, they are

[1] Since no transactions have been reported on January 2003, there are no data on loans or deposits issued in December 2002.

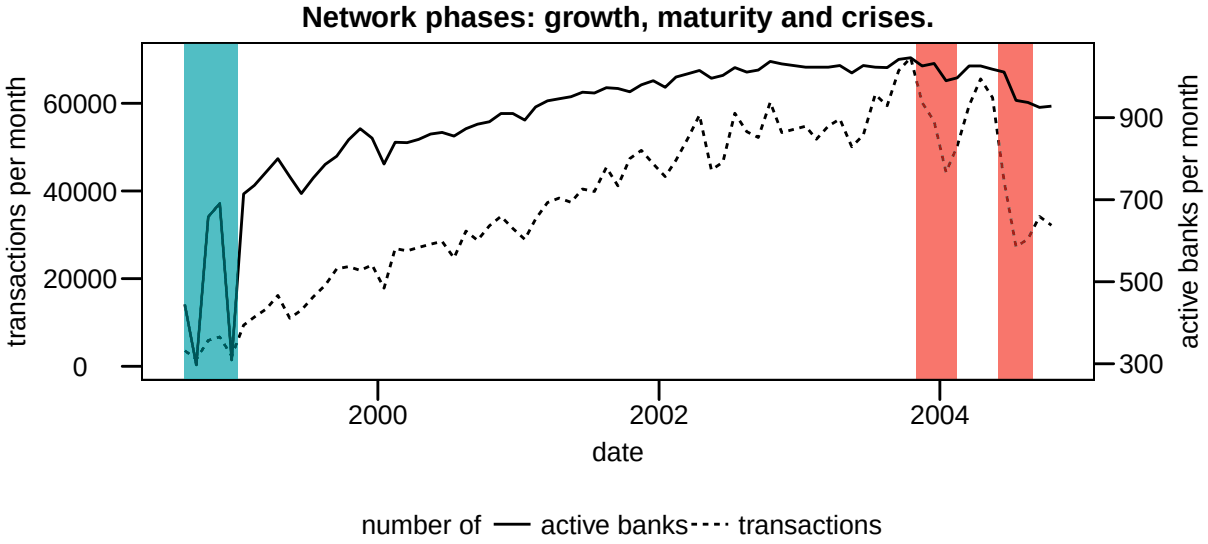


Figure 2.1: (based on Figure 1 in [39]) The evolution of the network is illustrated by two basic statistics: the monthly amount of transactions, i.e. loans and deposits, and banks active on the market. The issuance dates are deduced from the reporting date. Roughly until 2002, one sees a steady incline of issued loans and deposits and active banks. This buildup is referred to as the growth phase. After 2002, the number of banks stabilizes as the interbank market becomes saturated during the mature phase. Two large crises occur, shaded blue and red. The first crisis in August 1998 triggers two secondary network collapses next September and December. The trust crises are shaded in red. See Appendix A for a more detailed description of the crises through a list of selected events. Each crisis represents another phase in the network structure.

more valuable for analysis compared to the crisis of August 1998.

Apart from the crisis phases in the data, two tendencies on a longer time scale may be discerned in Figure 2.1. These reflect the growth and maturity phase of the network, respectively before and after 2002. During growth, the number of active banks and issued loans and deposits per month steadily increases. The Russian banking sector began to play a more prominent economic role in the 2000s thanks to favorable macroeconomic conditions and institutional development, particularly improvements in banking regulation [41]. In the mature phase, the interbank marketplace sees a stable amount of competitors while the transaction intensity increases. One can really sense this growth by comparing Figures 1.1 and 1.2 on pages 4 and 7 respectively; the former is a snapshot of the largest connected component during one week in April 2004, while the latter is the complete network in September 1998.

Turning to Figure 2.2, the progressively more sober policy of the Central

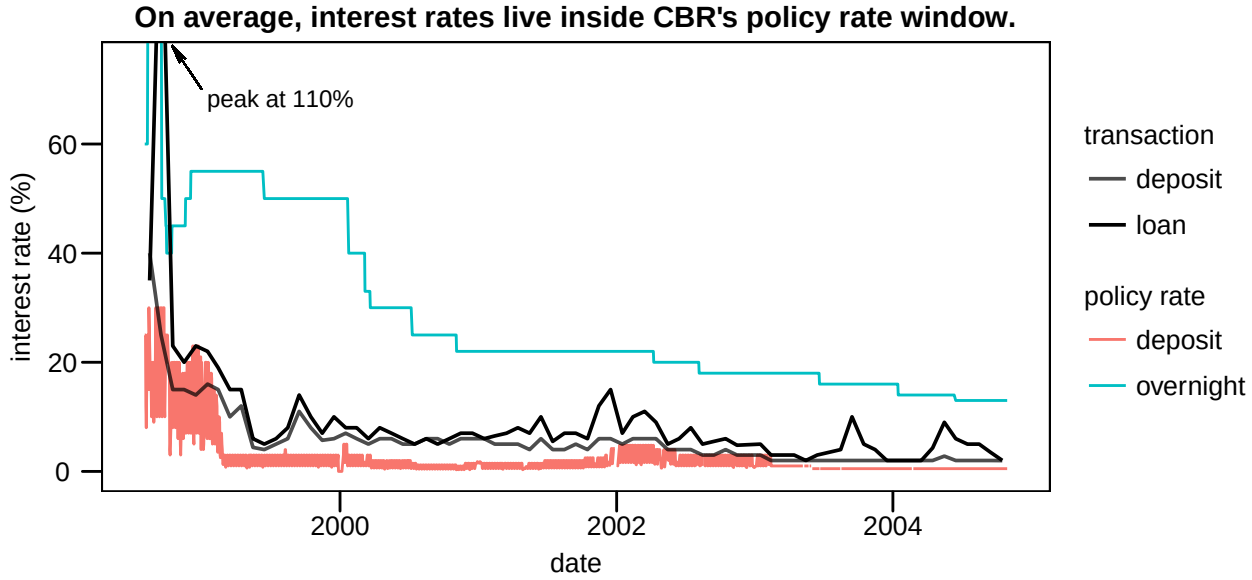


Figure 2.2: The median interest rate of loans and deposits is calculated over all terms and by monthly aggregation using the `reported` date. Because the interest rate distribution is skewed by several extremely large values ($> 50\sigma$), the median is chosen to represent the typical interest rate. In addition, transactions with zero interest rates have been excluded from the calculation (see Section 2.2). Central banks influence money market interest rates through their deposit and lending rates. In Russia, the Central Bank of Russia (CBR)'s monetary policy plays a key role in price-setting in the short-term segment of the money market [41]. Daily policy rates may be consulted on its website, www.cbr.ru [42, 43].

Bank of Russia (CBR) from 2001 onwards reveals a gradually maturing market. The strong policy of the Central Bank during the first crisis indicates its severity. The two spikes in the average loan interest rate near 2004 are reminiscent of the tense climate during the second crisis.

It is important to discriminate between different phases of the network, because they set natural domains for descriptive statistics and modeling. Statistics from an aggregation of different domains are often unclear – equivalently, from a complementary viewpoint, it is hard to separate distinct scales in aggregated data. For example, Vandermarliere et al. [39] find that the distributions of banks' lending and borrowing amount change before and after the start of the summer of 2004 crisis, regardless of the aggregation time scale (by week, month or quarter).

2.2 A closer look at the data

Several aspects of the data demand special care. About 14% of all entries exhibit a zero transaction volume. These transactions are repayments split

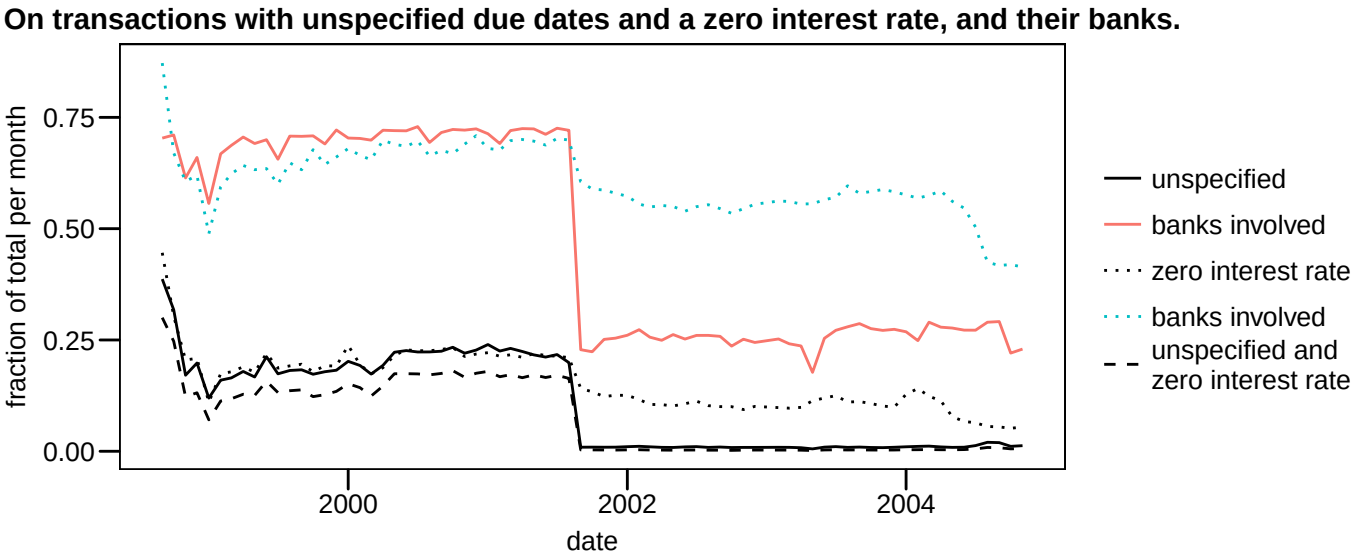


Figure 2.3: Loans and deposits with an unspecified due date and reported before September 1, 2001 account for a substantial and relatively stable fraction of the total of monthly interbank transactions. About 80% of these transactions are free `<1d` and `2-7d` loans, i.e. with a interest rate of zero. It can be seen that they are common usage among the active banks on the market, until after September 2001, their issuance is suddenly cut and only about one in five banks still engage in their trade. This is likely due to the Federal Law On Auditor's Activities of August 7, 2001, a piece of legislation designed to implement the Basel Committee recommendations on internal control in credit institutions and a legal mechanism against money laundering and terrorist financing [45]. In contrast, zero interest transactions remain important throughout the mature phase – though declining steadily.

up in several months and do not represent newly issued loans or deposits. Therefore, they may be safely discarded [44].

Roughly 3% of all remaining transactions have conflicting `term`, `reported` and `due` attributes. Consider for example the last row in Table 2.1. Since the term of this loan is two to seven days, it must have been issued between December 3 or December 7, the latter included – however, this is at odds with the reporting date. To retain such transactions, one could choose two attributes and deduce the third one. However, this would considerably complicate the matter of estimating the issuance date – and it is unclear which dating attribute is more trustworthy, `reported` or `due`. Therefore, transactions with conflicting date and term attributes are deleted from the data set.

The next issue are entries that do not specify due dates. Because the reporting date is known for all entries, an estimation of the date of issuance is at best of monthly resolution. Remarkably, the distribution of these entries over time follows a step function that jumps on September 1, 2001. Figure

2.3 shows that before September 2001, these transactions are almost all zero interest rate transactions. They comprise about 20% of all issued loans. After September 2001, their share becomes insignificant and they decouple from the zero loan and deposit rates. A similar pattern is seen in the share of active banks that trade these contracts: paying and receiving zero interest is a common practice and was linked with the due dates before September 2001.

Considering the seemingly economic absurdity of such loans and deposits, Figure 2.3 and a working paper from the central bank of Norway [46, p. 7], the author concludes that these transactions consist of loans that are repaid – and deposits that are redrawn – before the end of the working day[2].

These loans will not be included in the usual statistics on interbank loan interest rates, because only their initial price was of interest: it is assumed their interest rate on the monthly record had been changed to 0% as the loan was repaid on the same day. Section 2.6 includes an example of such an omission.

Nevertheless, it is obvious to consider transactions with zero interest rates in network representations of the data (see Chapter 3), where they capture the flow of liquidity and a sense of interaction between banks and their balance sheets. Previous work on this data set included these loans without exception [7, 8, 48].

2.3 Inferring the issuance date

Studying the system on different time scales is important. Different scales may exhibit different properties. This was shown explicitly by Finger et al. for interbank networks [24]. The statistics that live in a time scale are produced by aggregating over the correspondent time window. In finance literature, the time windows usually have lengths of days, weeks, months, quarters or years. Daily measures are too volatile and appear random for network analysis [8,24]. As one lengthens the time window, a sense of order and periodicity gradually emerges, at the cost of some detail. According to Finger et al., "[a]ggregating over a longer period is only preferable if it can reveal a non-random structure of the banking network" [24, p. 13]. Monthly compliance with regulatory requirements induces some periodicity into the data, which makes monthly aggregation a natural time scale [39].

For our data, Vandermarliere et al. have investigated whether the distributions of several financial and network measures change considerably with respect to the length of the time windows [39]. While finding that the types of distributions[3] stay unaffected, the authors emphasize that the distribution's parameters do not.

Previous work on the Russian interbank network [7, 8, 48] considered mainly monthly and weekly time windows and only loans with `<1d` and `2-7d` terms. We will use the same window widths, however we will explicitly deal

[2]Personal note: I contacted the CBR and Russian researchers on this issue, but received no reply. The authors of the data set [47] could not offer an interpretation either.

[3]They considered the (truncated) power law (stretched), exponential and log-normal distributions.

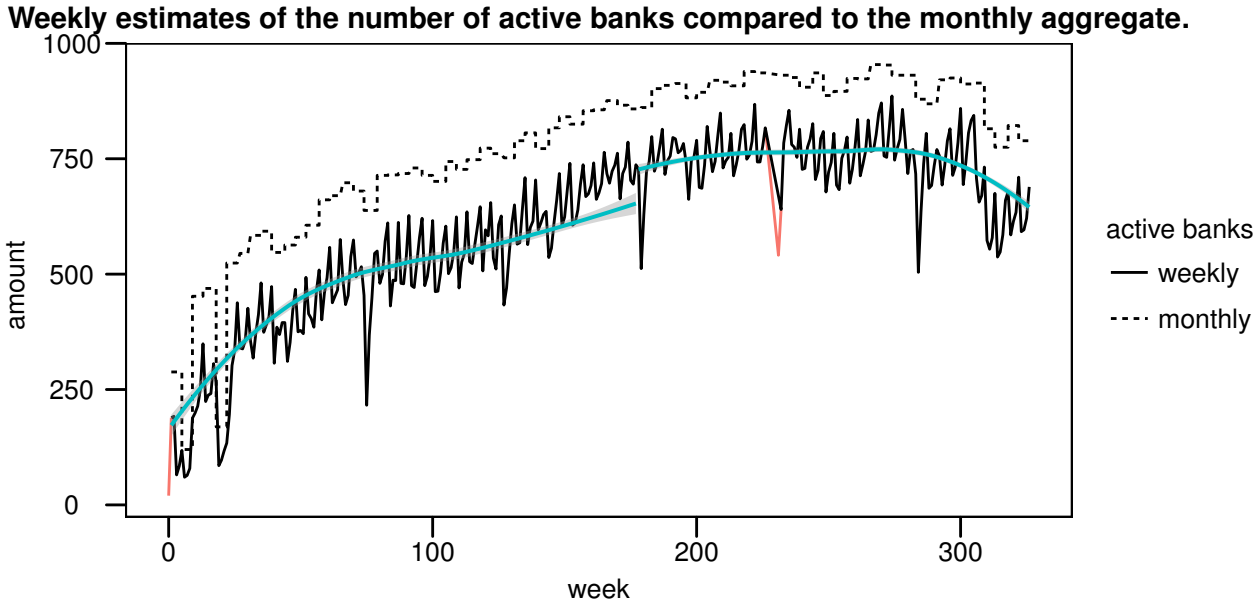


Figure 2.4: (see also Figure 2.1) The estimates of the weekly number of active banks are based on transactions with issuance dates that unambiguously fall in a certain week – about 75% of the total amount of transactions. Table 2.3 lists their composition. A bank is active in a certain time window if it has lent or borrowed at least once during this period. Active banks in a week are a subset of those in the corresponding month. Banks are forced by law to maintain certain reserves by the end of the month. The periodical surge of activity in the last week of each month shows this duty [7]. During the growth phase (`week` < 178), their number averages to 74% of the monthly total. This number naturally rises to 83% in the mature phase (`week` ≥ 178) as the number of transactions increase. These averages are marked by the smooth lines. The red lines are estimates of fragmented weeks that have been removed from the weekly data.

with several more terms to populate the layers in the interbank network. To derive time series of the network, the issuance dates of the transactions need to be reconstructed within accuracy determined by the time window. In the case of a weekly time window, this will cause some trouble, as we will soon see. However, we stress now that if we would content ourselves with only exactly datable transactions, the layers would largely melt away as the term lengths increase. This is contained in Table 2.3. Therefore, *an algorithm estimating the issuance dates of the transactions such as the one presented below is unavoidable for a layered description of the interbank network on a weekly time scale.*

Monthly aggregation is straightforward. A transaction is issued on some day in the month before the `reported` date (remember Table 2.1), thus estimating the issuance on this scale is always correct. Earlier work on our

Table 2.3: The term distribution of the transactions used for the weekly number of active banks in Figure 2.4. That is, the issuance dates of these transactions are dated within weekly precision with 100% confidence. The second row indicates their share relative to the total amount of transactions having the same term.

term	`overdraft`	`<1d`	`2-7d`	`8-30d`	`31-90d`	`91-180d`	`0.5-1y`	`1-3y`	`>3y`	`demand`
count	3448	1348823	670720	20495	4382	551	317	79	3	0
fraction	55%	94%	70%	8.5%	8.4%	5.7%	5.0%	2.5%	0.05%	0%

Table 2.4: Normalized weekday weights in % used for the weekly estimation algorithm. The distribution is sampled from issuance dates known on a daily time scale. In total, 50% of transactions qualify, of which 98% with `<1d` terms.

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
23.7	24.0	24.1	25.3	1.7	0.4	0.9

data shows that it is a reliable proxy [7, 8, 48]; in addition, banks tend to participate in the market at least once per month [39].

The calender week containing the issuance date is also estimated. To appreciate the gain in resolution, consider Figure 2.4. Averaging over both growth and maturity phase, the number of weekly active banks is 78% of those monthly active[4] Accordingly, some banks are only active during one or perhaps two weeks in the average month [39]. In addition, several spikes of low activity ($< 3\sigma$) appear.

The time interval in which a transaction could have been issued is deduced by subtracting the transaction's `due` and `term` attributes. The result is then intersected with the transaction's issuance month. In 8% of all cases, the due date or the term's length is unknown (i.e. `demand` loans and deposits) and the interval defaults to its issuance month. The length of the interval is inversely related to the estimation's precision. For each calender week the interval contains, the probability of the transaction being issued during that week is calculated. This probability depends on the overlapping weekdays of the interval and the calender week. Not all weekdays are equally probable: their weights are estimated by the weekdays' distribution of issuance dates that are known with daily resolution (see Table 2.4). It is assumed the weekday preference is independent of the transaction's term and type. The original entry is replaced by several weighted copies, one branch for each possible week. An example of the algorithm is laid out in Table 2.5.

Applying this algorithm to all transactions, roughly 25% expand to 64% extra loans and deposits; on average the weekly estimation of ambiguous

[4] In good agreement with Vandermarliere et al. (2015) [39].

Table 2.5: (see Table 2.1 for units) Estimating the issuance week of a transaction. Results are shaded in gray. The deposit in the upper table is issued on an unknown day belonging to the interval [1998-08-06, 1998-08-10]. This interval contains two calender weeks, namely the weeks starting on Monday August 3 (week 1) and August 10 (week 2). Consulting Table 2.4 yields an 54% resp. 46% probability of the deposit being issued in the first resp. second week. These probabilities serve as the branch weights of the expanded deposit.

lender	borrower	r	volume	term	type	reported	due	month
1	2402	46	50	2-7d	deposit	1998-09-01	1998-08-13	1

↓

lender	borrower	r	volume	term	type	reported	due	month	week	weight
1	2402	46	50	2-7d	deposit	1998-09-01	1998-08-13	1	1	0.54
1	2402	46	50	2-7d	deposit	1998-09-01	1998-08-13	1	2	0.46

Table 2.6: The distribution of the amount of expansions (branches) needed by the estimation algorithm, denoted in %. Transactions with an unspecified due date necessarily branch out for every (possibly incomplete) calender week the month of issuance contains.

branch amount	1	2	3	4	5	6
all transactions	74.6	10.2	2.64	4.20	6.59	1.78
known due date	80.0	10.9	2.83	4.41	1.48	0.25
unknown due date	0	0	0	1.25	76.2	22.5

entries takes 2.5 more, out of a maximum of 6 – Table 2.6 displays the general distribution. The remaining 75% are not expanded and have a weight of unity: their composition is listed in Table 2.3. These are the loans and deposits that were used to produce Figure 2.4.

As stated in the introduction, we study transactions issued on or between August 1, 1998 and October 31, 2004. There is no recording of any transaction issued in December 2002. The borders between recorded data and non-recorded (or non-observed) data fall on the frames of the monthly time windows. To the contrary, they arbitrarily divide weekly time windows. Because of this, weeks that are not fully included in the observed period[5] are not considered for weekly analysis.

[5]Weeks 0 and 231.

2.4 Statistics of weekly time scales

The formalism of statistical ensembles enables us to infer any statistic on a weekly time scale. Given the set of all transactions $\Omega = \{j\}$, we define a realization of the *complete* network as the assignment of an issuance week w to every element of $\{j\}$, i.e. $\{w(j)\}$. Thus $w(j)$ is a random variable associated with a set of possible *outcomes* w_j denoted by its *alphabet* $\mathcal{A}_j$. This means that we randomly choose one of j's branches, say $w^* \in \mathcal{A}_j$, with a probability $P\left(w(j) = w^*\right)$ that equals the branch's `weight`, from now on denoted in general as $\rho_{j,w}$. In symbols,

$$P\left(w(j) = w^*\right) = \rho_{j,w^*}. \tag{2.1}$$

In the language of statistical mechanics, such a realization is called a microstate $\mathcal{S}$ of the system. Each microstate has an equal amount of transactions n but differs with respect to their distribution over the weeks w. Because some calender weeks w[6] belong to two months m, m', monthly aggregation of microstates is ambiguous – but rather unnecessary, since we have the unexpanded data at hand. The phase space[7] of the complete network is defined by all possible microstates, i.e. combinations of branch choices $\{w(j)\} \in \mathcal{A}_1 \times \mathcal{A}_2 \times ... \times \mathcal{A}_{|\Omega|}$. We stress again that the phase space cannot be partitioned by month, because of the surjectivity of the map $w \mapsto m$.

The expected value or average of a statistic X of the whole network that takes into account the weekly granularity is then defined as

$$\langle X \rangle = \sum_{\substack{w(1)=w_1 \\ w(2)=w_2 \\ \vdots \\ w(n)=w_n}} X(1,2,...)\; P\left(w(1) = w_1 \text{ and } w(2) = w_2 \text{ and } ...\right). \tag{2.2}$$

Here X is a function of the transactions j and the set $\{w_j\}$ that covers every possible combination of branching choices. The number of possible w_j's $|\mathcal{A}_j|$ depends on j; remember that about 75% of the transactions branch only once, hence have a fixed w_j for every term in (2.12) because $|\mathcal{A}_j| = 1$. As a result, the largest part of Ω is fixed, and the degrees of freedom (dofs) in this problem are the issuance week of transactions with two or more branches. Because the assignment of one transaction to a week is independent of other such assignments, we can simplify the microstate probability as

$$\begin{aligned} \langle X \rangle &= \sum_{w(j)=w_j} X(1,2,...,j,...) \prod_j P\left(w(j) = w_j\right) \\ &= \sum_{w(j)=w_j} X(1,2,...,j,...) \prod_j \rho_{j,w_j}. \end{aligned} \tag{2.3}$$

[6] See Figure 2.5.

[7] Also called state space when the dofs are discrete, as in our case [49, 50].

Weeks in a box

Two flavors of X exist. We are interested in statistics on weekly time scales that provide the analogue to the descriptive statistics of the data reported with monthly resolution. These statistics may be *localized* or *correlating*. A localized statistic summarizes the transaction network contained strictly in one week w^* unaware of any week's network lying before or ahead of w^* in time. Because of this 'week in a box', equation (2.3) can be simplified under condition of a second assumption, that will be discussed below.

A correlating statistic describes the network in week w^* possibly taking into account any information in any other week $w \neq w^*$. To calculate a weekly $\langle X \rangle$, we generate an ensemble of microstates $\{\mathcal{S}\}$, determine X for each week w in $\mathcal{S}$ and then recollect these results in the vector $X(w)$. Then we average $X(w)$ over the microstates. Because (2.3) is essentially a weighted sum of X, the basic idea is to approximate its value efficiently by sampling microstates that have relatively heavy weights. We will elaborate on this procedure, called Monte Carlo integration [51–53], towards the end of this section. In the context of this thesis this way of obtaining weekly averages is called *ensemble sampling*. The price one pays for the possibility of having correlating statistics is the computational cost of overseeing the vast volume of the phase space, which is caused simply by the amount of dofs, $|\Omega| = 697\,409$. Therefore in the following section we will discuss several methods that reduce $|\Omega|$ by yielding different *views* of the complete network. However, in the case of localized statistics, we can already downsize the phase space *a priori* by decoupling Ω into sets of dofs Ω^* organized by week.

To calculate the average value of a localized statistic X for a week w^*, we restrict the phase space to the set of microstates $\mathcal{S}^*$ that have at least one $j^* : w(j^*) = w^*$. We can write this down in general as

$$\langle X \rangle_{w^*} = \frac{1}{Z} \sum_{w(j)=w_j} X(1,2,...) \big[1 - \prod_j (1 - \delta_{w(j)w^*})\big] \prod_j P\left(w(j) = w_j\right), \quad (2.4)$$

where δ is the Kronecker delta and $Z = \sum_{w(j)=w_j} \left(1 - \prod_j (1 - \delta_{w(j)w^*})\right) \prod_j P\left(w(j) = w_j\right)$ is a normalizing constant that depends on w^*.

This expression for $\langle X \rangle_{w^*}$ looks complicated. Luckily, it can be simplified considerably. By definition, X can only see dofs that have branched to week w^*. Figure 2.4 shows that an arbitrary week always contains at least one transaction. This means that X will always be defined for every possible microstate and the expression between square brackets in equation (2.4) and consequently Z reduce to unity. As such we can simplify the "at least one j^*" condition: the $\mathcal{S}^*$ microstates are determined exclusively by the set of transactions that have branches reaching w^*, i.e. $j^* : w^* \in \mathcal{A}_{j^*}$, according to the weekly issuance algorithm. Therefore all $\mathcal{S}^*$ are a function of the same set of elements we now relabel $\Omega^* = \{1^*, ..., i^*, ..., n^*\}$, so $|\Omega^*| = n^*$. We can therefore write that

$$\langle X \rangle_{w^*} = \sum_{i^* \in \Omega^*} X(1^*, 2^*, ..., i^*, ...) \prod_{i^*} P\left(w(i^*) = w_{i^*}\right). \quad (2.5)$$

The problem is that we have gained very little: the calculation of such $\langle X \rangle_{w^*}$ still requires expensive sampling from the distributions $\{w(i^*)\}$ while having the restriction of locality. So far, $\langle X \rangle_{w^*}$ is equivalent to a non-correlating statistic obtained by ensemble sampling. We should enhance the value of this procedure by reducing the cost of sampling the i^*'s. This can be achieved by using the fact that $\mathcal{S}^*$ is degenerated with respect to any $w(i^*) \neq w^*$. Consider the projection

$$p(i^*) = \begin{cases} 1 & \text{if } w(i^*) = w^*, \\ 0 & \text{otherwise.} \end{cases} \tag{2.6}$$

We can visit any of 2^{n^*} distinguishable microstates in the phase space belonging to week w^* by considering all possible realizations of $\{p(i^*)\}$. The proposed probability distribution of $p(i^*)$ is straightforward:

$$P\left(p\left(i^*\right)\right) = \begin{cases} \rho_{i^*,w^*} & \text{if } p\left(i^*\right) = 1, \\ 1 - \rho_{i^*,w^*} & \text{if } p\left(i^*\right) = 0. \end{cases} \tag{2.7}$$

The fact that $P\left(p\left(i^*\right) = 0\right) = \sum_{w \neq w^*} \rho_{i^*,w}$ indicates that $\mathcal{S}^*$ excludes the transaction i^* if it branches to any other week besides w^*. If we replace the probabilities in equation (2.5) by these computationally inexpensive Bernoulli distributions, the average of a statistic X in week w^* is now given by

$$\begin{aligned} \langle X \rangle_{w^*} &= \sum_{\substack{p(1^*)=0,1 \\ p(2^*)=0,1 \\ \vdots \\ p(n^*)=0,1}} X(1^*, 2^*, ...) P\left(p(1^*) \text{ and } p(2^*) \text{ and } ...\right) \\ &= \sum_{p(i^*)=0,1} X(1^*, 2^*, ..., i^*, ...) \prod_{i^*} P\left(p(i^*)\right). \end{aligned} \tag{2.8}$$

Let us now consider $X = v$ and $X = n$, respectively the average traded volume and amount of transactions, on a weekly basis. Denote the volume of transaction i^* as v_{i^*} and set $p(i^*) \equiv p_{i^*}$ for lighter notation. We have

$$\begin{aligned} \langle v \rangle_{w^*} &= \sum_{\substack{p_{1^*}=0,1 \\ p_{2^*}=0,1 \\ \vdots}} v\left(1^*, 2^*, ...\right) P\left(p_{1^*}\right) P\left(p_{2^*}\right) ... \\ &= \sum_{\substack{p_{1^*}=0,1 \\ p_{2^*}=0,1 \\ \vdots}} \left(p_{1^*} v_{1^*} + p_{2^*} v_{2^*} + ...\right) P\left(p_{1^*}\right) P\left(p_{2^*}\right) ... \\ &= \left(\sum_{p_{1^*}=0,1} p_{1^*} v_{1^*} P\left(p_{1^*}\right) \right) \sum_{\substack{p_{2^*}=0,1 \\ \vdots}} P\left(p_{2^*}\right) ... \end{aligned}$$

$$
\begin{aligned}
&+ \left(\sum_{p_{1^*}=0,1} P\left(p_{1^*}\right) \right) \sum_{p_{2^*}=0,1} \left(p_{2^*} v_{2^*} + ...\right) P\left(p_{2^*}\right) ... \\
&\qquad\qquad \vdots \\
&= \langle p_{1^*} \rangle v_{1^*} + \sum_{p_{2^*}=0,1} \left(p_{2^*} v_{2^*} + ...\right) P\left(p_{2^*}\right) ... \\
&\qquad\qquad \vdots \\
&= \langle p_{1^*} \rangle v_{1^*} + \langle p_{2^*} \rangle v_{2^*} + ... \\
&= \sum_{i^*} \rho_{i^*,w^*} v_{i^*}. \qquad (2.9)
\end{aligned}
$$

Setting all $v_{i^*} = 1$ allows us to recover the weekly average amount of transactions

$$\langle n \rangle_{w^*} = \sum_{i^*} \rho_{i^*,w^*}. \qquad (2.10)$$

This result agrees with the expectation value of a Poisson binomial distribution. Because v and n are linear combinations of the projections $p(i^*)$, they can be factorized in (2.8) into closed form expressions. For such statistics, their weekly localized form simplifies to a weighted sum over the week's branches' weights. In general, $\langle X \rangle_{w^*}$ must be computed by sampling the weighted sum (2.8) over the dofs contained in Ω^*. Repeating this independently for every week in $\mathcal{S}$ results in a set of measures describing the total network in weekly detail, a procedure we will name *decoupled sampling*. However, as stated before, this description is only approximate for correlating statistics. We will now illustrate this with an example.

Suppose we estimate the average amount of transactions $\langle n(m) \rangle$ in a certain month m by ensemble sampling. Since every j is known precisely at a monthly resolution, each sampled microstate will yield the same output: the number of transactions per month is constant, or $\langle n(m) \rangle \equiv n_m$. Decoupled sampling will nearly always overestimate n_m,

$$\sum_{w \in m} \langle n \rangle_w \gtrapprox n_m, \qquad (2.11)$$

because it is unclear which weeks w should be included in the summation. As noted, the set of months cannot be partitioned into weeks and the decoupling becomes apparent because the statistic $\sum_{w \in m} \langle n \rangle_w$ relates to more than one week. As the weights of the branches sum to unity only within one month m, degenerated weeks that belong to both m and m' bring excessive weight into the left hand side of (2.11), as can be seen after applying equation (2.10). The systematic error of estimating expectation values of correlating statistics with decoupled sampling is likely to be relatively small because many transactions are already determined with weekly precision. For example, estimates of the weekly amount of transactions $n(w)$ by $\langle n(w) \rangle$ and $\langle n \rangle_w$ deviate only by 2.4% on average – see Figure 2.5 for a comparison of the results.

The main advantage of decoupled sampling is that it allows any amount of sampling to be done since it can be implemented in a computationally

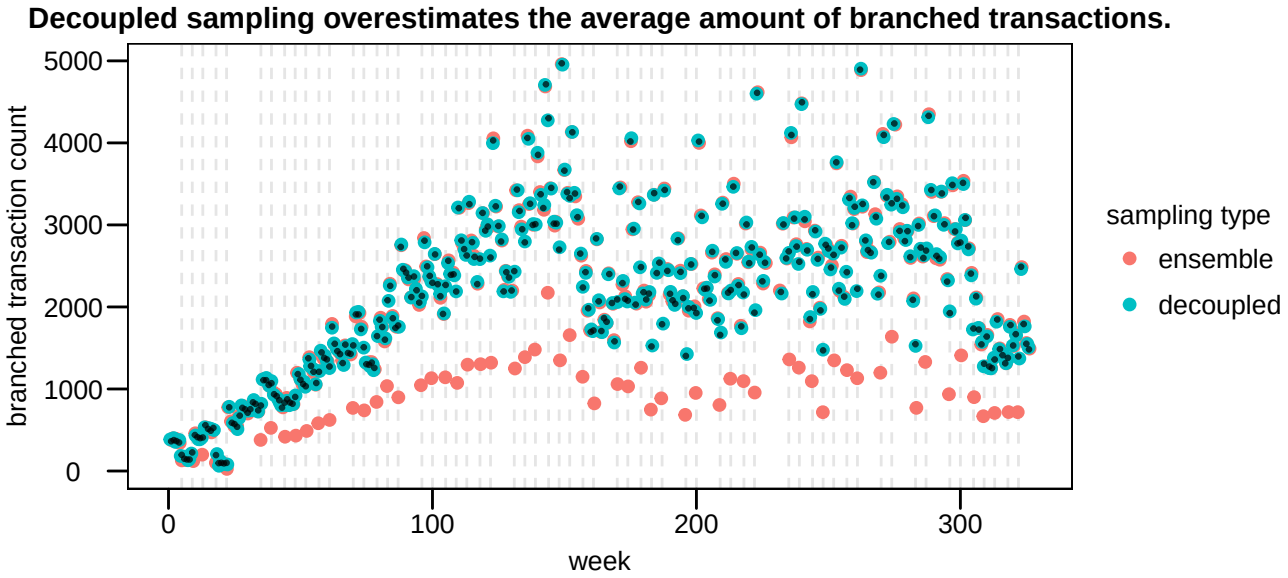


Figure 2.5: (points jittered slightly for clarity) Comparison of the estimates of the weekly amount of branched transactions $n(w)$ by ensemble and decoupled sampling. The transactions taken into account have at least two branches, i.e. have uncertainty about their issuance week. They represent about one fourth of the total amount. The coincidence of the black dots and the decoupled estimation expresses the equality of the right respectively left hand side of equation (2.10). Degenerated weeks are tagged by gray dashed lines that signal the overestimates of the decoupled $\langle n \rangle_w$ with respect to the ensemble $\langle n(w) \rangle$. The impression that banks seemingly trade less during weeks that happen to belong to two months is caused by the graph only showing branched transactions. Compared to ordinary weeks, degenerated weeks host transactions with smaller weights (typically twice as small) because the number of days available for a branched transaction to be assigned to is divided among two months. In other words, the sharp troughs during degenerated weeks are compensated by a corresponding increase in the amount of single branch transactions.

and memory-efficient algorithm, as opposed to ensemble sampling. A large amount of samples per week is necessary for measures X that are sensitive to the inclusion or exclusion of individual links, i.e. statistics with considerable variance over the state space. We will discuss this in detail soon. For simpler and more robust statistics, the estimates provided by ensemble sampling are usually sharp and there is no need for the decoupled approximation. Within the box approximation, decoupled sampling of a week w^* also samples the phase space $\mathcal{S}^*$ more efficiently simply because it has a smaller volume.

The following discussion and notation is based extensively on lectures given by prof. dr. Ryckebusch during the course "Computational Physics" [54] and

MacKay (2005) [53]

Monte Carlo integration

A Monte Carlo method is a computational technique that makes use of random numbers. *Monte Carlo integration* is a tool to approximately evaluate functions of random variables governed by complicated probability distributions – in particular high-dimensional ones. The problem we were facing is the calculation of the expected value and variance of a statistic X, for example

$$\langle X \rangle = \sum_{\substack{w(1)=w_1 \\ w(2)=w_2 \\ \vdots \\ w(n)=w_n}} X(1,2,...)\; P\left(w(1)=w_1 \text{ and } w(2)=w_2 \text{ and } ...\right). \quad (2.12)$$

Calculating (2.12) requires one to visit each possible state of the random variables $j = 1, ..., n$ and evaluate and multiply X and P accordingly; then the result must be summed to obtain $\langle X \rangle$. Given that the amount of dofs is approximately 2.75×10^6, we can estimate the amount of states using Table 2.6 to be $\sim 10^{353\,377}$.

How big is this number? It is certainly an unimaginable amount, comparing it with the number of electrons in the universe, $\sim 10^{80}$. On the other hand, it is of the same order of the number of states of a 2D Ising model with 600×600 spins[8]. Or, equivalently, a $70 \times 70 \times 70$ lattice in three dimensions. Either way, it is certainly impossible to visit the complete state space given reasonable computational resources in order to evaluate equation (2.12) accurately; instead, we will evaluate only a limited amount of terms in (2.12) simply by sampling states directly from the probability distributions proper to ensemble and decoupled sampling. In many applications of Monte Carlo integration, the target distribution is hard to sample from, or known only up to a normalizing constant. These problems have led to the development of Monte Carlo techniques such as importance sampling, rejection sampling and Metropolis-Hastings methods. In our case, however, the weekly joint distribution of transactions consists of many independent random variables that encode the chosen branch. This factorization into independent variables enables us to sample directly from the joint distribution; the appropriate sampling and post-processing algorithms were written in the `R` [56] and `Python` languages.

Before we continue, let us restate the problem in general terms. Consider a discrete joint probability distribution $P(\boldsymbol{x})$ defined over the vector of the

[8]The Ising model, 'likely the most exported exported model of physics' [54], defines a nearest-neighbor interaction on a lattice whereby (the spin of) each position can be in two states, i.e. up or down, possibly under the influence of an external (magnetic) field. The properties of such a system in equilibrium with a heat bath can be determined by the laws of statistical physics. The model was originally developed to explain spontaneous magnetization and has been shown capable of producing phase transitions in dimensions $D > 1$. More on the Ising model and its extensions can be found in many textbooks, for example [55].

outcomes of N random variables $\boldsymbol{x} = \{x_1, ..., x_N\}$. We wish to estimate the expectation of a function ϕ under this distribution,

$$\Phi \equiv \langle \phi \rangle = \sum_{\boldsymbol{x}} \phi(\boldsymbol{x}) P(\boldsymbol{x}) . \tag{2.13}$$

In the case of ensemble sampling, $\boldsymbol{x} = (w_1, w_2, ...)$ and $P(\boldsymbol{x}) = \rho_{1,w_1}\rho_{2,w_2}...$ Similar expressions apply to decoupled sampling, taking care that the relevant random variables depend on the week that is examined. The expectation and variance of any statistic X may be computed by setting two functions $\phi_1 = X$ and $\phi_2 = X^2$. Then $\langle X \rangle = \Phi_1$ and $\mathrm{var}(X) = \Phi_2 - \Phi_1^2$ so that the problem reduces entirely to approximating equation (2.13).

After obtaining R independent samples $\{\boldsymbol{x}^{(r)}\}$ from $P(\boldsymbol{x})$, an estimator of Φ is

$$\hat{\Phi} = \frac{1}{R} \sum_r \phi(\boldsymbol{x}^{(r)}), \tag{2.14}$$

Since

$$\langle \hat{\Phi} \rangle = \frac{1}{R} \times R \langle \phi \rangle = \Phi \tag{2.15}$$

$$\mathrm{var}(\hat{\Phi}) = \frac{1}{R^2} \times R \, \mathrm{var}(\phi) = \frac{\mathrm{var}(\phi)}{R}, \tag{2.16}$$

the Monte Carlo estimator (2.15) is unbiased and its accuracy $\epsilon \propto \sqrt{\mathrm{var}(\hat{\Phi})}$ depends only on the number of samples drawn as $\frac{1}{\sqrt{R}}$ and the variance of ϕ, *not on the dimensionality of the phase space*. So an estimate of a statistic X during a certain week will be accurate if the variance of the statistic is not too big, i.e. X is not overly sensitive to random shuffling of the links. Given that a large majority of the links are fixed, we expect this condition to hold quite often.

To get an idea of the convergence of the estimator (2.14), let us turn to a practical example. We want to estimate $\langle X \rangle$ and the error of the estimate. We set $X = d$, the *mean distance*, and sample it $R = 5120$ times in a random week, say week 323. The mean distance measures the average path length by calculating the shortest paths between all pairs of banks that have been active during this week (about 750). Links are directed and follow the flow of money, i.e. from lender to borrower. Thus the mean distance is a proxy for the number of loans it takes for excess reserves to disperse in the interbank network.

Table 2.7 displays the results that were obtained after taking 5120 samples $N = 9$ times and calculating the estimates for $\langle d \rangle$, $\mathrm{var}(d)$ (and $\mathrm{sd}(d) = \sqrt{\mathrm{var}(d)}$) for each group of samples[9]. In practice, we will only sample once, e.g. one distribution of $\{d(\boldsymbol{x})^{(r)}\}$, because its mean and variance provide the estimates for $\langle d \rangle$ and $\mathrm{var}(d)$, respectively. Via equation (2.16) we can

[9]In this thesis an explicit distinction between 'samples' and "samples consisting of equally sized groups of samples" is sometimes not made.

$\hat{\Phi} \to$	$\widehat{\langle d \rangle}$	$\widehat{\mathrm{var}(d)} \times 100^2$	$\widehat{\mathrm{sd}}$
	4.2054	7.7684	0.027872
	4.2058	7.6098	0.027586
	4.2044	7.6798	0.027712
	4.2057	7.7520	0.027842
	4.2059	7.8282	0.027979
	4.2050	7.8548	0.028026
	4.2061	7.8266	0.027976
	4.2059	7.7089	0.027765
	4.2060	7.8094	0.027945
$\langle \hat{\Phi} \rangle \to$	4.2055	7.7597	0.027856
$\mathrm{var}(\hat{\Phi}) \to$	3×10^{-7}	0.0065	2×10^{-8}

Table 2.7: Estimates (2.14) for the first and second moments of the mean distance d. The first nine rows represent $R = 5120$ samples each and summarize the distribution $\{d(\boldsymbol{x})^{(r)}\}_{r=1}^{5120}$ by mean, variance and standard deviation respectively. The last two rows contain statistics of their predecessors, providing the estimates (2.15)- 2.16). The consistently small standard deviations seem to indicate that the distributions have a well defined scale.

then calculate the expected error

$$\epsilon_1 \propto \sqrt{\frac{\widehat{\mathrm{var}(d)}}{R}}. \qquad \text{(one sample)} \qquad (2.17)$$

But the purpose of the table is to calculate the variance of the nine estimates of $\langle d \rangle$ directly:

$$\epsilon_N \propto \sqrt{\mathrm{var}\langle \hat{d} \rangle}. \qquad (N = 9 \text{ samples}) \qquad (2.18)$$

Comparing ϵ_1 and ϵ_N yields information about the number of samples R we would typically believe to suffice. To do that, we shall now reassign these samples at random into groups of $R = 2560, 1280, 640, ..., 10$ and make similar tables for each R – the tables becoming longer and longer (i.e. increasing N) as we make it to $R = 10$. In each table we select a row at random and calculate ϵ_1, then we calculate ϵ_N. The selected distributions for each R are plotted in the top panel of Figure 2.6. As expected by the central limit theorem, the distributions are bell-shaped. This means that roughly 95% of the $\hat{d}$ are within two standard deviations of the distribution's mean, so we set the proportionality constant in the ϵ-equations (2.17-2.18) to 2 to achieve 95% symmetrical confidence intervals $[\widehat{\langle d \rangle} - \epsilon_1, \widehat{\langle d \rangle} + \epsilon_1]$ for each sample with size R. They have been plotted in the bottom panel of Figure 2.6.

We can draw two conclusions from this exercise. First, the estimated mean distance in function of sample size converges rapidly. Second, the error bars of width $2\epsilon_1$ (the left bars in the bottom panel) based on the variance of the distributions in the top panel are reliable proxies for the 'true error' made by the estimators (the right bold bars – they span ϵ_N). It seems that $R = 100$ samples may be sufficient to get a very accurate estimation. In our case, the measure d displays small variance. Bear in mind that this variance indeed is a function of the week under examination.

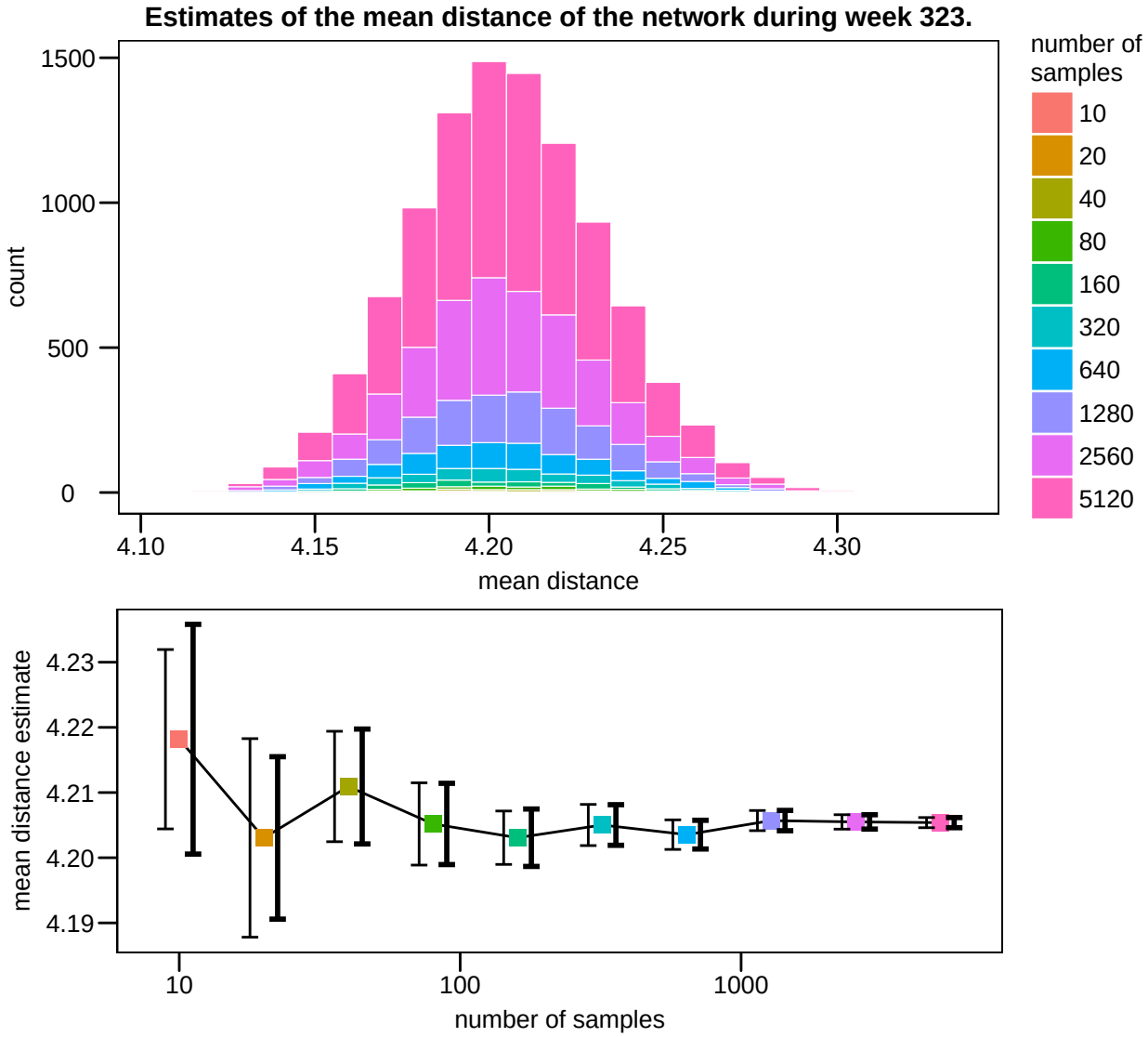


Figure 2.6: Ten estimates of the network's mean distance with standard errors obtained by ensemble sampling during the week of October 4, 2004 with increasing number of samples R. The network only contains loans and excludes the `overdraft` and `demand` terms (this still represents 95% of the original network, as will be explained in the following section). In the bottom panel, the left error bars designate ϵ_1, i.e. the estimated error based on one group of R samples, and the right bold bars represent ϵ_N. Taking into account all 10 230 (independent) samples, the mean distance is 4.20 ± 0.06 with 95% confidence. If we had only included precisely dated links, this distance would grow to be 4.58.

