ICT in Business and the Public Sector

Master thesis:

Impact of the intelligent lockdown on the Dutch inter-industry transaction network

Name: Martijn Vlak
Student-no: s1681729

Date: 25/3/2021

Supervisors:
dr. Frank Takes, dr. Fabian Jansen, dr. Carolina Mattsson

Leiden Institute of Advanced Computer Science (LIACS)
www.liacs.leidenuniv.nl
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

Abstract

We use company transaction data to empirically describe the effects of Covid-19 on the Dutch economy. By using transaction data, we distinguish ourselves from previous work: the transaction data spanning the years 2019 and 2020 provides the unique opportunity to empirically show the actual effect of Covid-19, without making estimations. In particular, we compare a network from the data during the intelligent lockdown period, with a network from the data over the same period in 2019.

We compare these networks based on their structure, specifically by analyzing the strength distributions. Furthermore, we compute the random-walk centrality of industries, to detect the industries that are most immediately impacted by a shock. Next, we describe the change of consumer spending behavior in the intelligent lockdown period. We then analyze how the full inter-industry network was affected by this change, both directly and indirectly, using the maximum flow metric.

The findings show that even though the economy was impacted during the intelligent lockdown, the structure of the network remained quite the same. Industries with high random-walk centrality are mostly industries that offer general services to a wide range of other industries, such as investment or administration services. Further patterns are hard to distill because during the intelligent lockdown, a variety of macroeconomic factors either positively or negatively influence industries in the network. By analyzing consumption, we focused on one of these factors. We find a decline in card payments here, which are in some cases substituted by the cashless payments channel. The change in consumer spending behavior affects the majority of industries negatively. Especially industries in the food and accommodation, and cultural sector, were directly impacted by a decline in consumer spendings. If there are positive effects, these tend to be bigger. Examples of industries that directly profit from the lockdown are supermarkets and webshops. The maximum flow measure indicates if industries' supply chains are affected in the intelligent lockdown period. In some cases, these effects are indeed primarily caused by a change in consumer spending behavior.

Acknowledgements

I would like to thank ING Wholesale Banking Advanced Analytics for giving me an opportunity to work with real-world data and to gain experience in the working environment. In particular, I would like to thank Fabian Jansen, for his continuous enthusiasm, guidance and creative ideas throughout the project.

Next to this, I would of course like to thank Frank Takes and Carolina Mattsson, for their practical insights and inspiring me to learn more about network science.

Contents

1 Introduction

In this section we first describe the background and relevancy of our problem and provide the context of this thesis. This includes discussing related work. After this, we define our problem in more detail and state the research questions.

1.1 Context

The impact of the recent Covid-19 pandemic on the modern day society is massive. In order to limit the spread of Covid-19, countries around the world introduce restrictions. Examples are the shutdown of national borders, closing of public facilities, such as schools or restaurants, and a limit on public gatherings. Also in many cases, cities, regions or even countries entered some sort of lockdown, closing down nearly all companies and public transport to these areas. In the Netherlands, an intelligent lockdown was introduced in March. The intelligent lockdown started the 23rd of March, after other measures such as working from home and closing catering establishments had already been announced. It was called intelligent as it is less strict than a full lockdown, relying on responsible and “intelligent” behavior [1]. The lockdown heavily impacted the Dutch economy: the GDP declined with 8.5% in the second quarter of 2020 relative to the first quarter [2]. Companies were affected due to severe drops in demand, especially in certain industries. Examples are companies in transport, culture and recreation, and food and accommodation sectors. This can for example be seen in an article by Statistics Netherlands (CBS) [3], which mentions the decrease in household spending during the intelligent lockdown in the Netherlands, directly impacting the revenue of companies. Although the revenues of these companies decline, expenditures of the public sector increase in order to prevent bankruptcies. The drop in household spending is one of the major causes for the negative trend in the Dutch economy. Other causes are trade impediments with foreign companies, and declines in investments by local businesses [4].

In an economy, industries depend on each other. Furthermore, economies are subject to *shocks*: unexpected or unpredictable events impacting the economy either negatively or positively. A shock refers to a change in macro economic factors, such as consumption or unemployment. During the intelligent lockdown period, certain industries were for example subject to demand shocks. The output of these industries might be affected by this: industries might sell less goods/services because of the drop in demand. Lower output due to a demand shock, means that these industries also require less inputs, which again impacts the output of connected industries. By inputs, we mean resources that are needed for production of goods or services.

To capture the effects of industry linkages, a network of industries can be created. By analyzing the topology and structure of an inter-industry network, we are able to learn about how that particular economy functions and changes over time. We can find important industries based on network properties and analyze the effects of Covid-19 on these industries, or on the economy as a whole. Linkages in these networks are represented by money flows. Money flows are in a sense equivalent to the flows of produced goods/services.

Network science enables us to better understand complex systems such as economies. Complex systems can be analyzed by encoding the interactions between the system's components. Examples of these are (1) *social networks*, where the links are determined by social ties, (2) *neural networks*, mapping connections between neurons and to understand how our brain functions, and (3) *communication networks*, such as the WWW, describing how wired or wireless communication devices are linked to each other [5]. In the following section, we give examples of how network science can be applied in the financial domain.