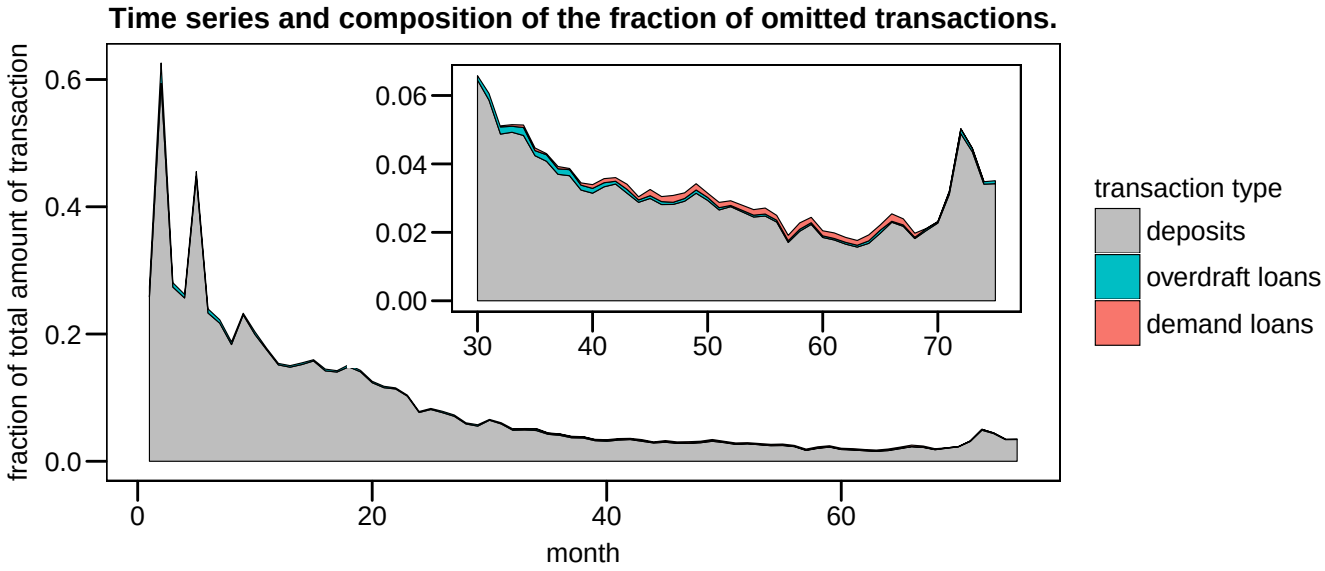


Figure 2.7: The time series show that the loans form a negligible portion of the excluded interbank transactions. After a turbulent start in the first crisis, we see a general decrease of deposits as the network tends to maturity (roughly in the year 2002 – around month 40) and beyond. The spike that occurs towards the end is caused by the second crisis in the summer of 2004. Figure 2.1 and 2.2 show a severe decline in the trade activity accompanied by a surge of loan interest rates; the trust crisis caused a fallback in the demand for loans, the price of borrowing increased and some lenders turned to deposits instead.

A more prudent general approach would be to take $R = 400$ samples for arbitrary statistics X, then to estimate X's variance and increase R as necessary – switching to decoupled sampling once a certain threshold of R is passed. The error on the variance estimate may also be computed if we take N observations of R samples, but it will be ignored in this thesis. The remaining part it shall use a sloppy notation, dropping all the hats and $\langle\cdot\rangle$'s: we will simply say "the mean distance d in week w is such and such" instead of "the estimate of the average mean distance $\widehat{\langle d \rangle}$ etc."

2.5 The dimensionality of the network

The data is a record of many different transaction types. In this thesis, we will dismiss deposits, regardless of term, and loans with `overdraft` and `demand` terms[10], transforming the original network into what will be called the *reduced network*. This simplification is motivated by several observations:

[10] See Table 2.2 for a complete list of the loan and deposit terms

(1) The majority of the interbank market models in the literature are concerned with interbank loans exclusively [13], including previous work on a smaller subset of our data [7, 8, 39]. Models that do take *interbank* deposits into account typically merge them with loans and other liabilities to form 'exposures' [36, 57, 58][11]. However, because the joint interest rate and transaction volume distributions of loans and deposits differ considerably[12], this thesis considers such aggregation into exposures a source of noise that impedes the Bayesian model selection and hierarchical clustering that will be performed in Chapter 4.

(2) Restricting our attention to only one type of interaction obviously simplifies the analysis considerably; thus the description of the interbank lending market shall be based on an approximation, i.e. the reduced network. Figure 2.8 shows that the *term structure*[13] of the interbank deposits and loans have different forms – the curve is well-behaved in the case of loans, with exception of the `overdraft` and `demand` terms. This is simply due to the legal essence of these terms. Here is a more detailed description in comparison with Table 2.2:

`overdraft`	Debt formed by a deficit in a bank account, caused by drawing more money than the deposit holds. In the rare case of interbank loans this term applies to credit lent by a bank that is in excess of the amount provided by the established credit line between the lending and the borrowing parties.
`demand`	A contract with variable term and possibly varying interest through time. `financialdictionary.net` defines a demand loan as follows: "a rare form of loan that can be called for complete repayment without any prior warning to the borrower. In other words when the lender demands the money, the borrower must pay it. So unlike a regular loan that is paid in installments and has a defined maturity date, demand loans work on the specific demand of the lender." [59]

By omitting these two terms we will only deal with 'classical' interbank loans. Coincidentally, loans with these terms are rare – see panel (d) of Figure 2.9.

(3) The simplification cuts the amount of transactions by 5%, reducing the computational cost. While we have seen that in theory weekly estimates are indifferent towards this cut, it is appreciated by many more advanced algorithms used in the following chapters.

[11] An explicit exception is [19]'s model of liquidity shocks caused by the sudden reallocation of deposits in a financial network.

[12] In the unsecured interbank market, the difference between loans and deposits are the legal terms. The lender (borrower) fixes the terms in the case of a loan (deposit). This generates different contracts, i.e. combinations of interest rate and volume.

[13] Also known as *yield curve*. This is the relation between bargained interest rate and the transaction term – Section 2.6 proposes a proper explanation.

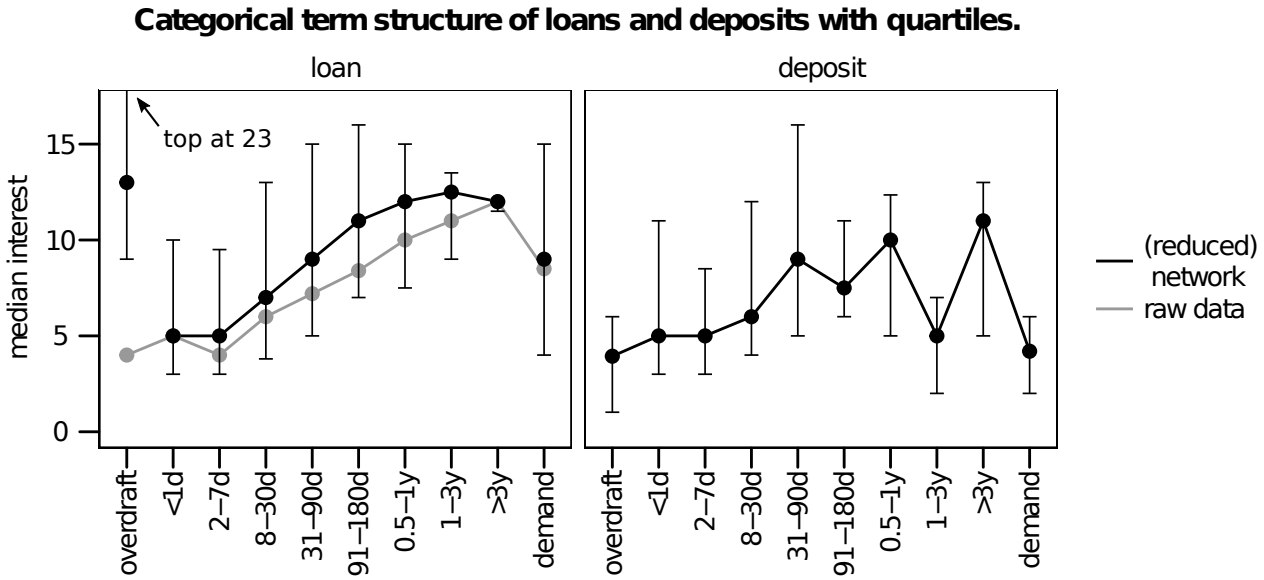


Figure 2.8: (see Section 2.6) The panels show the term structure of the interest rates of interbank loans and deposits, calculated per term category as time-integrated quartile interest rates. (left panel) The reduced network is indicated by the connected black dots. The raw data includes loans with zero-interest rates and conflicting `due`, `term` and `reported` triplets (Sections 2.2 and 2.3 motivate why these loans are omitted in the (reduced) network).

(4) Looking at Figure 2.7, deleting the deposits in the networks seems affordable from month 20 onward. Averaging over time, 95% of the original network is still represented by the simpler one, but it is clear that any description of the network during its growth phase needs deposit transactions.

We can remedy (4) in part by marginalizing links between banks according to two schemes: *directed* and *undirected aggregation*.

Two additional views of the reduced network

In the reduced network, a pair of banks may engage in multiple transactions during a certain period. Each transaction has five properties: issuance time, direction, term, interest rate r and volume v. So given a time window, any link exists in two discrete and two continuous dimensions, respectively. As in Vandermarliere et al. (2015) [39], we will consider two simpler views of the network that correspond to a reduction of the amount of information in the network by aggregating links *in the given time window* and *across sets of terms*. The latter is accomplished by dividing the terms into so-called classes.

For the sake of consistency with [39], the complete network is called the *multi-directed view*, as multiple directed links between banks may occur

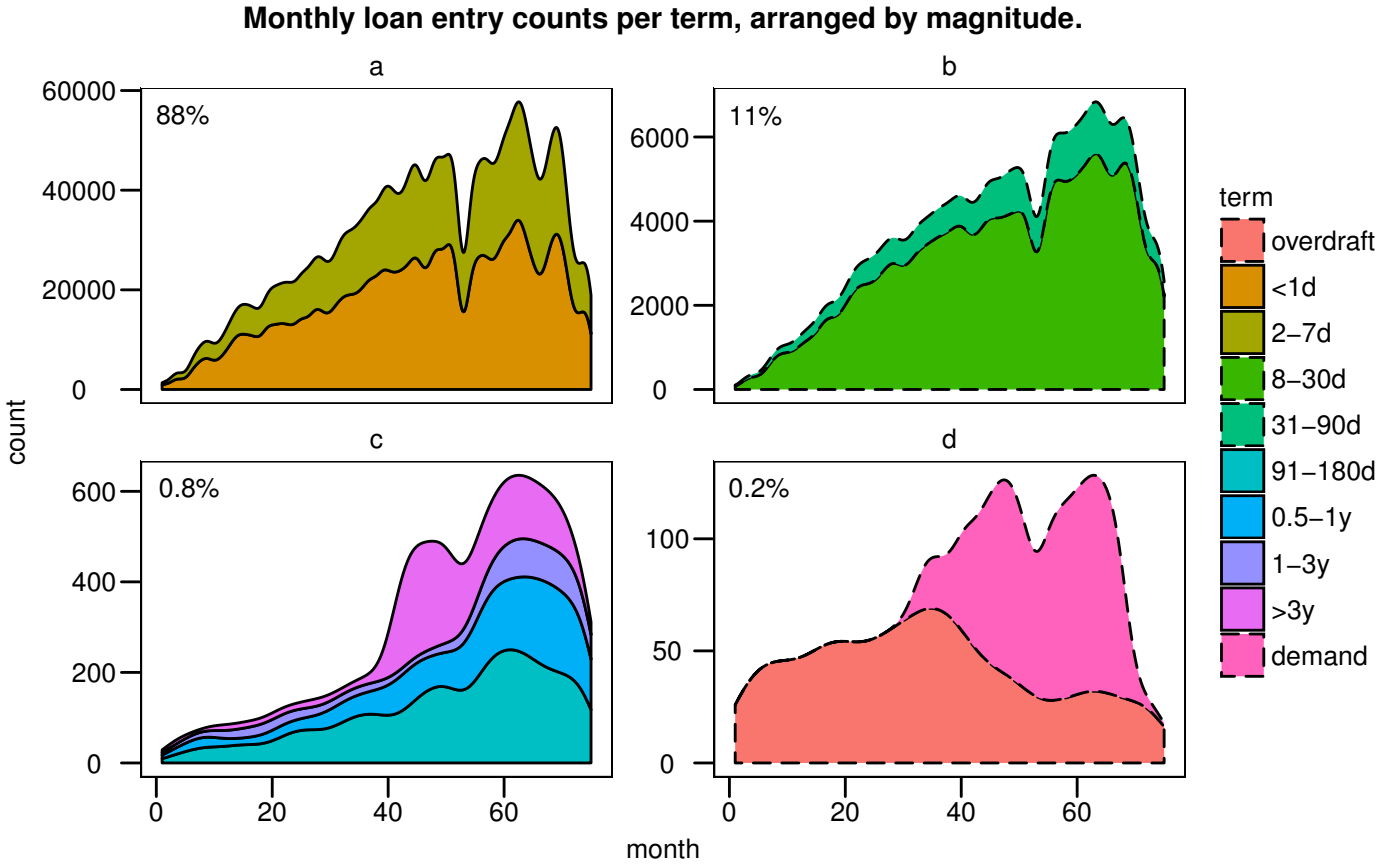


Figure 2.9: (smoothing effect for clarity) Pictorial time series of monthly amounts of issued loans categorized by term. Panels (a)-(c) make up the reduced network. Noting the vertical axes, observed terms come in different orders of magnitude. The percentages in the upper left corner of the panels equal the amount of loans contained in the panel relative to the total number of loans issued in the complete network.

during an observation. Unsurprisingly, the *directed view* and *undirected view* of the network are generated by the directed and undirected aggregation scheme, respectively.

The first, directed aggregation, consists of bundling two or more links that point in the same direction between two banks. So a given bank can have at most one lending and one borrowing contract with any other. If we label the links to be aggregated during a given period by i, the aggregated transaction is a loan of volume $v' = \sum_i v_i$ with an interest rate $r' = \frac{1}{v'} \sum_i v_i r_i$ that measures the interest rate charged, respectively paid per average unit volume[14]. This network view expresses the directed flow of money between banks, with averaged interest rates.

[14]Note that $r'v' = \sum_i r_i v_i$, so the total interest is indeed conserved.

liquid	semi-liquid	illiquid
<1d	8-30d	1-3y
2-7d	31-90d	>3y
	91-180d	
	0.5-1y	

Table 2.8: Classification of terms with respect to the loan's liquidity according to Berger and Bouwman (2006) [60].

In the second view, links are aggregated regardless of direction: the undirected edges merely establish whether a pair of banks have interacted during the time window.

As stated before, these aggregations may also be made across terms. For example, one can only speak of a general 'interaction' in the case of undirected aggregation across all terms. If we retain information about (classes of) terms, we can make more precise statements. An early illustration of such classification is given in Table 2.8 – we return to this matter in Chapter 4.

In the case of weekly time windows, the uncertainty with respect to the loans' issuance dates comes into play. In principle, we can employ ensemble sampling by aggregating each generated microstate and proceeding as usual. However, we can avoid this costly procedure with decoupled sampling, which allows us to aggregate the network *a priori*. Let us consider undirected aggregation first. As in Section 2.4, let us denote the set of loans that have a non-zero probability to occur in week w^* by Ω^*. The subset Ω^*_{AB} holds all loans between two banks A en B, regardless of direction. The weight of the aggregated link that replaces the edges $i^* \in \Omega^*_{\text{AB}}$,

$$\begin{aligned}\rho' &= P\left(\text{at least one of } i^* \text{ is realized}\right)\\ &= 1 - \prod_{i^* \in \Omega^*_{\text{AB}}} P\left(p(i^*) = 0\right)\\ &= 1 - \prod_{i^* \in \Omega^*_{\text{AB}}} (1 - \rho_{i^*,w^*}),\end{aligned} \tag{2.19}$$

can then be used to generate statistics describing the undirected view with weekly resolution using equations (2.6-2.8). Aggregating terms can be done by adding another label to Ω^*_{AB} that describes the class the links are being mapped to, which amounts to simple bookkeeping.

The same applies to directed aggregation, for which we need to discriminate between $\Omega^*_{\text{A}\to\text{B}}$ and $\Omega^*_{\text{A}\leftarrow\text{B}}$. To proceed, we shall focus on loans going from A to B, i.e. bank A lending to bank B, as the argument for the reverse case is completely symmetrical. As indicated, the weight of the aggregated loan is of equivalent form to (2.19),

$$\rho' = 1 - \prod_{i^* \in \Omega^*_{\text{A}\to\text{B}}} (1 - \rho_{i^*,w^*}). \tag{2.20}$$

Recall that the aggregated loan requires a volume v' and an interest rate r' in addition. To find v', we demand that the average volume that is lent by

bank A to bank B during week w^*,

$$\langle v \rangle_{w^*,\ \mathrm{A}\to\mathrm{B}}\,, \tag{2.21}$$

is equal in the directed and multi-directed view of the network. Applying equation (2.9), this is expressed by

$$\rho' v' = \sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*} v_{i^*}, \tag{2.22}$$

so

$$v' = \frac{1}{\rho'} \sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*} v_{i^*}. \tag{2.23}$$

To better understand (2.22), we can rewrite it as

$$\rho' v' = \frac{\sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*} v_{i^*}}{\sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*}} \times \sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*}$$

$$\overbrace{\text{loan volume} \times P\left(\text{loan appears}\right)}^{\text{directed view}} = \underbrace{\text{weighted average loan volume} \times \text{expected amount of loans to appear}}_{\text{multi-directed view}}. \tag{2.24}$$

We can find r' analogously by requiring that the average interest received by bank A is equal in both views:

$$\rho' r' v' = \sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*} r_{i^*} v_{i^*}, \tag{2.25}$$

since $\langle rv \rangle_{w^*}$ is a linear combination of the links' projections (2.7). Then

$$r' = \frac{1}{\rho' v'} \sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*} r_{i^*} v_{i^*} \tag{2.26}$$

$$= \frac{\sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*} r_{i^*} v_{i^*}}{\sum_{i^* \in \Omega^*_{\mathrm{A}\to\mathrm{B}}} \rho_{i^*,w^*} v_{i^*}}. \tag{2.27}$$

The interest rate of the aggregated loan is the average of the interest rates in the multi-directed view weighted by the loans' volume and probability of appearing in week w^*.

To summarize, we generate the (un)directed view first, using the recipes (2.19-2.27). We can then extract expected values from these views as usual. The use of decoupled sampling triggers a huge gain in efficiency as the amount of dofs contained by the multi-directed view of the network decreases by 56% and 60% for the directed and undirected views, respectively, causing the phase space volume to shrink considerably. Below we will quantify to which extent; Figure 2.12 displays the result. The efficiency gain is partly compensated by lacking support for correlating statistics and typically by a moderate increase in the required number of samples R needed for reliable estimates, i.e. with acceptable error intervals[15].

[15] Yet for small R, simulations comparing the estimates of $\langle v \rangle_{w^*}$ and $\langle rv \rangle_{w^*}$ by *a priori*

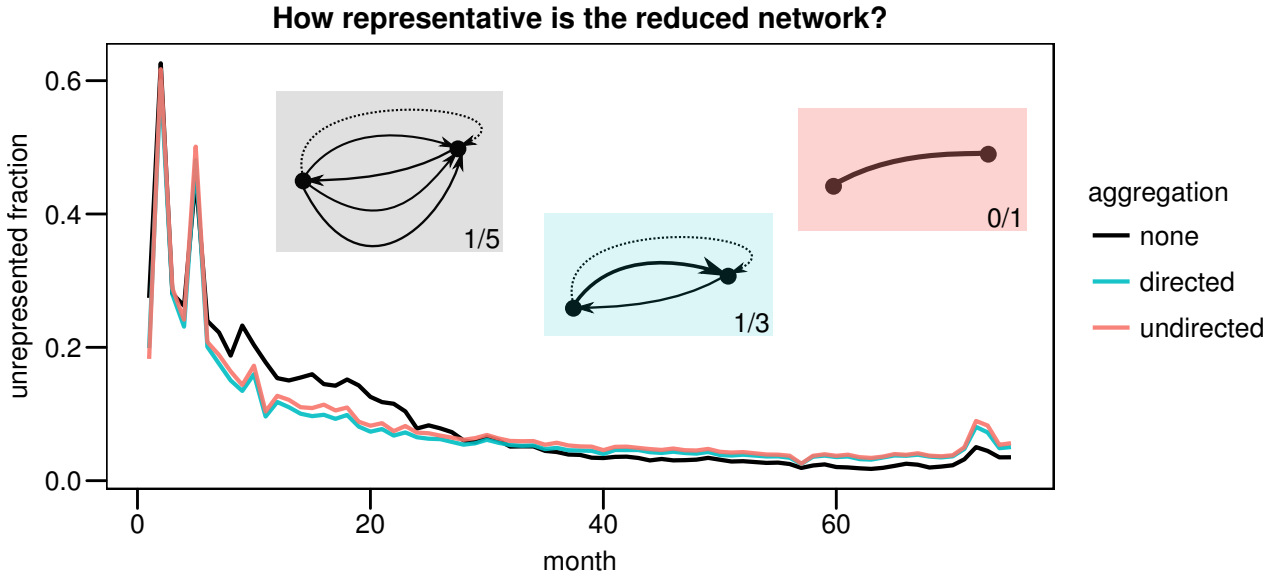


Figure 2.10: (see also Figure 2.7) The monthly unrepresented fraction equals the loans-to-transactions ratio of the reduced and complete network, respectively; both are aggregated beforehand by month and across all terms. The fact that this ratio is almost equal for the undirected and directed views indicates that nearly all deposit transactions are one-way. Indeed, one expects that typically only smaller agents accept the deposit terms of larger institutions. Roughly as of month 30, aggregating yields slightly more unrepresentative descriptions. This is also the case for the example configuration drawn in the graph. Loans and deposits are indicated by full and dotted edges, respectively. The unrepresented fraction is shown in the bottom-right corner.

Reducing the network's complexity

Now that we are familiar with the three distinct views of the network, the concern of observation (4) [page 30] may be readdressed. Figure 2.10 shows that in the growth phase, the complete network is better represented by the reduced network if both are viewed with the directed and the undirected representations. Nevertheless, we expect the reduced network's description of the first crisis to be unsatisfying due to the two peaks in the unrepresented fraction of loans and deposits, regardless of the view wielded.

Apart from the fraction of included transactions, the reduced network may also be judged on the basis of conserving *block lengths*. If a pair of banks trades on the interbank network for, say, three *consecutive* weeks, then these transaction form a block with a length of three. The average block length of a group of banks in the undirected view during an observation is a measure

---

and *a posteriori* (un)directed aggregation show agreement with the theoretical values within error intervals. So R is fixed only by the desired precision.

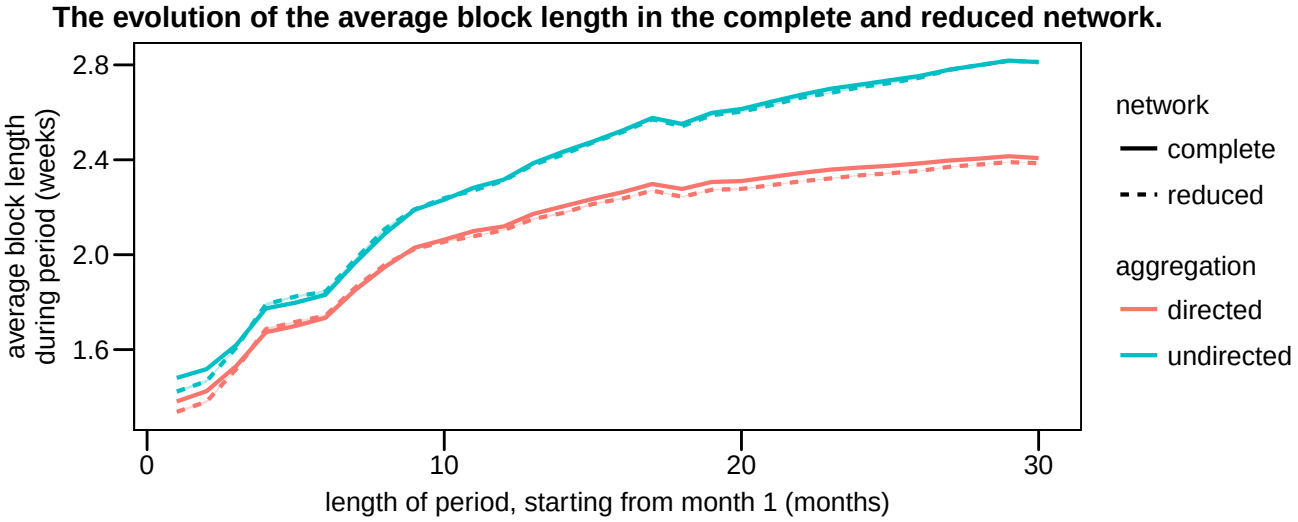


Figure 2.11: This graph shows the (un)directed average block length in the complete and reduced network during 30 periods. Each period starts on the first day of the first month and progressively includes an additional month. The steady increase in the average block length converges rapidly to 2.3 weeks and 2.7 weeks for the directed and undirected view, respectively; these are the average block lengths obtained over the complete lifetime of the networks.

of continuous interaction[16]. Employed in the directed view, it assesses the extent to which banks tend loan or borrow in uninterrupted streaks. Figure 2.11 shows that the reduced network preserves the average block lengths very well, so we expect that most week-to-week correlations in the complete network survive the reduced picture.

Considering the presented arguments *pro* and *contra* the reduced network, we will accept it and refer to it as 'the' network in subsequent chapters. It approximates the complete interbank lending market with increasing accuracy as the network grows mature. The main motif is the gain in simplicity – the cost is a poor description of the complete network during the first two dozen months. The (un)directed aggregation schemes and the reduced network all cut the dimensionality of the original problem, i.e. the amount of dofs involved. In exchange the analysis becomes easier.

Life in high-dimensional space

In the remainder of this section, we will discuss some aspects of dealing with highly dimensional data and introduce the concept of *entropy* along the way.

The reduced interbank lending network consists of 1470 banks that issue over 2.6 million loans in 75 months. In this thesis, these banks and loans become degrees of freedom in two distinct situations:

[16]Incidentally, it is an example of a correlating statistic (see Section 2.4).

(1) Describing the network on weekly time scales. Table 2.6 shows that 75% of all loans can be dated within weekly precision. The uncertain week of issuance of the remaining 25% was promoted to a random variable in Section 2.4. In the reduced network, then, we end up with 641 265 dofs.

(2) Inferring community structures in the network. The generative models in Section 4.2 sample random partitions of the network in blocks that fulfill similar roles by assigning each bank to a block. Thus the relevant dofs are the banks.

In both cases the amount of samples generated will be much smaller than the number of dofs involved. This large number of dofs, i.e. variables, is symptomatic of a major trend in today's data analysis [61]. The following excerpt is taken from a speech given by prof. dr. Donoho in the year 2000 in which he considered the mathematical challenges of the 21^{th} century:

> In broad overview, it has become much cheaper to gather data than to worry much about what data to gather. The ease with which we can now collect hyper-informative detail on individual units means that the standard situation for data analysis is where there are many variables [...] [61].

Some examples he discussed are consumer preference data, gene expression data and tick-by-tick financial data, where the dofs are object ratings, genes and exchange rates respectively. In our case, we started out with many observations of eight variables (see Table 2.1); yet it turns out that answering some questions about the network requires solving highly dimensional problems. More precisely, we are interested in characteristic properties of large collections of N random variables. One of these is the *concentration of measure*. In probabilistic terms, "a stochastic function that depends (in a smooth way) on the influence of many independent variables (but not too much on any of them) is essentially constant" [62, 63].

> Imagine a cloud of uniformly distributed points in a N-dimensional sphere with radius r whose volume $V(N)$ is proportional to r^N. Intuitively, this holds because $\frac{\mathrm{d}V(N)}{\mathrm{d}r} = S(N) \sim r^{N-1}$, the hypersphere's surface area which corresponds to some multiple of the volume of the sphere in $N-1$ dimensions. The fraction of the sphere's volume that lies in the outermost shell of thickness ϵ is
>
> $$\frac{r^N - (r-\epsilon)^N}{r^N} = 1 - \left(1 - \frac{\epsilon}{r}\right)^N .$$
>
> The expected amount of uniformly distributed points included in some volume is proportional to it, so essentially all the probability mass is in the surface shell for large N, regardless of ϵ. For example, a shell of $\frac{\epsilon}{r} = 0.01$ thickness in $N = 1000$ dimensions will contain 99.996% of the point cloud.

The above example observes a density function of N independent and identically distributed (iid) uniform distributions, but the "many dimensions effect" is not restricted to uniform distributions. Indeed, in the case of a N-dimensional Gaussian distribution of a vector $\boldsymbol{x} = (x_1, x_2, ..., x_N)$ that

has each component $x_i \sim \mathcal{N}(0, \sigma^2)$ independently it can be shown by calculating the first two moments of $r = |\boldsymbol{x}|$ that $\boldsymbol{x}$ is expected to lie in a shell, asymptotically $\sqrt{2}\sigma$ thick for large N, at a distance $\sqrt{N}\sigma$ – even though the maximum of the probability density lies at the origin[17]. These classic examples [53,61,62] apply to iid continuous distributions; the two situations outlined above deal with sets of discrete distributions that are not necessarily identical. The concentration of measure is equivalent to the concept of *typical sets* which can be derived from entropy considerations. We will see that it is at work in this thesis as well.

The following brief introduction is based on MacKay (2005) [53] and Jaynes (2003) [3].

Entropy and information content

Given a discrete random variable X[18] with outcomes $x_i \in \mathcal{A}_X$ of probability p_i, the *Shannon information content* of the outcome x_i [64] is defined by

$$h(X = x_i) = \log_2 \frac{1}{p_i}. \tag{2.28}$$

In information theory, the quantity h is related to the amount of information needed to describe the outcome of an experiment. The base b of the logarithm defines the unit of information – the Shannon information content ($b = 2$) measures in bits, a natural choice in information theory, but any convenient base may be chosen freely, corresponding to a multiplicative constant in h. Note that the information content of joint outcomes of independent variables is additive. Indeed, with somewhat lighter notation the information content of an outcome (x, y) of the ensemble (X, Y) where X and Y are independent is

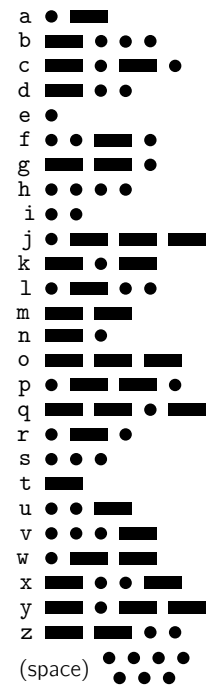
$$h(x, y) = \log \frac{1}{P(x, y)} = \log \frac{1}{P(x) P(y)} = \log \frac{1}{P(x)} + \log \frac{1}{P(y)}. \tag{2.29}$$

To indicate the concept underlying 'information content', imagine an engineer designing a noiseless channel, capable of sending binary data, through which a *source* X must transmit any conceivable string of *symbols* $x_i \in \mathcal{A}_X$, i.e. our outcomes. He can achieve this by mapping every possible outcome to an integer in the range $\{0, 1, ..., |\mathcal{A}_X| - 1\}$; the binary representation is then decoded by the receiver in possession of the inverse map. This way, every symbol will take $\log_2 |\mathcal{A}_X|$ bits to send, which we define as its codelength. Looking at (2.28), this corresponds to the information content of any outcome of an uniformly distributed random variable defined on an alphabet of equal size.

Now, from prior knowledge about the source, the engineer assigns probabilities p_i to the various possible symbols. For example, if the channel communicates (lowercase) English messages, the symbols are letters from the English alphabet, so $\mathcal{A}_X = \{\mathtt{a}, \mathtt{b}, ..., \mathtt{z}, (\text{space})\}$, and the engineer could

[17] This result is derived in Appendix B

[18] This symbol is used to denote a statistic and a random variable, which are alike.



consult a frequency table of English letters[19] to assign the probabilities[20]. By mapping symbols of higher probability to shorter codes, and *vice versa*, the channel can be used more optimally, i.e. the average number of bits per symbol needed to send a long string of symbols will be less than $\log_2 |\mathcal{A}_X|$. This is how the Morse telegraphic code was designed, in which the most frequently used letters are represented by the shortest codelengths (see the right margin [66]) – the exact prototype of what Shannon formalized and made precise a century later [3]. In effect, the engineer compresses the source symbols using prior knowledge about the source: if he knows little about the source, to him no outcome is more likely than the other. His state of knowledge is represented by the uniform distribution over the outcomes, and no compression is possible. It is in this sense that the probabilities of the outcomes determine their information contents (see the left panel of Figure 2.12). The *average* information content of the possible outcomes determines the theoretically optimal achievable compression by the *source coding theorem*, and with it the characterization of information content *of a source*. In theory, the optimal compression can be attained precisely by assigning codelengths equal to $h(x_i)$ bits; in practice, the codelengths must be integer, and various code schemes exist that approach it very well.

The average information content, or entropy, of a random variable X is given by

$$H(X) = \sum_i p_i \log_2 \frac{1}{p_i}. \tag{2.30}$$

Indeed, we have that $H(X) = \langle h(x_i) \rangle$. Note that the entropy of a collection of independent random variables is additive as well, and that for a given alphabet $\mathcal{A}_X$ the uniform distribution of probability along its members has the maximum entropy $H_0 = \log_2 |\mathcal{A}_X|$.

Entropy has several interpretations. In this context, it is important to distinguish between *information entropy* and *experimental entropy*. In this thesis only the former is used, so 'entropy' always denotes information entropy. The latter is a property of a physical system's macroscopic thermodynamic state which can be inferred by studying changes in measurable macroscopic quantities (i.e. state variables) using for example the Maxwell relations. It is defined as a measure of the distribution of energy over the many degrees of freedom associated with a loss of the ability to do useful work [67]. The interpretation in information theory is achieved by mapping random variables to sources and outcomes to symbols being sent over a channel; then the smallest sufficient amount of bits needed to communicate a string of N symbols coming from a source X converges to $NH(X)$ bits for asymptotically large N. Thus H is a sensible measure for *one* random variable's information content[21], but note that H depends on the observer

[19] Examples: online at `en.algoritmy.net/article/40379/Letter-frequency-English` or in [65].

[20] These are indeed context-dependent: it is rational that the engineer promptly increases $p_{\mathtt{x}}$, $p_{\mathtt{y}}$, $p_{\mathtt{z}}$ upon learning that the channel will be used by mathematicians! The point is that the more prior information the engineer uses to model the source X, the better will be the correspondence between the p_i and the observed frequencies.

[21] Other more sophisticated entropy measures take the information of the coding map

(our engineer in the thought experiment), simply because it depends on his prior information. In what has recently been termed objective Bayesian statistics [68] the information entropy is seen as a measure of uncertainty. More precisely, paraphrasing Jaynes (1957): if our partial understanding of the processes which determine the outcomes $(x_1, x_2, ...)$ of a variable X is represented by the corresponding probabilities $(p_1, p_2, ...)$, then $H(X)$ measures in an unique way (up to b) the amount of uncertainty represented by this probability distribution[22] [69]. Because of this, objective Bayesians may convert prior information I (expressible in a set of constraints) to probability distributions using the *maximum entropy principle* [70]: the distribution compatible with I that has the largest entropy is guaranteed to be the most objective (unbiased) representation of that knowledge. Because this principle maps prior beliefs to probability distributions, their entropy *will* be uniquely defined for observers with equal prior information. In other words, one can measure the information inherent in I by calculating the entropy of corresponding distribution via the maximum entropy principle.

According to some objective Bayesians, the three interpretations are connected in a profound way [3, 70], but this is far from being universally accepted [71] and evidently outside the scope of this thesis. Here we will simply use Shannon's concept of entropy to illustrate the behavior of probability distributions of a large number of independent random variables.

Life in high-dimensional space (bis)

Given an ensemble of N independent discrete random variables $X_1, X_2, ..., X_N$ with alphabets $\mathcal{A}_{X_1}, \mathcal{A}_{X_2}, ..., \mathcal{A}_{X_N}$. If we let $X = (X_1, X_2, ..., X_N)$, then outcomes $\boldsymbol{x} \in \mathcal{A}_X = \mathcal{A}_{X_1} \times ... \times \mathcal{A}_{X_N}$ are distributed according to the joint distribution of the X_i which is just the product of the marginal distributions.

The entropy of X, $H \equiv H(X) = \sum_i H(X_i)$, is the average information content $\langle h(\boldsymbol{x}) \rangle$. If the X_i had been iid, all $H(X_i)$ would be equal and the law of large numbers (many trials) would let $h(\boldsymbol{x})/N \to H(X_i)$ for large N. In our case however, we can use Chebyshev's inequality[23] to achieve a similar effect for $h(\boldsymbol{x})/N$. We have for $t \geq 0$ that $P(t \geq \epsilon) \leq \langle t \rangle / \epsilon$. Setting $t = (h(\boldsymbol{x})/N - \langle h(\boldsymbol{x}) \rangle / N)^2 \geq 0$ and taking $\epsilon \to \epsilon^2$ for convenience, we have that

$$P\left(\left(\frac{h(\boldsymbol{x})}{N} - \frac{\langle h(\boldsymbol{x}) \rangle}{N}\right)^2 \geq \epsilon^2\right) \leq \frac{1}{\epsilon^2} \mathrm{var}\left(\frac{h(\boldsymbol{x})}{N}\right) \leq \frac{\sigma_h^2}{\epsilon^2 N^2}. \tag{2.31}$$

The factor σ_h^2 is just the sum of the variances of the information content distributions of the independent X_i. If we take the largest variance to be

into account, as well as possible noise over the channel. But the basic concept remains the same.

[22] In fact it can be proven that equation (2.30) is the only acceptable form given "reasonable constraints on a measure of the amount of uncertainty" [3, p. 347].

[23] Also called Markov's inequality [72].

M, which is always finite for our purposes, then $\sigma_h^2 \leq NM$ and, introducing the entropy of the ensemble,

$$P\left(\left(\frac{h(\boldsymbol{x})}{N} - \frac{H}{N}\right)^2 \geq \epsilon^2\right) \leq \frac{M}{\epsilon^2 N}. \tag{2.32}$$

We see that for sufficiently large N, the distribution of $h(\boldsymbol{x})/N$ can be packed arbitrarily close to $\bar{H} \equiv H/N$, the entropy per dof. This is the concentration of measure we encountered before. In a nutshell, *one* sample $\boldsymbol{x} = (x_1, x_2, ..., x_N)$ of a collection of independent random variables contains progressively more outcomes x_i that lie close to their respective means $\langle X_i \rangle$ as N increases. It follows that the Shannon information contents of these outcomes tend to their average values, so their sum $h(\boldsymbol{x}) \to H$. One could say that for large N, once $\boldsymbol{x}$ has been drawn, we know that many of the individual outcomes $x_i \in \boldsymbol{x}$ must have yielded rather predictable results, i.e. have relatively low information outcome. As always in statistics, we know more about a larger group of random samples, even when they come from different independent sources.

Now let us introduce the typical set [53]

$$T_{N\beta} = \{\boldsymbol{x} \in \mathcal{A}_X : \left|\frac{h(\boldsymbol{x})}{N} - \bar{H}\right| < \beta\}, \tag{2.33}$$

which is equivalent to requiring that

$$2^{-N(\bar{H}+\beta)} < P(\boldsymbol{x}) < 2^{-N(\bar{H}-\beta)}. \tag{2.34}$$

Then we can rewrite (2.32) as

$$P(\boldsymbol{x} \in T_{N\beta}) > 1 - \frac{M}{\beta^2 N}. \tag{2.35}$$

For large enough N, almost all outcomes $\boldsymbol{x}$ will lie in the typical set, regardless of β. From (2.34) we see that if we take $\beta \ll \bar{H}$, the probabilities of the outcomes will then very roughly be equivalent to 2^{-H}. Since they are contained in the typical set that holds almost all the probability, the size of the equivalent set will be of order 2^H. Since H is always smaller than the maximum entropy $\log_2 |\mathcal{A}_X|$ over the same alphabet $\mathcal{A}_X$, *almost all outcomes of a large collection of non-uniform independent variables come from a region in phase space that is very small compared to the total volume.*

In practice, $\bar{H}$ and N are given and decide whether β can be taken small enough for the typical set to contain most of the probability. Only then the counting of $|T_{N\beta}|$ makes sense. If we make all X_i's iid, then $\bar{H}$ will remain constant for growing N and the 'squeezing effect' in equation (2.34) will be enhanced, enhancing the approximation for $|T_{N\beta}|$; this is called the *asymptotic equipartition principle* [53]. Intuitively, it explains why the optimal compression with a negligible chance of error of $N \to \infty$ symbols is exactly $N\bar{H} \equiv NH(X)$ bits: we can define a compression algorithm that gives a distinct name of length $NH(X)$ bits to each string of symbols in the typical set [53]. In origin, this is a consequence of measure concentration.

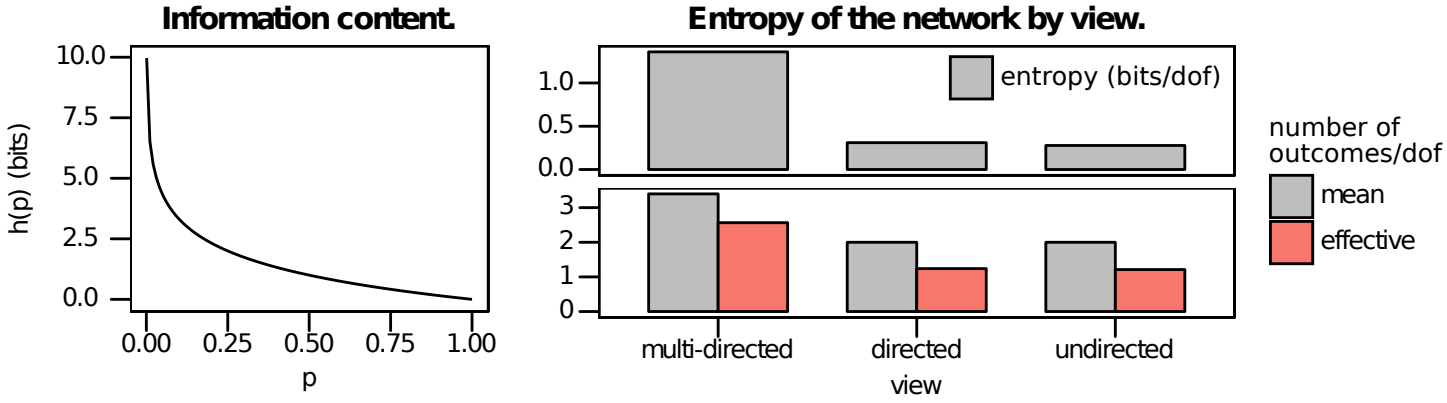


Figure 2.12: (left panel) The Shannon information content of an outcome having probability p. Note that no information is conveyed when one observes an event of $p = 1$. (right panel) Entropy and number of outcomes per dof. For the multi-directed view, the number of outcomes for a dof j equals $|\mathcal{A}_j|$ (see also Table 2.6). For the (un)directed views, the dof becomes $p(i^*)$ with alphabet size $|\mathcal{A}_{p(i^*)}| \equiv 2$ since we are using decoupled sampling. The effective number of outcomes per dof is defined as $2^{\bar{H}}$, in the sense that the total volume of the phase space of N random variables with $2^{\bar{H}}$ outcomes is $(2^{\bar{H}})^N = 2^H \approx |T_{N\beta}|$ for large N. The entropies of the (un)directed views were calculated using the weights derived in equations (2.19-2.20).

Let us apply this knowledge to situation (1) [page 36]. The top right panel of Figure 2.12 shows values of $\bar{H}$ for views of the (reduced) network. The entropy of the unreduced network is 1.41 bits/dof, which decreases to about 1.36 bits/dof for the reduced network in the multi-directed view[24]. The (un)directed view has (0.28) 0.31 bits/dof, a considerable difference which relates to the increased sampling efficiency of the decoupled approximation. This can be seen in the lower right panel: the effective number of outcomes $2^{\bar{H}}$ drops to about 1.2 'choices' on average per dof.

Finally, Figure 2.13 shows the typical set for the multi-directed network. In general, a rough estimate of an upper bound for $|T_{N\beta}|$ from (2.34) can be made by noting that the smallest probability of members of $T_{N\beta}$ is $2^{-N(\bar{H}+\beta)}$ and $P(\boldsymbol{x} \in T_{N\beta}) \lessapprox 1$, so $|T_{N\beta}|2^{-N(\bar{H}+\beta)} < 1$ or

$$|T_{N\beta}| < 2^{N(\bar{H}+\beta)}. \tag{2.36}$$

Denoting the maximum entropy as $H_0 = \log_2 |\mathcal{A}_X|$, with $|\mathcal{A}_X| \sim 10^{262\,375}$, then the fraction of the volume of the total phase space that contains almost all outcomes is

$$\frac{|T_{N\beta}|}{|\mathcal{A}_X|} < 2^{N(\bar{H}+\beta)-H_0} = 2^{H-H_0+N\beta}. \tag{2.37}$$

[24]For comparison, the English language has $\bar{H} \approx 1$ bit/letter [64].

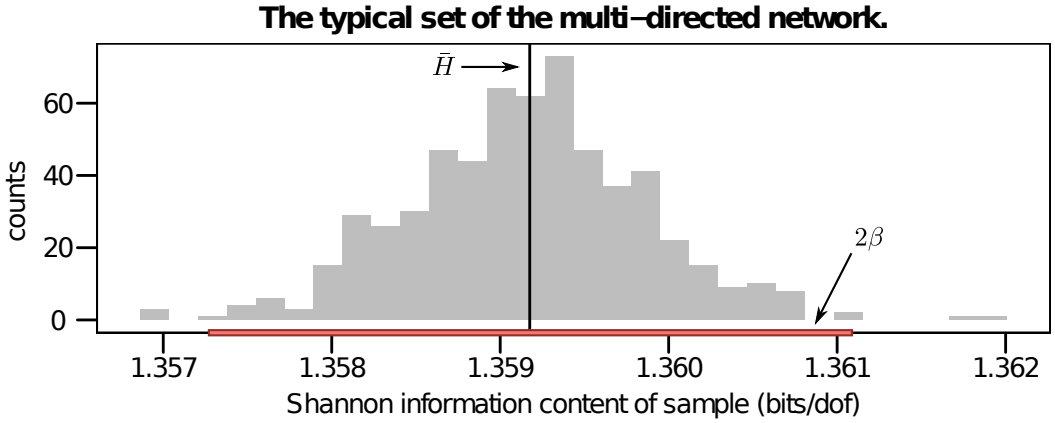


Figure 2.13: The information contents of 600 samples, i.e. outcomes, of the multi-directed network acquired with ensemble sampling. The number of dofs is 641 265. The estimated typical set $\bar{H} \pm \beta \approx 1.359 \pm 0.002$, indicated in red, contains 99% of the samples.

Plugging in the numbers from the caption of Figure 2.13 gives a reduction of the order $10^{180\,212}$ or more.