1.2 Related Work

In this section we will discuss related research on financial networks at different levels of aggregation. On one of the highest levels, a particular field of research is receiving much attention, namely that of international trade networks. In this line of research, the structure of the world-economy is analyzed, where nodes are countries and ties are international trade flows between these countries [6, 7, 8, 9]. On a lower level, economies within countries can be researched. Here, we can consider interactions between aggregated entities, such as sectors and industries. Most studies in this field analyze input-output tables of countries. These tables are constructed by national statistic agencies and describe how sectors depend on each other, based on money flows. Many studies build upon the input-output model [10], describing interdependencies between sectors. An example is showing how idiosyncratic shocks (shocks affecting individual firms or sectors) may lead to aggregate fluctuations in the presence of inter-sectoral input-output linkages [11]. The research by McNerney et al. [12] uses input-output tables of 45 national economies, to analyze the network structure of inter-industry relationships in these different countries.

On an even lower level, there is a field of study that considers the financial ties between individual organizations, rather than sectors or industries. Financial ties can be reflected by individual transactions. We distinguish different types of such transaction networks. The work by Iosifidis et al. [13] studies an intercompany transaction network. In particular, a community of businesses in Sardinia is analyzed, in which companies make transactions in so-called complementary currencies. A network analysis is performed to discover various network properties. The research is in particular focused on cyclic motifs in the network and transactions within these cycles. The aim is to analyze the stability and the performance of the inter-company system. Similar studies are the study by De la Torre et al. [14], analyzing the topology of an inter-firm payment network in Estonia and the work by Kichikawa et al. [15], studying the structure of a Japanese inter-firm transaction network. Other studies describe cascade effects, caused by idiosyncratic shocks due to natural disasters and epidemics such as Covid-19 [16, 17, 18]. By cascade effects, the propagation of a shock trough the network is meant. In other words, industries impacted by financial shocks affect other industries through their linkages.

A particular area of research is that of inter-bank networks. Some works study the topology of inter-bank payment networks [19, 20]. Instead of considering the transactions between banks, some studies analyze networks where banks have other financial interdependencies. These papers analyze inter-bank networks based on cross-holdings, debts and shares [21, 22, 23]. The aim is to see if there exists financial contagion as a consequence from a financial shock or particular defaults. This means that a small group of financial institutions infect other institutions in the economy, after being affect by a shock. Financial contagion is an example of cascade effects. A similar term is systemic risk, referring to how local disruptions can cause global effects in an economy. Equivalently, the term domino-effect is often used to describe similar mechanisms.

Company networks can also be constructed and analyzed using other interdependencies than transactions. An example of this is a global corporate ownership network, where the aim is to identify offshore financial centers, which often facilitate tax avoidance and lenient regulations network [24]. Furthermore, links can be represented by considering board interlocks [25].

Finally, on the highest level of granularity, various researches analyze transaction records in payment systems. An example is the study conducted by Mattsson [26]. In this paper, a network representation of money flows is constructed, based on the transactions made in a particular payment system. The goal is to discover the underlying network structure of such modern payment systems. This work does not consider inter-company transactions, but rather transactions between individual actors. Other research is devoted to the network topology of credit card transactions [27] and Bitcoin transactions [28]. The authors construct a transaction network, obtain the network topology and consider basic network characteristics over time.

In the following section we present our problem definition and explain how we are going to apply the different theories of financial networks. We also mention in what way we differentiate ourselves from previous work.

1.3 Problem definition

In this thesis, we use company transaction data to empirically describe the effects of Covid-19 on the Dutch economy. We construct an inter-industry network from this data. In particular, we form industry linkages, by aggregating payments between companies belonging to certain industries. Creating a sensible network from raw transaction data is not trivial, so we will carefully explain the different steps taken in this process. We distinguish ourselves from previous work in a sense that we use these inter-firm transactions in order to construct an inter-industry network. In other studies, either input-output tables were used to analyze industry interconnections, or these company transactions where used to construct and analyze an inter-firm network. Using company transactions allows us to construct and compare networks over specific periods. An advantage of this, is that we are able to describe the effect of Covid-19 on the network. In various studies, the impact of a shock is estimated, using some sort of model. The transaction data is already affected by a shock, giving us the unique opportunity to empirically show the actual effect of Covid-19, without having to make estimations.

To be able to quantify these effects, we compare the networks from the period during the intelligent lockdown in the Netherlands, with the network of the same period in 2019. During this thesis, we will call these networks the *2019 network* and the *2020 network*. The networks will be compared by their topologies and structure, using general network descriptives, as presented in [5]. We will focus on centrality measures, which unveil important nodes in our network. Also, we will analyze the distributions of these metrics, which allows us to learn about the general network structure. First of all, we will consider the strength of nodes, which indicates what the largest industries are in the network. If important nodes are impacted by a demand shock, the impact of this shock in the entire network may be bigger.

Important industries might also be more susceptible to shocks. In order to test this, we consider the random-walk centrality, proposed by Blöchl et al. [29]. This measures the closeness of nodes in the network, based on random-walks. We expect that industries to which other industries are close, have a higher probability of being affected by shocks.