2.6 Term structure of interest rates and loan volumes in the interbank loan market

In this section we discuss how the interest rates and loan volumes of the interbank market depend on the loan terms. The relationship between the terms of securities, e.g. loans, and their market rates of interest is known as the term structure of interest rates [73]. A plot of interest rates as a function of security maturities is also called a *yield curve*.

The global categorical yield curve

The yield curve is obtained after averaging interest rates r over time and by term. The averaging technique chosen is the median due to extreme outliers which disproportionately affect the mean: the standard deviation of the complete r population (including outliers) is $\sigma \approx 40\%$, yet 99% of the population is contained in $[0, \sigma]$. Within that interval the standard deviation is about 7%, which is a more sensible measure of the interest rate dispersion. By using the median, we avoid choosing any cutoff.

Yield curves are usually considered with continuous maturities ranging from one day to several years, but the available data only records term categories. Although methods exist to estimate the continuous yield curve from discrete data [74], we will only consider interest rates (and interests etc.) per category.

The solid line in the top panel of Figure 2.14 exhibits a typical stylized fact of the yield curve. It is upward sloping and has a convex shape, ex-

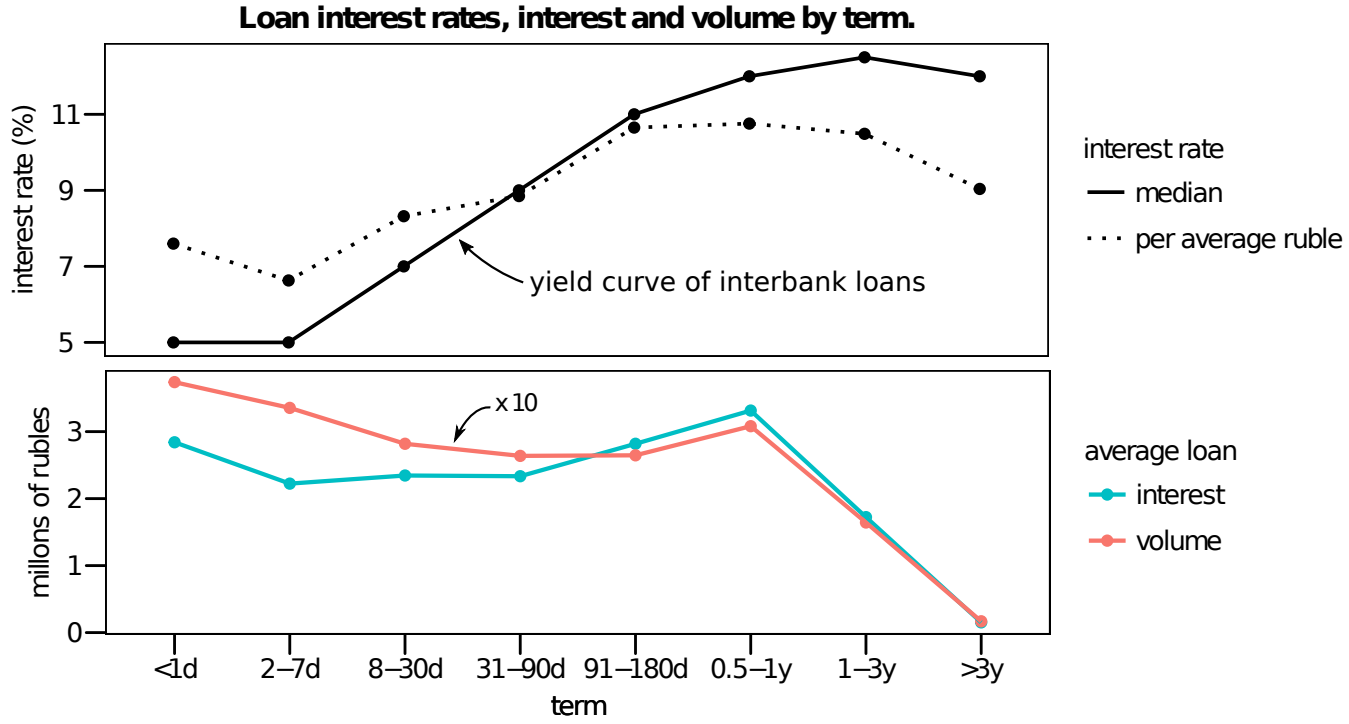


Figure 2.14: (top panel) The global categorical yield curve is the median interest rate over time per term. An alternative characterization is the interest rate per average ruble lent/borrowed, defined as the average interest weighed by the loan volumes. This statistic is robust to whether one excludes or includes the interest rate outliers. (lower panel) The average loan interest received/paid per term in millions of rubles and average loan volume (size) lent/borrowed in tens of millions of rubles.

cept for the `>3y` term, i.e. maturities three years or longer. The unusual steepness of the curve in Figure 2.14 is caused by the term categories; as they progress, they bucket a growing amount of maturities, so a continuous yield curve would be horizontally stretched with respect to the categorical yield curve. The upward slope is usually explained by classical *expectations theory* (ET). In a nutshell, interbank lending rates dynamics are determined by the structure of liquidity supply and demand [41]; the long-term interest rate is an average of expected future short-term rates, plus a term premium that increases with longer terms to compensate risk-averse lenders for the interest risk, which arises for lenders from fluctuating interest rates with respect to the base deposit policy rate (see Figure 2.2). One can also extend ET by including separate premiums for liquidity risk (selling loans on the secondary market tends to be harder as their maturities lengthen, as Table 2.8 reflects) and default risk[25], which is in theory governed by the credit rating of the borrower. Normally the short-end (long-end) of the yield curve

[25] Also called counterparty risk.

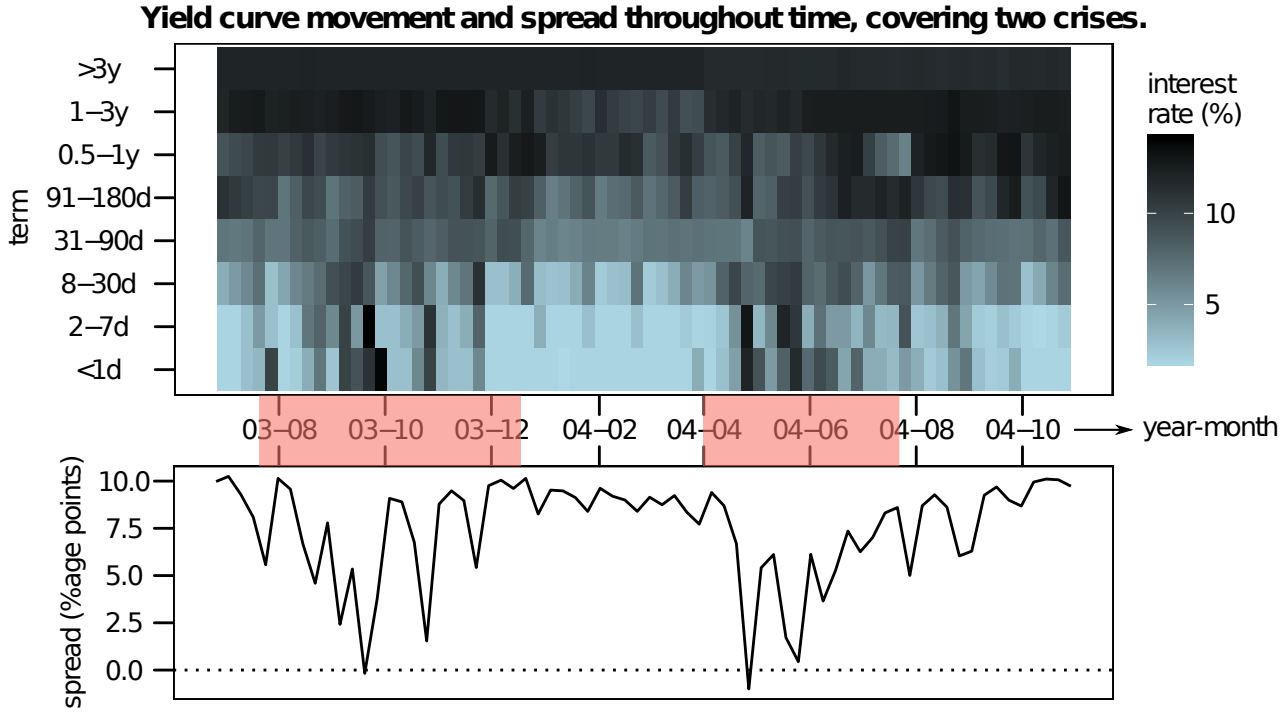


Figure 2.15: Weekly categorical yield curves and spread from July 2003 up until November 2004. This period is embedded in the maturity phase and includes the announcement and summer of 2004 crises (these are the trust crises – see Section A.2), both shaded in red on the shared timeline of the panels. The CBR lowered the overnight rate from 16 to 14% on January 15, 2004 and again to 13% on June 15, 2004 (see also Figure 2.2). The yield curve is the median interest rate per term, as in Figures 2.8 and 2.14; the spread is defined in the text. Since the loans with longer terms are increasingly sparse (see Figure 2.9), 200 samples were needed to get accurate estimates.

is dominated by the liquidity (default) risk; both are considered components of interbank lending risk [73, 75]. Each risk contributes to the upward slope of the yield curve given normal market conditions.

The interest per average ruble is also displayed in the top panel of Figure 2.14. Loans with longer maturities are the most profitable for lenders. The drop for the `>3y` term could point to the fact that these contracts may be made on more amicable (flexible) terms; this would imply lower perceived interbank risk which could explain the negative slope between the `1-3y` and `>3y` categories. The `>3y` interest rate quartiles in Figure 2.8 also indicate that the rates of this term barely change throughout time.

To illustrate the dynamics that underlie the global yield curve, Figure 2.15 plots the categorical yield curve for each of the last 70 weeks in the data. In this period the mature interbank network deals with the trust

crises which are explained in more detail in Appendix A. The lower panel contains the *(yield) spread*, which we define as the difference in average yield on long terms and short terms. We take these two classes to be (`1-3y`, `>3y`) and (`<1d`, `2-7d`), respectively. This classification agrees with Table 2.8 as well as [7, 8, 38] for the short terms and Basel recommendations together with [34, 35] for the long terms.

As of mid 2002 to November 2004, the weekly spread's trend is rising steadily from 7 to 10%age points, occasionally disturbed by sharp (i.e. short in time) troughs, in contrast to the structural decline in spread during the announcement and summer of 2004 crises.

In normal times, interbank markets are among the most liquid in the financial sector: banks prefer to lend out excess cash since the central bank's interest rate on excess reserves is smaller than rates available in interbank markets [76]. During trust crises, the perceived default risk grows, which inflates interest rates according to ET. Riskier banks, i.e. banks at risk of being in financial distress, exert an externality on safer banks who subsidize their liquidity. If the crisis gets worse, this externality on safer banks is so costly that they leave the unsecured market, and liquidity rich banks may prefer to hoard liquidity instead of lending it out to an adverse selection of borrowers; the interbank lending market dries up.

We see that this mechanism is indeed captured by the spread curve in the bottom panel of Figure 2.15: low spread seems to indicate abnormal market conditions with low liquidity because the short term interest rates suddenly explode. Note in the upper panel that the long term interest rates stay almost constant during the crises and drop during the recovery in between (roughly from January until April 2004). This suggests that we cannot make an analogy with the typical *inverted yield curves* of e.g. treasury securities that are associated with (predicting) recessions. Indeed, according to ET these are yield curves with negative slopes because investors have poor expectations of future interest rates. In contrast, we see for our data that the short term interest rates rise quickly during the trust crisis while the long term rates hardly change.

Lending activity by term

We define the *lending activity* simply as the amount of loans recorded during a certain period. Table 2.9 lists the yearly and monthly lending activities by term. A crucial observation is that the lending activity sharply decreases for longer terms, which is also reported for other interbank markets in [35]. We see that the overnight segment (`<1d`) is the most active, together with the `2-7d` term. The impacts of the crises in 1998 and 2004 have had clear impact on the lending activity, except for the longer terms; these seem relatively unaffected.

One is tempted to ask whether all terms need to be included for a satisfactory picture of the Russian interbank network as the `<1d` and `2-7d` terms make up 90% of lending activity. Indeed, previous research has used these two as a proxy for the whole network [7, 8, 39]. We will investigate the redundancy of the longer terms in the next Chapters.

Table 2.9: (top panel) Lending activity by term and per year. Note that 1998 and 2004 are incomplete years in the data, counting 5 and 10 months respectively. (lower panel) Monthly average lending activity by term. Time series of the lending activity can be found in Figure 2.9 on page 31.

	1998	1999	2000	2001	2002	2003	2004
<1d	7917	95 389	163 855	245 840	292 048	346 221	220 979
2-7d	4325	51 673	103 376	165 602	196 765	251 006	155 834
8-30d	958	12 914	29 250	42 067	45 204	62 190	42 181
31-90d	245	2770	6365	7939	10 606	14 184	8443
91-180d	82	460	742	1189	1754	2797	2010
0.5-1y	63	237	307	593	886	1901	1899
1-3y	52	207	279	202	180	914	845
>3y	55	128	180	367	2588	1712	890

	1998	1999	2000	2001	2002	2003	2004
<1d	1583	7949	13 655	20 487	26 550	28 852	22 098
2-7d	865	4306	8615	13 800	17 888	20 917	15 583
8-30d	192	1076	2438	3506	4109	5182	4218
31-90d	49	231	530	662	964	1182	844
91-180d	16	38	62	99	159	233	201
0.5-1y	13	20	26	49	81	158	190
1-3y	10	17	23	17	16	76	84
>3y	11	12	15	31	235	143	89

Characteristics of loan volumes and interests

The lower panel of Figure 2.14 shows the average interest and volume per term. The loan volumes are log-normally distributed [39] and this holds remarkably well for the interest too, especially for shorter terms. Because of the considerable variance on a linear scale, these averages may be interpreted only as rough order of magnitude estimates[26]. With this in mind, Figure 2.14 shows that interest rate and volume are roughly negatively correlated: except for the bump at 0.5-1y and the case >3y, the volumes decrease as the interest rates increase. This can also be seen in the slowly varying average interest for the first five term categories.

Table 2.10 lists the total loan volumes by term and per year. We observe that the relative importance of each term segment, as measured by the total volume of loans traded within it, follows the ranking of the loan terms. This, together with the typical volumes and interest rates in Figure 2.14 and the lending activity in Table 2.9, supports the conclusion that the Russian interbank network exhibits a distinct hierarchy with respect to the loan

[26] In general, log-normal distributions are reasonable descriptions for observers that only know the order of magnitude of the mean and the variance of a random variable [3].

Table 2.10: Total loan volumes by term and per year in billions of rubles.

	1998	1999	2000	2001	2002	2003	2004
<1d	226.3	2674.4	5984.3	8855.9	10 306.4	13 844.6	13 040.1
2-7d	83.1	1030.4	2910	5114.7	6104.6	9058.1	7880.2
8-30d	15.9	220.7	620.8	1103.1	1102.9	1871.8	1833.3
31-90d	4.4	31.9	105.2	205.3	318.3	415.4	253.6
91-180d	1.7	5.3	9.6	23.4	46.9	73.6	73.7
0.5-1y	3.2	2.4	8	20.2	24	50.8	69.9
1-3y	2.2	5.5	6	6	3.3	10.1	11
>3y	2.6	0.8	1.2	1.4	1.3	0.8	2.5

terms, which we could summarize by saying that banks lend greater volumes at lower interest rates more often for shorter loan terms.

Chapter 3

The Russian multiplex interbank lending network

In this Chapter we study the interbank data as a multi-layered network, or more precisely, a *multiplex network* [5]. The data presented in Chapter 2 is essentially a list of loan contracts, which we transform to an interbank loan network by drawing labeled nodes for banks and connecting them with edges directed from lender to borrower for each entry in the list. The resulting network contains a wealth of information fit for analysis using tools from network theory [26]. For example by studying the topology[1] of this network, one can gain insight into the principles of organization that emerge as banks lend and borrow on the interbank market [13].

The edges connecting lenders to borrowers can be equipped with the attributes from the previous Chapter, such as the estimated issuance date, the loan's term, its size and interest rate. Several approaches exist to take into account this additional information. For example, one can assign edge weights using the loan sizes. In a prototypical GPS network where nodes determine locations and edges possible routes between them, one could use the geodesic length of each route as its edge's weight. Shortest-path algorithms then take these weights as input to find the optimal route (i.e. an ordered set of edges) between two locations (i.e. nodes). Edge attributes are called *edge covariates* when their causal influence (possibly by construction) on the topology is emphasized. In the example of the GPS network, the geodesic distance is a covariate because two nodes are connected only if there is no shorter route via a third intermediate node.

Another way to incorporate the edge attributes is by placing all edges of the same type, induced by some attribute, in a separate layer. The nodes are globally indexed and present on each layer, where a node is said to be (in)active on a particular layer if it is (not) connected in that layer. This picture of the network is called a multiplex network, a concept that will be formalized in Section 3.2. Obviously, this approach works best for discrete attributes. In our case, we will study the multiplex interbank lending network by assigning each edge to a layer according to the designated loan's

[1] The 'wiring pattern' of the network [2].

term. This procedure leaves us with eight distinct layers, labeled `<1d` to `>3y`[2]; see Figure 3.1 for an illustration.

In this Chapter we examine each term layer separately[3] in order to characterize some of the organizational principles that shape the term's associated loan market. As argued in Section 1.2, we expect *a priori* that a term layer's topology depends significantly on the term because banks adopt different lending strategies according to the loans' terms involved. Therefore the term layers naturally introduce a new scale in the network.

We build upon this in the next Chapter by placing loans into layers defined by their term *and* issuance date. Two remarks are in place here. First, networks that live in different layers need not be independent; this is certainly not the case for the temporal layers in Chapter 4 that encode the time dynamics of the term multiplex network. Second, the time series of network statistics in *this* Chapter are generated using the same temporal layers, since aggregating a network to populate a time scale is equivalent to placing the network's edges in a temporal layer[4]. Thus one sees that the idea of a network consisting of several subnetworks distributed among layers, i.e. a multiplex network, is indeed quite broad and general. In fact, the epithet 'multiplex' has emerged only rather recently (in 2011) [77, 78], while the underlying idea of edge attributes is obviously older [5]. The difference between both is not formal, but suggestive.

The multiplex concept is usually employed when one wants to emphasize that different layers exhibit different *structure*, such that each type of interaction or interdependence between two nodes instantly defines a new layer. If one posits that the layer structure is generated by some mechanism, then each layer is characterized by its own dynamics, which justifies picturing the layers as a scale. Consequently, layers with similar dynamics could be aggregated to achieve a more compact model of the complete multiplex. This is precisely the subject of the next Chapter.

This Chapter, then, aims to characterize the interbank loan networks embedded in the term layers in order to study the lending dynamics associated with each term. We will employ both traditional [26] and multiplex [5,79,80] network measures. The former, applied to one layer, remain agnostic about the others, while the latter measure correlations between one or more layers. We extend previous work on the Russian interbank network [8] which only included loans of terms `<1d` and `2-7d`: in our formalism, these are but two of the eight term layers. These two layers account for nearly 90% of all loans (see Figure 2.9) however, so by Occam's razor it is natural to ask whether the other layers should be included in the first place.

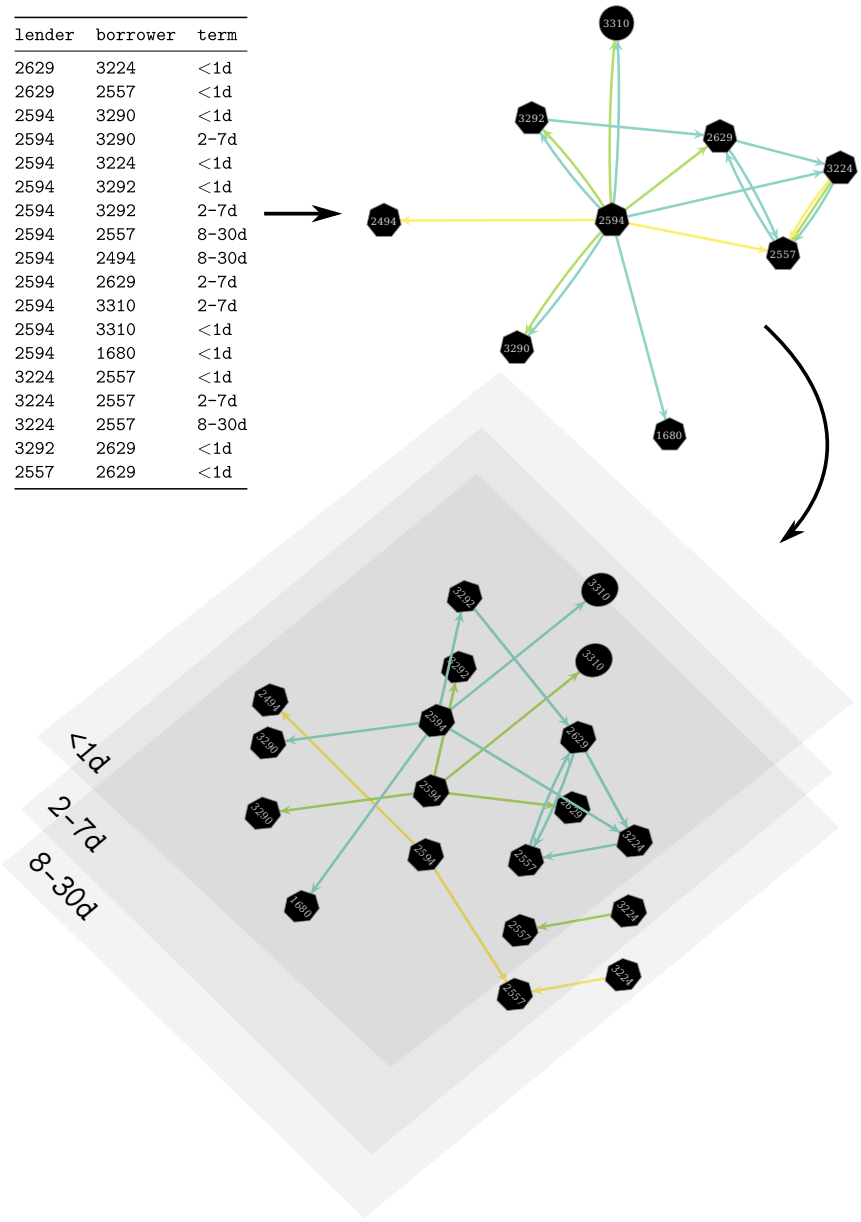


Figure 3.1: A list of loans is transformed into a network with edge attributes, which are then distributed across several layers, forming a multiplex network. This example displays the lending activity of bank `2594` during week 27, as well as the interaction between its borrowers. The edge attributes are limited to the loan terms. Edges representing loan with terms `<1d`, `2-7d` and `8-30d` are colored blue, green and yellow, respectively. The multiplex network shows a decreasing activity as the layers' terms lengthen.

Table 3.1: A list of stylized facts of interbank networks. Dashes are used to indicate several possible values. For example, for the clustering coefficients one reads "low clustering coefficients are reported by studies [8, 24]".

interbank network property	typical value	selected studies
density	sparse	[8, 13, 15, 24, 34, 35, 57]
degree distribution (in and out)	heavy-tailed	[13, 24, 27, 34, 35, 39]
topology	scale-free / core-periphery structure	[35, 81, 82]/ [13, 15, 27, 34, 57]
clustering coefficients	low / high	[8, 24]/ [13, 34, 35, 82]
average shortest path length	small / 'small world'	[8, 34, 57]/ [13, 24, 82]
bank size mixing	disassortative	[13, 24, 33, 34, 57, 81]
transaction volume	heavy-tailed	[24, 39]

3.1 Stylized facts of interbank networks

We argued in Section 1.1 that interbank lending markets have a natural network representation. The vast majority of empirical financial networks papers studies the overnight interbank market [13], which in our case coincides with `<1d` loans. Despite this limitation, numerous stylized facts of interbank networks have been established, which are listed in Table 3.1. The ultimate goal of the literature on interbank networks is to derive general results on the relation between the interbank network topology and financial stability [13]; these stylized facts summarize much of what is known today about the topology of a typical interbank network. As a side note, we note that much research has been directed towards contagion in interbank networks, i.e. cascading bank defaults, since the global crisis of 2007. That subject will not be treated in this thesis.

In the following paragraphs we will in general omit citations while discussing the properties listed in Table 3.1; instead we refer to the last column in that Table for backing studies. Most properties will be evaluated for each layer in the Russian multiplex interbank network in Section 3.4.

Density and degree distributions

Real-world interbank networks are typically *sparse*, meaning that in directed and undirected views of the network only a small fraction of all possible edges exist. One can then ask how the edges are distributed among the banks. To this purpose, we define the *degree* of a node. For undirected networks, this is the amount of edges connected with a given node, i.e. the number of the bank's counterparties. For directed networks, the indegree (outdegree) of a

[2] See Table 2.2 on page 11 for enumeration of the terms.

[3] This is possible because nodes are not connected between the layers in a multiplex network by definition.

[4] The network statistic may be correlating (see Section 2.4) because the layers are not required to be independent.

node is the number of incoming (outgoing) edges. The indegree (outdegree) of a bank is then simply the amount of loans borrowed (lent) by it.

Interbank networks, like many real-world networks [26], exhibit heavy-tailed degree distributions for each of the three types discussed above. In a nutshell, this means that *few nodes have many links and many nodes have few links*. In general, many typical distributions are heavy-tailed – in fact, they abound in descriptions of natural events like avalanches, earthquakes, turbulent flow and rainfall [83, 84]; an example in economics are the non-Gaussian return distributions in financial markets [67].

Several types of heavy-tailed degree distributions are reported in the interbank literature. These include power laws, stretched exponentials, log-normal and negative binomial distributions. One can also discriminate between the bulk and tail part, which may be described well by different distributions [39]. In what follows we will always consider both regimes jointly for simplicity.

Early work in the previous decade tended to postulate power laws for the degree distributions. Networks whose degree distributions follow a power law (at least asymptotically) are called *scale-free* networks. This important class of networks exhibits a remarkable – and very appealing – topology, which is illustrated in Figure 3.2. Scale-free networks are relatively robust to the random breakdown of nodes[5], the ability of their nodes to exchange messages being unaffected even by unrealistically high failure rates [85]. It has been argued that the World Wide Web is a scale-free network [86], which would explain why, despite frequent router problems, we rarely experience global network outages or, despite the temporary unavailability of many web pages, our ability to surf and locate information on the web is unaffected [85]. This robustness comes at a high cost, because the system is very vulnerable to the risk of the targeted removal of *hubs*, i.e. nodes with many connections. This characteristic has been coined *robust-yet-fragile*.

This concept is also used for interbank networks in the context of contagion and cascading defaults [4], although interbank networks are not communication networks. Hüser (2015) argues that if interbank networks are scale-free, identifying the hubs is a key policy objective [13].

However, while agreeing on the heavy-tail character, recent literature has cast serious doubt on the idea that power laws are the best candidate for the degree distributions, and thus on the scale-free character of interbank networks. In our case, power laws have been decisively rejected as best fit candidates for the heavy-tailed degree distributions of the `<1d` and `2-7d` by Vandermarliere et al. (2015) [39], while the network consisting of these loans has been found to be robust-yet-fragile [8].

Finally we note that heavy-tailed degree distributions are conveniently plotted by using *complementary cumulative distribution functions* (ccdfs) [84] with doubly logarithmic scales. Let us illustrate this for degree distributions. If we denote the degree and its distribution by k and $p(k)$ respectively,

[5]Note that random attacks are likely to land on nodes with small degree, since these are the most abundant.

the ccdf is given by

$$\text{ccdf}(k) \equiv 1 - \text{cdf}(k) = P\left(k \leq k'\right) = \int_k^\infty p(k')\,\mathrm{d}k', \tag{3.1}$$

where $\text{cdf}(k)$ is the usual cumulative distribution function. In Figure 3.2 we see that the ccdf of the rightmost Barabási-Albert network resembles a straight line, which is a necessary signature of a power law (but hardly more). By construction the ccdf is minimal (maximal) for nodes with large (small) degree, which directly translates to our phrase, "few nodes have many links and many nodes have few links".

Topology

It is easy to see that hubs emerge in any network with heavy-tailed degree distributions, i.e. they are certainly not exclusively tied to scale-free networks. Indeed, numerous empirical studies find that interbank markets have a *core-periphery structure*. The core-periphery structure is sometimes also called tiered or hierarchical structure in the network context [13]. Interestingly, it is currently an open question in the literature whether the core-periphery structure is robust or vulnerable to systemic risk [87]. The interbank market is tiered when some banks, called core banks, intermediate between periphery banks that hardly, or in the ideal case, do not extend credit among themselves. This implies that these 'money centers' [27] are hubs, while the periphery banks have relatively small degrees. The idealized core-periphery model for directed interbank networks can be summarized by four rules [13, 27]:

(1) core banks lend to each other,

(2) periphery banks do not lend to each other,

(3) core banks lend to periphery banks,

(4) core banks borrow from periphery banks.

The star graph with one core bank in Figure 3.2 has an idealized core-periphery structure if we substitute pairs of bidirectional edges[6] for one undirected edge. See the right panel of Figure 1.2 on page 7 for a qualitative picture of a more realistic situation.

Core banks tend to be large, with economies of scale and scope both playing a role in explaining network position. Network-related variables, such as systemic importance, also help identify which banks are in the core [27]. In addition, the set of core banks is found to be highly persistent over time.

Finally we return to the long-term trading relationships mentioned in Chapter 1, also known as preferential lending or relationship lending. Empirical research has shown that interbank lending is based on stable bilateral

[6] A pair of bidirectional edges are two edges connecting the same nodes pointing in opposite directions.

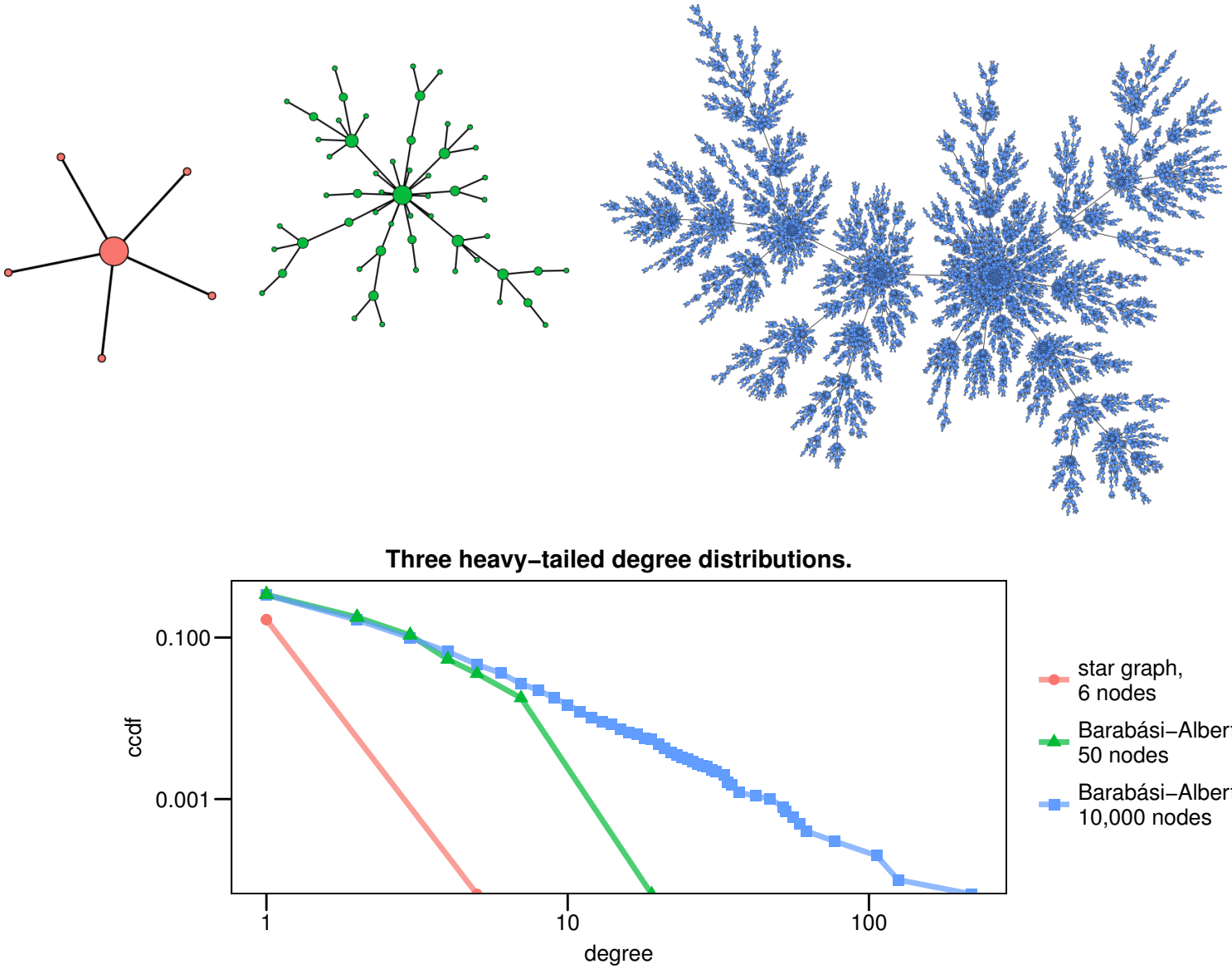


Figure 3.2: The ccdf of the degree distributions of three undirected networks with heavy-tailed degree distributions. The star graph is a stylized model of core-periphery structures, where the nodes in the periphery are unconnected among themselves but all depend on the core node. The Barabási-Albert networks are built from the star graph via the *preferential attachment* mechanism [86], which can generate power law degree distributions asymptotically if properly fine-tuned. We also mention that several other models for generating scale-free networks exist, such as the fitness model in [81].

trading relationships that facilitate access to credit and offer better loan con-

ditions [15]. It is interesting that several models which use various notions of 'trust' between banks are able to generate core-periphery structures [13].

Clustering coefficients and average shortest path length

Clustering coefficients measure to which extent banks form triangles (see the left panel of Figure 1.2 on page 7 for an illustration). Put differently, they measure the tendency of connected nodes to have common neighbors in undirected views of interbank networks.

According to Bargigli et al., an inverse relationship between the degree and the clustering of a node is observed quite commonly [35]. In the core-periphery picture, low clustering values of core nodes indicate that they essentially behave as star centers. A star graph (see the left panel of Figure 3.2) exhibits zero clustering, as the periphery nodes are unconnected amongst themselves. Consequently, deviations from the star graph, which has idealized core-periphery structure, can be probed by measuring the local clustering coefficients of the periphery nodes. One can identify these heuristically by simply considering the nodes with low degree, so the aforementioned inverse relationship hints that the core-periphery structure may exhibit considerable complexity. A second and more robust implication is that the clustering coefficient of the complete interbank network is dominated by the clustering of low-degree nodes if this inverse relationship is observed together with heavy-tailed degree distributions.

High clustering has obvious implications for systemic risk [39] and therefore clustering coefficients are of interest to the interbank network literature. Table 3.1 shows, however, that at least a few studies claim opposite 'typical' values for clustering coefficients. This may be due to the fact that the clustering tends to increase with longer time windows [35], or that the coefficients may not have been compared to the mean clustering coefficient obtained for random networks of the same size and number of edges.

The average shortest path length[7] D is defined for the largest connected component of the undirected view of interbank networks. It indicates the typical distance between two randomly chosen nodes, i.e. the smallest amount of edges needed to reach one node from the other. *Small-world networks* are characterized by high clustering and small average path length[8].

Most studies find that D is small for interbank networks, which indicates compact network structure, but not all conclude that they are small worlds. Hubs tend to lower the average path length; scale-free networks are ultra-small-worlds. D is a measure of the typical length of intermediation chains that are taking place among the market participants (at least in the largest connected component). Longer intermediation chains arise when D is large, which effectively contribute to slowing down the market transactions between participants and consequently harming the liquidity allocation between financial institutions. In contrast, when D is small, the information between the market participants flows quickly in the network, giving rise to a well-functioning liquidity allocation in the market [57].

[7] From now on we will usually write just 'average path length'.

[8] Where both should be compared to their mean values in an ensemble of random networks with the same amount of nodes and edges [88].

In the case of the Russian interbank network, the average path length peaks at times when the network is in crisis, and decreases gradually when the network is maturing. The clustering is volatile in times of crisis, and increases during the growing phase of the network [39]. The same patterns have been reported in other empirical literature, such as in [15].

Bank size mixing and transaction volume

Many studies point to *disassortative mixing* with respect to the bank size, meaning that small banks trade mainly with large banks and vice versa. Given the core-periphery picture, one expects that the core banks are the large banks. Indeed, it is found that total bank assets are significant in explaining core membership [27]. If we follow our intuition once more and let bank size correlate with total degree, we would also expect disassortative mixing of the banks with respect to the (total) degrees, i.e. high (low) degree nodes tend to be connected to low (high) degree nodes [26]. In fact several studies report this as an additional stylized fact [34, 35].

Lastly, heavy-tailed transaction volumes have been reported. In our case, the loan volumes are well described by log-normal distributions (see Section 2.6) [39].

3.2 Multiplex networks

The basic units of many real world systems are connected through a large variety of different relations. One of the new challenges in network theory is therefore to treat together ties of different kind preserving existing differences. The multiplex metaphor, which allows to distinguish the different kinds of relationships among a set of nodes, constitutes a promising framework to study and model multi-layer systems [79]. In this thesis, only a minimum amount of elementary mathematical concepts describing multiplex networks are introduced to keep matters simple. Their notation is mostly borrowed from [23]. For a full-fletched exposition we refer the reader to [5, 77, 78].

Basic definitions

Formally, a multiplex network[9] $\mathcal{M}$ with N nodes is a set of *adjacency matrices* $A^{\alpha}_{ij} \in \mathbb{R}^{N\times N}$ where $\alpha \in \{1, 2, \ldots, M\}$, with M being the number of layers in $\mathcal{M}$ [23]. Throughout the following, we will use Greek or named superscripts to denote the layer index when required. For example, an edge that is explicitly contained in layer α is denoted e^{α}. The matrix A^{α}_{ij} represents a graph $\mathcal{G}^{\alpha}$ with N nodes and E^{α} edges, which we also write as a (N, E^{α})-graph. For *simple graphs*, i.e. networks without parallel edges[10], edges can be uniquely identified by the pair of indices of the nodes they

[9] Networks are also called graphs in much of the mathematical literature [26]. We will use both terms interchangeably.

[10] Also called multi-edges.

Table 3.2: The bank classification (BC) from Vernikov (2016) [89] together with the amount of banks present in the network per class and the symbols used in the network graphics.

BC		description	# banks
S1	▲	Core banks controlled by federal executive authorities and CBR ('national champions')	3
S2	■	Other banks controlled by federal executive authorities	4
S3	⬟	Banks controlled by regional authorities and municipal authorities	17
S4	⬢	Banks majority-owned by state-owned companied and/or state-owned banks and/or 'state corporations'	45
F	⬢	Subsidiaries of foreign commercial banks in Russia	54
D	●	Domestic private banks	1347
			1470

connect: $e = (i, j)$. This tuple is (un)ordered for (un)directed graphs. Note that the directed and undirected view of the network introduced in Chapter 2 correspond to a directed and undirected simple graph, respectively.

The adjacency matrix A_{ij} encodes the topology of its associated network. For simple undirected graphs,

$$A_{ij} = \begin{cases} 1 & \text{if node } i \text{ and } j \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases} \tag{3.2}$$

Throughout this thesis we exclude self-loops, i.e. $A_{ii} \equiv 0$. One sees immediately that A_{ij} is symmetric. This does not hold in general for directed graphs, since $e = (i, j) \neq e' = (j, i)$. The above definition of A_{ij} can be extended to multigraphs, i.e. graphs where multiple edges between two nodes are allowed. In this case we adjust (3.2) to take the number of edges connecting node i and j into account, so $A_{ij} \in \mathbb{N}$. The multi-directed view of the network presented in the previous Chapter is an example of a multigraph. We can generalize this further by letting $A_{ij} \in \mathbb{R}$; we obtain a *weighted graph*.

Besides the adjacency matrix, a network may also hold edge or node attributes, or both. In this case we refer to a *labeled network* G. Multiplex networks are necessarily labeled because their nodes are indexed globally, i.e. each node can participate in any layer. In our case this is done by assigning a bank registration number and a *bank class* (BC) to each node; see Table 3.2 for a list of the latter.

To stress the importance of the edge and node attributes, we will denote the interbank lending multiplex network as $\{G^\alpha\}$. For example, the multiplex labeled network in Figure 3.1 consists of three layers $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$ and $G^{\texttt{8-30d}}$. The edge attribute `term` is identical in each of these layers.

The *collapsed graph* G^c associated with $\mathcal{M}$ corresponds to the merging of all edges in a single layer, with a resulting adjacency matrix $A^c_{ij} = \sum_\alpha A^\alpha_{ij}$

[25]. Returning to Figure 3.1, the collapsed graph is the network at the upper right. A related but crucially different concept is the *aggregated graph* G^a, in which edges connecting the same nodes in different layers are condensed. In symbols we have

$$A^a_{ij} = \begin{cases} 1 & \text{if node } i \text{ and } j \text{ are connected in at least one } A^\alpha_{ij}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

It follows that G^a is a simple graph. How the edge attributes are to be 'condensed' needs to be explicitely defined, unlike for G^c, where the collapsed edges simply coexist. The directed and undirected view of the interbank network are examples of aggregated graphs. The aggregating schemes for the edge attributes have been discussed in Section 2.5.

As a final remark, we mention that multiplex networks are a special case of *multilayered networks* [5, 77], in which nodes may have interlayer edges. As in this thesis, most of the work in the empirical literature is done within the multiplex framework [5].

Background and use

In Chapter 1 we have discussed several merits of network theory when it comes to understanding complex systems. These range from biological to technological, and social systems [5, 26]. Within the framework of classical network theory, scientists have focused primarily on network data sets that contain only one type of relation, thus (largely intentionally) neglecting the complexity of the connections between the various agents in these systems [90]. Taking advantage of the increasing resolution in real data sets[11], network scientists have directed their interest to the multiplex character of real-world systems [5, 78]. Indeed, recent literature on complexity has seen an upsurge of interest in multiplex networks [80, 90]. The multilayer idiom has been proven necessary to capture essential structural information in various social, biological and man-made systems, such as trade networks and interbank networks [2, 5, 20, 35, 36, 91].

As indicated before, multilayered networks allow a more detailed modeling of complex systems. In the case of multiplex networks, the layers typically constitute either different types of interaction between the agents, or distinct snapshots of the network throughout time, or sometimes both [31]. In the context of interbank lending networks, we can apply the multiplex formalism naturally by assigning loans to layers according to their term.

3.3 Empirical properties of multiplex interbank networks

The literature discussing multiplex interbank networks is definitely not large [35], mainly because of a lack of sufficient granular data [34]. Most of the extant literature on interbank networks has worked with the simplification of a

[11] Recall prof. dr. Donoho's speech on page 36.

Table 3.3: Classifications of loan terms used by the Central Bank of Russia and several studies.

	<1d	2-7d	8-30d	31-90d	91-180d	0.5-1y	1-3y	>3y
Central Bank of Russia [20]	instant	short term		long term				
Bargigli *et. al* (2015) [35]	overnight	short term					long term	
Aldasoro and Alves (2016) [34]	short term						long term	
Montagna and Kok (2013) [36]	short term				long term			

single layer of exposures; however, banks engage in a multiplicity of transactions [13]. Recent empirical studies [33–36,87,92] argue that a more realistic representation of the interbank market is a multiplex network, where layers can represent maturity, nature of the contract (secured versus unsecured), instruments, direct and indirect links [13]. In our case, we only discriminate between the maturities; all interactions are unsecured loans, since we use the reduced network (see Section 2.5). We now discuss some relevant findings, including results on contagion processes, of selected studies concerned with multiplex interbank *loan* networks composed of *term* layers[12].

Term layer topology and related properties

Bargigli et al. (2013) have studied the Italian interbank multiplex network [35]. They used three term layers, defined in Table 3.3. Each layer is further broken down by the secured or unsecured nature of the contract. While they find significant topological differences between the secured and unsecured networks, of interest to us are the reported *similarities* between the unsecured term layers in the Italian interbank market. Simultaneously they recover several stylized facts of general interbank networks: each layer is sparse and has small average path length, and layers exhibit similar disassortativity properties and heavy-tailed degree distributions. In contrast, they find that clustering is dependent on the layers, with the largest clustering observed in the unsecured overnight segment. The amount of banks and the lending activity in the layer decrease for longer terms, just as we have concluded for the collapsed Russian interbank network in Section 2.6. The inverse relationship between degree and clustering is also observed, albeit weaker for layers holding longer loan terms.

Although the layers share similar topological properties, it is found that they cannot be said to be *representative* of each other in the sense that the presence of an edge in one layer is not a good predictor of the presence of the same edge in another layer. Put differently, the topological similarity did not imply point-wise similarity. They conclude that this suggests significant complementarity between different segments of the interbank market. We add to this that finding evidence of layer-specific hubs would firmly support this idea.

[12]Reports that study multiplex interbank networks but do not or hardly discern term layers are [33, 87, 92].

A multiplex interbank network built out of 2245 exposures between 53 large European banks on December 31, 2011 was studied by Aldasoro and Alves (2016) [34,93]. The exposures were broken down by both maturity and instrument type. The instrument type of interest that corresponds to the loans in the Russian interbank network are the debt securities, and we only discuss the term layers applicable to their maturities – see Table 3.3. What is most interesting is that they find relative high point-wise similarity between the term layers, pointing in this case to a relative lack of complementarity. This is supported by the fact that the *centrality*[13] ranking of banks persists across the short term and long term layers. In addition, for each of the term layers they report similar degree distributions and high levels of reciprocity, i.e. the fractional amount of bidirectional edges.

Using the algorithm of Della Rossa et al. (2013) [1], a *coreness ranking* of the banks was inferred, a concept we will discuss at length in the next Section. In a nutshell, the coreness of a bank quantifies to which extent it resembles a core in a core-periphery structure. By comparing the ranking for the short term and long term layers it was found that these exhibit different core banks. Put another way, the largest hubs differ across the two layers. Additionally, they found that the short term layer has a stronger core-periphery structure than the other.

Implications for contagion and systemic risk

Montagna and Kok (2013) have developed an agent-based model on top of a multilayered interbank network [36], that includes short term and long term layers to reflect liquidity and counterparty risk, respectively (see Table 3.3 for the composition of the two layers). The model is calibrated on a data set of 50 large European banks. Their main finding is that the contagion effects of a shock to, say, the long term layer can be significantly amplified, compared to the situation where contagion risks are assumed to be confined within the short term layer where the initial shock arose. This finding points to the existence of important non-linearities in the way bank-specific shocks are propagated throughout the financial system. Thus they conclude that the contagion risk is likely to be underestimated by policy based on collapsed or aggregated forms of the interbank lending network. We add that in this regard the point-wise similarity between term layers is important, as the contagion must depend on it [35]; structural differences in these two layers immediately imply that counterparty credit risk and liquidity risk propagate in the interbank network by different contagion processes [13].

3.4 Term layer analysis[14]

We divide the Russian multiplex interbank network $\{G^\alpha\}$ by term, which defines eight term layers: $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$, $G^{\texttt{8-30d}}$, $G^{\texttt{31-90d}}$, $G^{\texttt{91-180d}}$, $G^{\texttt{.5-1y}}$,

[13]The centrality of a bank measures in some way the 'importance' of a bank. Aldasoro and Alves used various centrality measures to gauge the systemic importance of banks

[14]The work was done with two network `C++` libraries: `igraph` [94] and `graph-tool` [29], of which we used wrappers for `R` and `Python`, respectively.

Table 3.4: Note that the columns '1998' and '2004' are biased because they are incomplete in the data. (top panel) The amount of (unique) active banks involved in the trade of loans by term and per year. A total of 1040 banks are present in the data. (lower panel) The amount of loans per active bank by term and per year, obtained by dividing the top panel of Table 2.9 and the top panel of this Table element-wise. Care must be taken when interpreting; this is not indicative of an 'average' bank since the degree distributions are heavy-tailed.

	1998	1999	2000	2001	2002	2003	2004
<1d	610	810	878	954	1022	1027	983
2-7d	589	907	953	1026	1074	1098	1058
8-30d	375	849	920	980	1018	1054	986
31-90d	166	535	648	698	706	775	686
91-180d	63	203	265	310	321	429	378
0.5-1y	49	120	132	169	204	258	259
1-3y	24	64	80	60	58	113	107
>3y	11	45	29	38	33	23	20

	1998	1999	2000	2001	2002	2003	2004
<1d	13	118	187	258	286	337	225
2-7d	7	57	108	161	183	229	147
8-30d	3	15	32	43	44	59	43
31-90d	1	5	10	11	15	18	12
91-180d	1	2	3	4	5	7	5
0.5-1y	1	2	2	4	4	7	7
1-3y	2	3	3	3	3	8	8
>3y	5	3	6	10	78	74	44

$G^{1\text{-}3y}$ and $G^{>3y}$. Also considered are the collapsed (aggregated) graph G^c (G^a) to evaluate which layers are most similar to the complete interbank network. Depending on the context, time windows of weeks, months or years will be used. We will try to adhere to the structure of Section 3.1 as much as possible for easy comparison between the term layers and interbank network stylized facts; Section 3.5 summarizes the analysis done in this Section and concludes. Finally, we note that only the most relevant graphs are shown during the discussion. Additional and supporting findings can be found in Appendix C.

Density, directedness and reciprocity

The lending activity in the layers (Table 2.9) follows a clear pattern: less loan contracts exist for increasing terms, regardless of the time window. As

anticipated, this has strong implications for statistical analysis. Layers with longer terms, i.e. $G^{\texttt{91-180d}}$, $G^{\texttt{.5-1y}}$, $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$, sometimes have so little lending activity that monthly time windows yield bad statistics; in that case yearly time windows are used without further notice.

One expects that the number of banks participating in the market's segment as defined by a given term also decreases as the terms lengthen. Table 3.4 shows that this is approximately correct. More importantly, one sees that the number of active banks[15] becomes astonishingly small for $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$. For the latter, this results in an awkwardly high average lending activity per bank, comparing to shorter term layers. It seems that $G^{\texttt{>3y}}$ is occupied by a small club of banks that trade relatively intense.

This is reflected in the undirected density of the layers shown in Figure 3.3. All layers can be considered sparse, except for $G^{\texttt{>3y}}$, which has an extremely high [95] density. We also observe that the density in almost all layers grows steadily as the network matures, pointing to an increasing interconnectedness throughout the years [7]. Finally, we report that the directed density is approximately half of the undirected density everywhere (Figure C.1).

This hints that the network is rather strongly directed, i.e. bidirectional edges in directed views are not abundant. That implies that the directed view holds valuable information, as completely strongly directed networks can equally well be described by undirected networks. Figure C.2 confirms this suspicion, indicating reciprocity values of about 25% for the short term layers, and highly volatile lower values for the longer terms. The reciprocity is especially noisy for $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ during 1998 and 1999. Since the reciprocity r measures the probability that a random edge is reciprocated, one should compare it to the reciprocity expected in a random graph with the same number of nodes and edges in order to assess if mutual links occur more or less (or just as) often than expected by chance [26, 96]. This eliminates the bias incurred by r for small and dense networks, in our case $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$.

We will denote the corrected reciprocity for a simple directed (N, E) network as ρ-reciprocity [96],

$$\rho = \frac{r - \bar{r}}{1 - \bar{r}}, \tag{3.4}$$

where $\bar{r} = E/N(N-1)$ is the expected reciprocity in a uncorrelated random (N, E)-network. The ρ-reciprocity for the term layers is shown in Figure 3.4. It allows to distinguish between reciprocal ($\rho > 0$) and antireciprocal ($\rho < 0$) networks, with mutual links occuring more and less often than random respectively. The neutral or areciprocal case corresponds to $\rho = 0$ [96].

The fact that the layers are quite strongly directed from a more graph-theoretical perspective does not mean that the observed ρ-reciprocity is small. Indeed, we find large values of ρ for the short terms and the aggregated network, comparable to those found in e-mail networks [96]. The

[15] As a reminder, an active bank in a given time window and layer has lent or borrowed at least once. We will always consider active banks to deal with the death and birth of banks in Russian interbank market over the course of time [47].

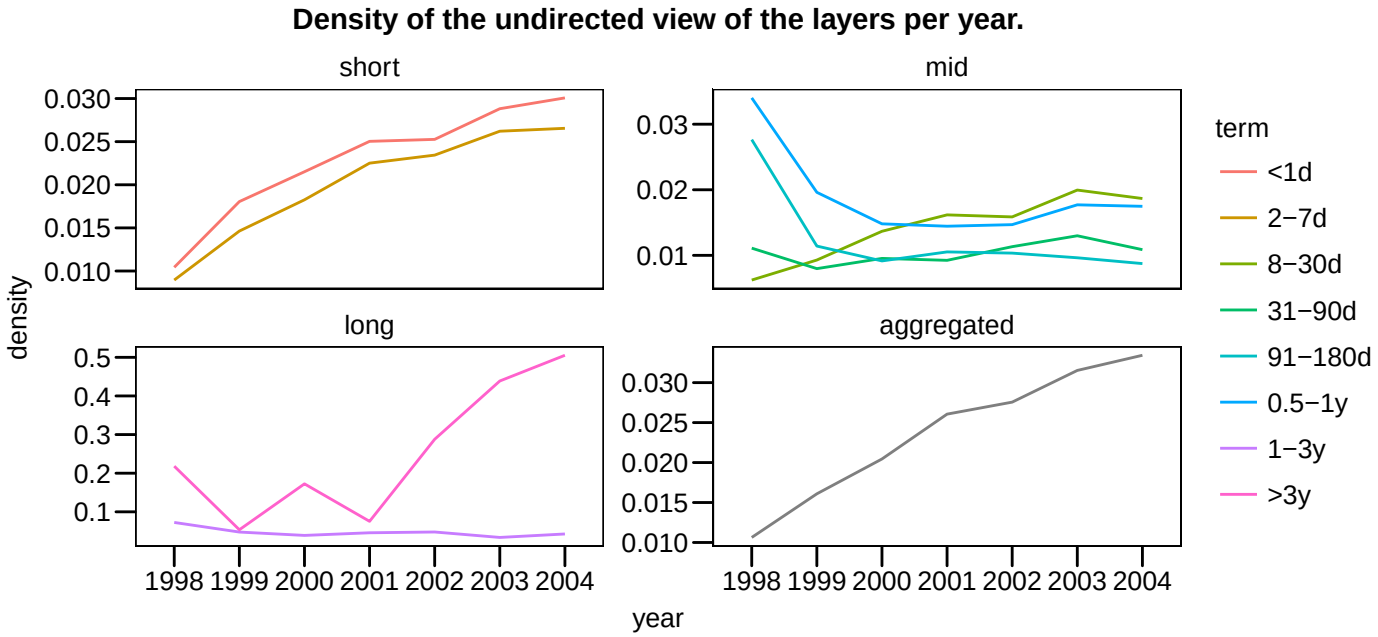


Figure 3.3: The density of an undirected *simple* graph with N nodes and E edges is $d = \frac{E}{N(N-1)/2}$, i.e. the fraction of observed edges to possible edges. One sees that the graph must be simple, for an infinite amount of edges can exist if parallel edges are allowed. Note the different individual density scales. The high density of the mid-term layers (topright panel) in 1998 are caused by incomplete data in that year.

ρ-reciprocity declines to zero for longer terms, even reaching slightly antireciprocal values for $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$. This decline in ρ-reciprocity is also reported for the Italian interbank network (IIN) by Bargigli et al. (2013) [35], although the reciprocity in the Russian network is notably smaller. This can be attributed to the larger amount of banks in our data (about 1500, 2.5 times the size of IIN) – this combined with a large geographical spread tends to generate community structures, which further pushes down the reciprocity. Furthermore, since the total trade volume in a layer is larger for shorter terms (see Section 2.6), the higher ρ-reciprocity may indicate that a high share of exposures are reciprocated by a corresponding counter-exposure, perhaps to compensate for the unsecured nature of the loan contracts[16]

Degree distributions

The degree distributions are shown in Figures 3.5 and C.3. Each layer exhibits fat-tailed in- and outdegree distributions, with the maximum degrees separated some 5 ($G^{\texttt{>3y}}$, indegree) to 18 ($G^{\texttt{8-30d}}$, indegree) standard devi-

[16]It is also suggested in [34] that high reciprocity can underestimate systemic risk in cases of high levels of aggregation since some exposures might be netted.

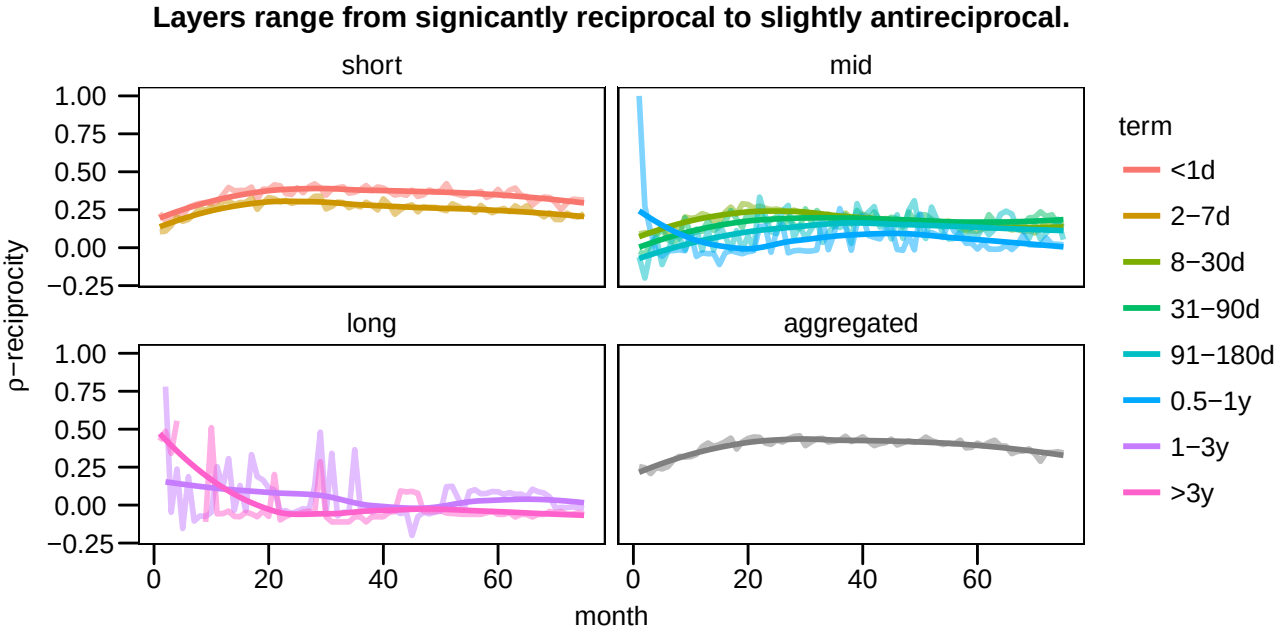


Figure 3.4: (smoothing added for clarity) ρ-reciprocity as defined in Garlaschelli (2004) [96] for the layers and the aggregated network by month.

ations from the mean. Vandermarliere et al. (2015) [39] have established that stretched exponentials of the form

$$f(d) = Cd^{\beta-1} \exp\left(-\lambda d\right)^{\beta} \tag{3.5}$$

provide the best overall fit for the bulk+tail multi-directed in- and outdegree for weekly, monthly and yearly time windows. In (3.5) d is the indegree or outdegree, C a normalization constant and λ, β distribution parameters. The stretched exponential can be understood as a Weibull distribution, with β being the shape parameter, and $1/\lambda$ the scale parameter. If $0 < \beta < 1$, the distribution has a fat tail, with a smaller β putting more weight towards smaller d. $1/\lambda$ widens the distribution, the mean of f being proportional to it.

We have fitted f to the degree distributions to see if this result can be extended to longer terms. The conclusions are identical for the in- and outdegrees. First, the stretched exponential fit does seem to describe the time-aggregated degree distributions of $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$ and $G^{\texttt{8-30d}}$. Then the leap in residual sum-of-squares suggests that $G^{\texttt{31-90d}}$ and longer terms can not be satisfactorily described by the stretched exponential. This is confirmed by visual inspection. While (parts of) the bulk distributions of the longer terms are reasonably well captured by f, the fits systematically underestimate the ccdf in the tail; put differently, they underestimate for a given large degree d how many nodes exists with an even larger degree d'.

Figure 3.5 also shows the significant variation of large degrees between the layers; this is listed in Table 3.5 to get a feel of the magnitude of the numbers involved. Except for the large differences between the maximum

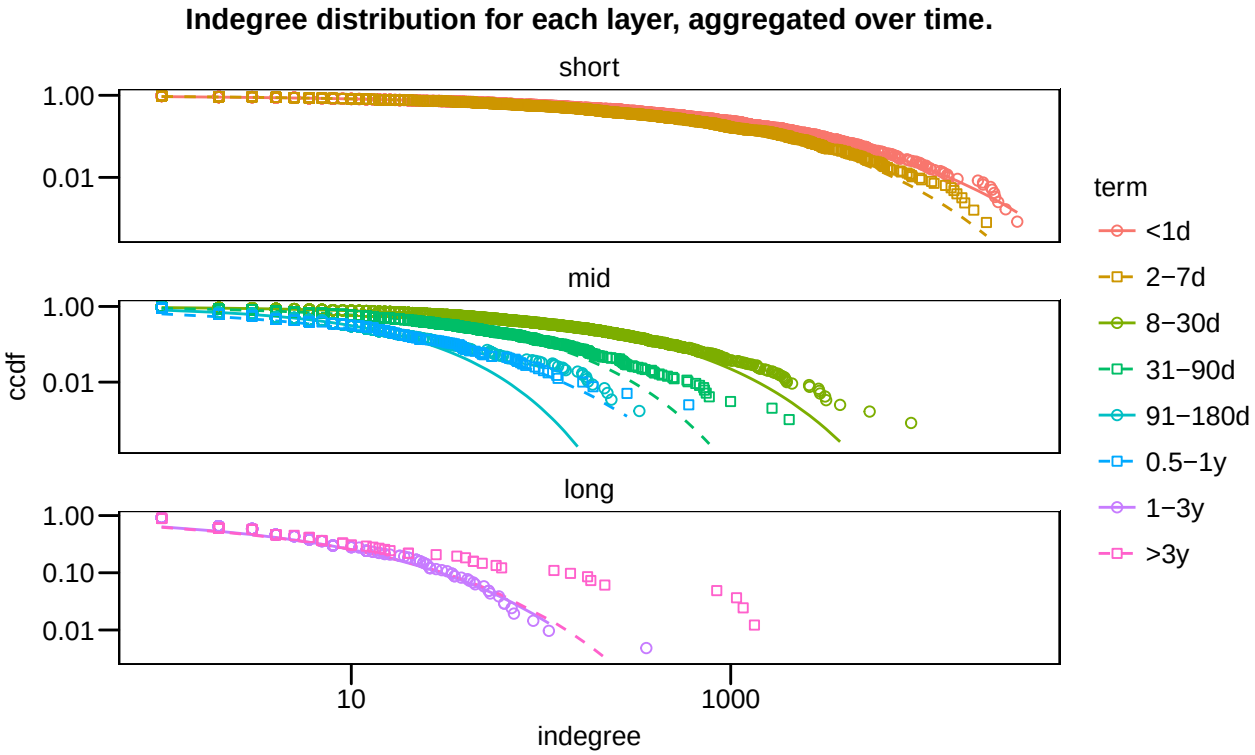


Figure 3.5: Indegree distribution with attempts to fit the bulk+tail with stretched exponentials, drawn as piecewise linear functions connecting each predicted value. The correspondence with this distribution seems to break down for terms longer than `8-30d`. Inferred (λ, β) parameters for $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$, $G^{\texttt{8-30d}}$ are $(2.6 \times 10^{-3}, 4.3 \times 10^{-1})$, $(3.4 \times 10^{-3}, 4.8 \times 10^{-1})$, $(1.0 \times 10^{-2}, 5.8 \times 10^{-1})$ respectively, with all standard errors below 1%.

and mean degrees, we also see that different banks dominate the lending and borrowing market[17] (except for the domestic bank `3279`), and that foreign banks seem the most frequent lenders for the long-end segment of the loan market. Again $G^{\texttt{>3y}}$ stands out when compared to other long terms; the 'foreign' lender `3338` completely dominates the lending segment, which can also be seen in Figure C.3, while the tail of the borrowers is more evenly distributed between four banks. In fact, the tails in Figure 3.5 and Figure C.3 reveal that the most active borrowers typically constitute a relatively large set of banks, instead of one super-dominant bank.

Finally, note that the 'mean' columns in Table 3.5 are comparable for in- and outdegrees. This raises the question whether frequent lenders are also frequent borrowers. To answer this, we first rank the banks with regard to in- and outdegree, giving the banks listed in Table 3.5 a rank of 1, etc. Then we calculate Kendall's coefficient of concordance W, which measures the rate

[17] At least in terms of lending and borrowing activity; we will investigate the weighted variants using loan sizes in the next paragraph.

Table 3.5: Mean and maximum in- and outdegree of the time-aggregated degree distributions, normalized by the total amount of months (74) in the data. For the maximum degrees, the bank registration number and class are also given. A list of bank classes (BC) is given in Table 3.2.

	monthly indegree				monthly outdegree			
term	mean	max	borrower	BC	mean	max	lender	BC
<1d	15.5	438	2412	D	15.2	383	1326	D
2-7d	10	301	3279	D	9.8	217	2361	D
8-30d	2.6	121	3279	D	2.6	60	2275	D
31-90d	0.7	27	3279	D	0.7	23	1326	D
91-180d	0.2	4	2361	D	0.2	9	2272	F
0.5-1y	0.2	8	1067	D	0.2	11	2272	F
1-3y	0.2	5	729	S4	0.2	6	3340	S4
>3y	1	18	605	D	1.5	68	3338	F

of agreement $0 < W < 1$ between two rankings; we obtain the extremely high value 0.96. So the answer is positive, but it is skewed towards shorter terms by their extreme overweight in terms of the amount of active banks (see Table 3.4). To get a more nuanced answer, we repeat the Kendall test for each layer separately. The significant results together with the joint degree distributions are shown in Figure 3.6. We conclude that lending and borrowing are highly correlated for $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$, $G^{\texttt{8-30d}}$ and $G^{\texttt{31-90d}}$; in other words, the dominant lenders are likely to be dominant borrowers, and vice versa. This correspondence breaks down markedly for the longer terms, with orderly decreasing W but insignificant test results. This can also be seen in the shapes of the contour plots in Figure 3.6: high indegree-outdegree correlation is marked by squeezed contours around the plot diagonal, which does not hold for terms longer than `31-90d`. $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ even display modest anticorrelation for the degree tails, meaning that the most frequent lenders (borrowers) area unlikely to be among the most frequent borrowers (lenders).

Tiering and intermediation

It is convenient to discuss *interbank intermediation* before dealing with core-periphery structure in the layers. In a seminal contribution Craig and von Peter (2014) introduced the concept of tiered banking in the interbank network literature [27]. As stated in Section 3.1, interbank markets are tiered when they operate in a hierarchical fashion where lower-tier banks deal with each other primarily through core banks, which typically behave as money centers. The core banks (c-banks) intermediate between the periphery banks (p-banks), which then lend to each other indirectly. It is in this sense that they are said to "keep the network together" [1, 13]. Non-random network tier structure must have been shaped through economical forces like search costs in OTC-markets for p-banks and economies of scale and scope for the

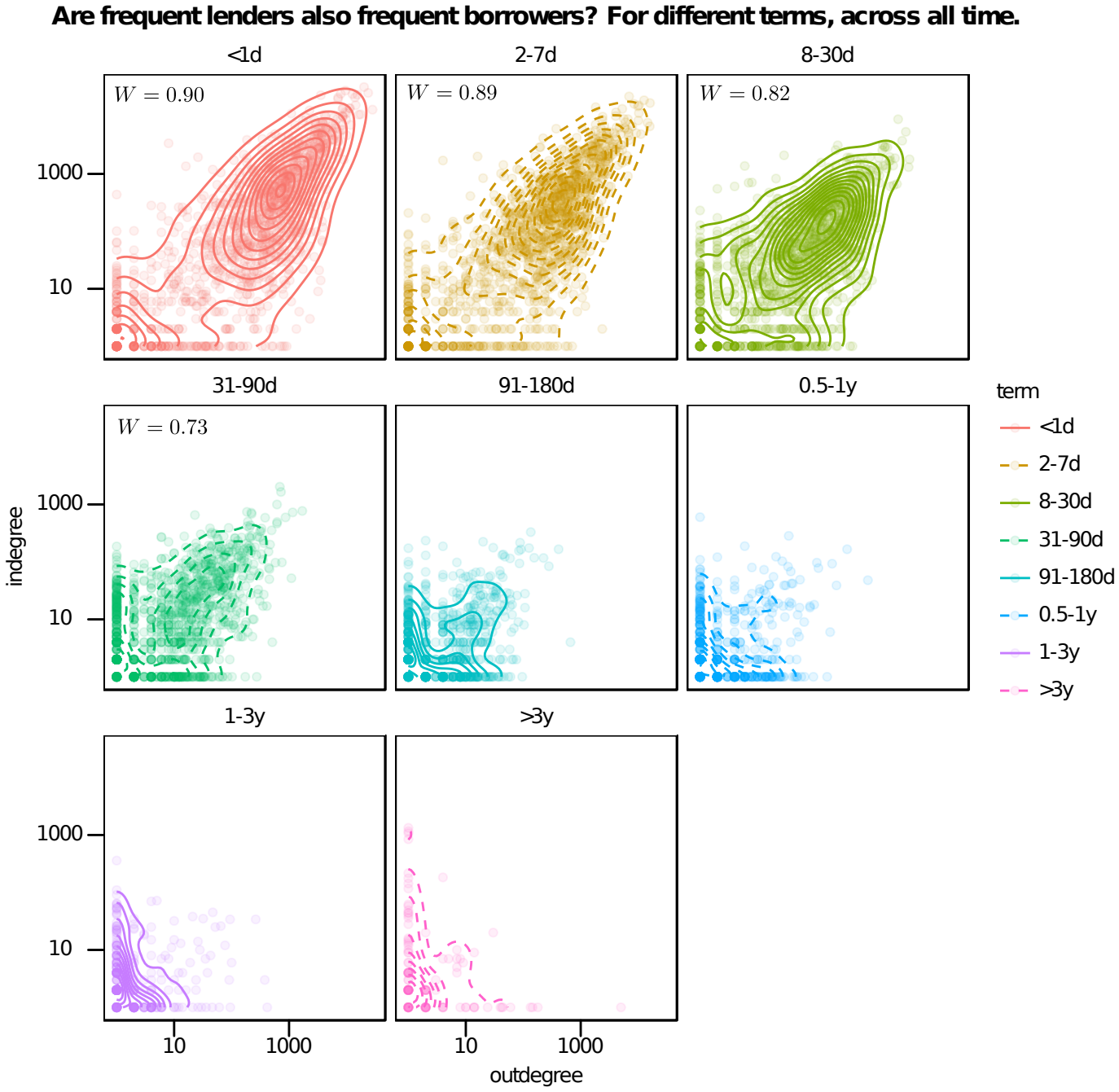


Figure 3.6: The time-aggregated joint degree distribution for each layer, drawn as density contour plots with almost transparent data points. For the first four terms, a highly significant Kendall W was obtained and put in the top left corner of the panel. Note that all scales are equal for easy comparison.

Table 3.6: Stylized exposure matrices between banks playing the same role. The arrows in the bottom left corners indicate the lending direction, e.g. the first matrix element in the left panel is the exposure of BO towards LO. The left matrix shows the naive exposure matrix, which after simplification is turned into a model of tiering structure.

LO	0	0	0	
IM	0	≥ 0	≥ 0	
BO	0	≥ 0	≥ 0	
$\nwarrow$	BO	IM	LO	

$\longrightarrow$

IM	> 0	> 0	
BO	≈ 0	> 0	
$\nwarrow$	LO	IM	

c-banks[18], strongly calling into question the fundamental assumption in the literature that random liquidity shocks form the basis for explaining interbank activity [27].

The main idea is that since core banks intermediate between the periphery banks, they must be a strict subset of the so-called *intermediaries*, i.e. banks that lend and borrow during a certain period. In general, given a time window, three bank *roles* can be distinguished: either banks only lend, only borrow, or do both; we will denote them by LO (lender only), BO (borrower only) and IM (intermediary). We emphasize that Craig and von Peter define interbank intermediation circularly as "that which IM banks do", i.e. any IM bank automatically intermediates between its LO, BO and IM counterparties, regardless of any maturity or risk transformation – indeed, regardless of any actual financial intermediation! While odd at first, this broad definition provides a working foundation that will finally enable us to to identify the core banks, of which most ultimately perform *real* financial intermediation, thus acting like money centers. In the following, then, we always use 'intermediation' to designate simply that which IM banks do, unless explicitly noted otherwise.

The division of banks into these three categories depends heavily on the term layer and the width of the time window (Figure C.4). In general, for any given time scale, the amount of IM banks varies continuously in size from dominant ($G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$) to insignificant ($G^{\texttt{1-3y}}$, $G^{\texttt{>3y}}$). Additionally, the amounts of LO and BO banks behave in the exact opposite way, and are roughly equal. For longer time scales, the amount of IM (LO and BO) banks grow (decline), as adding any edge to a random LO or BO bank is likely to 'flip it' into the IM role. In other words, weekly time scales are more appropriately to probe IM banks, while yearly time scales accurately identify LO and BO banks. Finally, we see that Figures 3.6 and C.4 point to the same conclusion, i.e. intermediation is more prominent in the shorter term layers.

The first step in exploring the core-periphery structure in the term layers (if any) is to investigate the aggregated exposures between banks that play the same role. The generic exposure matrix is displayed in the left panel of

[18]This can be probed by running regression tests with the dependent variable being core membership and economical explanatory variables such as bank size.

Table 3.6. We can immediately simplify the matrix using the definition of the bank roles – for example a BO bank cannot lend to any other bank, so the column above it must contain zero exposures, etc. The right panel of Table 3.6 contains the exposure matrix of a tiered interbank market, derived from the characterization of a perfectly tiered interbank market stated in Section 3.1 on page 53. We see that in this model the LO and BO banks are equivalent to p-banks, as they are mutually unconnected and only trade with IM banks. IM banks trade with other IM banks, as well as LO and BO banks. In general, only a relatively small subset of the IM banks actually function as cores, while the remaining IM banks are imperfect p-banks. By that we mean that they still trade with other p-banks, but heavily depend on the c-banks to both dissipate and acquire liquidity, according to their needs. Thus *core banks are special intermediaries that connect banks in the periphery* [27], which are expected to perform real financial intermediation between the periphery banks. Accordingly, observing a large amount of IM banks in a layer does not necessarily imply core-periphery structure – these two merely correlate [27] – but the exposure matrix sheds some light on the underlaying economical forces that shape tier structure and provides stringent constraints on possible c-banks, as they are required to be intermediaries: large exposure between LO and BO banks strongly suggests absence of core-periphery structure.

The exposure matrices for the term layers are given in Figure 3.8. Time aggregation is performed in two steps. First, a list of exposure matrices like the one in the right panel of Table 3.6 is generated for every week, month or year, and this is done for each layer. The exposure matrix holds for each cell the relative amount (in %) of exposure between a given pair of bank roles, normalized with respect to the total amount of exposures observed in a particular time window. Finally, the aforementioned list is averaged cell-wise across time for each layer separately. This means that the exposure percentages in a given matrix in Figure 3.8 need not sum up to 100%; indeed, this deviation is a measure of their volatility through time. Thus each cell must be interpreted as the typical portion of the total exposure two given bank roles take on. Figure 3.9 is constructed in identical fashion but with the weighted mean interest rate instead of the normalized exposure.

Figure 3.8 confirms our previous conclusion, which holds down to weekly time scales, that the `<1d` and `2-7d` layers exhibit pronounced interbank intermediation. When we combine Figures 3.8 and 3.9, we also find a trace of *financial* intermediation, if we adopt a simple model of liquidity dissipation that is displayed in Figure 3.7. Imagine an interbank market with fixed term `<1d` with only four banks, namely one LO, one BO and two identical IM banks. If the LO bank has excess liquidity, it can lend it to either a BO or an IM bank. According to the exposure matrix, the latter is chosen more frequently, even though the former offers a higher rate. This might be caused by a preferential lending relationship, shaped by search costs and lower perceived risk. The IM bank can then further dissipate a share of the liquidity at a profit, since the BO and the remaining IM bank both offer higher rates. This can be understood as a motive for financial intermediation. The same reasoning applies for an identical market that only trades `2-7d` loans. Of course, financial intermediation is more feasible here because

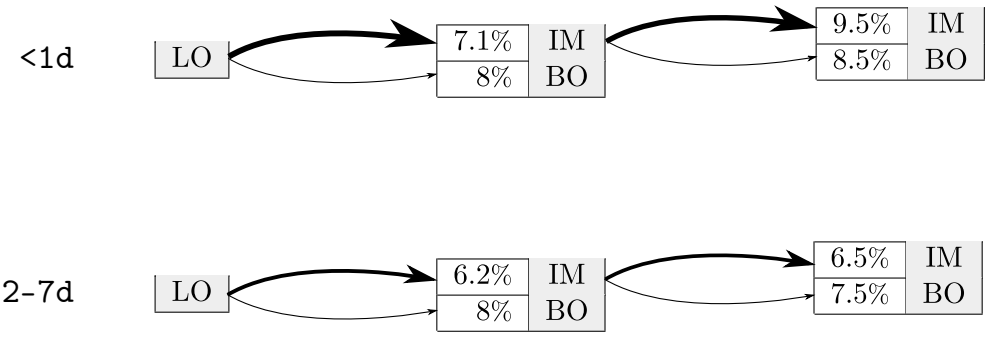


Figure 3.7: Stylized model of liquidity dissipation for $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$, taken from the weekly-averaged exposure and weighted interest rates tabulated in Figures 3.8 and 3.9. The arrows' thickness is proportional to the exposures, normalized by each column in the corresponding exposure matrix.

of the longer maturity involved.

The same pattern is found in the monthly aggregated `8-30d` layer, which is a more natural time scale for financial intermediation, but it breaks down for longer terms and time windows. Indeed, we find small or no tiering structure for $G^{\texttt{91-180d}}$, $G^{\texttt{.5-1y}}$, $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$, not even on yearly basis where the definition of LO and BO banks becomes very strict. The situation for $G^{\texttt{31-90d}}$ is rather unclear when inspecting the yearly time scale: on one hand, there is some tiering structure (Figure 3.8), but the incentive of financial intermediation appears to be lacking (Figure 3.9).

To conclude, our liquidity dissipation model tells us to expect a core-periphery structure in $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$ and $G^{\texttt{8-30d}}$; perhaps in $G^{\texttt{31-90d}}$; but not in $G^{\texttt{91-180d}}$, $G^{\texttt{.5-1y}}$, $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$. We acknowledge the simplicity of the model: it does not allow for maturity transformation across layers, and only incorporates liquidity transfers starting at LO banks. The question is now how to separate the IM banks into c- and p-banks in a more robust fashion.

Core-periphery and source-sink structure

Tiering inherently is a network concept, because it is based on the bilateral relations between the core and periphery, which are represented by the interbank network's topology (see Section 1.1). Craig and von Peter introduce a blockmodel that encodes the four rules on page 53. The optimal fit of the blockmodel to the network induces a partition of the set of banks into two blocks, i.e. the p- and c-nodes. Della Rossa et al. (2013) point out that their block model cannot incorporate edge weights and that determining which fit is 'optimal' relies on an ambiguous notion of distance[19] between a perfectly tiered structure and a proposed partitioned network [1]. They propose a heuristic method that ranks each bank i by assigning it a 'coreness' value $\alpha(i) \in [0, 1]$. The ordered set $\{\alpha(i)\}$ is then called the *core-periphery*

[19] We remark that their distance measure is rather forgiving because it is normalized by the amount of possible wirings, while interbank networks are known to be sparse.

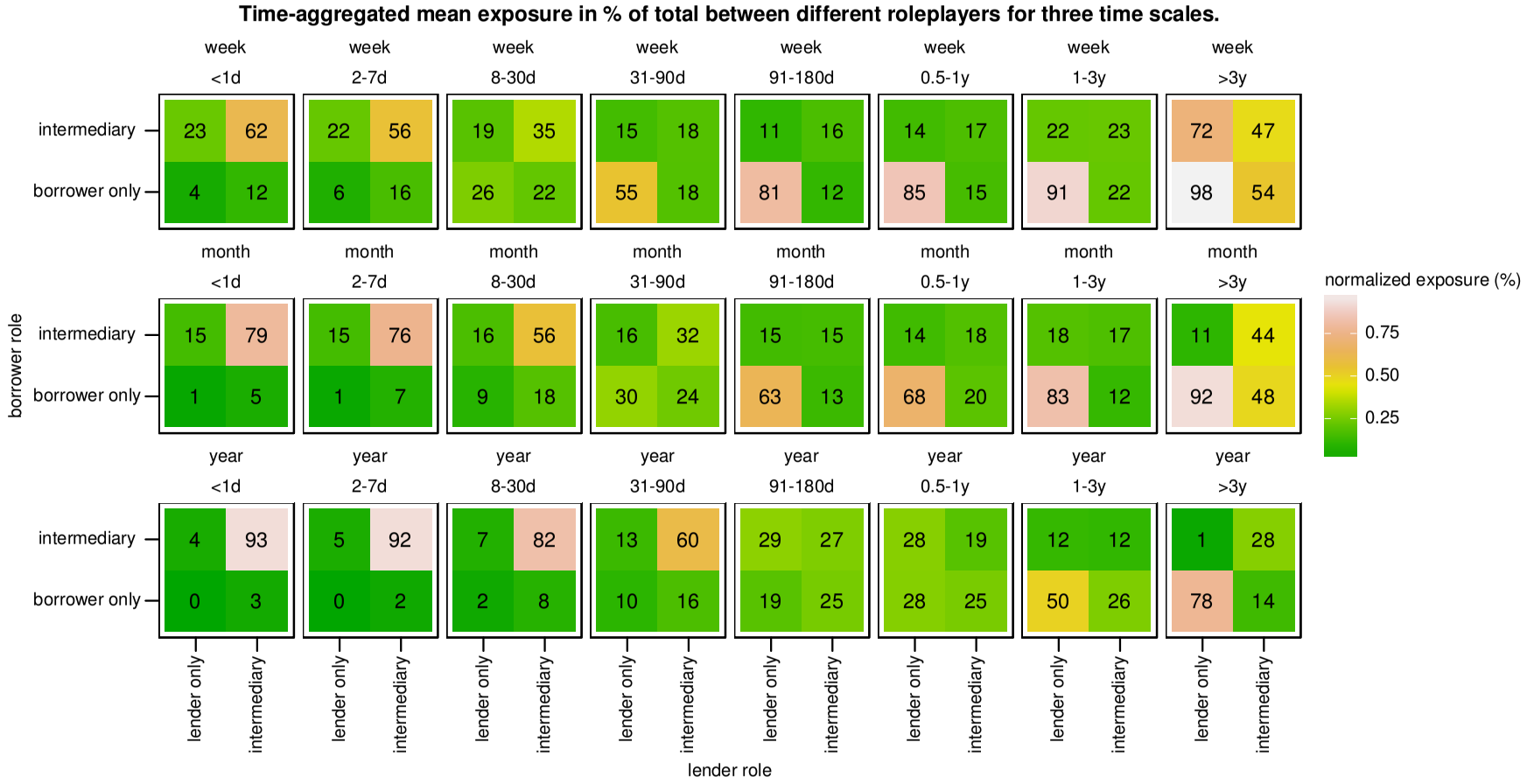


Figure 3.8: (realization discussed in text) Normalized exposure matrices for each layer, averaged through time and with weekly, monthly and yearly time windows.

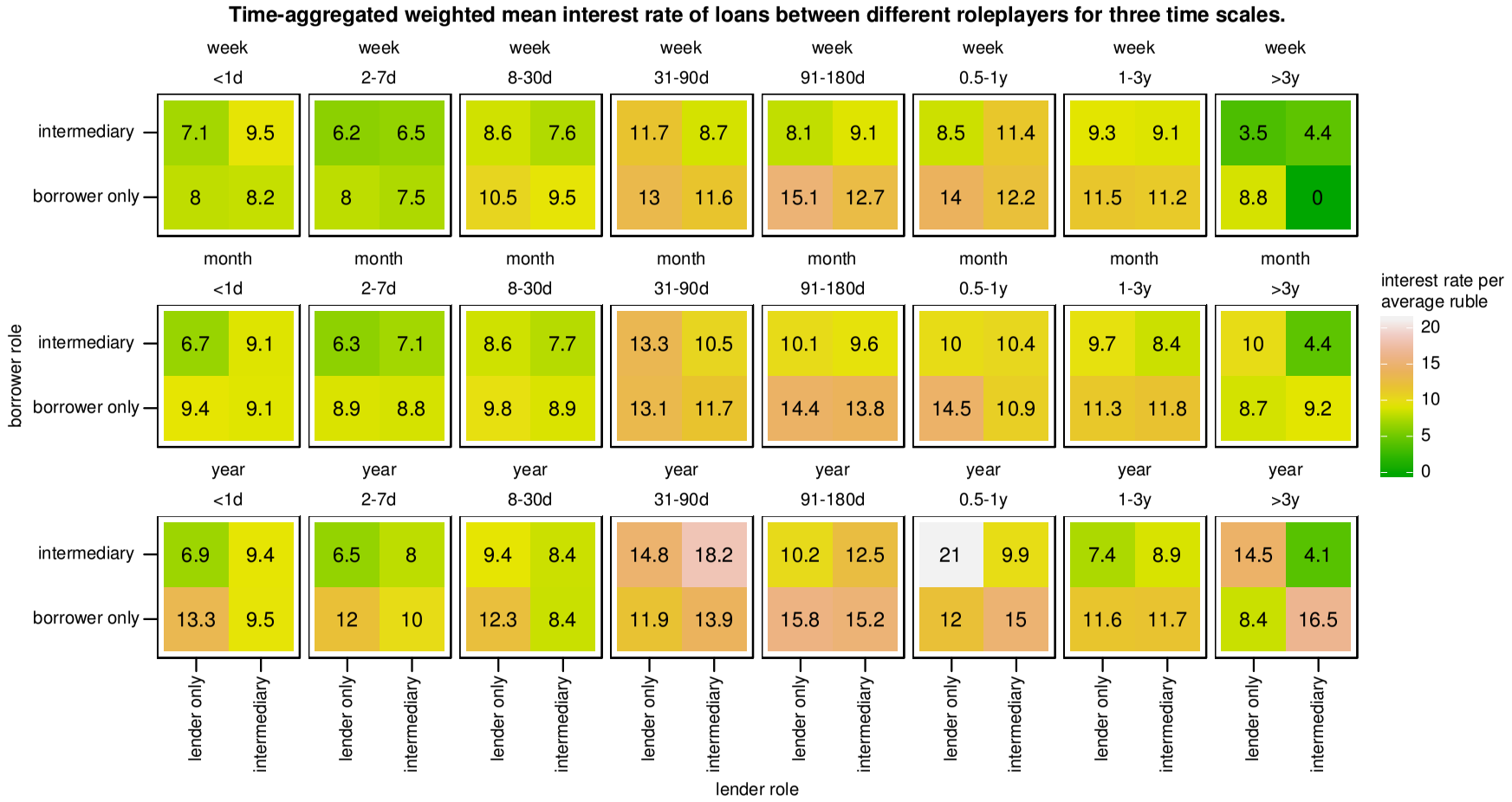


Figure 3.9: (realization discussed in text) The interest rates are weighted by the loan volumes, which corresponds to the interest rate per average ruble (see Section 2.6).

profile of the network. From this profile the *core-periphery centralization* of the network can be calculated, which is denoted C and abbreviated to *cp*-centralization. The network exhibits core-periphery structure if $C \in [0, 1]$ is large with respect to typical C-values for random networks of the same size. In that case, the α-ranking corresponds to a continuous characterization of the banks' coreness, with p-banks having small α and c-banks having α close to unity. While this ranking provides useful information about the importance of a bank with respect to its intermediation, one needs to manually define a cutoff α_c that defines the c-banks as having $\alpha(i) > \alpha_c$; consequently p-banks have $\alpha(i) \leq \alpha_c$. Particular choices of α_c can be motivated in several ways because one expects the coreness ranking to correlate with node centrality and measures of systemic importance; for this reason we accept the mandatory cutoff in return for the valuable coreness ranking.

However, the algorithm of Della Rossa et al. is unfit for our interbank network at hand for two reasons. First, it requires the network to be *strongly connected*. This means that every node must be reachable from any other by a random walker that traverses the network by hopping from node to node while respecting the connecting edges' directionality. Their method has been applied to interbank networks [34] by restricting the network to its largest strongly connected component. Unfortunately, the economic meaning of this restriction is unsound, especially for large networks which may have unconnected components. The layers in the Russian interbank network have a varying amount of unconnected components (see Figure C.8) and the relative size of the largest strongly connected component is definitely not large for individual term layers (Figure C.7).

Second, the coreness of a bank in [1] is based on the amount of liquidity it intermediates, in harmony with the tiering concepts of [27] described above. Indeed, a strongly connected interbank network implies that any dissipated liquidity is capable of flowing into all BO and IM banks. However, as stated before, when the network's *cp*-centralization C indicates absent core-periphery structure the α-ranking becomes irrelevant [1, p. 3]. As we anticipate that the longer term layers lack core-periphery structure, this poses a real problem since we cannot compare the α-rankings across layers.

To remedy these problems, we have modified the core-periphery profile algorithm of Della Rossa *et al.* In a nutshell, (1) the requirement of strong connectedness is abolished by using the *PageRank* (PR) algorithm [97] instead of a standard directed random walker: the network need not be strongly or weakly[20] connected; (2) the ranking is made relevant even for networks with insignificant *cp*-centralization. For these networks, banks with low rank play only a small part in the redistribution of liquidity, while highly-ranked banks *either*

(a) lend substantive amounts that disperse throughout the network: they act as a liquidity *sources*. Such banks receive higher ranks when the dissipation of their excess liquidity involves significant intermediation, compared to lending directly to BO banks or IM banks that intermediate only small amounts of the received loan.

[20] A graph is weakly connected when the underlying undirected graph has no unconnected components, i.e. a random walker can reach every node in the undirected graph.

(b) borrow substantive amounts that have been dispersed throughout the network, i.e. they are the *sinks* that receive the sources' liquidity. Analogously, the sink banks end up at the top of the ranking if they utilize liquidity that has passed through many IM banks, compared to sink banks that directly borrow from LO banks of only slightly active IM banks.

We acknowledge that the sink banks might seem less important at first sight, yet they are of equal importance as the source banks, being paying consumers of the sources' excess liquidity. Indeed, the interest paid by large sink banks generates cash flows in the opposite direction of the loan edges; the repaid loan and interest are in turn available for new loans.

The modified algorithm works in the following way. Imagine a network of banks that lend and borrow from each other, with each loan between banks i and j with volume v and interest rate r being represented by *two* edges that track the cash flow: one for the loan itself ($v_{ij} = v$), and another in the reversed direction that tracks the repayment ($v_{ji} = v(1+r)$). Now we mark one ruble by coloring it red, and we ask each bank to tally each time it sees the red ruble pass. The red ruble is randomly used by the bank who holds it for new loans or repayments. Furthermore, there exists a 15% chance, called the *damping factor*, that the red ruble is randomly swapped with another unmarked ruble belonging to a bank *that is unconnected* with the bank where it currently resides. This might be thought of as an interaction excluded by the preferential lending hypothesis outside the lending market.

This picture has several desirable properties. The red ruble is more likely to end up in banks that lend or borrow large volumes and that are connected to many banks, especially if they are intermediaries. Most importantly, every bank is reachable by the marked ruble: it cannot get stuck in LO or BO banks because of the repayment flows and it can hop between unconnected clusters because of the damping factor.

After a sufficiently long time T, we gather the tally of each bank and normalize it by dividing by the sum of all tallies. The result for each individual bank is then equal to its PageRank. The PR of a bank i, $\mathrm{PR}(i) \in [0, 1]$, measures the probability that the red ruble is in i after waiting T time units[21]. In the network, banks with high PageRanks must play important roles with respect to creating, absorbing and intermediating liquidity. This precisely why it is of interest to us.

The next step consists of obtaining the α-ranking. In the original algorithm in Della Rossa et al., banks that intermediate the most receive high α, which justifies calling α the coreness value in case the network has core-periphery structure. In our version, α also identifies periphery (small α) and 'core' (high α) nodes, but, crucially, *without requiring the 'cores' to be IM banks*. Therefore we will call them *hub banks*, who can play LO, BO and IM roles in the network. Thus hub banks can be cores, sinks or sources in the modified algorithm[22]. Additionally, we will use a simplified competition-like ranking scheme by giving the most prominent hub ($\alpha = 1$) rank 1, the

[21] For completeness, $\mathrm{PR}(i)$ is also the stationary probability distribution of a random directed walk on the weighted network with a 'teleportation' probability equal to the damping factor.

[22] This is achieved by applying the α-ranking scheme of the original algorithm to the

Table 3.7: Recapitulation of the used terminology in the context of core-periphery and source-sink structure. IM banks can be sinks, cores and sources. The intermediation is defined as the minimum of the instrength and outstrength [27].

bank role		instrength	outstrength	intermediation
BO		> 0	0	0
IM	sink	$\gg 0$	≈ 0	≈ 0
	core	$\gg 0$	$\gg 0$	$\gg 0$
	source	≈ 0	$\gg 0$	≈ 0
LO		0	> 0	0

second prominent central hub ($\alpha \lessapprox 1$) rank 2, etc. The periphery nodes ($\alpha = 0$) all share the last rank. In the following, banks referred to as highly ranked have a ranks close to the top bank, i.e. rank 1, and vice versa for lowly ranked banks. We will refer to this ranking as the *hub ranking.*

To summarize, we have modified the core-periphery profile to obtain a hub ranking of banks that is expected to work in interbank networks regardless of their posited core-periphery structure. More precisely, we aim to characterize 'important' banks by their high rank in order to compare the leading banks across the layers and time. We have achieved this by including source and sink banks in the ranking. If the top banks are IM banks, the network exhibits core-periphery structure. Conversely, if the top banks are source or sink banks, we say the network has *source-sink structure.*

The nature of the interbank network's structure can be determined from the *intermediation profile.* Figure 3.10 shows a typical source-sink and core-periphery network, together with the intermediation profiles. To continue, we define the *instrength* (*outstrength*) of a node as the total of received (issued) money in an aggregate period [8]. Note that the strengths reduce to the degrees for unweighted networks. Then the *intermediation* of a bank is defined as the minimum of the in- and outstrength [27]. Table 3.7 shows the relation between the bank's strength, intermediation and its role. It is then straightforward to determine the structure of the network: the network has a core-periphery structure if the highly ranked banks provide a substantial amount of intermediation. If this is not the case, we simply speak of a source-sink network. This can be easily inferred from the intermediation profile.

Returning to Figure 3.10, the intermediation profiles mirror the topological differences between the source-sink and the core-periphery structure. For the former, about 75% of the intermediation is conducted by banks ranked 30th or lower. In fact, only two banks (rank 30 and 33) really contribute; both of them are in the periphery. The top bank (rank 1) is the main source

network discussed above, with some minor modifications. We refer to the 'Method' section in [1, p. 6] and only discuss the results here. The key to decoupling core banks from IM banks is to include the repayment cash flows. This holistic approach to finding hubs or 'important' banks takes into account topology, loan volume *and* interest rates.

of the network, supplying many banks in its periphery. Thus we conclude that $G^{\texttt{1-3y}}$ during 2004 saw no tiering activity. This is in contrast to $G^{\texttt{8-30d}}$, where the top 3 banks generate roughly 50% of the total intermediation, clearly behaving as money centers. The largest part of the remaining intermediation comes from a single p-bank with rank 18 that is connected with the top bank. Note that the top 3 banks are state-controlled banks, as opposed to the top banks in $G^{\texttt{1-3y}}$, which are mainly foreign banks (see Table 3.2).