The second part of this research is specifically focused on consumer spending behavior during the intelligent lockdown. We are interested how this impacts different industries in the network. These effects can be split into two different categories: 1) direct effects and 2) indirect effects. By direct effects we mean that industries that highly depend on consumers, so that are downstream on the supply chain, have less output to consumers. This may also influence suppliers of these industries, which we refer to as indirect effects. As a measure to describe the effects that reduced consumer spending has, we propose to use the maximum flow value introduced by Ford and Fulkerson [30]. The corresponding research questions are mentioned in the upcoming section.

1.4 Research questions

This thesis can be divided into two overarching research questions:

- **RQ1**: **What are differences in topology and statistical properties between the 2019 and 2020 Dutch inter-industry transaction network?**
- **RQ2: To what extent does the change in consumer spending behavior influence the transaction network?**

These questions can be divided into subquestions. For our first question, these are:

- ***RQ1.1***: How do we construct a sensible aggregated economic network from disaggregated financial transaction data?
- ***RQ1.2***: To what extent is the 2019 network different from the 2020 network, when analyzing their structures?
- ***RQ1.3***: What are the most important industries based on random-walk centrality in both the 2019 and 2020 network?
- ***RQ1.4***: To what extent are industries with a high random-walk centrality more susceptible to shocks?

The second question is focused on the impact of the lockdown on consumer spendings. The different subquestions we can ask here are:

- ***RQ2.1***: To what extent can we use maximum flow to describe cascading effects?
- ***RQ2.2***: To what extent did consumer spending change during lockdown?

- ***RQ2.3***: What are indirect effects of the change in consumer spending?

We start with a brief description of our data in Section 2. The methods that we use to answer our research questions, are covered in Section 3. In this section we describe our data pre-processing steps and the methods for our network analysis. We focus on methods for analyzing the change in network structure, important industries based on random-walk centrality and their relationship to cascades, and on quantifying the effects of the change in consumer spending. In Section 4, we describe the results of these analyses. Finally, we draw conclusions in Section 5.

2 Data

The transaction data that we use, are payment records from business client bank accounts. The data is provided by ING Wholesale Banking, as part of this internship. We have access to transactions corresponding to the period of October 2018 until August 2020. We are able to see all the payments made from and received by ING-accounts. Every transaction consists of a payer, beneficiary, amount and date. Other attributes considering the payer and beneficiary are the country, name, company identifier, whether the party is a private individual or not. For companies, we have additional information, based on Dutch Chamber of Commerce data. This includes for example information about the sector, location and parent companies. Especially the sector code is relevant for our research, as we are interested in the industry in which the company is active. To identify this, the NAICS classification system is used [31]. NAICS is the North-American standard to classify companies by their industry. A NAICS code is a six-digit number that contains a hierarchy:

- 2-digit code: **sector**
- 3-digit code: **sub-sector**
- 4-digit code: **industry group**
- 5-digit code: **NAICS industry**
- 6-digit code: **national industry**

An example of NAICS code breakdown is shown in Figure 1.

Level	NAICS Code	Title
Sector	44-45	Retail Trade
Subsector	441	Motor Vehicle and Parts Dealer
Industry Group	4412	Other Motor Vehicle Dealers
NAICS Industry	44122	Motorcycle, Boat, and Other Motor Vehicle Dealers
National Industry	441221	Motorcycle, ATV, and Personal Watercraft Dealers

Figure 1: NAICS example
source: Economic census [32]

For this research, we want to give an insight in the Dutch economy, and how it was affected by Covid-19. An implication is that our dataset is limited to ING customers. Hence, we deal with a subsample of all company transactions in the Netherlands. As ING is the biggest bank in the Netherlands, it should still give a fair overview of the Dutch economy. A rough estimate is that half of the Dutch companies is an ING client. We must take into account that we use a sample, which might be biased towards certain companies/industries.

In the second part of this thesis, we analyze consumer spending behavior. We refer to consumers by using the term private individuals. In this category, we only want actual end-consumers. However, self-employed individuals, investors and even small to medium size enterprises often times are also seen as private individuals. We need to distinguish between these types of payments. Furthermore, private individual payments to companies are sometimes difficult to detect, as private individuals use a range of different payment channels. It is important to make a distinction between the different payment methods, as the type of payment method is closely related to the effects of the intelligent lockdown: in many cases, "physical" payments were ruled out as possible payment method. As the payment method is relevant, we briefly explain the different payment channels that private individuals might use.

With the help of various domain and data experts, we have been able to identify the main channels that consumers use to transfer money. We divide the private individual payments into two groups: (1) *card payments* and (2) *cashless payments*. Card payments consist of PIN transactions. This is the most used payment channel in physical stores. The cashless payments category contains the other cashless payment channels that consumers use. This category mostly consists of online payments, for example via iDEAL.

Card and iDEAL payments via ING are collected by ING internal accounts, transferring this money in batches to organizations. In this case, ING acts as intermediary party. In many cases, more intermediary parties are involved in the process of a private individual payment, making it hard to detect individual payments. This is called disintermediation. For companies with an ING account, we can see all the incoming transactions. Hence, we are also able to see when third parties pay these organization. By carefully analyzing the transaction data and consulting domain experts, we were able to detect a large fraction of these third party payments, which we can consider as private individual payments. Some of these are associated with card payments, others can be associated with cashless payments. We are not able to give examples due to client confidentiality.

When there is no intermediary party involved, private individuals directly transfer money to organizations. Most of these payments are online transactions, via a SepaCreditTransfer (SECT). We need to separate consumption and business payments in this category. In cases for which we are not certain whether a private individual is actually a consumer or not, we filter high value transactions, to filter out investment payments. The other remaining transactions are added to the cashless payments category.

3 Methods

In this section, we mention all different methods used in this thesis, from constructing the network to analyzing the network.

3.1 Data pre-processing and selection

We begin with addressing our first research question: *How do we construct a sensible aggregated economic network from disaggregated financial transaction data?* We mention the most important steps that we undertook to obtain an inter-industry network, starting with the complete transaction dataset. These steps are shown in Figure 2. The process involves using knowledge about the ING dataset and its domain, as well as using best practices in related network science papers.

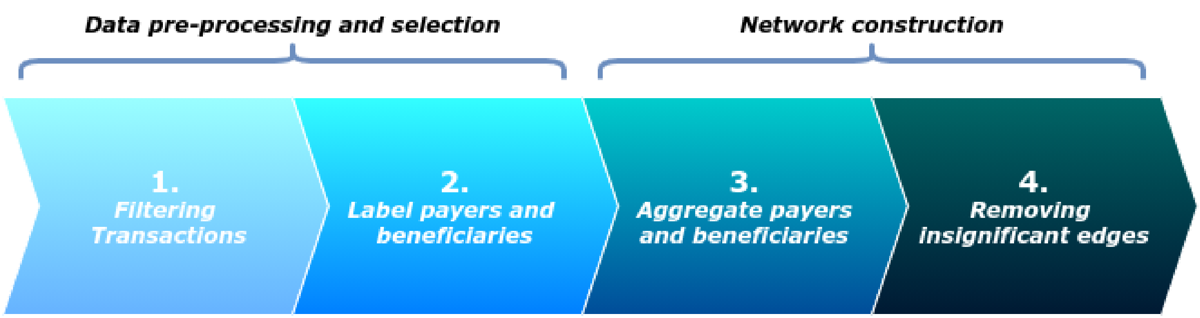

Figure 2: Different steps to take in order to achieve a industry network

3.1.1 Filtering transactions

We start with an initial filtering step. The full dataset contains transactions that are not relevant for our network: we are only interested in Dutch companies. If the payer or beneficiary is relevant, we keep all the transactions corresponding to this party. The criteria are:
The payer or beneficiary

1. has to be an company,
2. needs to have an ING account,
3. needs be registered in the Netherlands,
4. may not make a payment to itself/receive a payment from itself.

The reason that we want the payer or beneficiary to have an ING account, is that we have complete information over these accounts. We are able to see all the transactions corresponding to these payers or beneficiaries. For payers or beneficiaries with other bank accounts, we only see the transactions they make to ING accounts. This means that we will not get a fair picture of the transactions of these companies. Furthermore, we do not consider transactions where companies pay themselves. This is the case when companies have multiple accounts, and move their money from one account to another account. If we would consider these transactions, this would increase the outflow and inflow of these companies, whereas these companies did not receive or spend any money. This would leave to an overestimation of these companies' flows.

3.1.2 Label payers and beneficiaries

The last step we take before aggregating transactions, is labelling payers and beneficiaries based on their NAICS-code. If a party does not have a NAICS-code, it is automatically labeled as "external party". All external parties are mentioned in Table 1. External parties essentially are non-business entities. We can not remove flows from and towards external parties entirely from our data, as they contain information about the revenue and costs of industries in the internal network. This is why we model external parties by creating external nodes, which are connected to our internal network. The internal network consists of all industries that we analyze in this research. The external flows can be seen as exogenous forces which might influence the behavior of internal parties. Transactions between two external parties are removed entirely from our dataset. In the remainder of this section, we elaborate on why some parties are seen as external.

External Party
Public Administration
Non-business Organizations
Agriculture, Forestry, Fishing and Hunting
ING
Parent Companies
Private Individuals
Self Employed Individuals
Foreign Counterparties
iDEAL Payments (processed by ING)
Card Payments
Payment Service Providers
Other

Table 1: Different external parties

Sometimes if a company contains a NAICS-code, we still might label it as external party. This is in particular the case for *Public Administration* industries. The first reason for this is that the organizations in this sector are funded by the government. Their revenue does not depend on economical circumstances. Another reason is that the public sector accounts for the largest amount of money flow in this network, making it more difficult to find effects for other industries in the network. We also will not consider other industries that are government led or non-business. To determine this, we use only the NAICS-codes that are included by the Economic Census Bureau. This bureau provides data about the economy in the United States [32]. For this, we assume that the economy of the United States and the Netherlands are similar, hence we can exclude the same industries. Next, we filter out transactions from and to ING bank. The reason for this, is that there is a bias towards ING; we see payments from the bank's transaction system. First of all, these withdrawals and deposits are not interesting, as money is not moved from one company to another, meaning it does not contribute to economic activity. Secondly, we might see relatively more transactions with banks, than these companies perform in reality when we would consider all transactions. This is because we consider payments from the bank's perspective.

Finally, We also consider parent companies as external parties. Larger organizations often consist of holding and subsidiary companies. The money flows between these companies do not represent industry dependencies and can affect the results. This is because these flows tend to be large, as they are associated with the biggest corporations. A company is attributed as parent company, when it is either a direct parent or *higher in the same branch of the corresponding organization tree*.