The objective now is to see which layers exhibit core-periphery structure and whether the most prominent hubs persist through term layers and time. Figure 3.11 shows the *persistency* of the top 100 hubs. In general, call $T_n(t)$ the set of the top n hubs (i.e. the banks with ranks $1, 2, \ldots, n$) at time t. Given an appropriate time window, the persistency $p(t)$ is then defined as

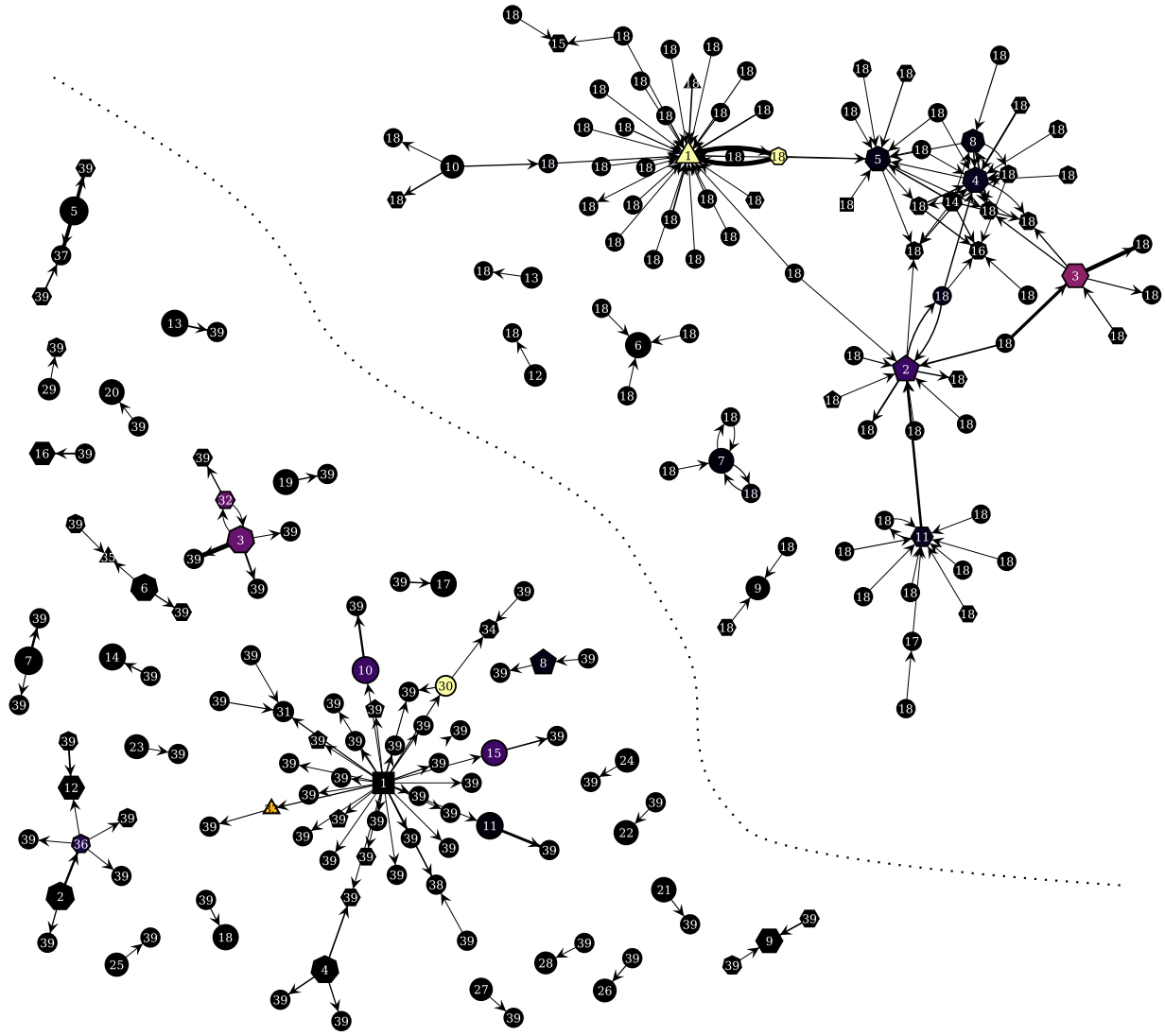
$$p_n(t) = \frac{|T_n(t-1) \bigcap T_n(t)|}{(|T_n(t-1)| + |T_n(t)|)/2}, \tag{3.6}$$

with $|\cdot|$ denoting the cardinality of a set. Ranks may be degenerate[23], but the denominator in (3.6) will almost always equal n. Figure 3.11 reveals a relatively high persistency, which grows as the network matures, especially for longer time windows; this implies that the hub ranking is well defined, which seems to hold quite well even on weekly time scales. The persistence is highest within monthly time, supporting the view that monthly time windows are natural time scales to describe the network (see Section 2.3). Finally, we report that $p_{50}(t)$ and $p_{10}(t)$ follow roughly the same pattern as $p_{100}(t)$ and are even more persistent, their average monthly persistency being roughly 80%. Naturally, increasing n has a stabilizing effect on the volatility of $p_n(t)$.

Looking at the persistency by layer for $n = 10$, a more nuanced image appears in Figure 3.12. As usual, the `<1d` and `2-7d` terms drive the result for the aggregated network, since they account for 90% of the loans. The persistency of the top 10 hub banks is seen to decrease for longer terms. This indicates more volatile, less consolidated market segments, in contrast with $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$ and $G^{\texttt{8-30d}}$, which seem to be dominated by a small set of banks. In their case, the consistently high monthly persistence indicates that banks join and leave the top hubs in a rather slow and controlled fashion. Interestingly, for $G^{\texttt{31-90d}}$, $G^{\texttt{91-180d}}$ and $G^{\texttt{0.5-1y}}$ the monthly persistency is smaller (on average) than its yearly counterpart, which points to a more diverse set of hub banks that compete monthly, yet not devoid of regularity on a longer term. Finally, the greater monthly persistency is restored again for $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$, which have the smallest hub persistency on yearly basis.

We now ask whether the same hubs operate across different layers, and if they are intermediating cores or rather sinks or sources. If financial intermediation is really being done by core hubs, we expect to see some evidence for maturity transformation by finding similar hubs across term layers. The hub similarity is displayed in Figure 3.13 and the core-periphery structure can be deduced from Figure 3.14. While we could in principle study the intermediation profile for each layer and time window separately, in practice

[23]Banks with equal α are assigned the same rank. This is also called dense ranking.



Intermediation profiles of a source–sink (left) and a core–periphery network (right).

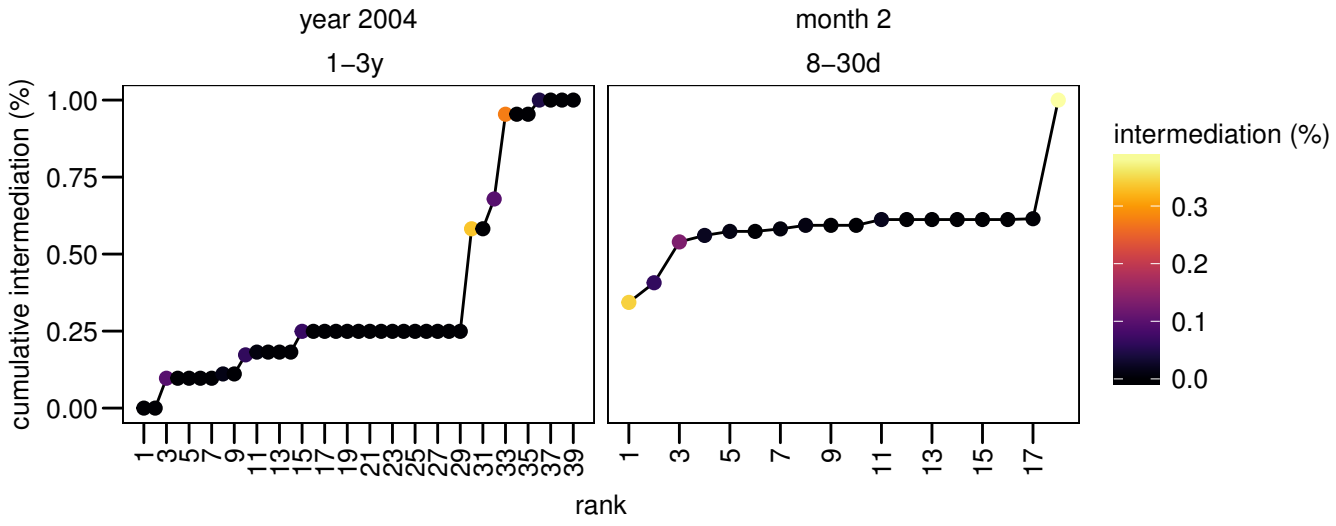


Figure 3.10: Two example intermediation profiles together with the directed view of the underlying networks. The label and size (color) of the nodes indicates the rank (intermediation) and the edges' width is proportional to their loan volumes.

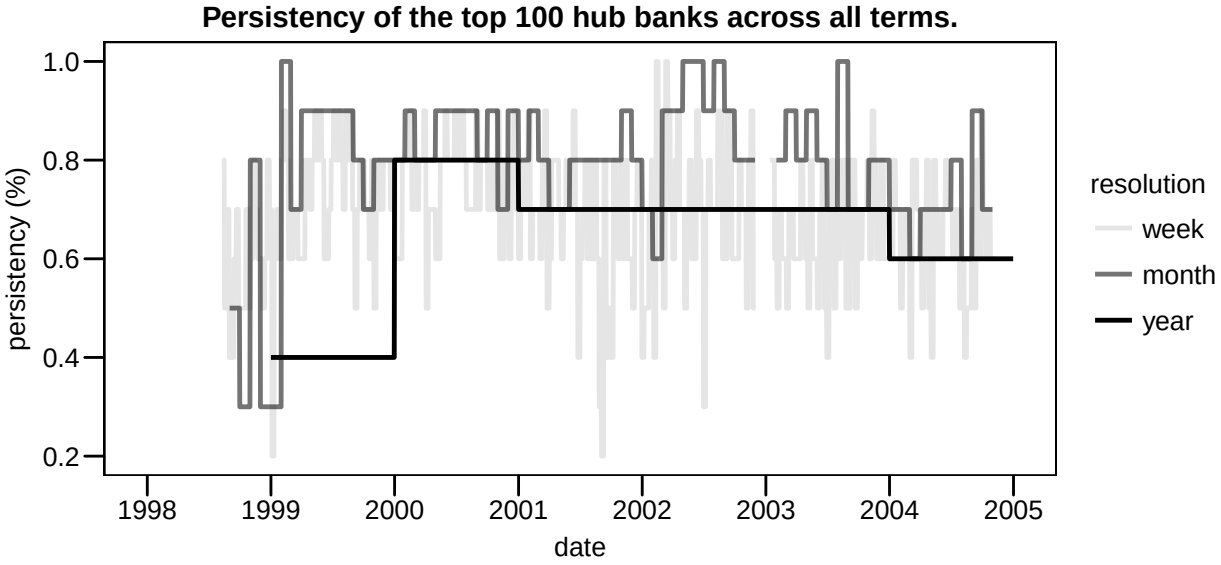


Figure 3.11: The yearly persistency for the 1998 is omitted because of incomplete data.

this is feasible only for the yearly time windows (indeed – see Figure C.6). Therefore we take a more heuristic approach: the quantity measured per layer is the amount of top-ranked hubs needed to account for at least 50% of the total intermediation during the period spanned by the time window, divided by the last rank. In symbols, if the rank of a particular bank i is $r(i)$, then $k(i) = r(i)/\max_i r(i)$ denotes its relative rank. Furthermore, let the intermediation by a *bank* i be $\mathrm{im}(i) = \min(\mathrm{instrength}(i), \mathrm{outstrength}(i))$. Then the total intermediation associated with the relative *rank* k' is given by

$$\mathrm{im}(k') = \sum_{i:\ k(i)=k'} \mathrm{im}(i). \tag{3.7}$$

We then formally define the (normalized) cumulative intermediation

$$I(k) = \frac{\sum_{k' \le k} \mathrm{im}(k')}{\sum_{k'} \mathrm{im}(k')}. \tag{3.8}$$

Thus the quantity we measure for each layer and for the monthly and yearly time windows is $k_{0.50} = \mathrm{argmin}_{k'}\ I(k') \ge 0.50$. If $k_{0.50}$ is *consistently* small throughout time, then half of all intermediation is located in only a few hubs, which implies a core-periphery structure. Of course, one could devise more sophisticated measures of core-periphery structure using the same information. What is convenient about our approach is that $k_{0.50}$ can be read directly from the intermediation profile: for example, in Figure 3.10 we have $k_{0.50} = 30/39 \approx 0.77$ and $k_{0.50} = 3/18 \approx 0.17$ for the source-sink and core-periphery network, respectively.

Returning to Figures 3.13 and 3.14, one sees immediately that the first three term layers ($G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$ and $G^{\texttt{8-30d}}$) and the aggregated network –

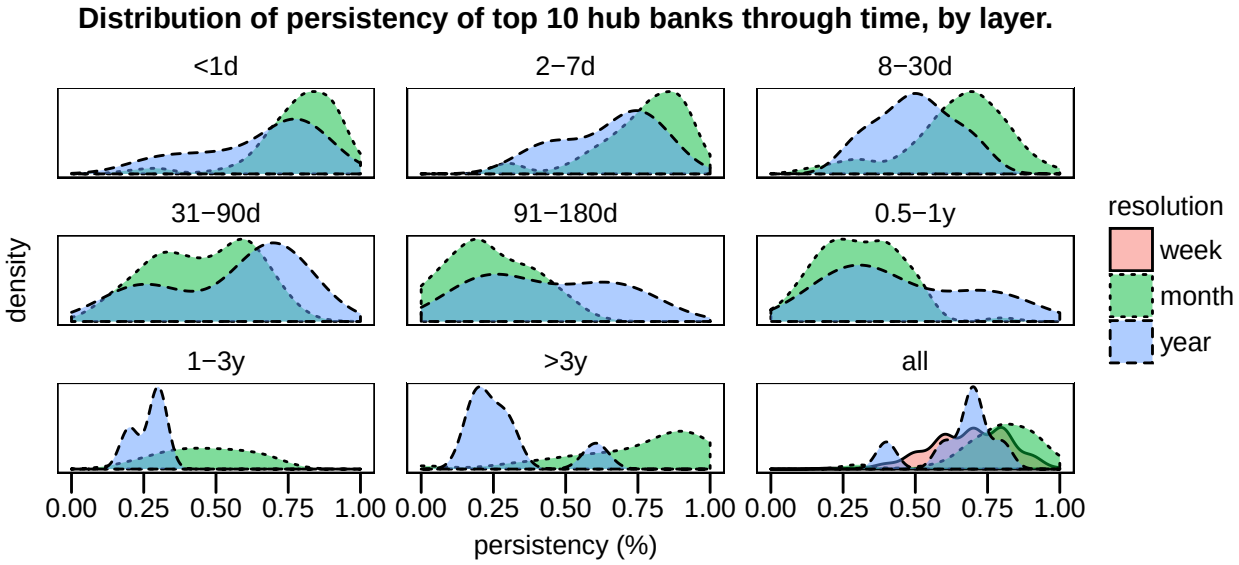


Figure 3.12: (smoothing effect for clarity) Weekly resolution is only available for aggregated terms because of the limited computing time available. $p_{50}(t)$ and $p_{100}(t)$ exhibit roughly the same peaks, though considerably more localized. See Figure C.5 for the time series underlying these distributions.

which allows for maturity transformation – exhibit a strongly pronounced core-periphery structure, that is disturbed only during the first crisis. Furthermore, the top 10 cores in these layers overlap strongly, especially for $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$, and have a high persistency, both by monthly and yearly standards. The top 10 cores in the aggregated network are nearly identical to the core banks in the $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$, and gradually make way for other hubs as the terms lengthen. The top row for the yearly time scale in Figure 3.13 reveals that the overall core banks specialize in short-term liquidity intermediation, hardly playing any role of significance in the long-end of the term spectrum. We also report that the indegree strength and high α-ranking correlate significantly more than the outdegree strength, which points to the fact that core banks are characterized by extensive intermediation and even larger borrowing. Langfield et al. (2014) explain this as follows: "[...] core banks tend to absorb funding from peripheral banks. In other words, core banks are net borrowers in the funding network – in part because all market-makers have a latent demand for liquidity, which must therefore be provided by non-market-makers" [87, p. 32]. In addition, the core banks in the aggregated network exhibit high persistence on weekly, monthly and yearly time scales (Figure 3.12). Finally, we mention that the top 10 core banks in the aggregated network is typically composed of 3 state-owned banks, 1 foreign bank and 6 domestic banks.

The core-periphery structure of the next three layers, $G^{\texttt{31-90d}}$, $G^{\texttt{91-180d}}$ and $G^{\texttt{0.5-1y}}$, is much less pronounced, and only (partially) present on a yearly time scale. Indeed, for these maturities intermediation activity cannot be captured within a monthly time window. Because the persistency of the

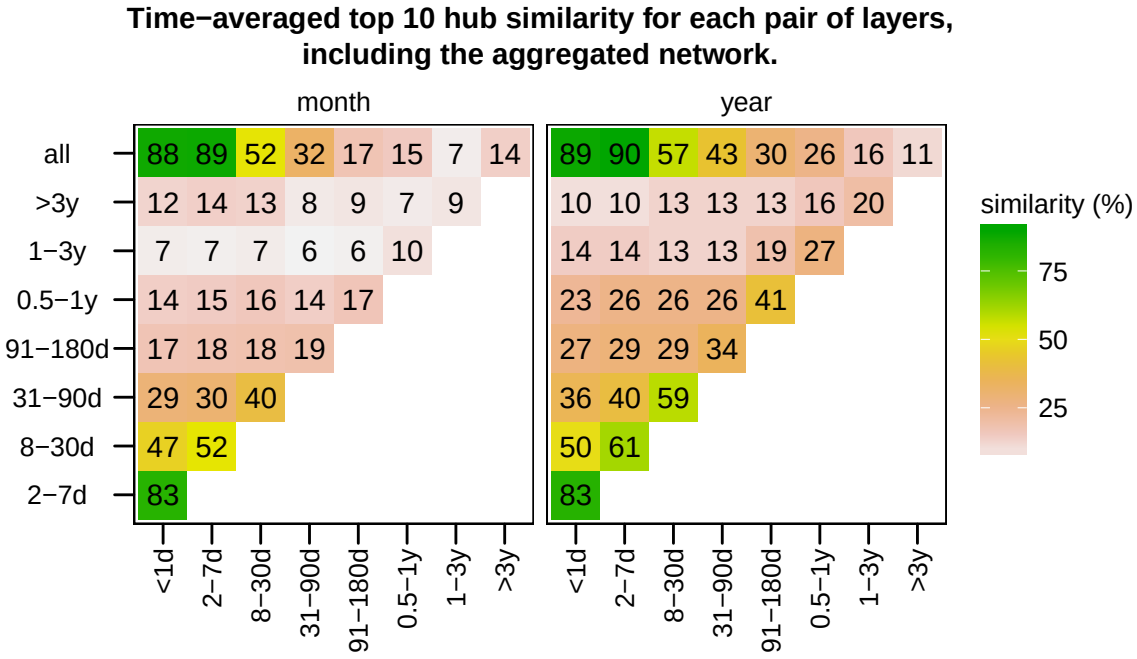


Figure 3.13: The similarity between two lists of the top 10 hubs is defined as the number of common hubs, divided by 10. The similarity matrix is averaged over time for each cell separately. Typical standard deviations are of the order of 5%. The aggregated network is indicated by 'all' (from 'all terms'). Note that for the eight term layers the diagonal similarity is almost always the largest value in its row and column, and that the rows (columns) tend to increase (decrease) from left to right (bottom to top). This implies that, given that a bank i is a hub in layer α, the more 'closeby' another layer β is, the more likely that i is a hub bank again. Consequently, though several banks are present on each term layer (the top row), most of the sink and source hubs in the long end of the term spectrum operate specifically in 'nearby' layers, pointing to some amount of specialization.

top 10 hubs is larger (on average) on a yearly basis, we are inclined to conclude that the persistency is only significant when the hubs are, in fact, intermediating cores. On monthly time scales the network in these layers tend to alternate rapidly between source-sink and core-periphery structure, with typically half ($G^{\texttt{31-90d}}$) to one-fourth ($G^{\texttt{91-180d}}$, $G^{\texttt{0.5-1y}}$) of the banks remaining in the top hub ranking. Inspecting the intermediation profiles in Figure C.6, it is clear that only $G^{\texttt{31-90d}}$ exhibits almost impeccable core-periphery structure, while large parts of the total intermediation are still concentrated in the periphery banks in $G^{\texttt{91-180d}}$ and $G^{\texttt{0.5-1y}}$. Furthermore, $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ have predominantly source-sink structures, thus we conclude that *only the four first term layers exhibit true core-periphery structure*. This agrees with our liquidity dissipation model, thus indicating tiering structure, i.e. real financial intermediation by the core banks. Therefore from now

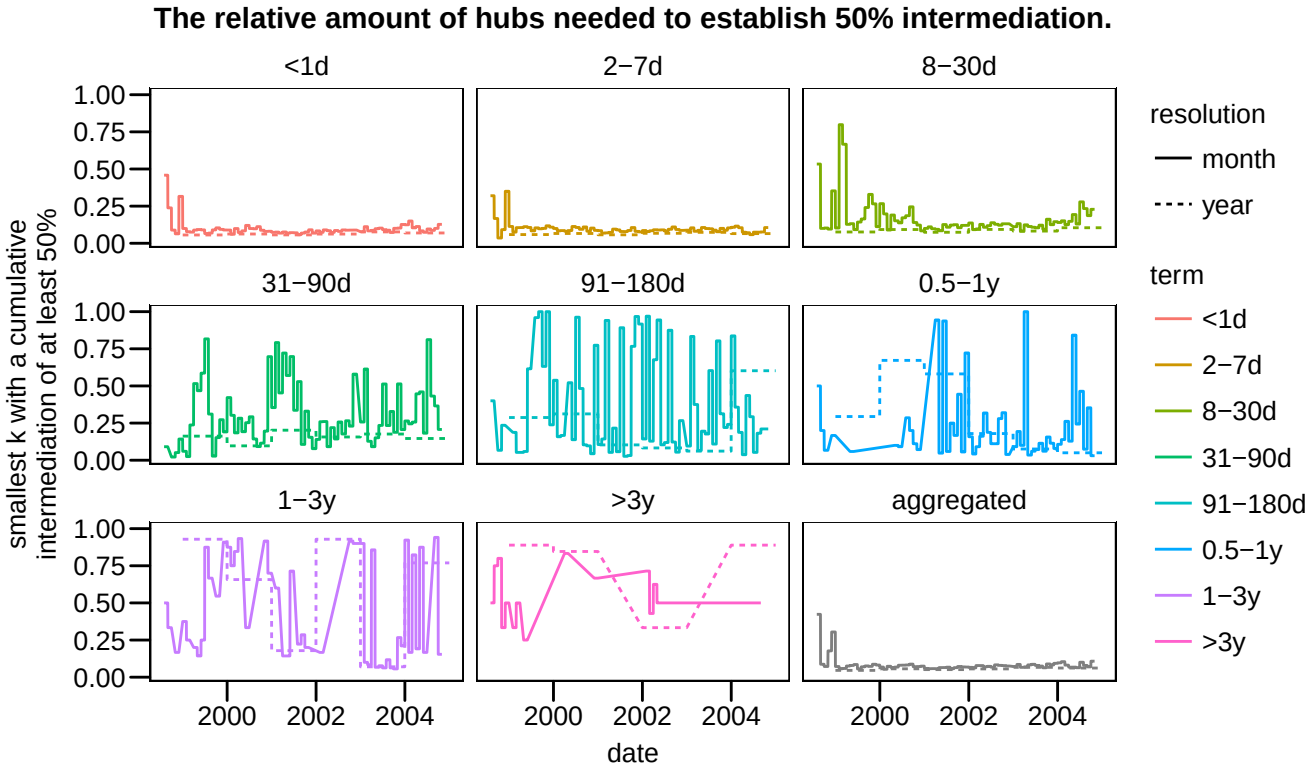


Figure 3.14: Time series of $k_{0.50}$ for the term layers and monthly and yearly time windows. $k_{0.50}$ for 1998 has been omitted due to incomplete data.

on we will refer to $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$, $G^{\texttt{8-30d}}$ and $G^{\texttt{31-90d}}$ as the *tiering layers*. Furthermore, to a good extent the top 10 cores are identical and persistent through time in these layers. The same is true for the aggregated network, which is almost completely determined by these four layers, which account for 99% of all issued loans.

The core banks do not translate entirely to the hubs in the remaining layers, although a small amount of them seem to be active in each layer, acting as intermediaries in some and sources or sinks in others. It seems that the layers in the long end of the term spectrum do not have a small and persistent set of dominating hub banks; the number of active banks in this market segment may be too small for hub-like structure to emerge.

Clustering coefficients

The *local* clustering coefficient [26] of a node i is defined in function of triangle motifs in undirected networks:

$$C_i = \frac{\text{number of of triangles connected to } i}{\text{number of triples on } i}, \tag{3.9}$$

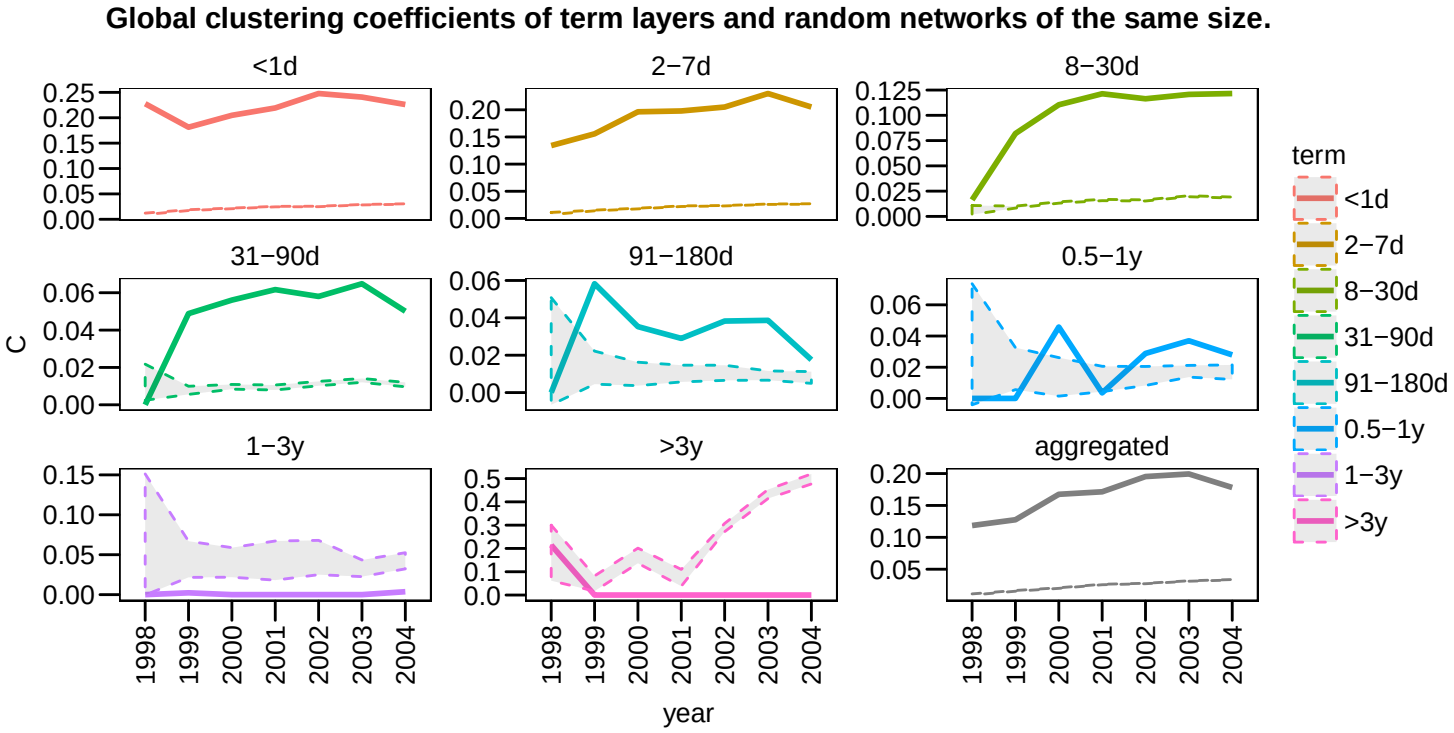


Figure 3.15: Time series of clustering coefficients for the term layers and the aggregated network, together with a ribbon centered around $\langle C_{\text{rand}} \rangle$ of total width $2 \times \text{sd}(C_{\text{rand}})$. Following the methodology in [35], a total of 100 Erdős–Rényi networks were generated to test the significance of the observed C's.

where a triple on i means an unordered pair of nodes connected via i, and possibly connected directly. If that is the case, the triple is counted as a triangle, ending up in the nominator of (3.9) as well. Thus C_i expresses the degree of connectedness among the neighbors of i [8]. The (*global*) clustering coefficient of an undirected (N, E) network is then simply the average of the local coefficients:

$$C = \frac{1}{N} \sum_i C_i. \tag{3.10}$$

To assess whether an observed network possesses a non-random clustering structure and thus a significant clustering coefficient C_0, we generate n random networks of the same size, i.e. the same amount of nodes and edges, having clustering coefficients $C_{\text{rand}} = \{C_1, C_2, \ldots, C_n\}$. Then we can calculate the z-score of the observed C_0 as

$$z = \frac{C_0 - \langle C_{\text{rand}} \rangle}{\text{sd}(C_{\text{rand}})}. \tag{3.11}$$

Large z indicates a significant value of C_0, the sign indicating more (+) or less (−) clustering than a random network of the same size.

Because of bad statistics for monthly time scales, we have calculated clustering coefficients and z-scores for the undirected view of the term layers using yearly time windows which are displayed in Figure 3.15. All results are highly significant, except for $G^{\texttt{91-180d}}$ and $G^{\texttt{0.5-1y}}$: these layers do not possess any structural clustering structure. As always, the clustering in the aggregated network resembles mainly the first two layers.

A clear pattern emerges from Figure 3.15: the clustering varies from significantly high (the tiering layers), via moderately low and insignificant ($G^{\texttt{91-180d}}$ and $G^{\texttt{0.5-1y}}$), to significantly low ($G^{\texttt{91-180d}}$ and $G^{\texttt{0.5-1y}}$). Since all layers exhibit heavy-tailed degree distributions, the sum in Equation (3.10) is dominated by the (local) clustering coefficients of banks with small degree. Thus we use C to proxy the clustering of the periphery banks, i.e. banks at the bottom of the α-ranking, on a yearly time scale.

Within that approximation, one sees that the core-periphery structure of the tiering layers deviates considerably from the idealized star network, at least when looked at with yearly resolution. Notwithstanding the strong tiering mechanism, the lower-tier banks still trade extensively with each other, indicating that contagion risk is not located entirely in the high-tier banks. Additionally, this motivates the use of the α-ranking algorithm to find the periphery nodes, since alternative simpler degree-based approaches would likely underestimate the size of the periphery.

In contrast, the last two layers exhibit considerable star-like structure, low degree nodes and periphery banks being equivalent in most cases. An example of this can be seen in the left panel of Figure 3.10. In addition to the low clustering, we recall the more antireciprocal nature of their lending relationships and the modest anticorrelation between in- and outdegree; these observations lead us to refer to $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ as the *source-sink star layers*. These layers have starlike hub structure, but hardly any intermediation within yearly time windows occurs in the hubs. They act simply as sources and sinks, generating and dissipating excess liquidity in the interbank network. Of course, one could argue that for loans with maturities of at least one year, no intermediation is possible within the scope of one year. Looking at Figure 3.6, which completely aggregates time, however, we see that the banks with the largest degrees tend to be *either* sinks or sources, especially in $G^{\texttt{>3y}}$. Thus the tendency of the hubs to be either sink or source, but not an intermediary, holds for time scales longer than one year. The next question is whether the star centers in the source-sink star layers are connected, or rather at the center of disconnected components. These are plentiful compared to the number of active banks per year (see Figure C.8 and Table 3.4 on page 61), and less than half of the latter participate in the largest (weakly) connected component, as Figure C.7 shows. Furthermore, we report[24] that in the two source-sink star layers the average shortest path length is between 2 and 3, which, together with the proven existence of hubs, indicates a compact starlike structure. In a nutshell, the source-sink star layers $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ are composed in general of many disconnected 'islands', of which the largest exhibit an almost perfect starlike structure around a sink or source hub.

[24] Figure 3.16 plots the average path length for $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ on a *monthly* basis, which is typically 2.

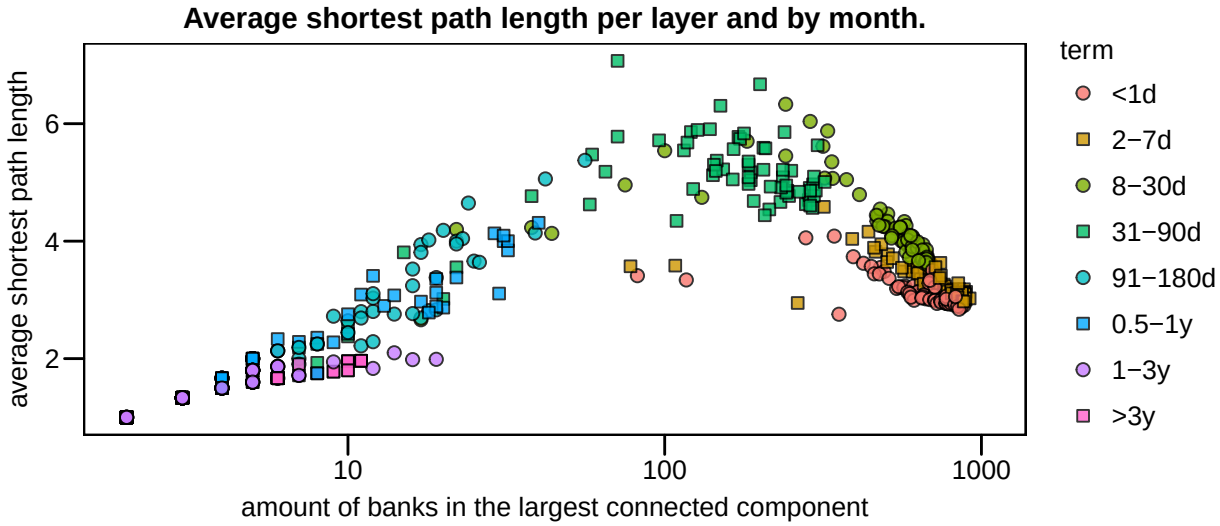


Figure 3.16: The average (shortest) path in function of the size of the largest (weakly) weakly connected component. For the four tier layers, the points lying astray from the dense cloud all occur during the first five months. Note that the amounts of banks in the largest connected component for the tier layers is notably larger than those of the remaining layers, which have source-sink structure. Due to limited computing time, we were unable to compute z-scores (3.11) for the data in Figure 3.16, thus we will focus rather on the rate of change of D in the text.

Average shortest path length

Figure 3.16 shows that exclusively for the tier layers the average path length D *decreases* roughly linearly when the size of the largest connected component[25] increases *logarithmically*. Thus we see that core-periphery structure is extremely effective in channeling and intermediating liquidity; of course this comes at the cost of systemic risk in the core hubs. We cannot formally state that the tier layers are small worlds[26], though the significantly high global clustering combined with the small D for large connected components support this claim[27].

The remaining layers display the familiar pattern: the size of the largest connected component shrinks for longer terms, while simultaneously the network becomes increasingly fragmented as the number of unconnected components increase, relative to the amount of active banks (Figure C.8) – both tend to suppress D.

[25] Which covers the vast majority of the active banks within yearly time windows, see Figure C.7.

[26] In fact this is only feasible for network models, as one needs to investigate the growing behavior of D for increasing network size while keeping the average degree fixed [8, 26].

[27] We note that in [38] the small world property was empirically rejected for the complete Russian interbank dataset, i.e. with deposits and other contract types.

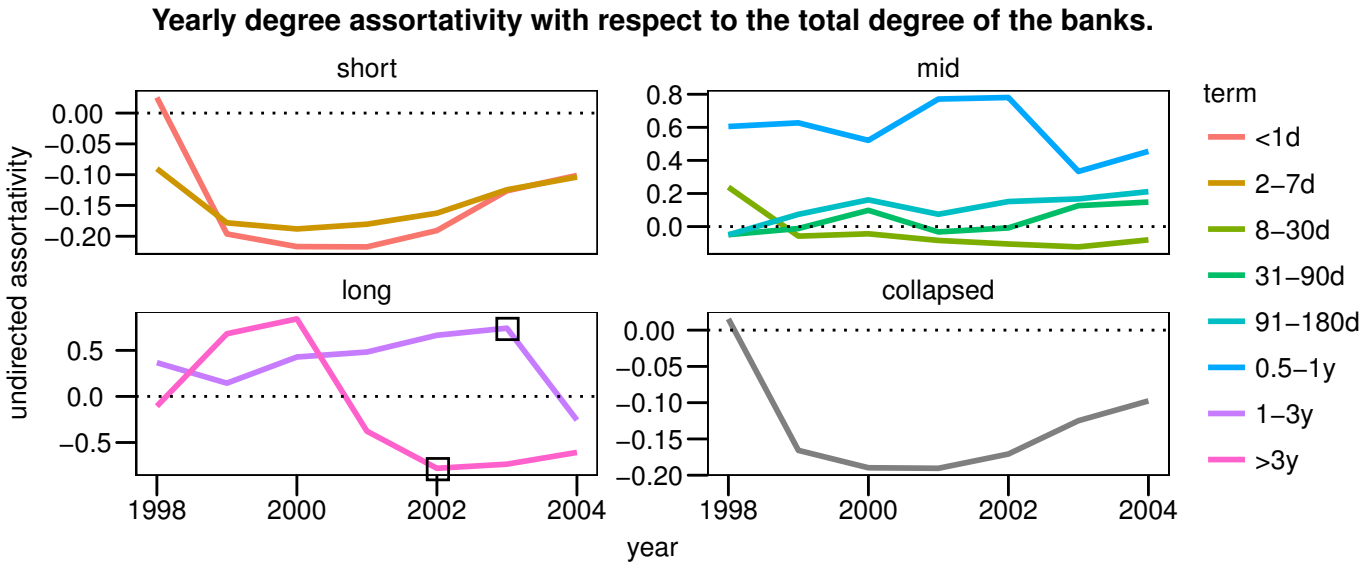


Figure 3.17: The degree assortativity measures to which extent nodes with a given degree associate preferentially with other nodes of similar degree – see Equation (2) in [98] for the formal definition. In this case the type of degree considered is the total degree, i.e. the sum of the in- and outdegree. A yearly resolution was chosen because monthly aggregation suffered from bad statistics starting from $G^{\texttt{91-180d}}$. The underlying networks of the boxed data points in the lower left panel are shown in Figure 3.18.

Total degree correlation

Figure 3.17 shows the assortativity with respect to the total node degrees on yearly basis, which is also called the total degree correlation. The short term layers show a clear preference for the low-degree nodes to attach to the high-degree ones, which is indicative of their strong core-periphery structure. This preference weakens when we look at longer terms. Although the graphs might suggest that no starlike structure is possible for the source-sink star layers because of the large positive assortativity, our conclusion of the last paragraph still holds. The assortativity is pushed up by the large amount of disconnected clusters, in most cases simply isolated pairs of trading banks, as Figure 3.18 shows. The disassortative degree mixing in the larger clusters is present, caused by their starlike structure. As the amount of unconnected components grows drastically for the four source-sink layers (which include as the last two the source-sink star layers), we would expect that the same mechanism is behind the high assortativity for $G^{\texttt{91-180d}}$ and especially for $G^{\texttt{.5-1y}}$.

Interestingly, this expectation turns out to be false: evaluating the degree assortativity only in the largest connected component results in lower values for all layers (notably in the source-sink star layers), but it remains strongly positive for $G^{\texttt{91-180d}}$ and $G^{\texttt{.5-1y}}$ during the mature phase of the network

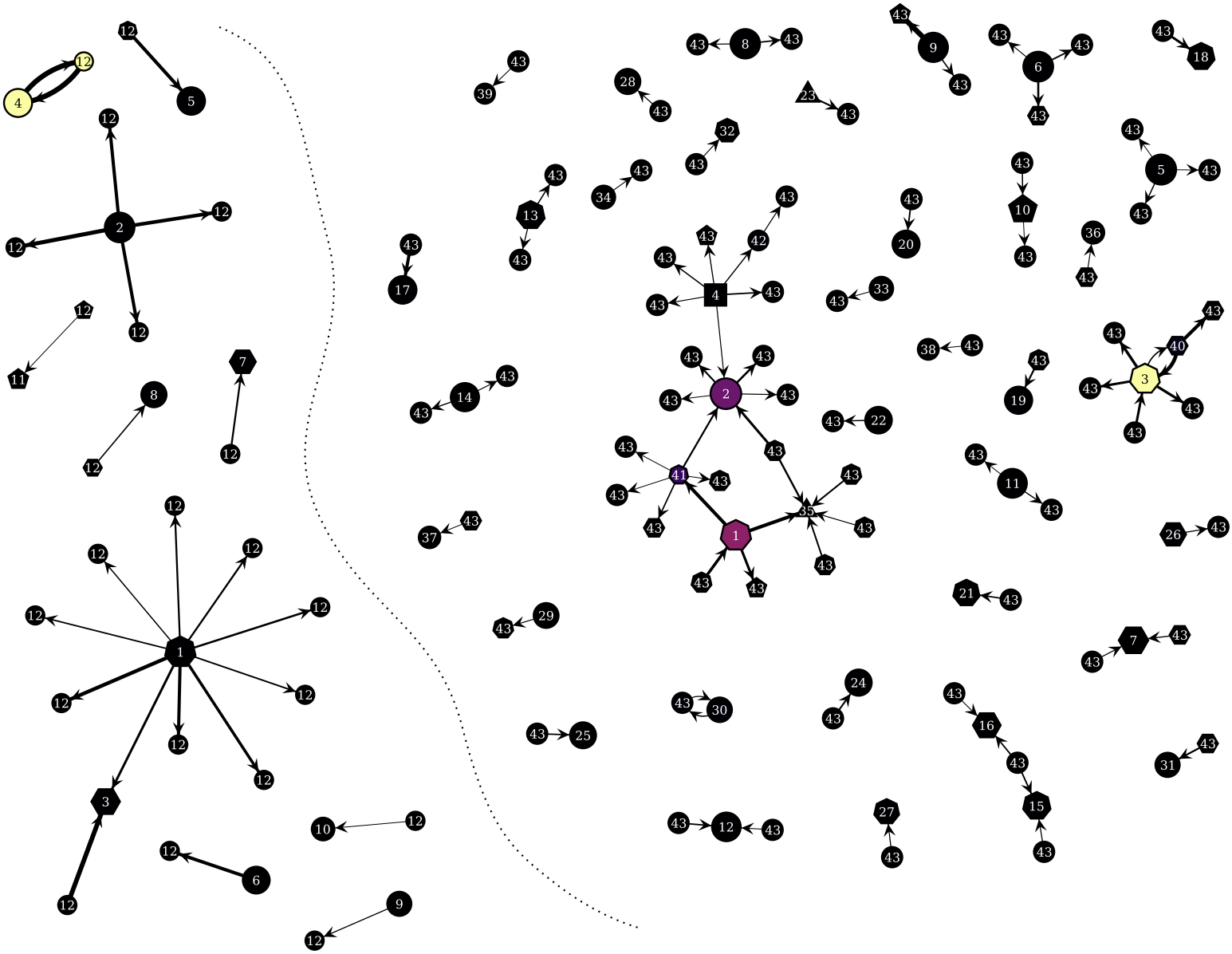


Figure 3.18: Two source-sink star layers. The node ranking and color and the edge width are indicated as in Figure 3.10. (left) $G^{\texttt{1-3y}}$ in 2002. (right) $G^{\texttt{>3y}}$ in 2003.

(Figure C.9). This is strong evidence for a more flat network structure: instead of a starlike structure – which amounts to core-periphery structure if the hubs are intermediaries, or a source-sink star structure if the cores are sources or sinks – with sparse connectivity between the periphery nodes, in a flat network the periphery ceases to exist in a clear form, since the non-hub banks also trade extensively. We know that all layers have some hub

structure because of the heavy-tailed degree distributions. The tier layers and source-sink layers have a strong hub-periphery structure, where the hubs are cores and sources or sinks, respectively, but the two 'intermediate' term layers seem to evolve to a more flat structure where the banks connected to the hubs – that are mostly intermediaries – are connected among themselves significantly. The resulting higher density remains hidden in Figure 3.3 because the networks are split in many unconnected components, while the relative size of their largest connected components increases roughly from 10% to 50% during maturity (Figure C.7). To emphasize this feature of $G^{\texttt{91-180d}}$ and $G^{\texttt{1-3y}}$, we will refer to them as the *flat layers*, completing our layer classification.

We point out that this flattening, which occurs during the maturing of the interbank network, occurs for the short term layers and $G^{\texttt{31-90d}}$ as well. In this last case, the flattening is a destructive force that disrupts the core-periphery structure. The distinction between core-periphery structure and intermediary hubs with a more interconnected periphery is not sharp, and we see a continuous transition from the former ($G^{\texttt{<1d}}$) to the latter ($G^{\texttt{.5-1y}}$). However, the global clustering coefficients and the average shortest distance do mark discontinuous behavior between $G^{\texttt{31-90d}}$ and $G^{\texttt{91-180d}}$, firmly supporting our distinction between the tier layers and the flat layers. Of course, any classification of such complex matter as interbank term layers needs to be taken with a grain of salt.

Bank class modularity

The *bank class modularity* quantifies to which extent banks of the same class lend only to each other [99]. If we think of the bank classes as bank communities, the modularity is large (approaching unity) if lending happens almost exclusively within the communities, with no liquidity seeping out. The classic example of highly modular communities are social networks [26, 99], which tend to form community structure based on race, with social interactions restricted mainly within communities.

The overall bank class modularity is about 0.05, which indicates that in general, the classes play very little role in doing business. Interestingly, during the August 1998 ruble crisis and the summer of 2004 crisis, the modularity increases significantly, and apparently the bank class matters again. We suspect that the moratorium on all private foreign liabilities, declared on August 17, 1998, severely impacted interbank trade with subsidiaries of foreign commercial banks in Russia (the F-class). The sudden increase in the bank class modularity during the second trust crisis in the summer of 2004 seems to indicate that the distrust between a given pair of banks can be partly explained by their bank class.

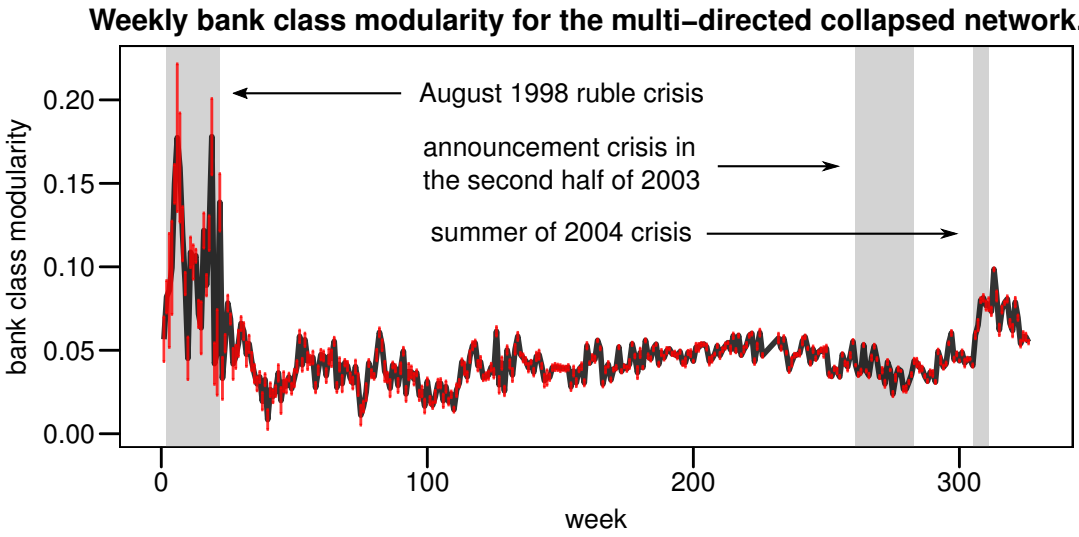


Figure 3.19: The bank classes are defined in Table 3.2. A precise definition of modularity is stated by Equation (1) in [99]. We have used the unweighted version, being only interested in the amount of contracts (if any) between two different bank classes. Each week was sampled 200 times, and error bars with a total height of four standard deviations are indicated in red. See Appendix A for more information about the three crises.

Bank activity

A node i is defined as active on layer α if it has at least one connection on this layer, i.e. its degree $k_i^\alpha > 0$. In symbols,

$$b_i^\alpha = \begin{cases} 1 & \text{if } k_i^\alpha > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.12}$$

Then the *total activity* B_i measures the amount of layers the node participates in, i.e. $B_i = \sum_\alpha b_i^\alpha$ [80]. Figure 3.20 shows the distribution of B_i on monthly time scales. We see that the average total activity grows steadily as the network matures, finally settling around at a value of about three. Almost no banks make use of more than six terms on a monthly basis. The distribution of the total activity is quite broad and relatively unpeaked, especially during the mature phase of the network. This has been reported as a typical quality of real-world multiplex networks [80].