3.2 Network construction

In this section, we describe the process of inter-industry network construction, over the pre-processed dataset. We start with the most important step: aggregating parties and creating the actual network. After that we explain how and why we remove insignificant edges.

3.2.1 Aggregate payers and beneficiaries

Constructing the network is done by aggregating transactions. A network is a set of nodes connected by links. In our case, the set of nodes contains either industries or external parties. Industries are represented by their 6-digit NAICS-codes. We use the 6-digit NAICS-codes to obtain the maximum level of granularity. The different external parties are mentioned in Table 1.

As mentioned in the introduction, we consider transactions over two periods: the months March, April and May in 2019 and the same months in 2020. We need to consider the same months in 2019 and 2020 for a fair comparison, as the data is subject to seasonality effects. An edge between two industries is formed, *if one or more companies from a certain industry make at least one payment to one or more companies from another industry, in either the 2019 period or the 2020 period*. The same principle holds for links between industries and external parties. Transactions from one node to another node are modeled by one single weighted directed edge. For every edge, we consider the 2019 and the 2020 weights: *transaction count* and *sum of transaction values* over the 2019 period and over the 2020 period. The weights reflect monthly averages: we compute the transaction count and sum for every individual month (March, April and May) and take the average over these months, for the years 2019 and 2020. In the end, we have two networks: the 2019 network and the 2020 network, with the same edges and with different values for the edge weights. Having the same edges, makes it easier to compare the networks with each other.

3.2.2 Removing insignificant edges

The linkages in the network that we construct, must represent dependencies between industries. If a link is formed by two companies making a small amount of minor payments to each other, this does not indicate that the industries of these companies have a dependency. Such linkages do not reflect representative behavior of the industries as a whole. Next to this, most smaller edges have no transactions associated with them in one of the two periods, or the relative change is much bigger compared to edges with larger weights. We remove insignificant links between industries based on the count of transactions and the sum of the values, corresponding to these links. In Figure 3, we see that by far the most edges are positioned in the bottom corner (the vertical axes of the histograms are logarithmic). This means that most links between industries represent only a few transactions with a low sum.

Edge weight scatterplots 2019 and 2020

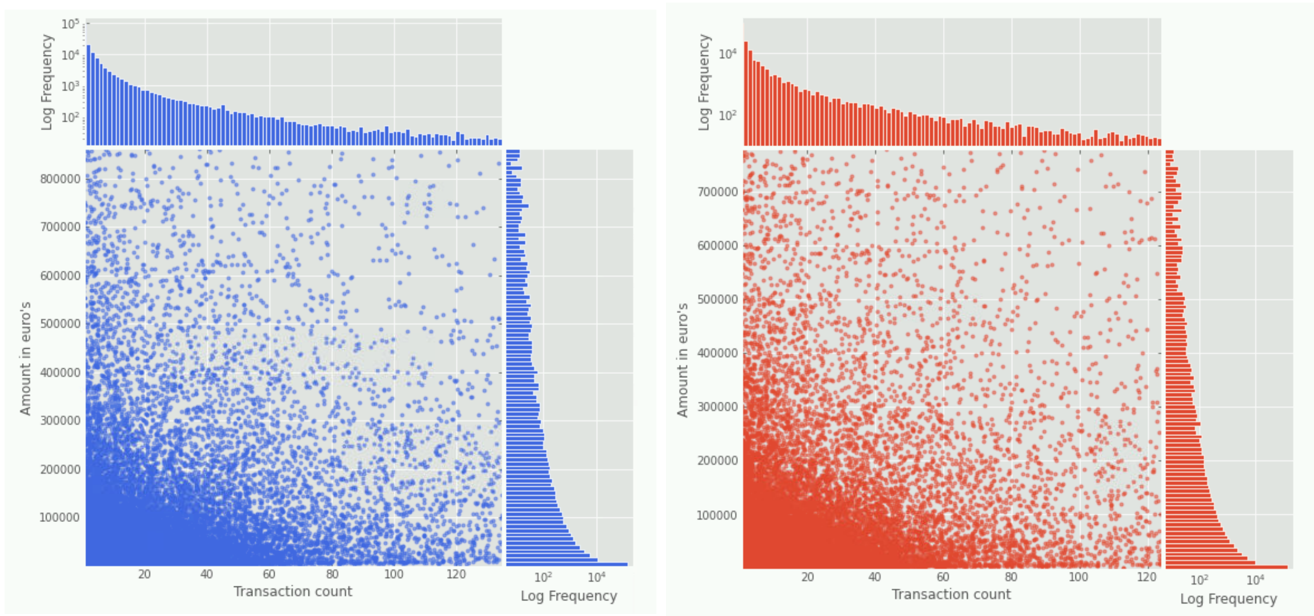

Figure 3: Scatterplots of the transaction count vs. the sum, of the 2019 and the 2020 network, before removing insignificant edges. The plots show 99% of the data, not displaying 1% of the outliers, with higher transaction count and/or sum.

To filter insignificant linkages, we use a threshold for both the transaction count and the sum. In particular, we prune edges from the left bottom corner: we keep edges if they have a transaction count of bigger than 5 or a sum bigger than 10,000, either in 2019 or in 2020. In this way, we have the same edges for both the 2019 and 2020 network, making it easier to compare both networks. Using these "hard thresholds" has the advantage that apart from removing insignificant edges, we also filter out the smallest industries. In this way, we do not see industries that virtually have no impact on the economy. A disadvantage is that we will consider links that are relatively small for larger industries. An approach to counter this, would be to filter links proportionally to the size of industries. In Section 4.2.1, we show that we remove most edges using our filtering approach. However, as we only remove low value edges, we remove only a small fraction in terms of the total sum of edge values. This means that the findings in terms of absolute flow will remain similar compared to before the removal of edges.

3.3 Methods for analyzing structure of the 2019 and 2020 network

In this section we explain the different methods that we use to analyze the structure of our inter-industry networks, as constructed in Section 3.2. We start by introducing network notation used throughout this thesis.

3.3.1 Network notation

We represent both the 2019 and 2020 network, using two adjacency matrices: C and V. C denotes the adjacency matrix, where weights reflect the average monthly transaction count between industries.

Matrix V represents the average monthly sum of transaction values between industries. This means that C_{ij} shows the amount of transactions that industry j made to i and V_{ij} represents a money flow from industry j to industry i. A dummy example of adjacency matrices is depicted in Figure 4. As we can see, C and V contain the same edges, but with different values. We do not take linkages between external parties into account. We do allow for self-loops as self-loops are a major part of the economy: companies often times interact with companies from the same industry. Furthermore, we deal with external flows. These consist of inflows and outflows: flows coming into the network, and flows leaving the network. In Figure 4, the external parties are “Private Individuals” and “Public Administration”. The corresponding network representation of these adjacency matrices is depicted in Figure 5.

C Transaction count	Full-Service Restaurants (722511)	Supermarkets and Other Gocery Stores (445110)	Fish and Seafood Markets (445220)	Fish and Seafood Merchant Wholesalers (424460)	Offices of Notaries (541120)	Private Individuals	Public Administration
Full-Service Restaurants (722511)	10	0	3	0	5	30	20
Supermarkets and Other Gocery Stores (445110)	0	5	0	15	15	10	30
Fish and Seafood Markets (445220)	0	0	2	5	2	30	7
Fish and Seafood Merchant Wholesalers (424460)	0	2	0	8	8	0	5
Offices of Notaries (541120)	0	0	0	0	10	0	30
Private Individuals	200	400	40	0	0		
Public Administration	20	30	10	15	25		

V Sum of transaction values	Full-Service Restaurants (722511)	Supermarkets and Other Gocery Stores (445110)	Fish and Seafood Markets (445220)	Fish and Seafood Merchant Wholesalers (424460)	Offices of Notaries (541120)	Private Individuals	Public Administration
Full-Service Restaurants (722511)	100	0	40	0	60	30	200
Supermarkets and Other Gocery Stores (445110)	0	100	0	200	300	100	800
Fish and Seafood Markets (445220)	0	0	10	50	20	30	70
Fish and Seafood Merchant Wholesalers (424460)	0	10	0	50	50	0	50
Offices of Notaries (541120)	0	0	0	0	500	0	700
Private Individuals	200	700	40	0	0		
Public Administration	200	800	100	150	250		

Figure 4: We represent a network by using two adjacency matrices C and V.