The question is now how the banks distribute their loan activity with respect to the terms. In multiplex terms, the heterogeneity of the number of neighbors of a node i across the M layers can be measured through the *multiplex participation coefficient* [80]

$$P_i = \frac{M}{M-1}\left[1 - \sum_\alpha \left(\frac{k_i^\alpha}{o_i}\right)^2\right], \tag{3.13}$$

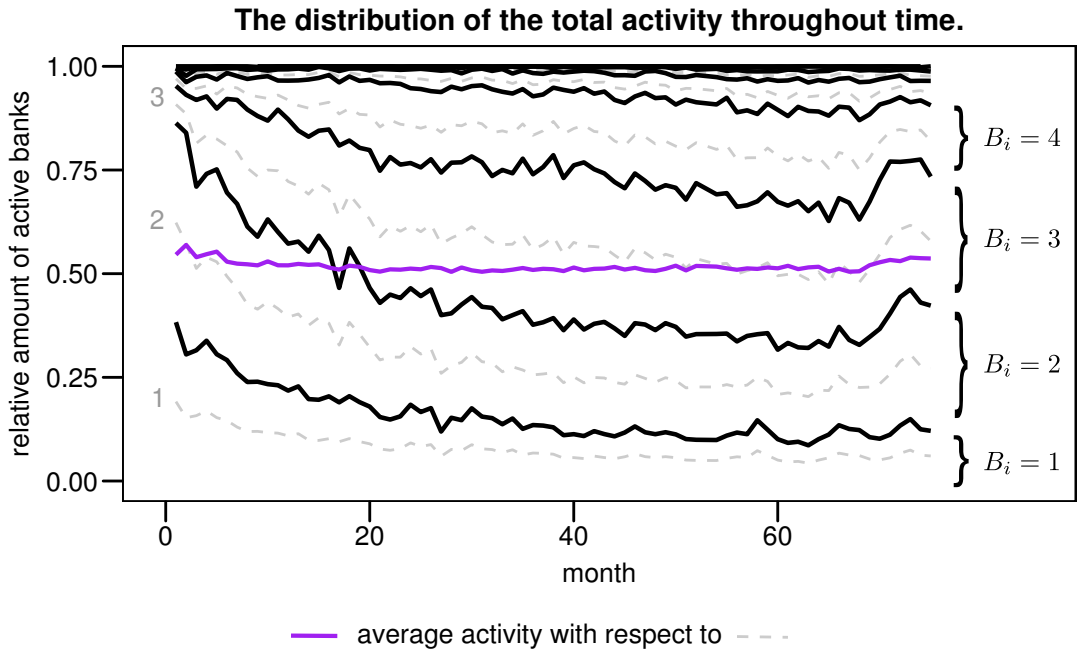


Figure 3.20: For a given month, the area between the thick lines is proportional to the relative amount of active banks with a certain B_i value, which is indicated on the right. The dashed lines make up the moving coordinate system of $\langle B_i \rangle$, in relation to which the average total activity indicated by the thick purple line must be understood. For example, $\langle B_i \rangle$ is about 2 (3) for month 1 (75). The summer of 2004 crisis is mirrored by the drop in $\langle B_i \rangle$ during the last months.

where $o_i = \sum_\alpha k_i^\alpha$ is the total degree across layers. $P_i = 1$ when the edges incident on node i are equally distributed across the layers, and $P_i = 0$ when a node is only active on one layer. The distribution of (P_i, o_i) is pictured in Figure 3.21. Only a few banks specialize in a small amount of terms, most of them having a rather small degree (with the notable exception of one foreign bank). Likewise, the number of banks with large P_i is limited. Since in most cases the bank degrees will be much larger in $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$, compared to the other layers, Equation (3.13) is heavily biased towards values of 0.5 or greater, measuring actually to a large extent the degree of homogeneous participation in $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$. However, we still see significant variation in P_i; clearly not all banks limit the terms in the majority of their loan contracts to the shortest two.

3.5 Conclusion

The main result of our survey in thus Chapter is the classification of the term layers in three categories, based on their topology and the function of the hubs. Each layer contains hubs as is evident from the ubiquitous

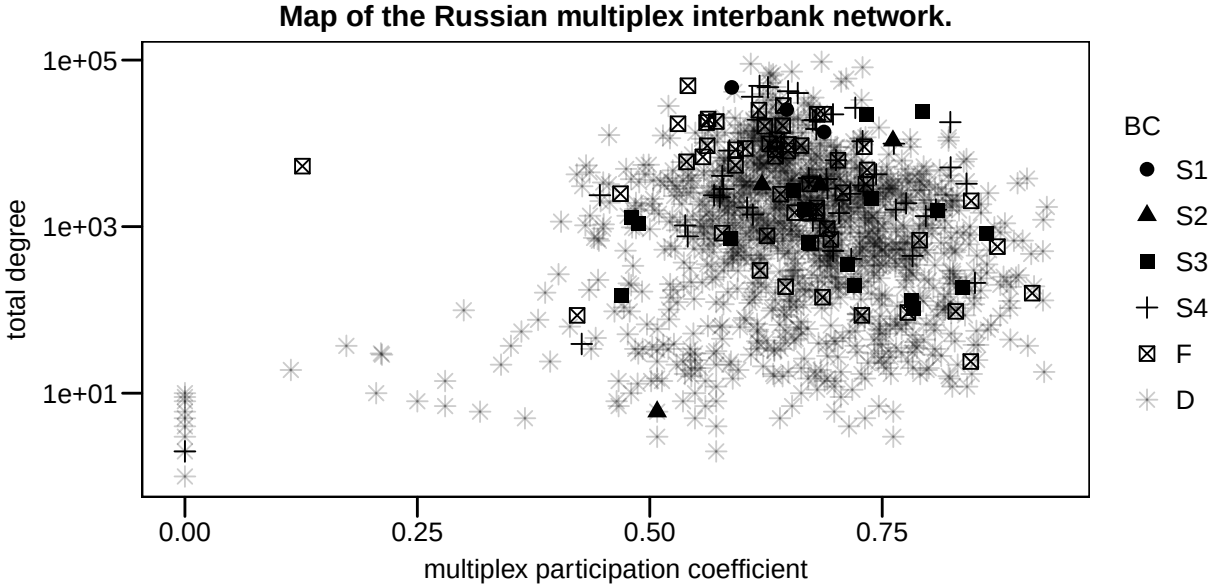


Figure 3.21: The time-aggregated distribution of the total degree o_i and the multiplex participation coefficient P_i. For a description of the bank classes (BC), see Table 3.2 on page 57.

heavy-tailed degree distributions. The three categories are:

Tiering layers: $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$, $G^{\texttt{8-30d}}$, $G^{\texttt{31-90d}}$**.** The interbank lending network is organized in a core-periphery structure. The hubs are the largest intermediaries in the network, acting as money centers that dissipate, generate and redistribute liquidity. This tiering structure causes systemic risk associated with high contagion risk should the core banks malfunction or default. The core banks tend to lend more than they borrow. The periphery banks exhibit significant clustering but do not act in general as intermediators.

Flat layers: $G^{\texttt{91-180d}}$, $G^{\texttt{.5-1y}}$**.** The hubs still perform the largest part of the intermediation in the network, but the periphery is significantly more interconnected compared to the tiering layers. This causes less steep intermediation profiles. The networks in the flat layers consist of many unconnected components, the majority of which only count several banks; the largest components are dense and harbor the hubs.

Source-sink star layers: $G^{\texttt{1-3y}}$, $G^{\texttt{>3y}}$**.** The hubs are sinks or sources, connected to periphery banks which are themselves almost unconnected. The networks consist of star-like hub structures and a number of small unconnected components.

Chapter 4

Layer reducibility

In this Chapter we investigate whether some or all of the term layers can be aggregated without loss of crucial information. More specifically, we seek optimal partitions of the term layers into layer bins in Section 4.1, while in Section 4.2 we test if aggregating all layers into the collapsed graph obscures the mesoscale in the network. As in the previous Chapter, additional Figures can be found in Appendix D.

The interbank literature typically posits *ad hoc* classifications of the loan terms of which a few examples are listed in Table 3.3 on page 59. In the Section 4.1 we try to find an optimal classification in a more quantitative manner. That such classification may be feasible is evident from Chapter 3, where we have established common ground between particular sets of layers, i.e. the tiering layers, the flat layers and the source-sink star layers. For example, the short term layers $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$ share many characteristics. Is there enough gain in keeping them separate in the analysis? The first step in answering this question is to quantify the redundancy between two layers in some way.

The identification of epochs in the network can be accomplished within the same framework. As stated in Section 1.2, one can think of time windows as temporal layers. Then similar temporal layers can be aggregated in time bins, i.e. epochs. Epochs designate phases in the network, each phase associated with different dynamics that shape the network's topology.

4.1 Term layer and temporal layer aggregation based on the von Neumann entropy of a graph

The *von Neumann entropy* (VN entropy) of a graph $\mathcal{G}$, which we denote $H(\mathcal{G})$, is a relatively young concept in network theory. In this Section we will use the methodology of De Domenico et al. (2015) to quantify the reducibility of a multilayer network [2], which is founded upon the VN entropy of a *multiplex* network $\mathcal{M}$ i.e. $H(\mathcal{M})$.

Von Neumann entropy of undirected graphs

The first article that linked the VN entropy – a concept from quantum mechanics – to network theory is Braunstein et al. (2006) [100]. In quantum mechanics the VN entropy is associated with the density operator ρ, a positive semi-definite Hermitian matrix with eigenvalues η_i summing up to unity, which encodes all the information about the statistical ensemble of pure states of the system [2]. Then the VN entropy is given by[1]

$$\begin{aligned} H(\rho) &= \mathrm{Tr}[\rho \log \rho] \\ &= \sum_i \eta_i \log \frac{1}{\eta_i}, \end{aligned} \tag{4.1}$$

where we have used the spectral decomposition $\rho = \Phi \,\mathrm{diag}(\{\eta_i\})\Phi^T$ and the cyclic property of Tr. This definition of the VN entropy has the same form as Equation (2.30), hence its name. Note that since $0 \leq \eta_i \leq 1$ and $\sum_i \eta_i = 1$, the eigenvalues of ρ take on similar roles as the probabilities p_i in (2.30).

Braunstein et al. and subsequent work by other authors [2, 32, 100–102] showed that by defining an appropriate density matrix of a *graph*, the VN entropy can be used as an useful measure that can characterize the structure of graphs. Given the *combinatorial Laplacian* of an undirected simple graph $\mathcal{G}$ with adjacency matrix A_{ij},

$$\mathcal{L} = D - A, \tag{4.2}$$

where D_{ij} is called the *degree matrix* of $\mathcal{G}$, defined as

$$D_{ij} = \begin{cases} k_i & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

Then the density matrix of a (N, E) graph $\mathcal{G}$ [100] is defined as

$$\rho_{\mathcal{G}} = \frac{1}{\sum_i k_i}\mathcal{L}, \tag{4.4}$$

which is a symmetric and positive definite $N \times N$ matrix; in addition, the rescaling of $\mathcal{L}$ ensures that $\rho_{\mathcal{G}}$ has trace 1, therefore justifying its use as a density matrix. Denote the ordered set of eigenvalues (i.e. the spectrum) of $\rho_{\mathcal{G}}$ as $\lambda_1, \lambda_2, \ldots, \lambda_N$, with $\lambda_i \leq \lambda_{i+1}$. Then the VN entropy of $\mathcal{G}$ is

$$\begin{aligned} H(\rho_{\mathcal{G}}) &= \mathrm{Tr}[\rho_{\mathcal{G}} \log \rho_{\mathcal{G}}] \\ &= \sum_{i=1}^{N} \lambda_i \log \frac{1}{\lambda_i}. \end{aligned} \tag{4.5}$$

The requirement that $\mathcal{G}$ is undirected and simple (i.e. has no self-loops or parallel edges) is important, because it guarantees that $\rho_{\mathcal{G}}$ has the properties

[1] In the following we set $0 \log \frac{1}{0} = 0$ by convention.

of a density matrix. Note that we have used the symbol $\mathcal{G}$ instead of G to emphasize that only topological information is being used by $H(\rho_{\mathcal{G}})$.

The spectrum of $\mathcal{L}$ has many properties that have been extensively studied in the algebraic graph theory literature [103]; for example its smallest eigenvalue is zero – thus $\lambda_1 = 0$ – whose multiplicity is equal to the number of connected components in $\mathcal{G}$ [100]. The combinatorial Laplacian is the discretized version of the Laplace differential operator Δ acting on a *graph*. The interpretation of the spectrum of $\mathcal{L}$ depends on the topology; for example for a lattice graph the eigenvalues correspond to the characteristic frequencies of a membrane that is modeled by nearest-neighbour interactions (edges) between atoms (nodes) [104]. Another example is a graph that models a network of communicating pipes conveying some kind of current (a liquid, electric current, ...); then the eigenvalues determine the kinetic behavior of the flow and the eigenvalue gap (the second smallest unique eigenvalue) fixes the basic nature of the flow, i.e. whether it is periodic or aperiodic [104]. It is worth noting that the large eigenvalues of $\mathcal{L}$, that determine the fast temporal scales of the diffusion dynamics on the network, give the largest contribution to the Von Neumann entropy [105].

These examples convey that the VN entropy is yet another application of the spectrum of $\mathcal{L}$ at which some effort has been directed for a better understanding of precisely how $H(\rho_{\mathcal{G}})$ quantifies the structure of a graph [101, 103, 105]. In spite of a full understanding, applications of $H(\rho_{\mathcal{G}})$ have been found that include image and pattern recognition [102], machine learning and thermodynamical descriptions of time evolution processes on graphs [32], and, finally, a framework for layer reducibility [2] that will be applied in this Section.

To get a feel of the nature of $H(\rho_{\mathcal{G}})$, consider the set of simple undirected graphs $\mathcal{G}_N$ with N nodes and at least one edge, $\mathcal{G}_N = \{\mathcal{G}(N, E)\}$, where $N \geq 2$ and $1 \leq E \leq N(N-1)/2$. We now discuss the two extremal cases of E where we will write $H(\mathcal{G}(N, E))$ instead of $H(\rho_{\mathcal{G}(N,E)})$ for the sake of brevity.

First, $\mathcal{G}(N, 1)$ is a graph with only two connected nodes and $N - 2$ unconnected nodes. It has $N - 1$ connected components, so the first $N - 1$ eigenvalues of $\rho_{\mathcal{G}(N,1)}$ $\lambda_1, \lambda_2, \ldots, \lambda_{N-1}$ are equal to zero. It is easy to check that the second unique eigenvalue is unity, which implies that $H(\mathcal{G}(N, 1)) = 0$ for all $N > 2$, i.e. unconnected nodes do not contribute to the VN entropy of the graph.

Second, $\mathcal{G}(N, E_{\max})$ is the complete graph K_N. Thus the multiplicity of $\lambda_1 = 0$ is one. One can prove that the other unique eigenvalue of ρ_{K_N} is $1/(N-1)$ with a multiplicity of $N - 1$; in addition, the VN entropy attains its maximal value[2] $H(K_N) = \log(N - 1)$ on $\mathcal{G}_N$ for K_N [100]. Thus

$$H(\mathcal{G}(N, 1)) = 0 \leq H(\mathcal{G}(N, E)) \leq H(\mathcal{G}(N, E_{\max})) = \log(N - 1). \qquad (4.6)$$

Now let us imagine a discrete time process where we add for each step one edge to $\mathcal{G}(N, E)$, starting from $\mathcal{G}(N, 1)$ and eventually ending up at

[2]Remarkably, for asymptotically large N the VN entropy of a (N, p) Erdös–Rényi graph attains the maximum value as well, independent of the choice of $0 < p < 1$. This shows that both randomness and regularity have similar roles with respect to the VN entropy, at least for sufficiently large graphs [103].

$\mathcal{G}(N, E_{\max}) = K_N$. Then at each step i the VN entropy increases:

$$\Delta H(i) = H(\mathcal{G}(N, i+1)) - H(\mathcal{G}(N, i)) > 0. \tag{4.7}$$

We can minimize $\Delta H(i)$ if we at step i add the edge in a 'local' fashion, i.e. by increasing the number of complete subgraphs (also called cliques) or by homogenizing the degrees in a connected component. In contrast, $\Delta H(i)$ is the largest for edges that when added contribute to the amount of long paths, nontrivial symmetries and unconnected components [105].

This shows that $H(\mathcal{G})$ may be seen as a measure of *regularity* in networks, i.e. regular graphs with an equal number of neighbors for all nodes tend to display a higher entropy than those with heterogeneous degree distributions and the same number of links [105]. This seems odd at first, since maximal entropy tends to be associated with complete disorder, while precisely the opposite is true for the VN entropy of a graph. We can resolve this anomaly by returning to the quantum mechanical origin of $\rho_{\mathcal{G}}$. The VN entropy is a quantitative measure of the mixedness of a quantum system with N states. For general density matrices, $H(\rho) = 0$ if ρ is a pure state, $H(\rho) > 0$ is called in general a mixed state, for which the maximum VN entropy is reached when $H(\rho) = \log N$. This happens when $\rho = \frac{1}{N} I_N$, i.e. a completely random state [101]. The graph analogon of the most *random* state of ρ is the maximally *regular* complete graph K_N[3].

Although a general interpretation of the VN entropy of a graph $\mathcal{G}(N, E)$ still is an open problem [103], it has been suggested that (1) a pure state corresponds with $\mathcal{G}(N, 1)$ and (2) a mixed state can be interpreted as the entanglement of the statistical ensemble of pure states where each pure state is one of the edges of the graph [2]. In this Section we shall not dwell further on the interpretation of the VN entropy with respect to its quantum mechanical origin; encouraged by previous successful applications of the VN entropy, we put this matter aside and focus on the results.

The Jensen-Shannon distance between graphs

Passerini and Severini (2012) suggested that the *quantum relative entropy* between the density matrices $\rho \equiv \rho_{\mathcal{G}_1}$ and $\sigma \equiv \sigma_{\mathcal{G}_2}$ of two graphs $\mathcal{G}_1$ and $\mathcal{G}_2$ could be a used as a measure of distinguishability for graphs [101]. Note that this requires that the density matrices have the same dimensions, thus $\mathcal{G}_1$ and $\mathcal{G}_2$ must have the same amount of nodes. The quantum relative entropy, which is also known as the Kullback–Liebler divergence, of ρ with respect to σ is defined as [101]

$$\mathcal{D}_{KL}(\rho||\sigma) = \mathrm{Tr}[\rho(\log \rho - \log \sigma)]. \tag{4.8}$$

De Domenico et al. (2015) note that $\mathcal{D}_{KL}$ is not a proper metric because it is not symmetric with respect to its arguments (in general, $\mathcal{D}_{KL}(\rho||\sigma) \neq \mathcal{D}_{KL}(\sigma||\rho)$) and does not satisfy the triangle inequality [2]. They propose the *Jensen-Shannon divergence* $\mathcal{D}_{JS}$ as a more suitable dissimilarity measure.

[3] Or indeed, in the asymptotic limit of large N, the Erdös–Rényi random graph with $0 < p < 1$.

Let $\mu = \frac{1}{2}(\rho + \sigma)$ be the new density matrix obtained as the mixture of the original graph density matrices. Then

$$\mathcal{D}_{JS}(\rho||\sigma) = \frac{1}{2}\mathcal{D}_{KL}(\rho||\mu) + \frac{1}{2}\mathcal{D}_{KL}(\sigma||\mu). \tag{4.9}$$

Then using Equation (4.5) results in

$$\mathcal{D}_{JS}(\rho||\sigma) = H(\mu) - \frac{1}{2}(H(\rho) + H(\sigma)). \tag{4.10}$$

Now by construction $\mathcal{D}_{JS}$ is reflexive and symmetric. The *Jensen-Shannon distance* (JS distance) $d_{JS} = \sqrt{\mathcal{D}_{JS}}$ takes values in $[0, 1]$ and behaves as similar to a metric when applied to any pair of density matrices, i.e. mixed quantum states, as recent numerical arguments have shown [2]. Thus for three graphs $\mathcal{G}_1$, $\mathcal{G}_2$, $\mathcal{G}_3$, we say that $\mathcal{G}_1$ is more similar to $\mathcal{G}_2$ than it is to $\mathcal{G}_3$ if $d_{JS}(\rho_{\mathcal{G}_1}, \rho_{\mathcal{G}_2}) < d_{JS}(\rho_{\mathcal{G}_1}, \rho_{\mathcal{G}_3})$.

As a last remark on the JS distance, we point out a very important implication of Equation (4.10). The VN entropy of a graph $\mathcal{G}$ is based on the spectrum of the combinatorial Laplacian, which is a graph invariant [106]; thus $H(\rho_{\mathcal{G}})$ is a graph invariant as well. This means that it is invariant under a relabeling of the nodes – d_{JS}, however, is not. The invariance is broken by the first term $H(\mu)$ in (4.10). Given that $\mu = \frac{1}{2}(\rho+\sigma)$, the graph $\mathcal{G}_\mu$ is a *weighted* undirected simple graph that aggregates $\mathcal{G}_\rho$ and $\mathcal{G}_\sigma$. The weight A^μ_{ij} of an edge (i, j) in $\mathcal{G}_\mu$ indicates whether the edge is present in either $\mathcal{G}_\rho$ or $\mathcal{G}_\sigma$ ($A^\mu_{ij} = \frac{1}{2}$) or in both ($A^\mu_{ij} = 1$). We can assemble $\mathcal{G}_\mu$ in our imagination by laying $\mathcal{G}_\rho$ and $\mathcal{G}_\sigma$ on top of each other, node by node, with the doubly-present edges producing one thick edge with a weight of unity. This implies that for a meaningful comparison between $\mathcal{G}_\rho$ and $\mathcal{G}_\sigma$ a far stronger requirement must be met than an equal amount of nodes in each graph alone: the nodes in the graph must have the same labeling because the VN entropy of the weighted aggregated graph contributes to the JS distance. In our case, when we compare term layers G^α and G^β by calculating the JS distance between them, it is always made sure that the bank labeling agrees and the same banks are put on top of each other. To emphasize that d_{JS} implicitly uses a node attribute (namely its index i), we will note the networks as G again instead of $\mathcal{G}$; additionally we will omit the density matrices for the sake of brevity. Thus in the following we write $d_{JS}(G_1, G_2)$ instead of $d_{JS}(\rho_{\mathcal{G}_1}, \rho_{\mathcal{G}_2})$.

It is instructive to apply d_{JS} to an example before we turn to the multiplex interbank network; we illustrate how the JS distance can be used for image recognition in Figure 4.1. We start with a set of four glyphs $\{\alpha, \alpha, \alpha, \beta\}$. Using Delaunay triangulation the images of the glyphs are transformed into four glyph graphs. This is done in the following way: first the glyphs are converted to black-and-white bitmaps with a dimension of 10×15 pixels. Each pixel is either black or white. The glyph graphs are lattices that consist of 10×15 nodes, one for each pixel. Two *neighboring*[4]

[4]Various schemes exist to define the neighbors of a given node. We chose the parsimonious Delaunay triangulation.

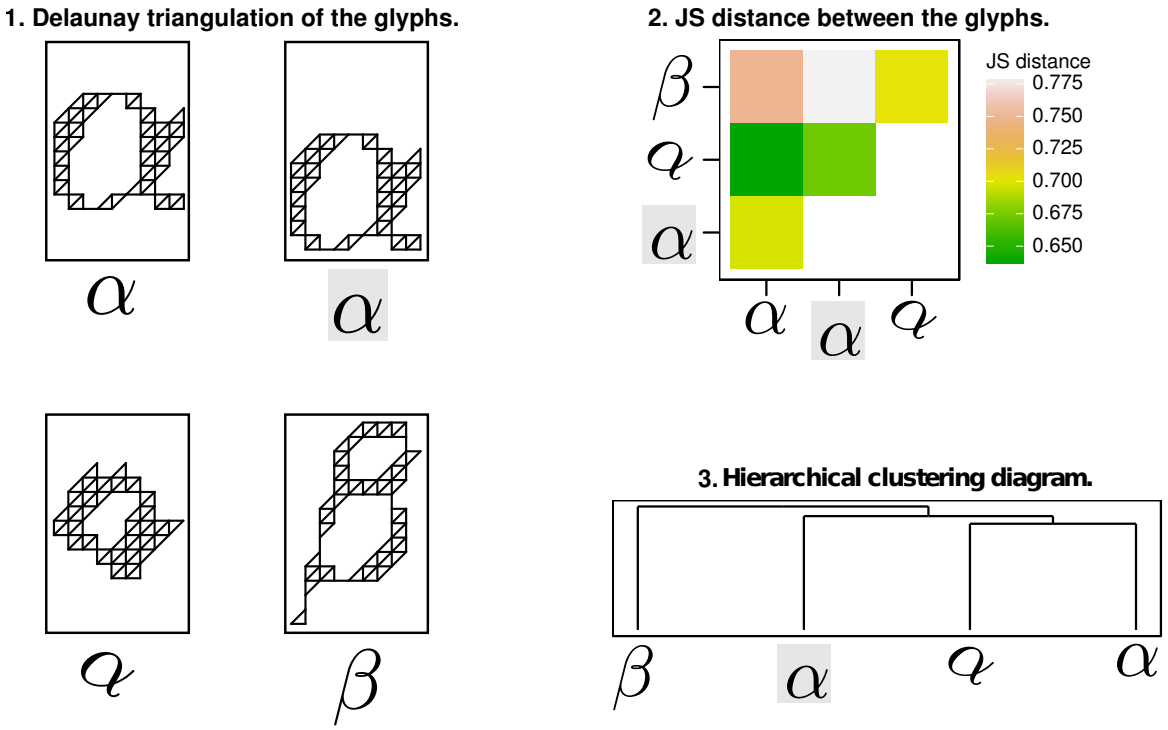


Figure 4.1: An example of using the JS distance to quantify the similarity between the four glyphs showed in the first panel. The α-glyph is put in a shaded box for clarity. In the hierarchical clustering diagram, the length of the vertical lines is proportional to the distance between the clusters that are joined at the next step.

nodes are then connected in the glyph graph if the corresponding pixels in the glyph bitmap are both black. We emphasize that in panel 1 of Figure 4.1 only the edges of the glyph graphs are drawn; thus most of the nodes in the glyph graphs are unconnected.

The glyph images have been chosen specifically to test three qualities of the d_{JS} metric[5]. First, α and β are two glyphs with markedly different structure. Second, α and α are a vertical translation and a rotation+scaling of α, respectively; together with α we will call them the α-glyphs. These glyphs model (kind) noise on α, which could for example correspond to differences in handwriting if the α and α are input images for a program that tries to recognize user-drawn LaTeX characters[6]. Third, α and α are graph isomorphisms, i.e. they are equal modulo node index permutation. This is also seen in the VN entropy ranking of the glyph graphs,

$$H(\alpha) = H(\alpha) < H(\alpha) < H(\beta), \tag{4.11}$$

[5]Note that in this case the glyph graphs are all connected; this is not required for the VN entropy.

[6]Like `http://detexify.kirelabs.org`.

where we have relaxed the notation for the sake of readability, i.e. $H(G_\alpha) \equiv H(\alpha)$.

The JS distance between the glyph graphs are displayed in panel 2 in Figure 4.1. The distance between pairs belonging to the α-glyphs is indeed relatively small, except for α and $_\alpha$, which is caused by the fact that d_{JS} is indeed remarkably sensitive to the location of the edges which is encoded by the node indices of the glyph graphs. This sensitivity ensures that common patterns (like triangles or higher-order motifs) exhibited by different sets of banks do not unrighteously decrease the JS distance between two term layers. Continuing, the distance between the β and α-glyphs is indeed relatively large, except for β and $_\alpha$, which is probably due to the considerable overlap in the aggregated glyph $\not\beta$.

The similarities can be summarized by performing an hierarchical clustering algorithm on the distance matrix in panel 2, which results in the hierarchical clustering diagram (also called a *dendrogram*) displayed in panel 3 in Figure 4.1. The hierarchical diagram classifies the glyphs correctly, in an hierarchical order that bins the most similar glyphs first. In this Chapter we aways use *average linking clustering* [107], mainly because of its simplicity and straightforward interpretation. Clustering algorithms take as input the $n(n-1)/2$ distances $d(i,j)$ between a list of n objects $(1, 2, \ldots, n)$. d can have any form, but it must comply to the properties of a distance metric, as otherwise the clustering algorithm will infer unintuitive distances between clusters – this is why we insisted on a graph similarity measure that behaves as a distance metric. The distance matrix $d(C_i, C_j)$ defines at each step the distance between clusters C_i and C_j, which initially contain only the individual objects. At the first step, we initialize the list of clusters,

$$L_C = (C_1 = \{1\}, C_2 = \{2\}, \ldots, C_n = \{n\}), \tag{4.12}$$

and the distance matrix

$$d(C_i, C_j) = d(i, j). \tag{4.13}$$

Then the pair of objects that are nearest to each other are joined in a new cluster. Without loss of generality we can take these to be $(1, 2)$. Then clusters C_1 and C_2 are removed from L_C and replaced by a new cluster C',

$$C' = C_1 \cup C_2 = \{1, 2\}, \tag{4.14}$$

and the list of clusters is updated:

$$L_C \to L_C = (C', C_3, \ldots, C_n). \tag{4.15}$$

Now the distance matrix can be generalized to

$$d(C_i, C_j) = \frac{1}{|C_i||C_j|} \sum_{i \in C_i} \sum_{j \in C_j} d(i, j). \tag{4.16}$$

We can repeat these steps until each object is in the same cluster. At each step the clusters closest by are joined according to (4.16), thus the order

of clusters gives a preference for putting objects in the same bin. In other words, the hierarchical clustering diagram proposes a partition of the objects L_C at each step, where the objects within each cluster belong to the same bin. Applying the hierarchical clustering to a set of layers $\{G^\alpha\}$, then, yields a configuration of aggregated layers at each step, based on the JS distance between the layers of the multiplex.

Methodology of De Domenico et al. (2015)

Consider the a multiplex network $\mathcal{M} = \{G^\alpha\}$ with M layers, $\alpha \in [1, 2, \ldots, M]$. If, as in the case of the Russian multiplex interbank network, $\mathcal{M}$ is a general multi-directed graph, we redefine $\mathcal{M}$ to be the *undirected view* of itself. Then the VN entropy of the aggregated graph is denoted $H(G^a)$. In general, if we aggregate *some* of the layers, we obtain a reduced multiplex network $\mathcal{C} = \{C^\beta\}$ with $X \leq M$ layers, where a layer C^β, $\beta \in [1, 2, \ldots, X]$, is either one of the original layers G^α in $\mathcal{M}$ or the aggregation of two or more of them[7].

Then the VN *entropy per layer* [2] of $\mathcal{C}$ is

$$\bar{H}(\mathcal{C}) = \frac{H(\mathcal{C})}{X} = \frac{\sum_\beta H(C^\beta)}{X}. \tag{4.17}$$

To define the VN entropy of a multilayer network, De Domenico et al. suggest that each layer represents one possible state of the system, that is, a network state, extending the concept discussed above that a graph is in a pure state if it contains only one edge, and in mixed state otherwise, as captured by the graph density matrix. Informally, the VN entropy of a *graph* measures the entanglement of its edges, while the VN entropy of a *multilayer network* measures the entanglement of its layers. In the special case that the multilayer network is a multiplex, i.e. no interlayer edges exist, the VN entropy reduces simply to the sum of the VN entropies of its layers [2]; hence the last equality in Equation (4.17).

Returning to the VN entropy per layer of $\mathcal{C}$, De Domenico et al. then propose to quantify the distinguishability between the multiplex network $\mathcal{C}$ and the aggregated multiplex G^a through the *relative entropy*

$$q(\mathcal{C}) = 1 - \frac{\bar{H}(\mathcal{C})}{H(G^a)}. \tag{4.18}$$

The larger $q(\mathcal{C})$, the more distinguishable is the configuration of aggregated layers in $\mathcal{C}$ from the corresponding aggregated graph G^a. In the special case that all the layers have been aggregated, $\mathcal{C} = G^a$ and $q(\mathcal{C}) = 0$. De Domenico et al. show that aggregating another layer in $\mathcal{C}$, obtaining a further reduced configuration of $X - 1$ layers $\mathcal{C}'$, can both increase and decrease $\bar{H}$, depending on the structure of the layers being aggregated, thus $q(\mathcal{C}')$ can be greater or smaller than $q(\mathcal{C})$. A decrease in q (increase in $\bar{H}$) is observed when one aggregates layers with very different edge densities or

[7] Refer to Equation (3.3) in Section 3.2 for the definition of the aggregation of two graphs.

when the aggregation would create structural patterns that did not exist in any of the two original layers, while an increase q (decrease in $\bar{H}$) usually corresponds to the merging of layers having very similar structure. Hence, by maximizing q one tends to avoid layer configurations that might contain spurious structural patterns or redundant layers [2].

Thus the optimal partition of the M layers G^α into X bins corresponds to $\mathcal{C}^*$, where $\mathcal{C}^*$ is the multiplex network with layers C^β that aggregate one or more G^α and that maximizes $q(\mathcal{C})$ over all possible layer partitions $\mathcal{C}$. Note that the extremal cases are $\mathcal{C}^* = \mathcal{M}$ if the layers are too different to be aggregated (which is the case for Figure 4.1 for example) and $\mathcal{C}^* = G^c$ if all layers are sufficiently redundant.

Enumerating all possible partitions of a set of M layers becomes quickly intractable, thus a heuristic method must be used. The idea is to use the proposed partitions by the hierarchical clustering diagram, as they cluster similar layers, which are likely to increase q when they are aggregated. For each proposed partition L_C we build $\mathcal{C}$ by aggregating the layers belonging to the same cluster for each of the $X \leq M$ clusters; then we calculate $q(\mathcal{C})$ according to Equation (4.18). Our heuristic approximation to the optimal partition $\mathcal{C}^*$ is then the partition given by the hierarchical clustering diagram with maximal q.

Note on the von Neumann entropy of directed graphs

Before we continue to the results, we mention that recently a formalism has been developed by Ye et al. (2014) [108] that extends the VN entropy to directed simple graphs, along with additional applications in [32]. However, the VN entropy in this work is only properly defined for strongly connected graphs – we argued in Section 3.4 while discussing core-periphery and source-sink structure that limiting the interbank network to the strongly connected component has no clear economical interpretation.

Furthermore it is unclear whether the VN entropy for directed graphs allows a distance metric similar to d_{JS} because it is based on a different Laplacian, i.e. one for directed graphs [109]. For these reasons we only use the VN entropy on the undirected view of the Russian multiplex interbank network.

Results for term layer aggregation

Figure 4.2 displays the result of applying the methodology of De Domenico et al. The first striking feature is that the layer bins seem to widen as the network matures, indicating that the tiering layers and flat layers become increasingly homogeneous in structure. In other words, the multiplex description of the interbank network becomes increasingly redundant; this is confirmed by Figure 4.3. Secondly, except for month 5, all term bins are continuous, following the natural lengthening maturity pattern that is ubiquitous in our survey. We emphasize that in no way this ordering is inherent in the JS distance, thus it constitutes strong proof for a natural ordering with respect to the term layers. Third, $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ are never binned, while $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$ are always the most similar layers as measured by d_{JS}. An

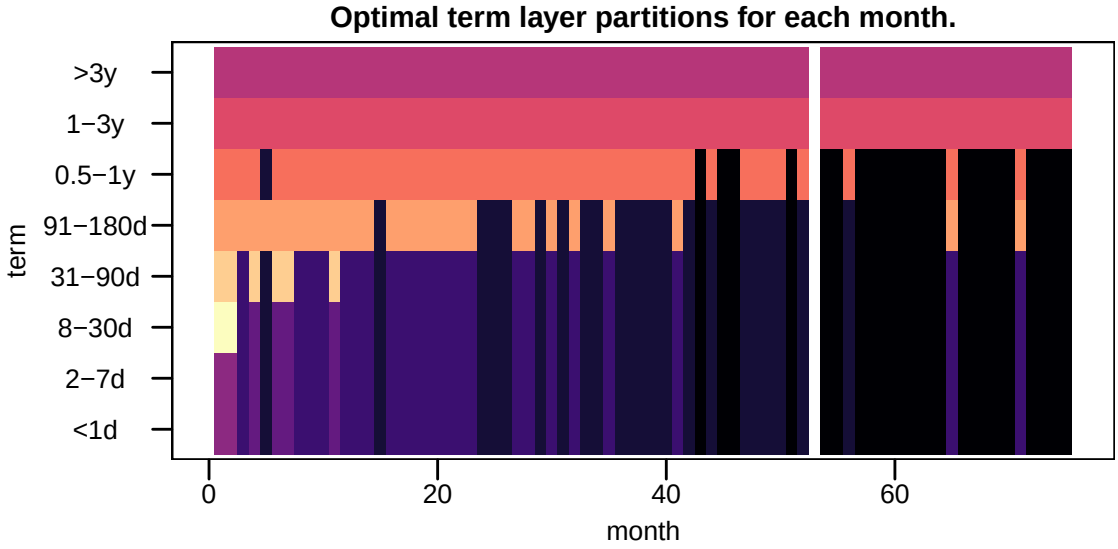


Figure 4.2: Each column represents a partition of the term layers. Cells with equal color belong to the same layer bin.

example of this is displayed in Figure 4.4, that illustrates the layer binning procedure for a prototypical month for the mature network, i.e. month 59, June 2003. The distance matrix shows that the distance between $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$ is small. The two source-sink star layers do not in particular resemble any other layer, themselves included. Furthermore, it is interesting that the hierarchical clustering diagram complies to our layer classification of the first six terms, i.e. layers are binned until first the tiering layers and then the flat layers are completed; in contrast, the relative entropy q falls off fast for the proposed merge of $G^{\texttt{1-3y}}$ and the tiering and flat layers, indicating a strong structural difference between them.

Results for time layer aggregation

As an experimental application, we repeat the above exercise for time layers. Each time layer G^t is the undirected view of the term-aggregated network at month t. The procedure is illustrated in Figure 4.5. The JS distance matrix holds some clues with respect to the temporal correlations between the monthly time scales. In general, during the growth period until month 59 (see Section 2.1), we see an accelerating 'settling down' tendency, which is indicated by the fact that for a given month t_1 in the growth period, the lagged month $t_2 < t_1$ for which the JS distance $d_{JS}(G^{t_1}, G^{t_2})$ is roughly twice $d_{JS}(G^{t_1}, G^{t_1-1}) \approx d_{JS}(G^{t1}, G^{t_1+1})$ lies increasingly further in the past, i.e. during this period period $t_1 - t_2$ grows at an accelerating pace. Halfway during the network maturity (starting from month 60), however, $t_1 - t_2$ more or less stays constant, indicating some sort of saturation with respect to the stabilizing force. In addition, the August 1998 ruble crisis clearly produced anomalous network wiring patters, whereas the trust crisis of the Summer of 2004 (month 70) causes only a ripple in the JS distance.

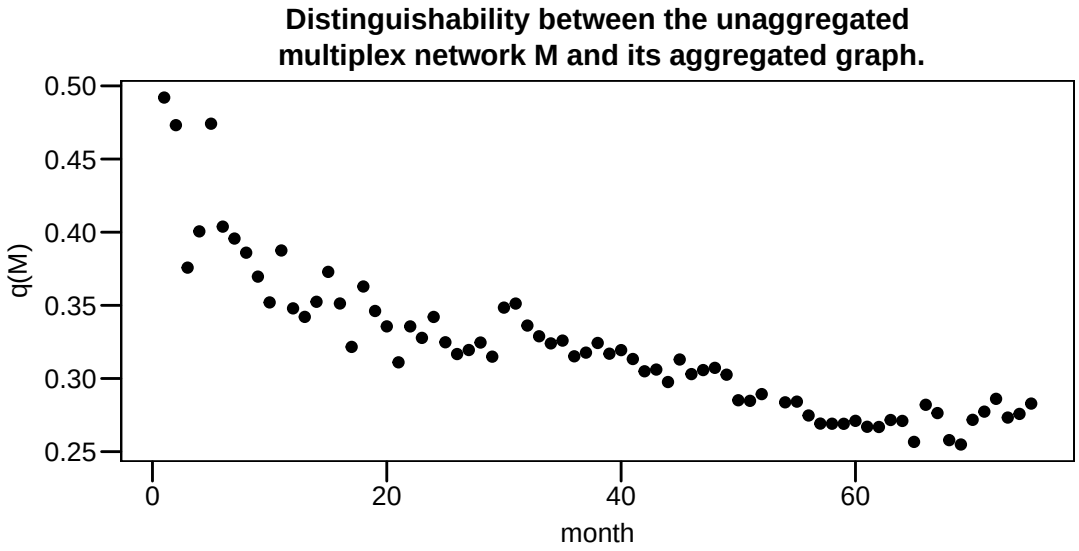


Figure 4.3: For each month, the relative entropy $q(\mathcal{M})$ between $\mathcal{M}$ and G^a is calculated, which is a measure of the distinguishability between both. Thus $q(\mathcal{M})$ is measure of the redundancy of the term layers.

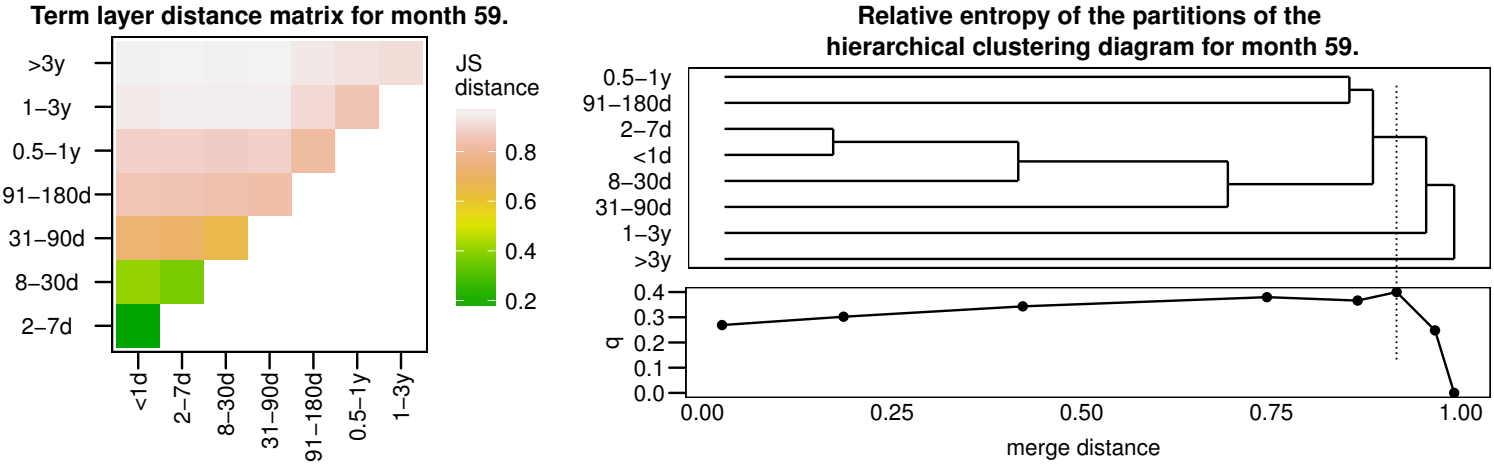


Figure 4.4: The optimal partition for which q attains a maximum is indicated by the dotted vertical line in the right panel; in full it is $(\{G^{\texttt{<1d}}, G^{\texttt{2-7d}}, G^{\texttt{8-30d}}, G^{\texttt{31-90d}}, G^{\texttt{91-180d}}, G^{\texttt{0.5-1y}}\}, G^{\texttt{1-3y}}, G^{\texttt{>3y}})$.

The relative entropy peaks sharply for a certain partition configuration at merging step 70 that we will call the *optimal partition*, which is indicated in blue in Figure 4.5. The optimal partition aggregates all months, except for months 1, 2 and 5, i.e. August, September and December, 1998. q deteriorates heavily after merging months 2 and 5 in step 71, which indicates that the banking activity on the interbank market during is structurally differ-

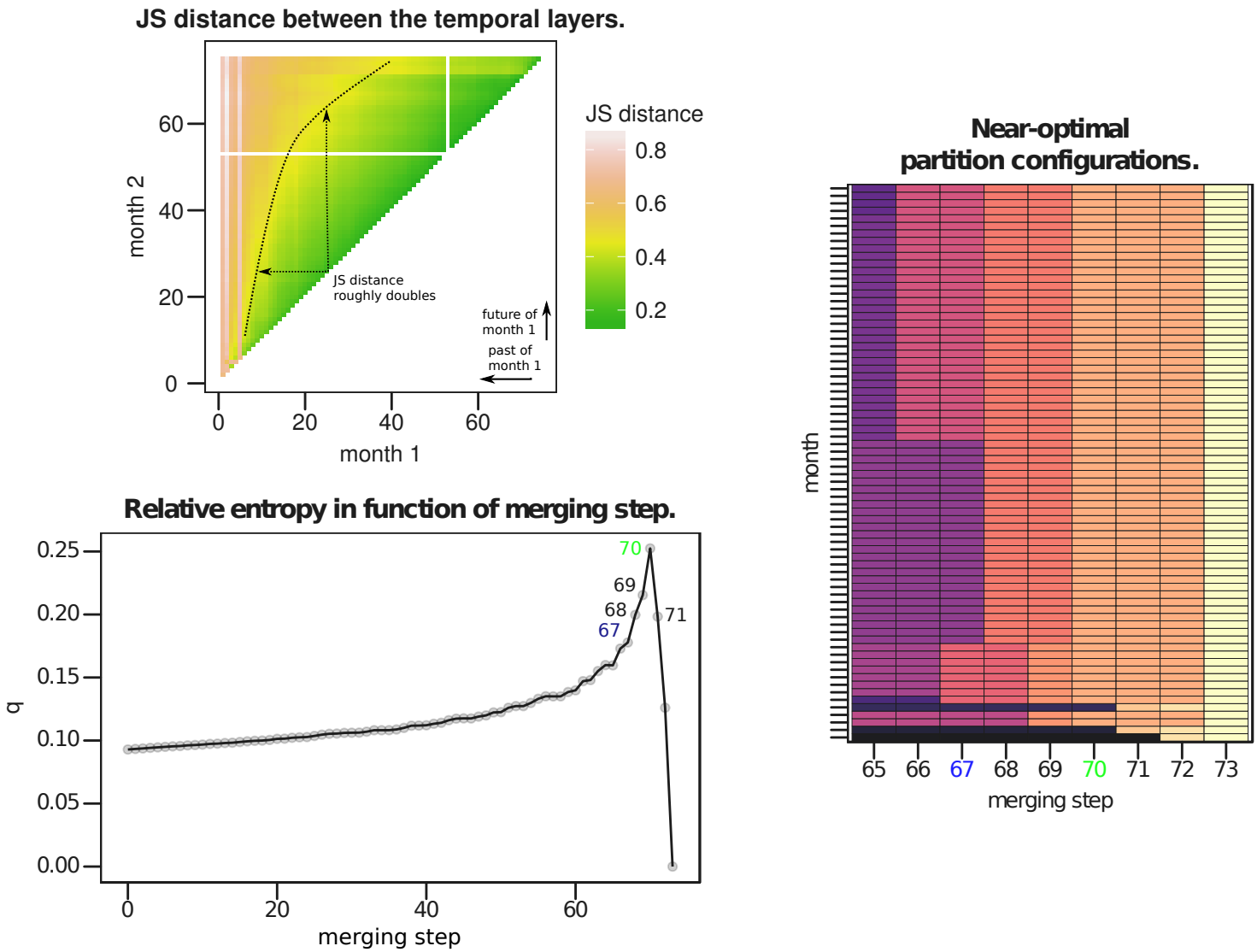


Figure 4.5: The JS distance can be used informally to proxy correlations between the monthly structural lending patterns. (top left panel) For a given month ('month 1'), the curve stipulates the time ('month 2') in the past and in the future where the JS distance roughly doubles. (bottom left panel) The relative entropy $q(\mathcal{C})$ for each merging step, where $\mathcal{C}$ is a multiplex graph that aggregates the original temporal layers according to the partition configuration proposed by the hierarchical clustering diagram in Figure D.2. The most optimal configuration, i.e. where q attains its maximum, is achieved at step 70. (right panel) The partition configurations for steps 65–73. The epoch (optimal) partition is indicated in blue (green).

ent during these months. The hierarchical diagram in Figure 4.5 shows that months 3, 4, and 6 are clustered relatively late (around step 60), again showing the graveness of the ruble crisis in terms of unprecedented turbulence on the interbank market.