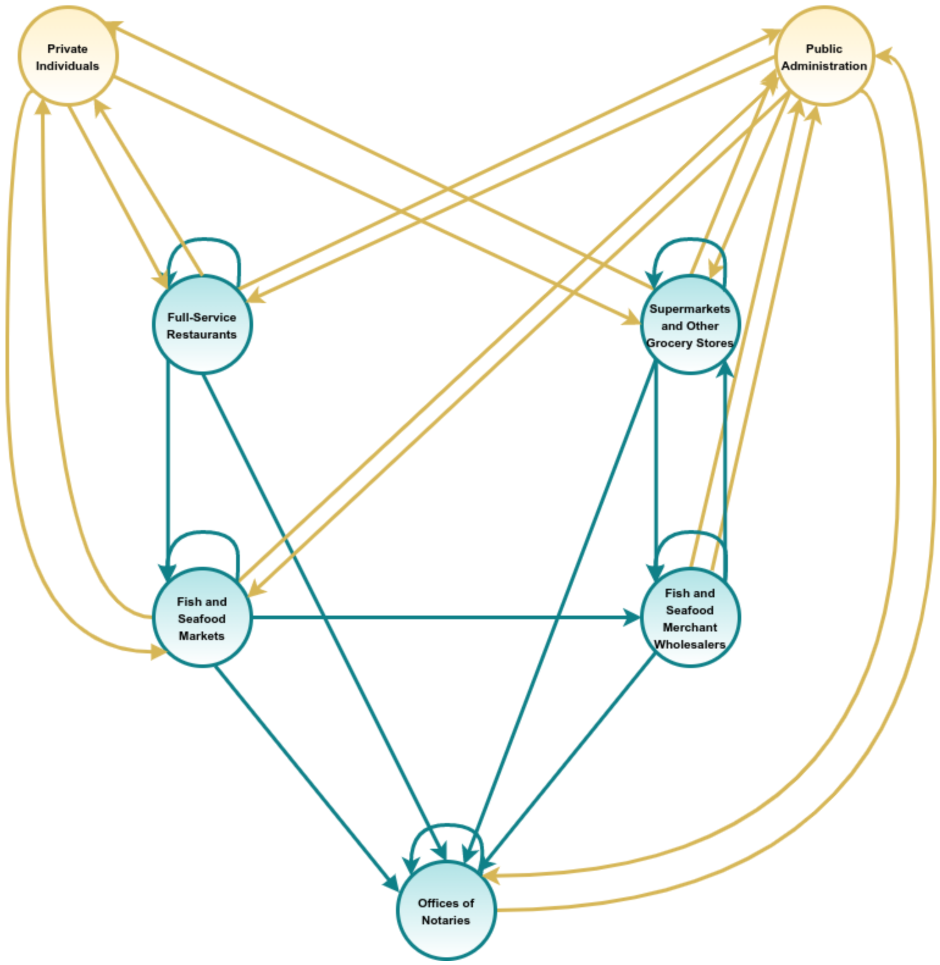

Figure 5: Corresponding network representation of the matrices C and V from the dummy example in Figure 4, without edge weights that are stored in these matrices.

To give an idea of what the actual network looks like, a snapshot of the network is depicted in Figure 6. The visualization was created in Gephi, during the pre-processing stage and is not representative for the final network. To make the image more clear, the smallest edges were filtered based on weight. In a network without filtering, the visualization would become very unclear.

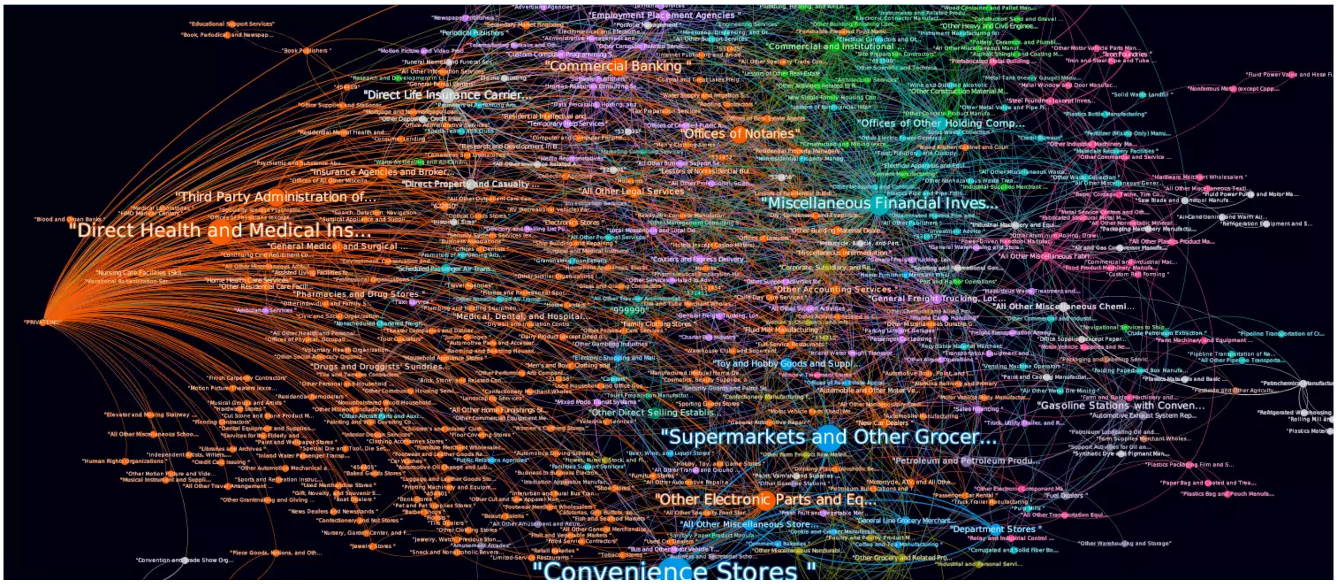

Figure 6: An example visualization of the inter-industry transaction network, created during the pre-processing stage

3.3.2 Density

Network density describes what fraction of the potential edges in a network are actual edges. Real-world networks often are sparse, meaning that the number of edges that are present in the network is much lower than the total number of edges that this network can possibly have. If a network contains all possible edges, it is *complete*. In complete directed networks, the number of edges equals: $n(n-1)$, where n is the number of nodes in the network. To compute the density of a network, we divide the actual number of edges m by the maximum number of edges in the case of a complete network: $\frac{m}{n(n-1)}$ [5].

3.3.3 Components

In a network, nodes are *connected* if there exists a path between them. A network is connected if all pairs of nodes in the network are connected. A network is *disconnected* if this is not the case, meaning there exists at least one pair of nodes without a path between them. Components are subgraphs for which there is a path between all the node pairs belonging to this subgraph [5]. We can differentiate between the undirected and directed case. The largest component where all nodes are connected to each other via undirected paths, is called the Giant Weakly Connected Component (GWCC). We can break the GWCC up into smaller components. First of all, the nodes in the Giant Strongly Connected Component (GSCC) can all reach each other via directed paths. The Giant In-Component (GIN) has paths towards the GSCC, the Giant Out-Component has paths from the GSCC. Then there is also tendrils that do not have directed paths to or from the GSCC, but to GIN or from GOUT. Apart from the GWCC, networks often contain smaller Disconnected Components (DC's). We show the different components in Figure 7 [14].By considering connected components, we learn how well the network is connected. The GSCC forms the foundation of the economy. Industries not in the GSCC do not play a central role in the economy. These industries may for example have a role of money creator (GIN) or money sink (GOUT).