The optimal configuration bins months 3 and 4 together with months 6–75, thus aggregates non-adjacent temporal layers, which is undesirable for further analysis because it does not respect time ordering. Therefore we will choose a partition with a larger precision, i.e. the *epoch partition* which is proposed at step 67, indicated in blue in Figure 4.5. Looking at the right panel of Figure 4.5 again, we see that the epoch partition strikes a good balance between detail (more bins) and compactness (less bins). The epoch partition consists of 7 epochs, i.e. time bins, denoted as E1, E2, etc. These epochs, together with some historical comment, are:

E1 (month 1), E2 (month 2), E3 (months 3–4), E4 (month 5). The first four epochs span the 1998 ruble crisis rom August to the end of December (see Appendix A).

E5 (months 6–13). E5 starts at January 1999, when the interbank market stability hindered by the 1998 turmoil is largely restored.

E6 (months 14–40). E6 starts in September 1999, but the underlying event that might have changed the interbank market's structure, if any, is unknown to us. One possibility is the fact that Jeltsin, who resigned on December 31 later that year, fired his prime minister and his entire cabinet on August 9, 1999, i.e. the last month of E5 .

E7 (months 41–75). E7 starts at the beginning of the maturity period (roughly around January 1, 2002) and includes the announcement crisis and the Summer of 2004 trust crisis.

The optimal term layer partitions for each *epoch* are displayed in Figure D.1, which is essentially a straightforward simplification of Figure 4.2. As of E3, the tiering layers arise, and by E6 the tiering layers and flat layers are preferred to be aggregated by the methodology of De Domenico et al. In E5, $G^{\texttt{0.5-1y}}$ is not aggregated, while in E6 and E7 it joins the 'bulk bin' consisting of the tier layers and $G^{\texttt{91-180d}}$, completing the merging of the tier layers and the flat layers. As expected, no epoch exists where the source-sink star layers are aggregated. In general, we see a pattern of continuous expansion of the bin that initially consists only of $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$, until only $G^{\texttt{1-3y}}$ and $G^{\texttt{>3y}}$ are left over. This corroborates our finding in Figure 4.3 that the multiplex network as a whole becomes increasingly redundant.

Finally, we compare the results on term layer and temporal layer aggregation with the hierarchical clustering diagram of a frequently used similarity measure, i.e. the *Jaccard similarity* J. When applied to directed or undirected unweighted simple graphs, the Jaccard similarity can be interpreted as the probability of observing an edge in a network conditional on the observation of the same edge in the other network [35].

For two networks with N nodes – or, equivalently, two labeled layers of the same multiplex – G^α and G^β, we flatten their adjacency matrices to a

vector of length $N(N-1)/2$ in the undirected case, and of length $N(N-1)$ in the directed case; denote this vector as a (b) for G^α (G^β). Thus each entry in the ordered vectors corresponds to some edge e which is either (1) present in G^α and G^β, (2) present in G^α or G^β, or (3) not present in G^α, nor in G^β. Then

$$J(G^\alpha, G^\beta) = J(a, b) = \frac{|a \cap b|}{|a \cup b|}, \tag{4.19}$$

with $J \in [0, 1]$. The limiting cases $J = 0$ ($J = 1$) occur when no (all) edges occur in G^α and G^β simultaneously. Writing $e \in a$ ($e \in b$) for "edge e is present in G^α (G^β)", we have that

$$a \cap b = \{e : e \in a \wedge e \in b\} \tag{4.20}$$

$$a \cup b = \{e : e \in a \vee e \in b\}, \tag{4.21}$$

which implies trivially that $a \cap b \subseteq a \cup b$. Inspired by the Venn diagram picture of probability theory, we can write

$$\begin{aligned} P\left(e \in a \wedge e \in b | e \in a \vee e \in b\right) &= \frac{P\left((e \in a \wedge e \in b) \wedge (e \in a \vee e \in b)\right)}{P\left(e \in a \vee e \in b\right)} \\ &= \frac{P\left(e \in a \wedge e \in b\right)}{P\left(e \in a \vee e \in b\right)} \\ &= J(a, b). \end{aligned} \tag{4.22}$$

Thus if we know that a given random edge in G^α is present, we have a probability of $J(G^\alpha, G^\beta)$ that the same edge is also present in G^β.

The *Jaccard distance* $d_J(G^\alpha, G^\beta) = 1 - J(G^\alpha, G^\beta)$ behaves as a metric, enabling one to perform hierarchical clustering on a Jaccard distance matrix between the layers of a multiplex network. Unlike d_{JS}, d_J is defined for directed networks, too.

Figure 4.6 compares the obtained epochs to the hierarchical clustering diagram produced by the Jaccard distance between the *undirected* views of the monthly temporal layers. We see that the clusters in the final merging steps agree almost perfectly with the epoch partition. Figure D.3 does the same but for the *directed* views of the monthly temporal layers. There is still considerable agreement with the epochs, though evidently to a lesser extent as compared to the undirected view, as the Jaccard distance in the latter case does not discriminate between lending and borrowing. The Jaccard clustering seems to indicate that the separation of E5 and E6 at month 14 is artificial from the directed point of view, suggesting month 23 instead, which might be related to the fact that we found no historical event that could explain what is special about month 14.

To conclude this Section, we report that the Jaccard hierarchical clustering diagrams both for the undirected and directed view of the term layers largely comply with our results. We find that the tiering layers are consistently aggregated first, while starting from E3 the tiering layers and flat layers are always clustered in the natural order we have observed for the optimal term layer partitions, precisely as in Figure D.1.

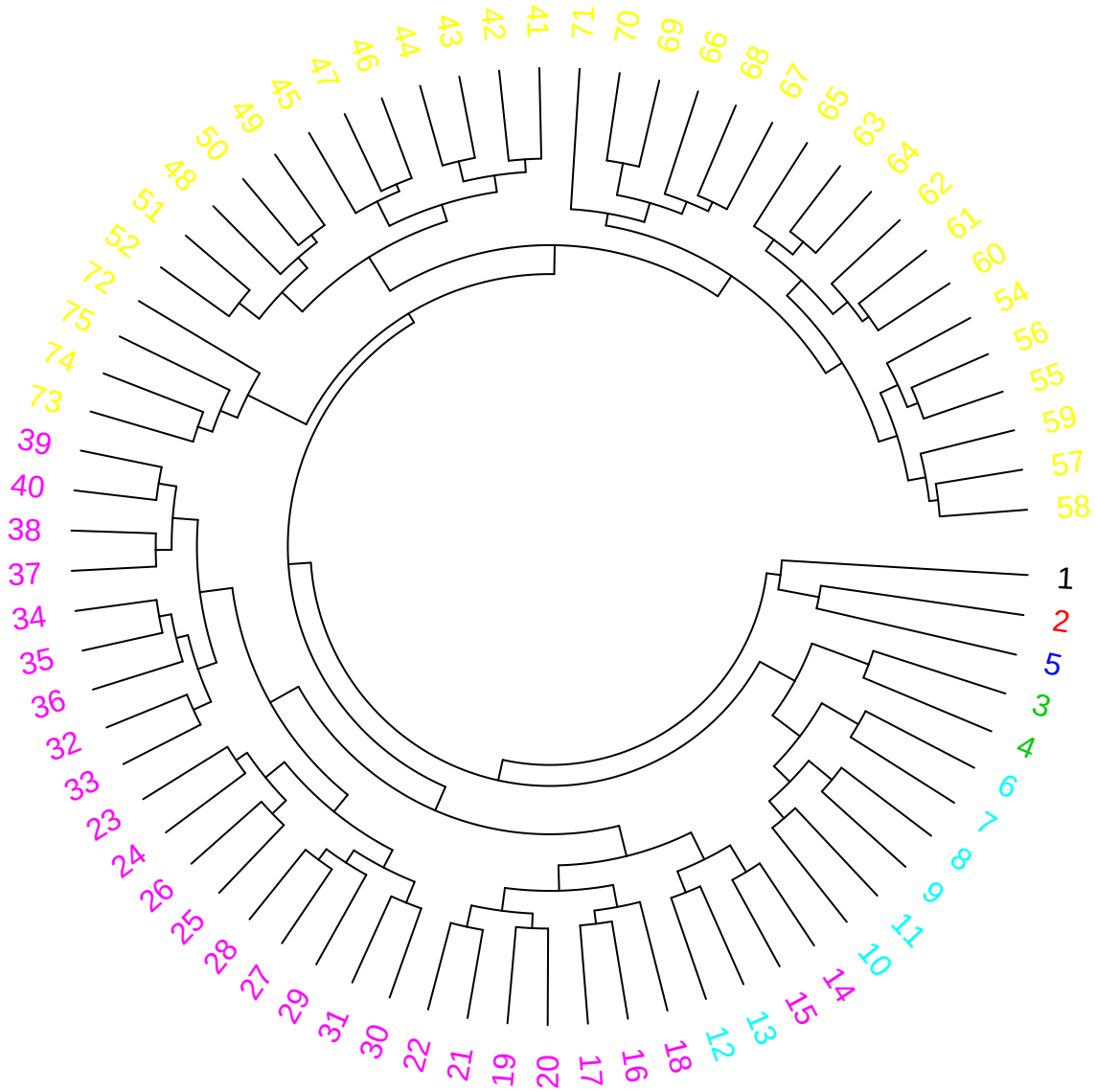


Figure 4.6: Hierarchical clustering diagram obtained from the Jaccard distance between the undirected views of the temporal layers. Each month is colored according to the epoch it belongs to.

4.2 Testing the significance of the layers based on generative models of layered networks

We use the methodology of Peixoto (2015) [25] to test whether the term layers are significant in explaining the mesoscale structure of the complete

multiplex interbank network. This is done by model selection between a null model where the edges are distributed in an entirely random fashion and a generative layered model of the interbank network using the *minimum description length* (MDL) principle. The `graph-tool` library [29] provides a complete framework to perform model fitting and selection. Before turning to the results, we focus on the main ideas underlying layer binning in the context of model selection and refer to [23,25] for a discussion of the inference algorithm.

How necessary are the term layers?

As stated several times before, the $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$ layers host 90% of the total amount of loans recorded in our data set. In Chapter 3 it was demonstrated that these two layers have a very similar structure and function which is confirmed by the results in the previous Section regarding the term layer aggregation. Thus for the sake of compactness, one could collapse $G^{\texttt{<1d}}$ and $G^{\texttt{2-7d}}$ into a single network and simply discard the other term layers. Indeed, this was done in most of the previous work on the Russian interbank data [7, 8, 39]. Less drastically, we could consider only the collapsed graph G^c in our analysis without removing any loan from the data.

The question is whether the term layers are a natural and necessary characterization of the interbank network. For example, does the choice of the counterparty depend sensitively on the term of the loan contract? Do banks adopt different lending and borrowing strategies based on the contract terms? If so, this must emerge in the topology of the interbank network – more specifically, and crucially, then the topology of the term layers must differ considerably, and the mesoscale structure *of the complete multiplex network* can only described satisfactorily by explicitly taking into account the term layers.

So far, we have found mixed evidence as to whether the term layers should or should not be included in a compact description of the interbank network. The results of the previous Section suggest considerable redundancy is introduced by analyzing separate layers instead of aggregated configurations. In Chapter 3 we found that the term layers share similar properties, such as the existence of hubs, but differ in other aspects, of which the most important is the intermediation function of these hubs with respect to loans of the same term. As indicated in the previous Chapter, the maturity transformation inherent to financial intermediation can necessarily be captured only in a limited way when one studies each term layer separately, even when they themselves allow a range of loan contract maturities (neglecting $G^{\texttt{<1d}}$).

For example, consider a bank in $G^{\texttt{2-7d}}$ that has received a loan, evidently with a term `2-7d`. In this layer it can only intermediate the borrowed cash by extending another `2-7d` loan with a limited possibility of maturity transformation, since the maturity of the second loan should be short enough to repay the first loan in time. Of course, this is a simplistic model of financial intermediation, but it constitutes a concrete example of how incorporating layers in a model – thus adding more detail – may actually *obscure* information; in our case we would underestimate the amount of intermediation

in the interbank network.

Another example is community inference (see Section 1.2): if the communities are similar across the term layers, i.e. if the lending and borrowing strategies of banks depends only weakly on the loan terms, then a SBM model that infers the communities in each layer separately will perform worse than a SBM model inferring the community structure in the collapsed graph. This is caused by the fact that the layers make the original available data increasingly sparse by dividing the edges across the layers, thus making it harder for the SBM model to reliably separate noise from actual structure, again obstructing the extraction of information by including too much detail in a model.

The minimum description length (MDL) principle

The mesostructure of the multiplex network $\mathcal{M}$ is modeled by a generative model of layered networks $\mathcal{H}$ that, once fitted to $\mathcal{M}$, entails the community structure that describes the topology of each layer at the mesoscale, as argued in Section 1.2. More precisely, let $\mathcal{M}$ have N nodes and M layers. Then the community structure at each layer can be described by a $M \times N$ matrix b_i^α, where i and α denote the node and layer, respectively, and $b_i^\alpha \in [0, B^\alpha - 1]$ is the group membership of node i in layer α, which has B^α groups. We define the model $\mathcal{H}(\theta, b_i^\alpha)$ by (1) the probability with which a multiplex network $\mathcal{M}'$ is generated,

$$P\left(\mathcal{M}'|\theta, b_i^\alpha\right), \tag{4.23}$$

where θ are additional parameters for the model, and (2) the prior likelihood of the model parameters

$$P\left(\theta, b_i^\alpha\right). \tag{4.24}$$

Given our observation of $\mathcal{M}$, the likelihood that it was generated by a partition b_i^α with model parameters θ is obtained via the Bayesian posterior

$$P\left(\theta, b_i^\alpha|\mathcal{M}\right) = \frac{P\left(\mathcal{M}|\theta, b_i^\alpha\right) P\left(\theta, b_i^\alpha\right)}{P\left(\mathcal{M}\right)}, \tag{4.25}$$

where $P\left(\mathcal{M}\right)$ is the model evidence, i.e.

$$P\left(\mathcal{M}\right) = \sum_{\theta, b_i^\alpha} P\left(\mathcal{M}|\theta, b_i^\alpha\right) P\left(\theta, b_i^\alpha\right). \tag{4.26}$$

By marginalizing out θ in Equation (4.25), we obtain

$$\begin{aligned} P\left(b_i^\alpha|\mathcal{M}\right) &= \sum_\theta P\left(\theta, b_i^\alpha|\mathcal{M}\right) \\ &= \frac{\sum_\theta P\left(\mathcal{M}|\theta, b_i^\alpha\right) P\left(\theta, b_i^\alpha\right)}{P\left(\mathcal{M}\right)}, \end{aligned} \tag{4.27}$$

which is the probability distribution on all possible node partitions – i.e. on all possible community structures – that can be used for further analysis

of $\mathcal{M}$ that depends on the community structure. Take for example the community modularity $Q(b_i^\alpha)$ [110]. The distribution of Q can be derived from the Bayesian posterior:

$$P\left(Q'|\mathcal{M}\right) = \sum_{b_i^\alpha:\ Q'=Q(b_i^\alpha)} P\left(b_i^\alpha|\mathcal{M}\right), \tag{4.28}$$

from which we can calculate the expected community modularity, etc.

Usually one hopes that $P\left(b_i^\alpha|\mathcal{M}\right)$ is sharply peaked; in this case one can bypass having to deal with a probability distribution of community partitions by simply selecting the community partition $\hat{b}_i^\alpha$ for which the distribution reaches its maximum value, i.e.

$$\hat{b}_i^\alpha = \mathrm{argmax}_{b_i^\alpha} P\left(b_i^\alpha|\mathcal{M}\right), \tag{4.29}$$

which amounts to the Fisherian maximum likelihood estimation[8] (MLE) of b_i^α. In general, *model fitting* is done by finding some appropriate value of $\hat{b}_i^\alpha$ which encodes the inferred community structure given $\mathcal{H}$ and θ.

Model selection is performed when one needs to decide which of two models describes the observed data D best. Model comparison is mainly an exercise in preventing overfitting, since given a general model $\mathcal{H}$ with parameters x and θ and maximum likelihood $P\left(\hat{x}|D,\mathcal{H},\theta\right)$, we can (almost) always find a model $\mathcal{H}'$ with an equal or larger maximum likelihood $P\left(\hat{x}'|D,\mathcal{H}',\theta'\right)$ simply by adding a degree of freedom d to θ, i.e. $\theta' = \theta \cup \{d\}$. Thus more complex models must be penalized appropriately. The Bayesian approach addresses this by integrating over the model parameter space, which in effect acts to automatically penalize overly-complex models. Bayesian model selection thus has a built-in safeguard against overfitting, effectively embodying Occam's razor [53].

Returning to our case, if we have two generative models of layered networks $\mathcal{H}_1$ and $\mathcal{H}_2$ with parameters $\{b_1 \equiv b_{i,1}^\alpha, \theta_1\}$ and $\{b_2 \equiv b_{i,2}^\alpha, \theta_2\}$, the evidence for the models according to the observed multiplex network $\mathcal{M}$ can be compared by evaluating the *(posterior) odds ratio*

$$\begin{aligned}\Lambda &= \frac{P\left(\mathcal{H}_2|\mathcal{M}\right)}{P\left(\mathcal{H}_1|\mathcal{M}\right)} \\ &= \frac{P\left(\mathcal{M}|\mathcal{H}_2\right)}{P\left(\mathcal{M}|\mathcal{H}_1\right)}\frac{P\left(\mathcal{H}_2\right)}{P\left(\mathcal{H}_1\right)} \\ &= \frac{P\left(\mathcal{M}|\mathcal{H}_2\right)}{P\left(\mathcal{M}|\mathcal{H}_1\right)},\end{aligned} \tag{4.30}$$

where we have used that $P\left(\mathcal{H}_2\right) = P\left(\mathcal{H}_1\right)$, i.e. we remain ignorant whether one of the models is more plausible than the other *a priori*. We prefer model $\mathcal{H}_2$ over $\mathcal{H}_1$ if $\Lambda > 1$, with increasing confidence for increasing values of Λ.

[8] Other approximations are possible; for example we could use the mean value of b_i^α, i.e. $\hat{b}_i^\alpha = \langle b_i^\alpha \rangle = \sum_{b_i^\alpha} P\left(b_i^\alpha|\mathcal{M}\right)$.

Using Equation (4.26) we obtain

$$\Lambda = \frac{\sum_{\theta_2,b_2} P\left(\mathcal{M}|\mathcal{H}_2,\theta_2,b_2\right) P\left(\theta_2,b_2|\mathcal{H}_2\right)}{\sum_{\theta_1,b_1} P\left(\mathcal{M}|\mathcal{H}_1,\theta_1,b_1\right) P\left(\theta_1,b_1|\mathcal{H}_1\right)}. \tag{4.31}$$

Evaluating Λ is very costly computationally if the phase space of $\{\theta_i, b_i\}$ is large, which it is in our case. Therefore we can approximate (4.31) in exactly the same spirit as the MLE in Equation (4.29) by finding for each model $\mathcal{H}_i$ the set of parameters $\{\hat{\theta}_i, \hat{b}_i\}$ that maximizes the expression

$$P\left(\mathcal{M}|\mathcal{H}_i,\theta_i,b_i\right) P\left(\theta_i,b_i|\mathcal{H}_i\right), \tag{4.32}$$

thus we write

$$\begin{aligned}\{\hat{\theta}_i, \hat{b}_i\} &= \text{argmax}_{\{\theta_i,b_i\}} P\left(\mathcal{M}|\mathcal{H}_i,\theta_i,b_i\right) P\left(\theta_i,b_i|\mathcal{H}_i\right) \\ &= \text{argmax}_{\{\theta_i,b_i\}} P\left(\theta_i,b_i|\mathcal{M}\right),\end{aligned} \tag{4.33}$$

where we used that the denominator in Equation (4.25) is independent from θ_i and b_i, acting in this respect only as a normalization constant. Thus we see that we seek the model parameters that maximize the Bayesian posterior probability: we call $\hat{\theta}_i$ and $\hat{b}_i$ the inferred community structure and the inferred additional model parameters of model $\mathcal{H}_i$, respectively. These can be found approximately by employing Markov chain Monte Carlo (MCMC) algorithms that traverse the phase space of $\{\theta_i, b_i\}$ in an intelligent way, searching for the global optimum of $P\left(\mathcal{M}|\mathcal{H}_i,\theta_i,b_i\right) P\left(\theta_i,b_i|\mathcal{H}_i\right)$ [23].

The expression in Equation (4.32) can also be cast in the following form:

$$\Sigma_i = -\log P\left(\mathcal{M}|\mathcal{H}_i,\theta_i,b_i\right) - \log P\left(\theta_i,b_i|\mathcal{H}_i\right). \tag{4.34}$$

Σ_i is called the *description length* (DL) of $\mathcal{M}$ according to the model $\mathcal{H}_i(\theta_i, b_i)$. It measures the amount of information required to describe the data, if we encode it using the particular parametrization of the generative model given by $\{\theta_i, b_i\}$, as well as these parameters themselves [25]. Thus fitting the parameters of a model $\mathcal{H}_i$ according to Equation (4.33) is equivalent to minimizing the DL of the model,

$$\{\hat{\theta}_i, \hat{b}_i\} = \text{argmin}_{\{\theta_i,b_i\}} \Sigma_i, \tag{4.35}$$

which constitutes the MDL principle. Thus the Bayesian inference of the 'optimal' parameters (optimal only in the sense of Equation 4.33) corresponds to minimizing the DL of the observed data; hence this approach amounts to finding the model $\{\mathcal{H}_i, \theta_i, b_i\}$ that most compresses the observed multiplex network $\mathcal{M}$ [25].

Returning to model comparison, we finally have

$$\Lambda \approx \frac{P(\mathcal{M}|\mathcal{H}_2,\hat{\theta}_2,\hat{b}_2)P(\hat{\theta}_2,\hat{b}_2)}{P(\mathcal{M}|\mathcal{H}_1,\hat{\theta}_1,\hat{b}_1)P(\hat{\theta}_1,\hat{b}_1)}. \tag{4.36}$$

Denoting the MDL of the models as $\Sigma_i^{\min}$, this can be stated equivalently as

$$\log\Lambda \approx \Sigma_1^{\min} - \Sigma_2^{\min}. \tag{4.37}$$

Thus if $\Lambda > 1$, then $\log\Lambda > 0$ and $\Sigma_1^{\min} > \Sigma_2^{\min}$; in the MDL picture this indicates that $\mathcal{H}_2$ compresses the data better than $\mathcal{H}_1$, hence should be preferred to it.

Generative models of layered networks

The generative models available in the `graph-tool` library [29] are the stochastic blockmodel SBM, the degree-corrected SBM (DCSBM) and two layered variants, called the layered (DC)SBM with edge covariates (EC) and the (DC)SBM with independent layers (IL).

A small remark is in place before discussing the generative models. What is meant by 'generative' is that during the MCMC algorithm new configurations of the parameters are generated stochastically at each step, based on the previous parameter configuration. More specifically, denote the general parameter set at step k as θ_k. Then a newly proposed configuration θ_{k+1} is more likely to be accepted if it has a considerably smaller DL than that of θ_k. If the Markov chain satisfies certain criteria, then after a sufficient amount of steps, known as the mixing time, each observed set of parameters θ occurs with a probability $\propto P(D|\theta)$, where we have denoted the data to which the model $\mathcal{H}(\theta)$ is fitted as D. Thus the parameter set with the MDL $\hat{\theta}$ is likely to be observed, as $P(D|\hat{\theta})$ attains its maximum value for $\theta = \hat{\theta}$.

SBM and DCSBM. A stochastic blockmodel (SBM) is a generative model for communities in networks. In the simplest variant, each of N nodes is assigned to one of B groups via $b_i \in [0, B-1]$ and undirected edges are placed independently between node pairs (i, j) with probabilities ψ_{rs} with $r = b_i$ and $s = b_j$ [110]. Thus in the terminology of the above paragraphs, θ is the $B \times B$ matrix ψ_{rs}. The SBM can describe a wide variety of different network structures from simple modular structures (ψ_M) to core-periphery structure (ψ_{CP}, where banks i with $b_i = 0$ (1) are the core (periphery) banks). The prototypical ψ_{rs} matrices in undirected networks with $B = 2$ communities for these cases are

$$\psi_M \propto \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \ \psi_{CP} \propto \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}. \tag{4.38}$$

The former is also referred to as the traditional community structure – a set of communities with dense internal connections and sparse external ones, or even none, as is the case for ψ_M [26].

A limitation of the traditional SBM is that every node in a group has the same expected degree, as the edges are distributed randomly between between the groups. This is circumvented by the DCSBM by conditioning the edge placement on a new parameter, the expected degree of the nodes [110]. Although adding a degree of freedom increases the DL *a priori*, the DCSBM often provides a smaller MDL than that of the SBM for networks that feature significant degree variability, which justifies the increased modeling complexity in these cases.

Layered (DC)SBM with EC. The observed multiplex network $\mathcal{M}$ has M layers, N nodes and $E = \sum_\alpha E^\alpha$ edges; in addition we suppose it is an undirected simple multiplex network. Denote the parameters of the layered SBM model with edge covariates as θ. Then the process that generates $\mathcal{M}$ with a probability $P(\mathcal{M}|\theta)$ is the following. First a collapsed graph from the traditional SBM ensemble is generated, i.e.

the nodes are randomly assigned to B communities via the b_i vector and E edges are placed randomly between the communities, according to the ψ_{rs} matrix.

Let $e_{rs} = E\psi_{rs}$ be the number of edges placed between groups r and s. These edges are then assigned randomly to the layers; the number of edges that have been placed in layer α is denoted by e_{rs}^{α}. The following identity summarizes the partition of the edges:

$$2E = \sum_{rs} e_{rs} = \sum_{rs}\sum_{\alpha} e_{rs}^{\alpha}, \tag{4.39}$$

where by convention we count the number of edges twice in e_{rs} and e_{rs}^{α} if $r = s$. For a given pair of groups (r, s), each assignment of the edges to the layers e_{rs}^{α} is equally probable, and the probability is given by [25]

$$\frac{\prod_{\alpha} e_{rs}^{\alpha}!}{e_{rs}!}, \tag{4.40}$$

since the edges are indistinguishable. Then, with $\theta = \{b_i, e_{rs}^{\alpha}\}$, the likelihood of observing $\mathcal{M}$ is [25]

$$P(\mathcal{M}|\theta) = P(G^c|\theta)\prod_{r\leq s}\frac{\prod_{\alpha} e_{rs}^{\alpha}!}{e_{rs}!}, \tag{4.41}$$

where G^c is the collapsed graph associated with $\mathcal{M}$. The generative process of the layered DCSBM is completely analogous; the only difference is that in the first step the DCSBM model is used to generate a collapsed graph. Because the edges are distributed across the layers in an entirely random manner, nodes with a large collapsed degree will also tend to possess uniformly larger degrees in all layers; thus this model does not allow for (considerable) degree variability across layers.

Layered (DC)SBM with IL. In this model the layers are generated according to the (DC)SBM independently, with the constraint that the group memberships of the nodes are identical across all layers. In addition, nodes are allowed to be inactive in some layers, since the DCSBM allows for nodes to have a zero (expected) degree. The layered DCSBM with IL requires for each layer α the expected degrees of the nodes and ψ_{rs}^{α}, while the node group memberships b_i are globally applied. Thus this model allows for degree variability across layers, i.e. a node with a large degree in layer α may have a much lower degree in layer β, but requires M additional degree sequences compared to the layered DCSBM with EC.

Results for the undirected and directed view of the Russian multiplex interbank network

To test whether the term layers in the multiplex interbank network $\mathcal{M}$ are significant with respect to the mesoscale of the complete interbank network,

i.e. the collapsed graph G^c, we compare the MDL of two models for each epoch. We emphasize that we only consider the undirected and directed views of $\mathcal{M}$ and G^c for simplicity. As before, the number of nodes and edges in $\mathcal{M}$ and G^c are N and E, respectively.

The first model $\mathcal{H}_l$ is a layered DCSBM with EC that is fitted to the multiplex interbank network with the $M = 8$ term layers. The second model is a null model $\mathcal{H}_0$ where the edges are deposited randomly among the term layers. The distribution of the indistinguishable edges is parametrized by the number of edges in each layer E^α. Accordingly, such a distribution has a probability

$$\frac{\prod_\alpha E^\alpha!}{E!} \tag{4.42}$$

of occurring. Then the likelihood of observing $\mathcal{M}$ according to this model is

$$P(\mathcal{M}|\theta, E^\alpha) = P(G^c|\theta)\ \frac{\prod_\alpha E^\alpha!}{E!}, \tag{4.43}$$

where $P(G^c|\theta)$ is the likelihood of observing G^c according to the (DC)SBM.

After fitting $\mathcal{H}_l$ and $\mathcal{H}_0$ to $\mathcal{M}$ and obtaining their MDLs $\Sigma_l^{\min}$ and $\Sigma_0^{\min}$, the models are compared as in Equation (4.37), i.e.

$$\log \Lambda = \Sigma_l^{\min} - \Sigma_0^{\min}. \tag{4.44}$$

If $\log \Lambda \ll 0$, the incorporation of the term layers is correlated with the group structure and the layered structure is very informative on the network structure. In the opposite case, if $\log \Lambda \gg$, then we conclude that the term layers are not informative with respect to the mesoscale; indeed, in the null model the edges are distributed among the layers in a manner that is entirely independent of the group structure, thus the layers may be considered to be redundant and the collapsed graph alone suffices to represent the mesoscale structure of the multiplex interbank network.

Before comparing the two models, one still needs to find the optimal models for $\mathcal{M}$ and G^c, according to the model evidence (see Equation 4.26). We report that for the collapsed graph in each month and epoch the DCSBM model was preferred, for both the directed and undirected view; thus for $\mathcal{H}_0$ we calculate $P(G^c|\theta)$ in (4.43) using the DCSBM. Note that this was expected since the degree distributions are heavy-tailed in the aggregated network, as was shown in Chapter 3. For the layered models, again for each month and for each epoch the layered DCSBM with EC strongly rejected the IL variant. Consequently $\mathcal{H}_l$ is the layered DCSBM.

The results in Figure 4.7 are unambiguous: $\mathcal{H}_0$ is significantly rejected by $\mathcal{H}_l$ for each epoch and for both views. The fact that the layered DCBSM with EC outperforms the null model indicates that the interactions between banking communities differs significantly across the layers. This suggests (1) that the lending and borrowing behavior between bank communities depends on the term of the loan contract, which agrees with the results obtained in Chapter 3, and (2) that aggregating all terms in the Russian interbank network obscures structural features of the interbank topology.

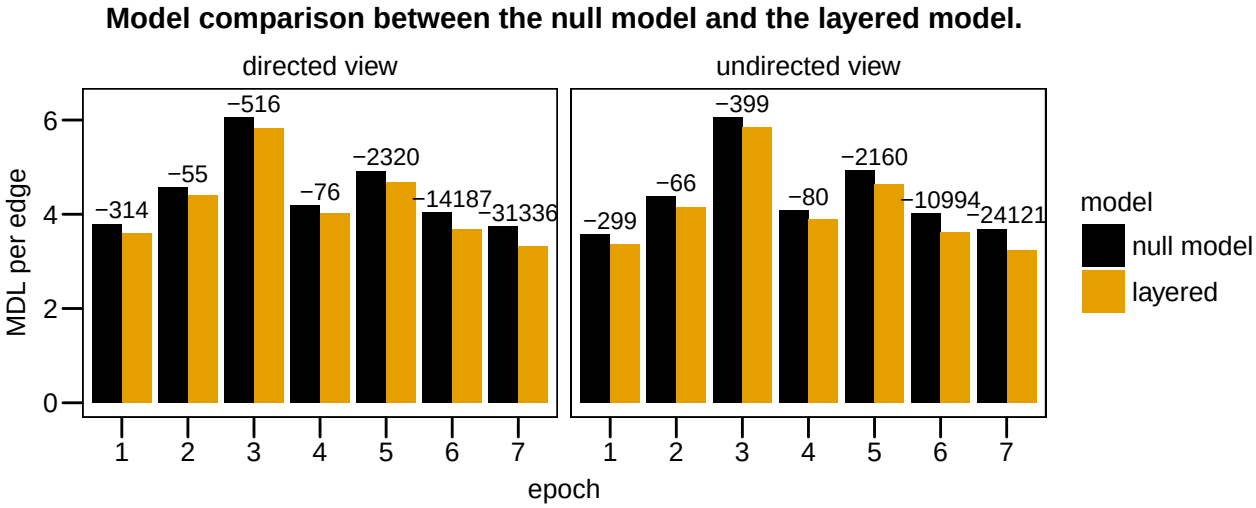


Figure 4.7: Comparisons of the MDLs $\Sigma_l^{\min}$ and $\Sigma_0^{\min}$ of $\mathcal{H}_l$ and $\mathcal{H}_0$ respectively, normalized by the amount of edges in the (un)directed view of each epoch. The value of $\log \Lambda = \Sigma_l^{\min} - \Sigma_0^{\min}$ is displayed on top of the bars.

Chapter 5

Summary

The goal of this thesis is to contribute a comprehensive methodological survey of the Russian interbank lending network to the relatively undeveloped literature on multiplex interbank networks. The key theme throughout the thesis is the study, concept and relevance of scales in complex networks. We argue that interbank lending networks are complex systems with emergent features which can be captured in an elegant manner with network theory. In line with the recent developments therein, we apply the multiplex formalism on the interbank loan market by artificially placing all loans contracts with the same term in separate layers. This breaking up is done in order to study and characterize the topology and structural features of the term layers. Finally, we investigate whether the term layer representation of the interbank network is worth the effort, and to which extent layers may be aggregated to achieve a more compact description of the Russian multiplex interbank network.

Results

Our first achievement is the tackling of inferring issuance dates of interbank transactions from the due date and a partial knowledge of the contract's maturity. This was done in order to access the weekly time scales in the data set; although the issuance week for shorter terms is well defined (a major reason why earlier surveys of this data set opted to treat only the first two terms), this becomes increasingly tricky for longer terms. The issuance date algorithm provides a way to still include these loans in analysis on a weekly time scale. It is available in the form of a `R` and `Python` library – well documented and with plenty of examples – in the hope that it may be of use for future research.

A second result is the characterization of the topological structure of the term layers. After listing stylized facts of (single–layered) interbank networks, each term layer was carefully scrutinized. We have found that although all term layers exhibit hub structure, i.e. centralized banks that lend to and borrow from a large amount of counterparties, the financial function of these hubs is layer-specific. Using a simplistic model of liquidity dissipation we showed that the hubs in the layers with terms `<1d`, `2-7d`

and `8-30d` behave as tiering banks that intermediate liquidity between periphery banks. Then we have modified and extended the core-periphery algorithm of Della Rossa et al. (2013) [1] in order to achieve two things. First, the requirement of strongly connectedness was resolved by using a PageRank centrality measure, which was desired as we have shown that the strongly connected component is unacceptably small in layer with longer terms, and the economical interpretation of taking only these banks into account is rather unclear. Second, the original core-periphery algorithm could only infer core-periphery structure. We have replaced the α-profile of Della Rossa et al. by intermediation profiles, which allowed us to characterize the intermediation activity of the hub in each term layer.

Doing that, we found that three classes of terms can be discriminated based on the underlying topology of the term layers. These are the tiering layers, the flat layers and the source-sink star layers. The tier layers consist of loans with terms `<1d`, `2-7d`, `8-30d` and `31-90d`. Their interbank lending network is organized in a core-periphery structure. The hubs are the largest intermediaries in the network, acting as money centers that dissipate, generate and redistribute liquidity. This tiering structure causes systemic risk associated with high contagion risk should the core banks malfunction or default. The flat layers consist of loans with terms `91-180d` and `.5-1y`. While the hubs still perform the largest part of the intermediation in the network, the periphery is significantly more interconnected compared to the tiering layers. This causes less steep intermediation profiles. The networks in the flat layers consist of many unconnected components, the majority of which only count several banks; the largest components are dense and harbor the hubs. Finally, the hubs of the source-sink star layers with terms `1-3y`, `>3y` are sinks or sources of liquidity, connected to periphery banks which are themselves almost unconnected. The networks consist of star-like hub structures and a number of small unconnected components.

Our final important results relate to the reducibility of the term layers. The above classification of the layers was largely reproduced by the optimal layer bins according to the Jensen-Shannon distance. We implemented the methodology of De Domenic et al. (2015) [2] to find the optimal layer partition of the term layers according to maximal relative entropy between the aggregated graph and a multiplex network with some of the layers binned. Furthermore, we showed how the Jensen-Shannon distance can be used for image recognition and for inferring epochs in temporal layers. The temporal change points between the epochs have mostly been found to agree with underlying historical events and the maturity and growth periods of the network. Lastly, we found considerable agreement between an established similarity measure, the Jaccard distance, and the epoch partition, which is an encouraging sign for the rather poorly understood von Neumann entropy of a graph. To test the layer significance in a more robust framework, we have performed a Bayesian model comparison to have the last word on whether the layers are informative of the general mesoscale structure of the interbank network, and found that they significantly are.

Outlook

We hope that this work can be a valuable source for further research on the multiplex representations of interbank lending markets, since it both provides tools to do so and an in-depth case study of the Russian interbank network. Furthermore, many interesting concepts have been deliberately presented in detailed fashion to make this work more accessible.

In any sense, much of the work done in this thesis can be improved. For example, more research is needed to fully characterize our extension of the core-periphery algorithm of Della Rossa et al, especially regarding the selection of the most important banks. The inferring of epochs can be improved considerably by using more sophisticated methods, such as [25, 30, 31]. Likewise, the Bayesian model selection can be extended to infer optimal layer bins and not just check whether the term layers correlate with the network mesoscale. In fact, some progress on the latter was made during research for this thesis. In the same sphere, nested SBMs [28] allow one to describe the community structure in network at arbitrary scales; surely an interesting application for any interbank network. Lastly, we have discarded the deposit transactions in the beginning of the analysis. In future work these could be absorbed in the analysis as an additional layer. Furthermore, we did not use balance sheet information such as bank size etc. in our survey. Such valuable information can be used to complement and refine further research into the topological properties of the Russian multiplex interbank network.

Appendix A

Selected events during interbank market crises

These selected events serve to illustrate the forces shaping the interbank network during distress. This Appendix has been compiled from [7, 38, 111, 112].

A.1 August 1998 ruble crisis

A severe banking, currency and sovereign debt crisis that completely paralyzed the interbank market.

`August 13`	The Russian stock, bond and currency markets collapse as a result of fears for a ruble devaluation and a default on domestic debt.
`August 17`	Emergency measures are announced: abandoning of the exchange rate regime, default on domestic public debt and a moratorium on all private foreign liabilities.
`September 2`	The ruble becomes a freely floating currency.
following weeks	Sharp depreciation of ruble, strong inflation. Social unrest and protesting.
`November 20`	Government declares that it can pay back 60% of foreign debt.
following weeks	Russian bank deposits decrease by 15% compared to August.
`January 1999`	Interbank market stability hindered by the 1998 turmoil largely restored.

A.2 Trust crises

The trust crises consist of the announcement crisis in the second half of 2003 and the summer of 2004 crisis.

Investigations of money laundering led the CBR to deprive Sodbusinessbank of its license in May 2004. The following mutual suspicion led to a drying up of liquidity on the interbank market in the summer of 2004. Roughly one year earlier, this investigation was announced, which caused a smaller trust crisis (i.e. the announcement crisis).

end of July 2003	Start of announcement crisis. Decline in reciprocal bank interactions.
September 2003	The network starts to fall apart because of the distrust among banks.
December 2003 - January 2004	Back to normal liquidity.
March	Peak of interbank lending market recovery from early trust crisis. The bank lending reciprocity, however, does not regain pre-crisis levels.
following weeks	Three stages of the summer of 2004 crisis.
First stage: April	Volatility on the interbank market with even higher lending rates than later stages but without significant outflow of individual's deposits. Demand for liquidity was caused by policy changes and statements of the CBR; no perception of crisis by the banks themselves, yet this financial instability undoubtedly impacted the crisis to come.
Second stage: May	CBR deprives Sodbusinessbank of its license.
May 19-21	Several conflicting statements from authorities about the deposit insurance for the clients of Sodbusinessbank make the depositors increasingly uneasy.
Third stage: June 3	Crisis now definitively developed beyond Sodbusinessbank alone. Banks start to introduce additional control measures; several suspend lending activities on the interbank market.
June 11	CBR changes policy rates to accommodate the banks, officially for the low liquidity on the interbank short-term market.
June 21-22	Peak of the crisis.
July 13	Interbank market starts to stabilize.
July 16	The 'crisis of confidence' is declared to be at an end.

Appendix B

Measure concentration for N independent Gaussian distributions

The following derivation has been taken directly from MacKay (2005) [53, p. 124] and is restated here for the reader's convenience.

Consider the vector $\boldsymbol{x} = (x_1, x_2, ..., x_N)$ with distribution

$$P(\boldsymbol{x}) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{\sum_i x_i^2}{2\sigma^2}\right). \tag{B.1}$$

This is a multivariate Gaussian distribution of N independent variables x_i with each $x_i \sim \mathcal{N}(0, \sigma^2)$ such that $P(x_i) = (2\pi\sigma^2)^{-1/2} \exp\left(-x_i^2/2\sigma^2\right)$ and $P(\boldsymbol{x}) = \prod_i P(x_i)$.

We can illustrate the measure concentration by examining the volume likely to contain a sample $\boldsymbol{x}$ drawn according to (B.1). Then $\boldsymbol{x}$ is a point in N-dimensional space whose radius is $r = |\boldsymbol{x}|$ and by spherical symmetry this volume is characterized by the mean and standard deviation of r, denoted by $\langle r \rangle$ and δr, respectively.

Since $r^2 = \sum_i x_i^2$ is a sum of N independent variables and $\langle x_i^2 \rangle = \mathrm{var}(x_i) + \langle x_i \rangle^2 = \sigma^2$, we have that $\langle r^2 \rangle = N\sigma^2$ or

$$\langle r \rangle = \sqrt{N}\sigma. \tag{B.2}$$

Given that $\langle x_i^4 \rangle = 3\sigma^4$, by a similar argument $\mathrm{var}(x_i^2) = \langle x_i^4 \rangle - \langle x_i^2 \rangle^2 = 2\sigma^4$ thus $\mathrm{var}(r^2) = 2N\sigma^4$.

For large N, the central-limit theorem indicates that r^2 has a Gaussian distribution with mean $N\sigma^2$ and standard deviation $\delta r^2 = \sqrt{2N}\sigma^2$, so the probability density of r must similarly be concentrated about $r \simeq \sqrt{N}\sigma$.

The thickness of this shell is given by turning the standard deviation of r^2 into a standard deviation on r: for small $\delta r/r$, we have

$$\frac{\delta r}{r} \approx \frac{1}{2}\frac{\delta r^2}{r^2} = \frac{1}{2}\frac{\sqrt{2N}\sigma^2}{r^2}, \tag{B.3}$$

so that with $r \approx \sqrt{N}\sigma$

$$\delta r \approx \frac{\sigma}{\sqrt{2}}. \tag{B.4}$$

We conclude that for sufficiently large N almost all the probability measure lies in a volume with radius $\sqrt{N}\sigma$ and thickness $2 \times \delta r \approx \sqrt{2}\sigma$.

Appendix C

Additional figures for Chapter 3

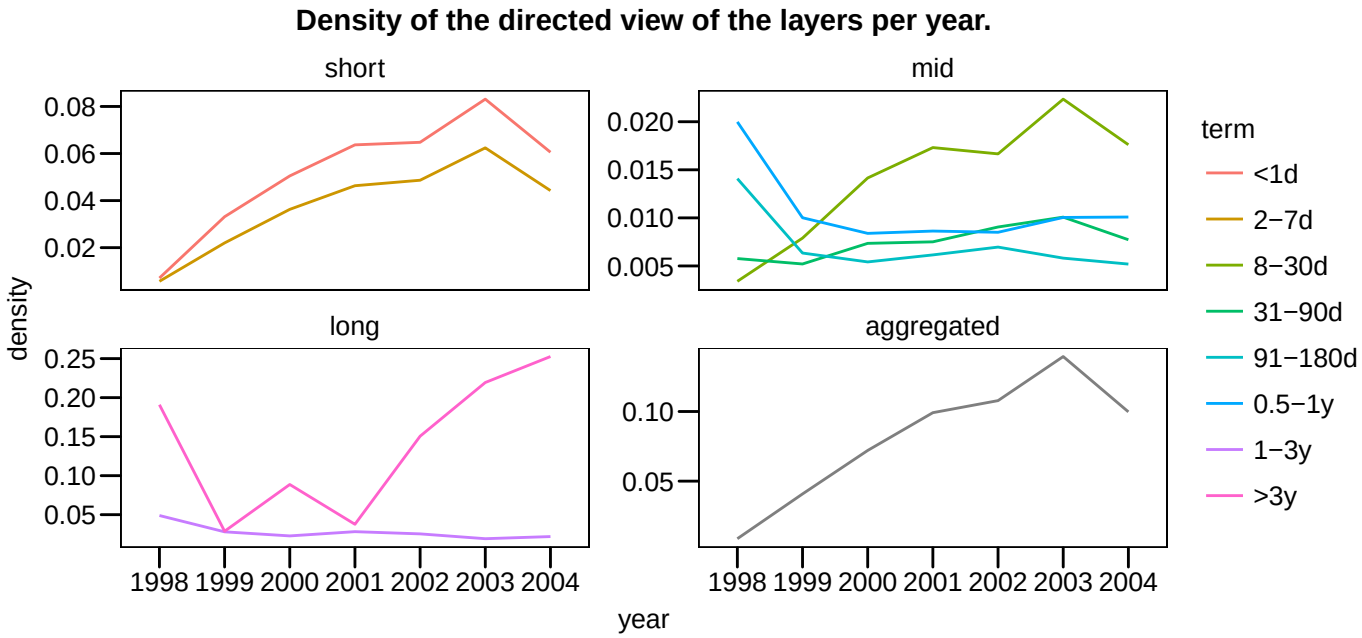


Figure C.1: The directed density for a simple graph with N nodes and E edges is defined as $d = \frac{E}{N(N-1)}$.

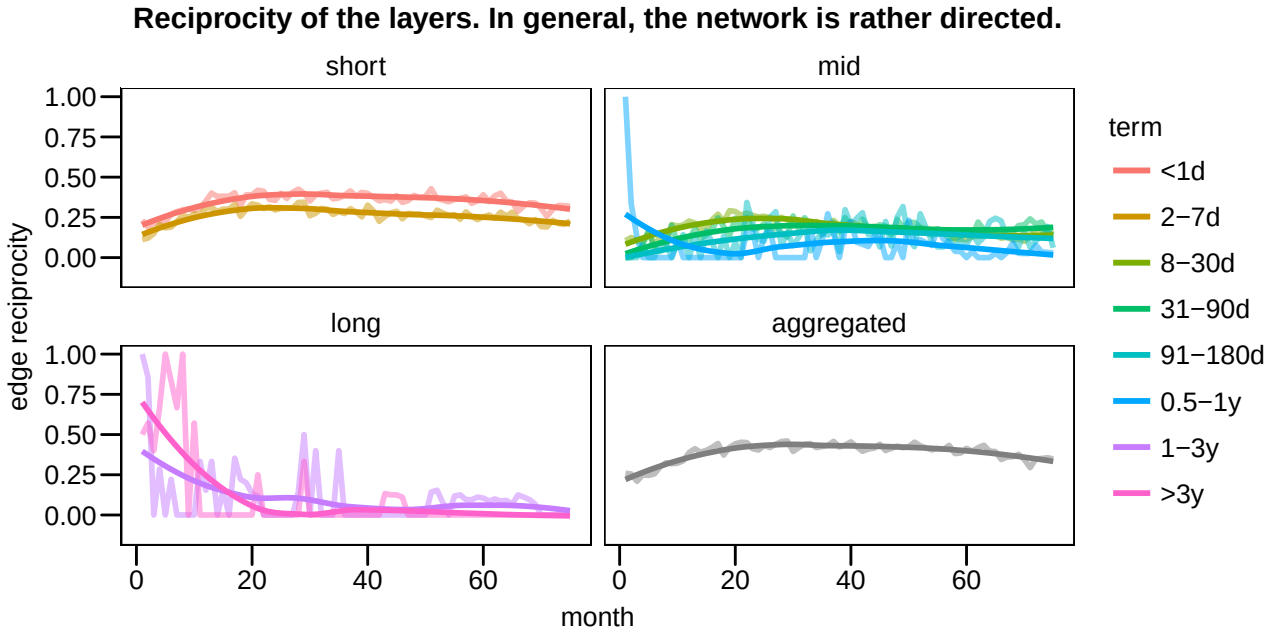


Figure C.2: Reciprocity for directed views of the layers. The reciprocity $r = E_{\leftrightarrow}/E$ is defined as the fractional amount of bidirectional edges. An edge is included in $E_{\leftrightarrow}$ if at *least one* parallel adjacent to it exists. This implies that the directed view typically yields lower reciprocity values than the multi-directed view.

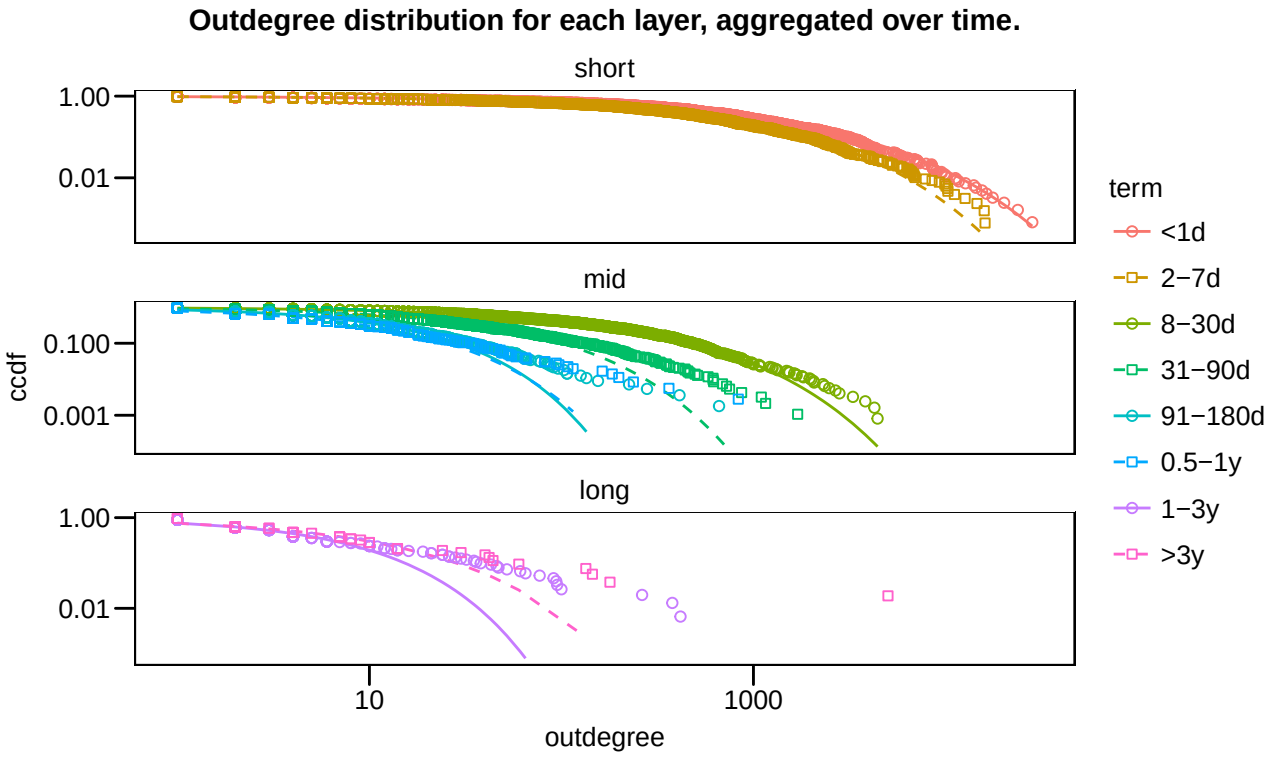


Figure C.3: Outdegree distribution with fitted stretched exponentials, drawn as piecewise linear functions connecting each predicted value. It is clear the correspondence with this distribution breaks down for terms longer than `8-30d`. Inferred (λ, β) parameters for $G^{\texttt{<1d}}$, $G^{\texttt{2-7d}}$, $G^{\texttt{8-30d}}$ are $(1.5 \times 10^{-3}, 5.3 \times 10^{-1})$, $(2.3 \times 10^{-3}, 5.7 \times 10^{-1})$, $(8.3 \times 10^{-3}, 6.1 \times 10^{-1})$ respectively, with all standard errors below 1%.

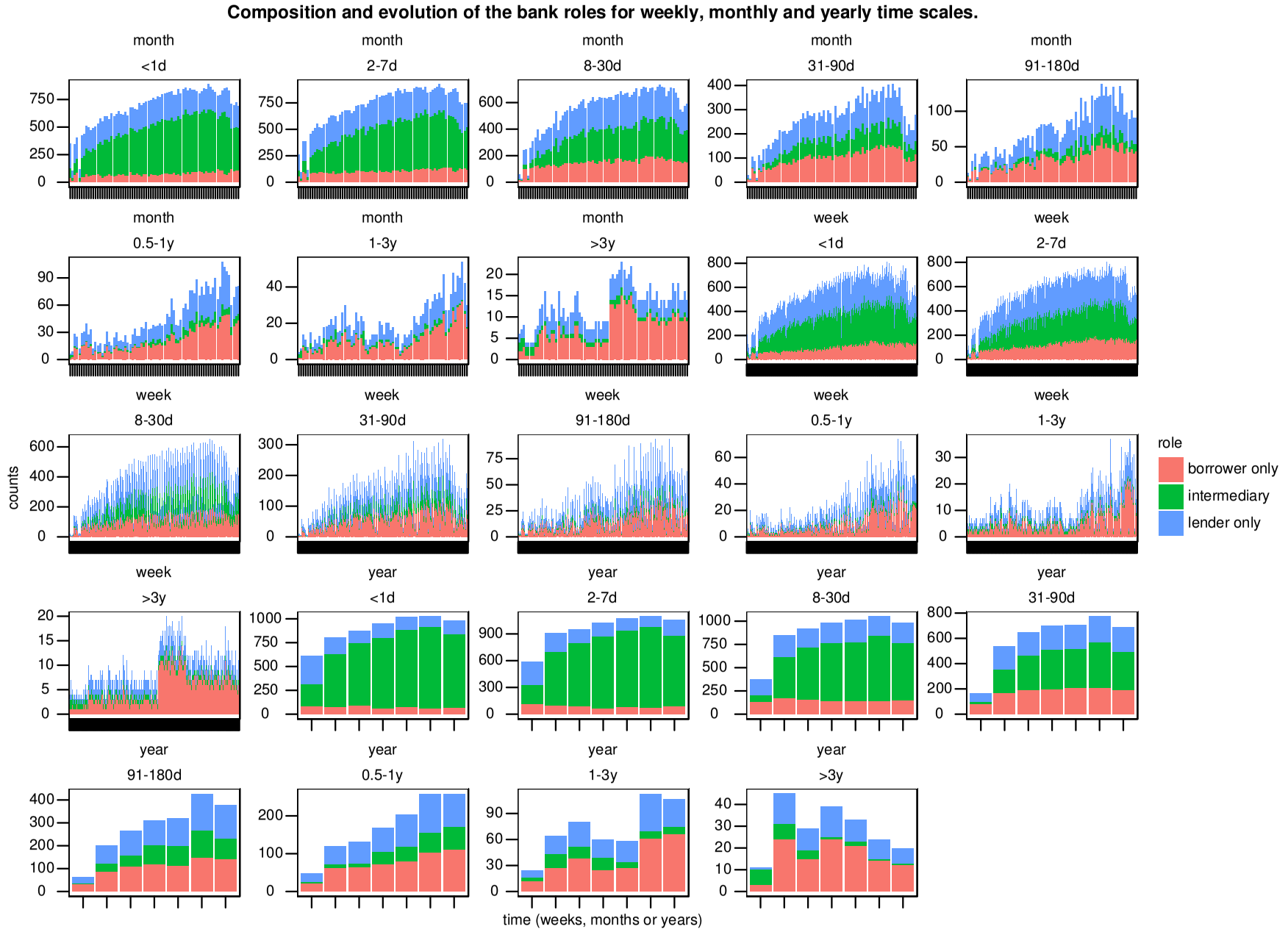


Figure C.4: Time series of the bank roles' composition for each term layer. Note the individual scales for each panel. The weekly time windows display a monthly periodicity with respect to hikes in the total amount of active banks in the market. This is caused by the banks having to fulfill capital requirements at the end of the month; many banks lend extra cash in the last week to comply with policy [8].

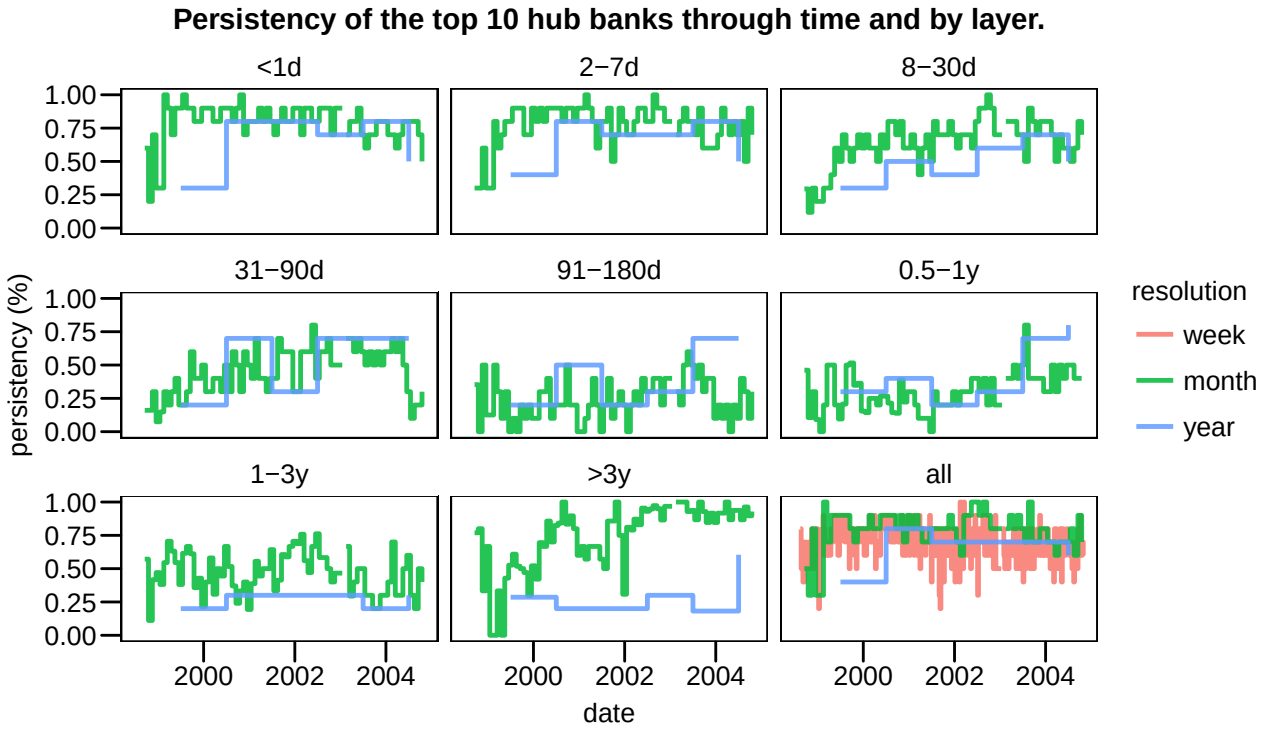


Figure C.5: Weekly resolution only available for the term-aggregated network because of limited computing time.

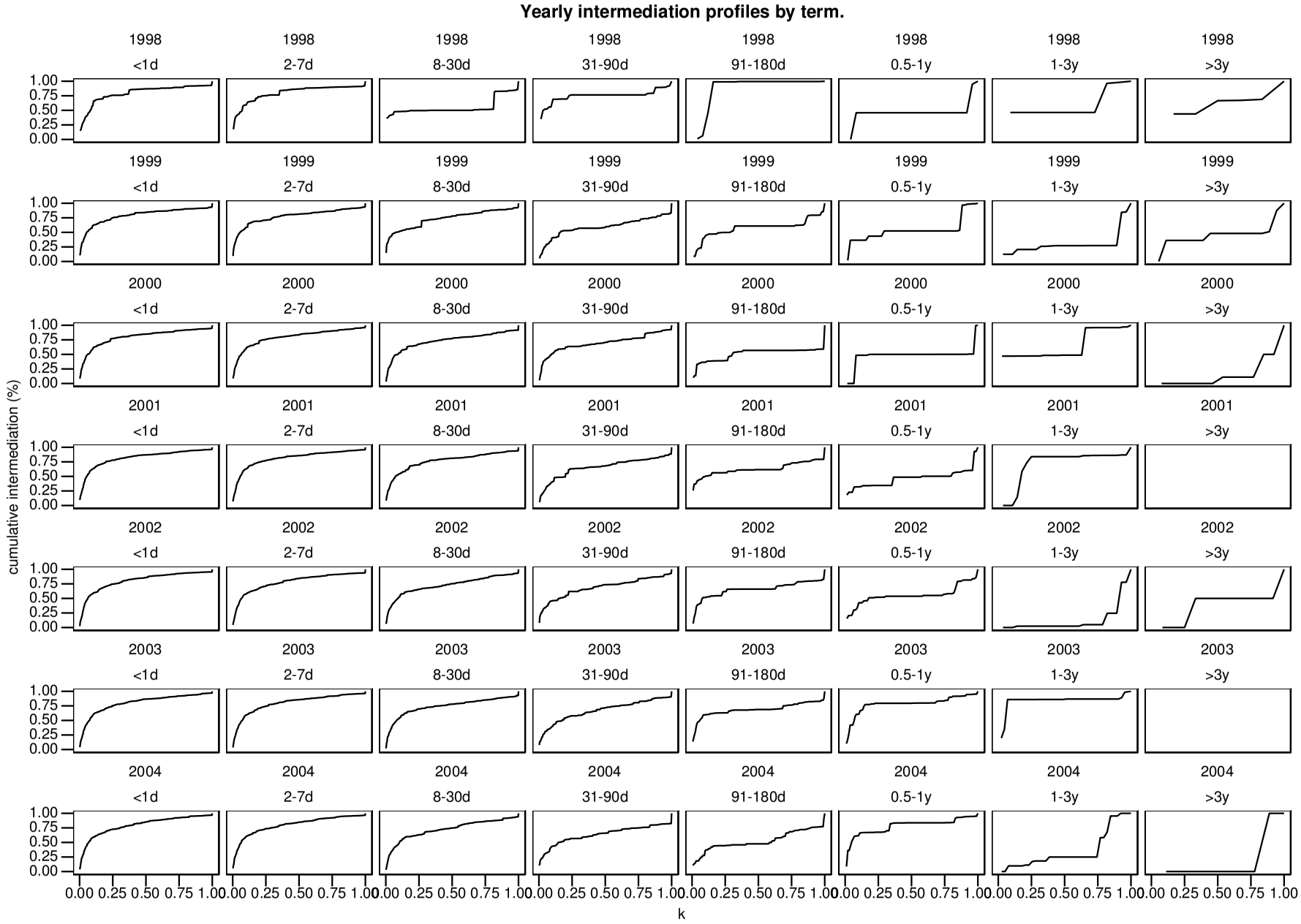


Figure C.6: k denotes the fraction of the total ranks traversed, e.g. $k = 0.5$ indicates the rank halfway, while for $k = 1$ these are the lastly ranked bank(s). No intermediation happened in $G^{1\text{-}3y}$ in 2001.

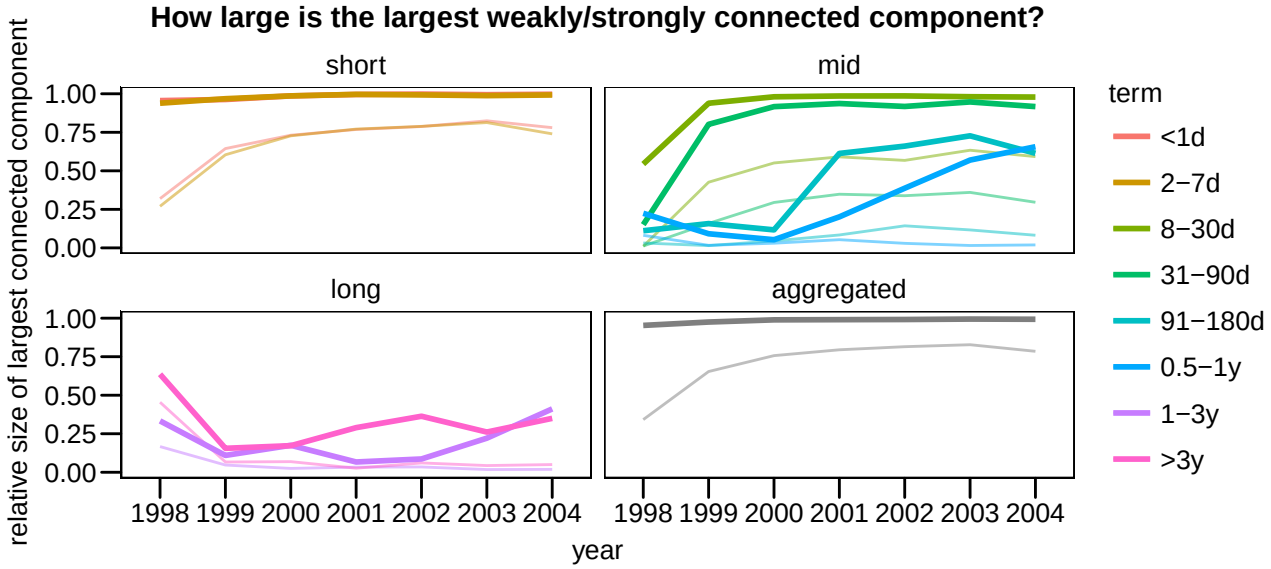


Figure C.7: The relative size of the largest connected component, given the weak or strong requirement, is calculated by dividing the amount of banks in the largest connected component by the total amount of active banks in a specific layer and year. The fat (thin) lines indicate the weak (strong) version.

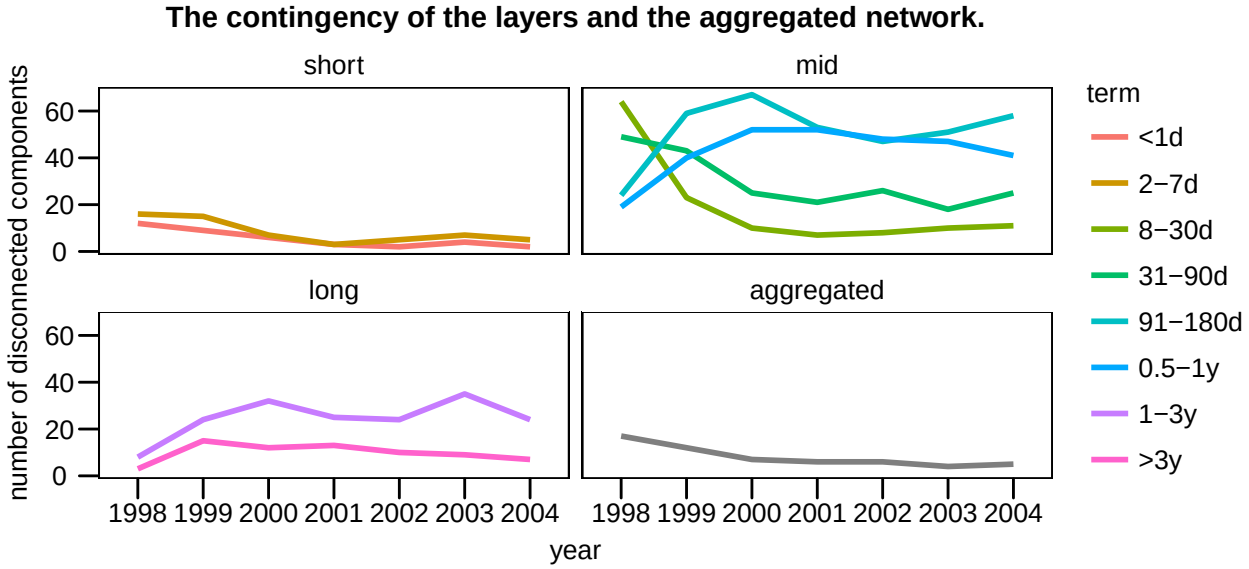


Figure C.8: To compare the number of (weakly) disconnected components with the amount of active banks per year, consult Table 3.4 on page 61. The relative size of the largest weakly connected can be found in Figure C.7.

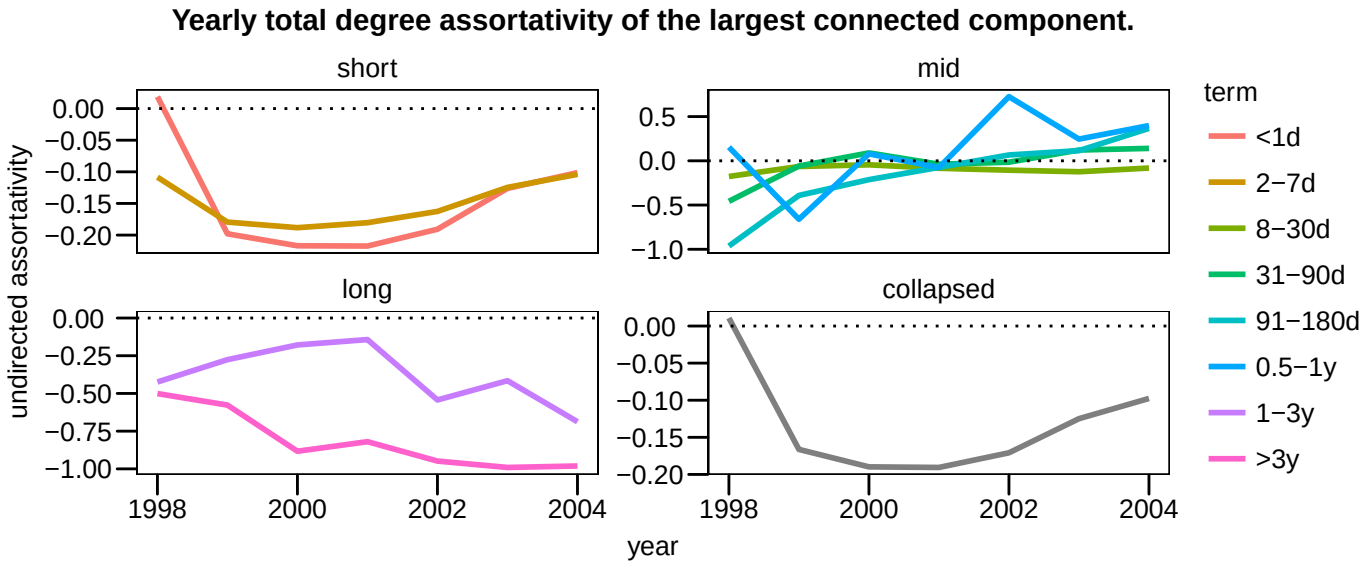


Figure C.9: The degree assortativity for the total degrees throughout the years. The degree assortativity is also called degree correlation [26].

Appendix D

Additional figures for Chapter 4

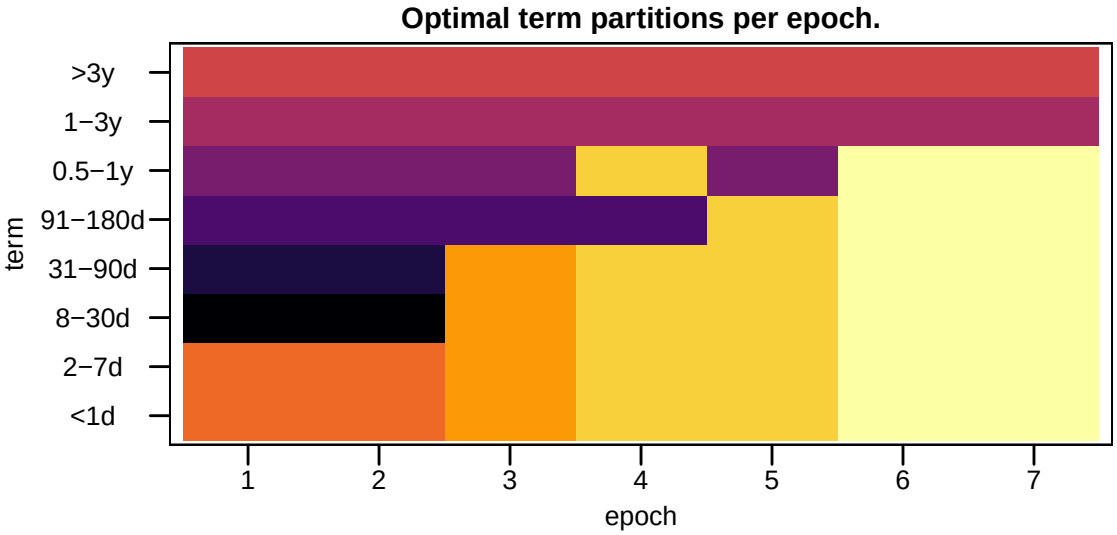


Figure D.1: This Figure is completely analogous to Figure 4.2, but the optimal term partition is found for each epoch rather than for each month. The epoch layer is built by collapsing the month layers that belong to it.

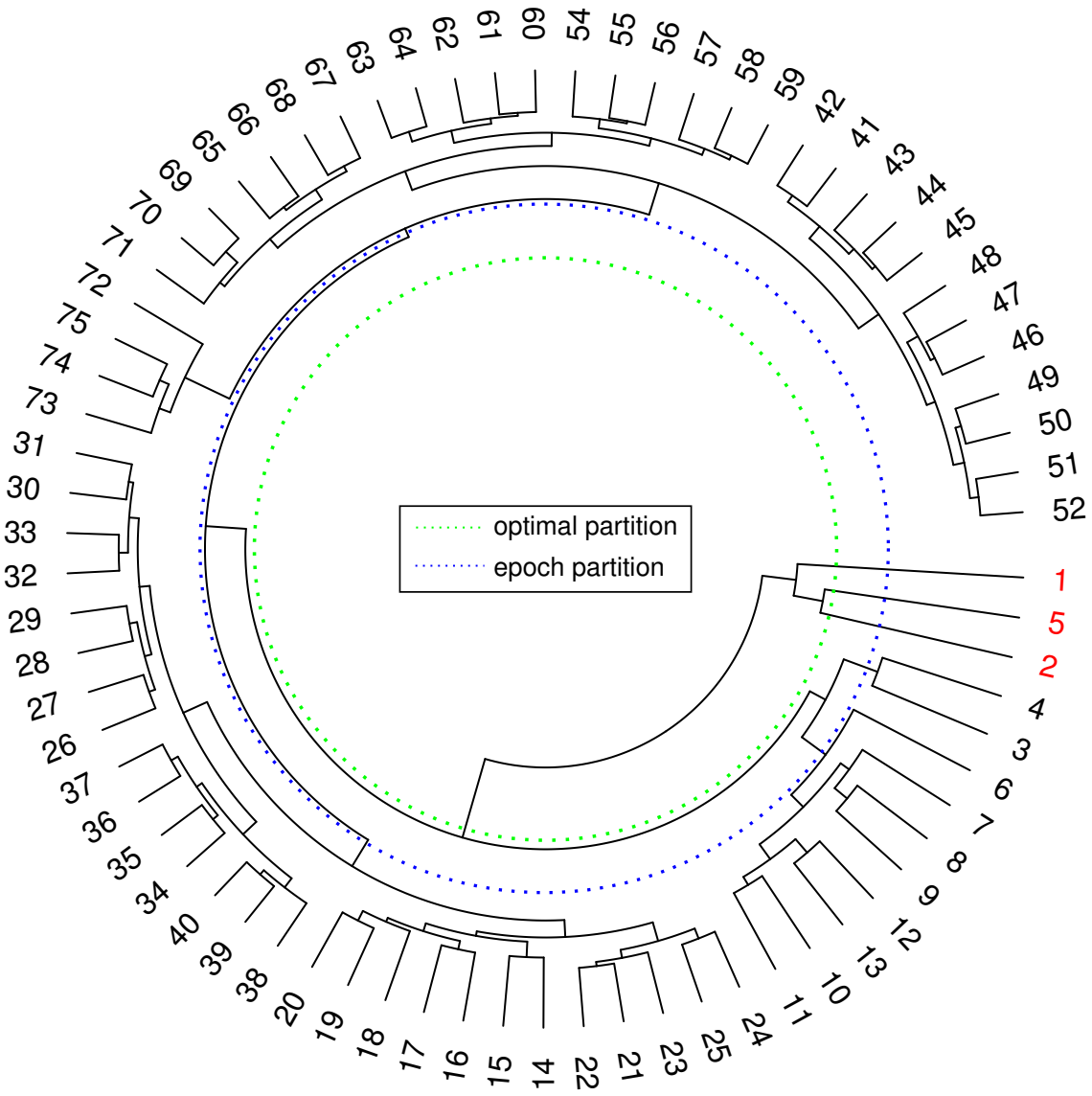


Figure D.2: Circular dendrogram for the monthly temporal layers in undirected view of the term-aggregated Russian interbank network. Non-adjacent binned months have been colored red.

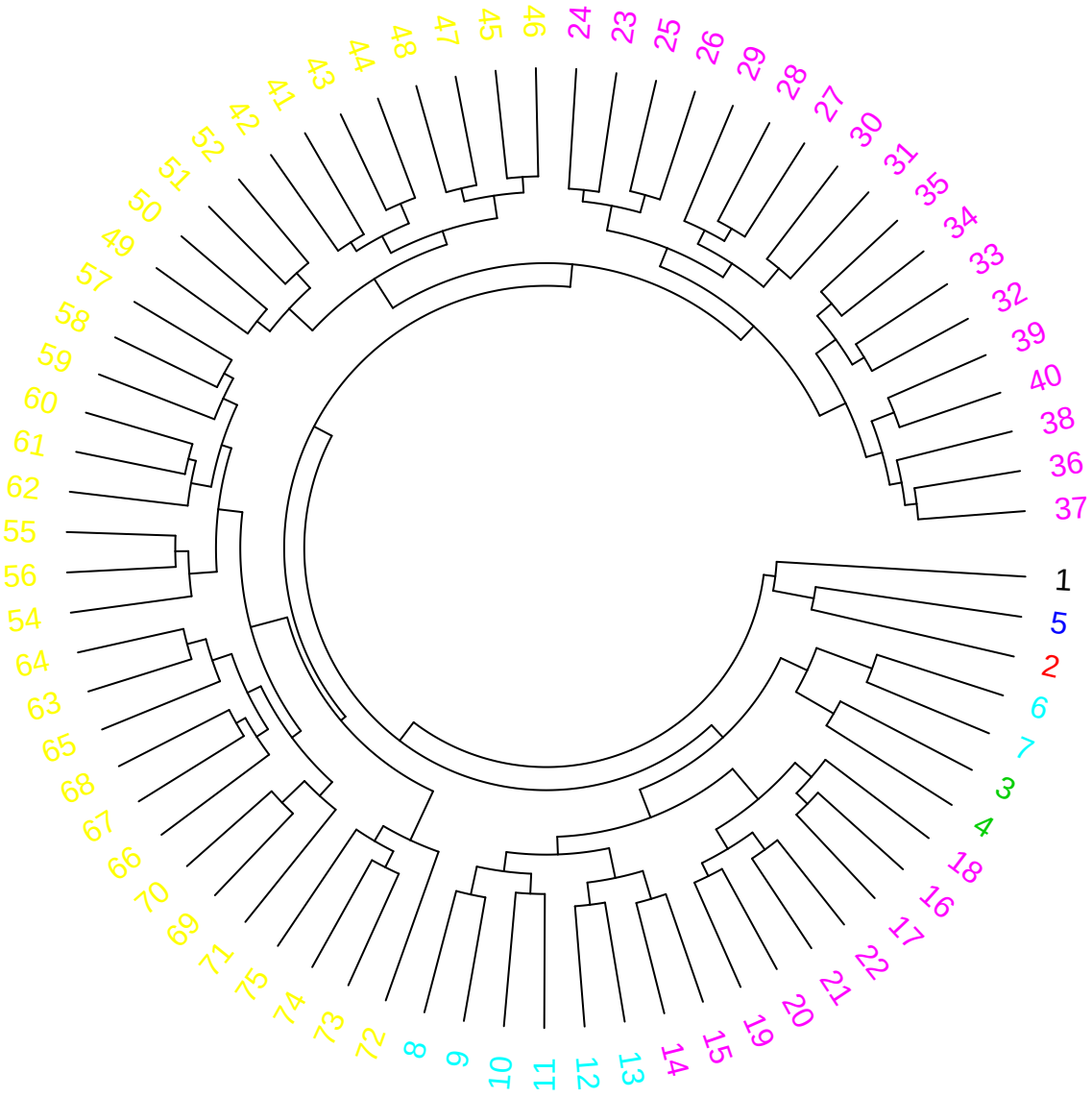


Figure D.3: Hierarchical clustering diagram obtained from the Jaccard distance between the directed views of the temporal layers. Each month is colored according to the epoch it belongs to.

Bibliography

[1] F. D. Rossa, F. Dercole, and C. Piccardi, "Profiling core-periphery network structure by random walkers," *Scientific Reports*, vol. 3, Mar. 2013.

[2] M. De Domenico, V. Nicosia, A. Arenas, and V. Latora, "Structural reducibility of multilayer networks," *Nature communications*, vol. 6, 2015.

[3] E. T. Jaynes, *Probability Theory: The Logic of Science*. Cambridge University Press, Apr. 2003.

[4] A. G. Haldane, "Rethinking the financial network," in *Fragile stabilität–stabile fragilität*, pp. 243–278, Springer, 2013.

[5] S. Boccaletti, G. Bianconi, R. Criado, C. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña Nadal, Z. Wang, and M. Zanin, "The structure and dynamics of multilayer networks," *Physics Reports*, vol. 544, pp. 1–122, Nov. 2014.

[6] J. Ladyman, J. Lambert, and K. Wiesner, "What is a complex system?," *European Journal for Philosophy of Science*, vol. 3, no. 1, pp. 33–67, 2013.

[7] M. Van den Heuvel, "Journey through the controllability of interbank credit networks," 2015.

[8] B. Vandermarliere, J. Ryckebusch, and K. Schoors, "Network analysis of the russian interbank system," 2012.

[9] U. Witt, "Self-organization and economics–what is new?," *Structural Change and Economic Dynamics*, vol. 8, pp. 489–507, Oct. 1997.

[10] D. L. Veld, *Complex Systems in Financial Economics: Applications to Interbank and Stock Markets*. Universiteit van Amsterdam [Host], 2014.

[11] T. P. Peixoto and S. Bornholdt, "No need for conspiracy: Self-organized cartel formation in a modified trust game," *Physical review letters*, vol. 108, no. 21, p. 218702, 2012.

[12] C. Sewell, "Is Ben Bernanke Having Fun Yet?," May 2010.

[13] A.-C. Hüser, "Too interconnected to fail: A survey of the interbank networks literature," SAFE Working Paper Series 91, Research Center SAFE - Sustainable Architecture for Finance in Europe, Goethe University Frankfurt, 2015.

[14] J. Wiemers and U. Neyer, "Why do we have an interbank money market?," tech. rep., IWH Discussion Papers, 2003.

[15] F. Blasques, F. Bräuning, and I. ¡!¿van Lelyveld, "A dynamic network model of the unsecured interbank lending market," *Social Science Research Network Working Paper Series*, Feb. 2015.

[16] K. Khashanah and T. Alsulaiman, "Network theory and behavioral finance in a heterogeneous market environment," *Complexity*, vol. 21, no. S2, pp. 530–554, 2016.

[17] F. Allen and D. Gale, "Financial contagion," *Journal of Political Economy*, vol. 108, pp. 1–33, Feb. 2000.

[18] F. Allen and A. Babus, "Networks in finance," 2008.

[19] F. Castiglionesi and M. Eboli, "Liquidity flows in interbank networks," 2015.

[20] V. Y. Guleva, M. V. Skvorcova, and A. V. Boukhanovsky, "Using multiplex networks for banking systems dynamics modelling," *Procedia Computer Science*, vol. 66, pp. 257–266, 2015.

[21] S. Battiston, M. Puliga, R. Kaushik, P. Tasca, and G. Caldarelli, "DebtRank: Too central to fail? financial networks, the FED and systemic risk," *Scientific Reports*, vol. 2, Aug. 2012.

[22] M. P. Rombach, M. A. Porter, J. H. Fowler, and P. J. Mucha, "Core-Periphery structure in networks," *SIAM Journal on Applied Mathematics*, vol. 74, pp. 167–190, Feb. 2014.

[23] T. P. Peixoto, "Efficient monte carlo and greedy heuristic for the inference of stochastic block models," *Physical Review E*, vol. 89, no. 1, p. 012804, 2014.

[24] K. Finger, D. Fricke, and T. Lux, "Network analysis of the e-mid overnight money market: the informational value of different aggregation levels for intrinsic dynamic processes," *Computational Management Science*, vol. 10, no. 2, pp. 187–211, 2013.

[25] T. P. Peixoto, "Inferring the mesoscale structure of layered, edge-valued, and time-varying networks," *Physical Review E*, vol. 92, no. 4, p. 042807, 2015.

[26] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, vol. 45, pp. 167–256, Jan. 2003.

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

[28] T. P. Peixoto, "Hierarchical block structures and high-resolution model selection in large networks," *Physical Review X*, vol. 4, no. 1, p. 011047, 2014.

[29] T. P. Peixoto, "The graph-tool python library," *figshare*, 2014.

[30] S. D. Ridder, B. Vandermarliere, and J. Ryckebusch, "Detection and localization of change points in temporal networks with the aid of stochastic block models," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2016, no. 11, p. 113302, 2016.

[31] T. P. Peixoto and M. Rosvall, "Modeling sequences and temporal networks with dynamic community structures," *ArXiv e-prints*, Sept. 2015.

[32] C. Ye, *Entropic Characterization and Time Evolution of Complex Networks*. PhD thesis, University of York, 2016.

[33] J. L. Molina-Borboa, F. López-Gallo, and M. van der Leij, "A multiplex network analysis of the mexican banking system: link persistence, overlap and waiting times," 2015.

[34] I. Aldasoro and I. Alves, "Multiplex interbank networks and systemic importance: An application to european data," 2016.

[35] L. Bargigli, G. Di Iasio, L. Infante, F. Lillo, and F. Pierobon, "The multiplex structure of interbank networks," *Quantitative Finance*, vol. 15, no. 4, pp. 673–691, 2015.

[36] M. Montagna, C. Kok, *et al.*, "Multi-layered interbank model for assessing systemic risk," 2013.

[37] A. Karas and K. Schoors, "A guide to russian banks data," *SSRN: http://ssrn.com/paper-1658468*, 2010.

[38] A. Karas, K. Schoors, *et al.*, "Bank networks, interbank liquidity runs and the identification of banks that are too interconnected to fail," *Second CInST Banking Workshop, Moscow*, 2012.

[39] B. Vandermarliere, A. Karas, J. Ryckebusch, and K. Schoors, "Beyond the power law: Uncovering stylized facts in interbank networks," *Physica A Statistical Mechanics and its Applications*, vol. 428, pp. 443–457, June 2015.

[40] O. E. Dictionary, *"overdraft, n.1"*. Oxford University Press, 2016.

[41] A. Egorov and O. Kovalenko, "Structural features and interest-rate dynamics of russia's interbank lending market," 2013.

[42] C. B. of Russia, "Overnight credit rates (one day settlement credit)," 2016. [Online; accessed 4-September-2016].

[43] C. B. of Russia, "The bank of russia deposit rates," 2016. [Online; accessed 3-September-2016].

[44] A. Karas, "Russian interbank market," 2008.

[45] C. B. of Russia, "Banking supervision report," 2002.

[46] Q. F. Akram and C. Christophersen, "Interbank overnight interest rates-gains from systemic importance," 2010.

[47] A. Karas and K. J. Schoors, "Heracles or sisyfus? finding, cleaning and reconstructing a database of russian banks," 2005.

[48] V. Cremers, J. Ryckebusch, and K. Schoors, "Default cascades in interbank networks," 2014.

[49] J.-F. Richard and W. Zhang, "Efficient high-dimensional importance sampling," *Journal of Econometrics*, vol. 141, no. 2, pp. 1385–1411, 2007.

[50] S. T. Tokdar and R. E. Kass, "Importance sampling: a review," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 1, pp. 54–60, 2010.

[51] J. M. Thijssen, *Computational Physics*. New York, NY, USA: Cambridge University Press, 1999.

[52] H. J. Herrmann, *Introduction to Computational Physics*. Zürich, Switzerland: Swiss Federal Institute of Technology ETH, 2009.

[53] D. J. MacKay, *Information theory, inference and learning algorithms*. Cambridge university press, 2005.

[54] J. Ryckebusch, "Computational physics." UGent lectures, 2014-2015.

[55] H. Gould and J. Tobochnik, *Statistical and thermal physics: with computer applications*. Princeton University Press, 2010.

[56] R Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[57] T. C. Silva, M. S. da Silva, B. M. Tabak, *et al.*, "Liquidity performance evaluation of the brazilian interbank market using a network-based approach," tech. rep., 2015.

[58] P. Gai and S. Kapadia, "Contagion in financial networks," in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, p. rspa20090410, The Royal Society, 2010.

[59] "Demand loan." `http://financialdictionary.net/define/Demand+Loan`. Accessed: 2016-11-10.

[60] A. Berger, "The measurement of bank liquidity creation and the effect of capital," 2006.

[61] D. L. Donoho *et al.*, "High-dimensional data analysis: The curses and blessings of dimensionality," 2000.

[62] M. Ledoux, *The concentration of measure phenomenon*. No. 89, American Mathematical Soc., 2005.

[63] M. Talagrand, "Concentration of measure and isoperimetric inequalities in product spaces," *Publications Mathématiques de l'Institut des Hautes Études Scientifiques*, vol. 81, no. 1, pp. 73–205, 1995.

[64] C. E. Shannon, "A mathematical theory of communication," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 5, no. 1, pp. 3–55, 2001.

[65] R. Lewand, *Cryptological mathematics*. 2000.

[66] R. T. Snodgrass and V. F. Camp, "International morse code," 1922. Edited by the author.

[67] J. Ryckebusch, "Statistical physics II." UGent lectures, 2014-2015.

[68] J. Berger *et al.*, "The case for objective bayesian analysis," *Bayesian analysis*, vol. 1, no. 3, pp. 385–402, 2006.

[69] E. T. Jaynes, "Information theory and statistical mechanics," *Physical review*, vol. 106, no. 4, p. 620, 1957.

[70] A. Giffin, "Maximum Entropy: The Universal Method for Inference," *ArXiv e-prints*, Jan. 2009.

[71] D. Mumford, "The dawning of the age of stochasticity," *Mathematics: frontiers and perspectives*, pp. 197–218, 2000.

[72] E. W. Weisstein, "Markov's inequality. From MathWorld—A Wolfram Web Resource." Accessed: 2017-2-24.

[73] S. H. Russell, "Understanding the term structure of interest rates: The expectations theory," *Review*, vol. 74, 1992.

[74] S. Anatolyev and S. Korepanov, "The term structure of russian interest rates," *Applied Economics Letters*, vol. 10, no. 13, pp. 867–870, 2003.

[75] B. M. Tabak and S. C. Andrade, "Testing the expectations hypothesis in the brazilian term structure of interest rates," 2001.

[76] F. Heider, M. Hoerova, and C. Holthausen, "Liquidity hoarding and interbank market spreads: The role of counterparty risk," 2009.

[77] M. De Domenico, A. S. Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, "Mathematical formulation of multilayer networks," *Phys. Rev. X*, vol. 3, pp. 041022+, Dec. 2013.

[78] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of Complex Networks*, vol. 2, pp. 203–271, Sept. 2014.

[79] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Physical Review E*, vol. 89, p. 032804, Mar. 2014.

[80] F. Battiston, V. Nicosia, and V. Latora, "The new challenges of multiplex networks: measures and models," *ArXiv e-prints*, June 2016.

[81] G. De Masi, G. Iori, and G. Caldarelli, "Fitness model for the italian interbank money market," *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, vol. 74, no. 6, pp. 066112+, 2006.

[82] K. Soramäki, M. L. Bech, J. Arnold, R. J. Glass, and W. E. Beyeler, "The topology of interbank payment flows," *Physica A: Statistical Mechanics and its Applications*, vol. 379, pp. 317–333, June 2007.

[83] K. Christensen and N. R. Moloney, *Complexity and criticality*, vol. 1. Imperial College Press, 2005.

[84] M. E. J. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary Physics*, vol. 46, pp. 323–351, Sept. 2005.

[85] R. Albert, H. Jeong, and A.-L. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, pp. 378–382, July 2000.

[86] A.-L. Barabasi, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world-wide web," *Physica A: Statistical Mechanics and its Applications*, vol. 281, pp. 69–77, June 2000.

[87] S. Langfield, Z. Liu, and T. Ota, "Mapping the uk interbank system," *Journal of Banking & Finance*, vol. 45, pp. 288–303, 2014.

[88] D. J. Watts and S. H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440–442, June 1998.

[89] A. Vernikov, "A guide to russian bank data: Breaking down the sample of banks," *Available at SSRN 2600738*, 2016.

[90] A. Halu, R. J. Mondragon, P. Panzarasa, and G. Bianconi, "Multiplex PageRank," June 2013.

[91] M. De Domenico, A. Lancichinetti, A. Arenas, and M. Rosvall, "Identifying modular flows on multilayer networks reveals highly overlapping organization in interconnected systems," *Physical Review X*, vol. 5, no. 1, p. 011027, 2015.

[92] S. Poledna, J. L. Molina-Borboa, S. Martínez-Jaramillo, M. Van Der Leij, and S. Thurner, "The multi-layer network nature of systemic risk and its implications for the costs of financial crises," *Journal of Financial Stability*, vol. 20, pp. 70–81, 2015.

[93] I. Alves, S. Ferrari, P. Franchini, J.-C. Heam, P. Jurca, *et al.*, "The structure and resilience of the european interbank market," tech. rep., European Systemic Risk Board, 2013.

[94] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006.

[95] V. Zlatic, G. Bianconi, A. Díaz-Guilera, D. Garlaschelli, F. Rao, and G. Caldarelli, "On the rich-club effect in dense and weighted networks," *The European Physical Journal B - Condensed Matter and Complex Systems*, vol. 67, pp. 271–275, Jan. 2009.

[96] D. Garlaschelli and M. I. Loffredo, "Patterns of link reciprocity in directed networks," *Physical Review Letters*, vol. 93, Dec. 2004.

[97] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web," tech. rep., Stanford Digital Library Technologies Project, 1998.

[98] M. E. J. Newman, "Mixing patterns in networks," *Physical Review E*, vol. 67, pp. 026126+, Feb. 2003.

[99] M. E. J. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, pp. 8577–8582, June 2006.

[100] S. L. Braunstein, S. Ghosh, and S. Severini, "The laplacian of a graph as a density matrix: a basic combinatorial approach to separability of mixed states," *Annals of Combinatorics*, vol. 10, no. 3, pp. 291–317, 2006.

[101] F. Passerini and S. Severini, "The von neumann entropy of networks," Apr. 2012.

[102] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, "A quantum jensen–shannon graph kernel for unattributed graphs," *Pattern Recognition*, vol. 48, no. 2, pp. 344–355, 2015.

[103] W. Du, X. Li, Y. Li, and S. Severini, "A note on the von neumann entropy of random graphs," *Linear Algebra and its Applications*, vol. 433, no. 11, pp. 1722 – 1725, 2010.

[104] B. Mohar, "The laplacian spectrum of graphs," in *Graph Theory, Combinatorics, and Applications*, vol. 2, pp. 871–898, 1991.

[105] K. Anand, G. Bianconi, and S. Severini, "Shannon and von neumann entropy of random networks with heterogeneous expected degree," *Physical Review E*, vol. 83, pp. 036109+, Mar. 2011.

[106] A. Banerjee and J. Jost, "On the spectrum of the normalized graph laplacian," May 2007.

[107] R. R. Sokal and C. D. Michener, "A statistical method for evaluating systematic relationships," *University of Kansas Scientific Bulletin*, vol. 28, pp. 1409–1438, 1958.

[108] C. Ye, R. C. Wilson, C. H. Comin, L. d. F. Costa, and E. R. Hancock, "Approximate von neumann entropy for directed graphs," *Physical Review E*, vol. 89, no. 5, p. 052804, 2014.

[109] F. Chung, "Laplacians and the cheeger inequality for directed graphs," *Annals of Combinatorics*, vol. 9, pp. 1–19, Apr. 2005.

[110] B. Karrer and M. E. J. Newman, "Stochastic blockmodels and community structure in networks," *Phys. Rev. E*, vol. 83, p. 016107, Jan 2011.

[111] B. Porfiriev, *Crises in Russia: Contemporary Management Policy and Practice From A Historical Perspective.* New York, NY, USA: Routledge, 2012.

[112] I. van de Wiel, "The 1998 Russian Crisis," 2013.