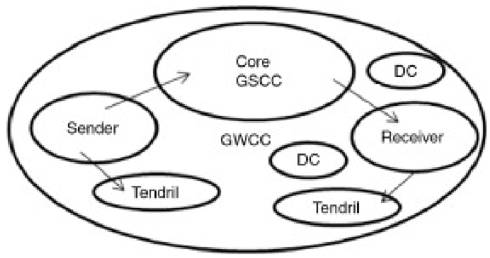

Figure 7: Different components in a directed network [14]

3.3.4 Degree and strength

We first consider two relevant properties of our network the degree and strength. Because we deal with a directed network we distinguish in- and out-degree. The in-degree for a node i, k_i^{in}, is the number of incoming links of a node. The out-degree for a node i, k^{out}, equals the number of outgoing links. The in and out-strength are the weighted equivalent of degree; we sum the weights of incoming and going transactions to obtain the in and out-strength. The degree or strength can provide us with information about the importance of different industries in the economy: if you have a high degree, you interact with many other industries, meaning that many industries depend on you, or vice versa. The strength is important, to see to what extent an industry contributes to economic activity. For a network with n nodes, we define the following strengths for a node i:

Total number of incoming transactions: $c_i^{in} = \sum_{k=1}^{n} C_{ki}$
Total number of outgoing transactions: $c_i^{out} = \sum_{k=1}^{n} C_{ik}$
Total incoming flow: $v_i^{in} = \sum_{k=1}^{n} V_{ki}$
Total outgoing flow: $v_i^{out} = \sum_{k=1}^{n} V_{ik}$

The average degree or strength is an important property of a network. For a directed network, the average degree equals the number of links m, divided by the number of nodes n. The number of links is namely equivalent to the sum of in-degree or out-degree over all nodes. Hence, the average degree equals:

$\langle k^{in} \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i^{in} = \langle k^{out} \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i^{out} = \frac{m}{n}$

We can use a similar definition for the average strengths:

$\langle c^{in} \rangle = \frac{1}{n} \sum_{i=1}^{n} c_i^{in} = \langle c^{out} \rangle = \frac{1}{n} \sum_{i=1}^{n} c_i^{out}$

$\langle v^{in} \rangle = \frac{1}{n} \sum_{i=1}^{n} v_i^{in} = \langle v^{out} \rangle = \frac{1}{n} \sum_{i=1}^{n} v_i^{out}$

3.3.5 Degree and strength distributions

Much research has been conducted in the field of degree distributions. The shape of the distribution determines many network related phenomena, such as epidemical spread and robustness. A property of many real world networks, is that they are scale-free, indicating a power-law degree distribution. In practice, empirical power-law distributions are not always detected. According Broido et al. [33], log-normal distributions fit the data better than the power-law, in most cases.

For a power-law distribution, the logarithm of the probability $\ln p_k$ of a certain degree k is expected to depend linearly on $\ln k$, multiplied by the degree exponent γ, indicating the slope of the line. In other words, we see a straight line when plotting a power-law on a log-log scale: $p_k \sim k^{-y}$.

Power-law: $f(x) = Ax^{-y}$

In the case a network has a power-law degree distributions, it is called scale-free. A property of the power-law degree distribution is that compared to random networks, there tend to be some hubs (large degree nodes) and relatively many small degree nodes. This means that in general, nodes have only a small influence, leading to a robust structure. However, the large degree nodes can impact the network on larger scale. In practice, distributions may have an exponential bound. In this case, a power-law with exponential cutoff is used:

Power-law with exponential cutoff: $f(x) = Ax^{-y}e^{-\lambda x}$

In many cases, distributions are not bounded, but decay faster, compared to a power-law. These distributions are still characterized by having a fat-tail. To describe this phenomenon, two types of distributions are commonly used: a *stretched exponential distribution* (*Weibull*) and *log-normal* distributions.

Weibull: $f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}\exp[-(\frac{x}{\lambda})^k]$

In this case, k is the stretching component, indicating the "fatness" of the tail of the distribution.

The log-normal degree distribution is found when $\ln k$, resembles a normal distribution. A variable typically follows such a distribution, when it is the product of many independent random (positive) numbers. This is for example the case for returns on stock market, which is assumed to be random, and is a multiplication of different bets.

Log-normal: $f(x) = \frac{1}{\sqrt{2\pi}\sigma_x}\exp[-\frac{(\ln x-\mu)^2}{2\sigma^2}]$

The function for a log-normal distribution is the same as the function for a normal distribution, except that the variable in the exponential term is $\ln x$ and not x.

Although degree distributions have been studied more elaborately, we expect the same phenomena for strength distributions. When edge weights come at play, log-normal distributions might still occur, assuming that wealth of entities is a product of random numbers, similar to the stock market returns example. In real-world inter-industry networks, the distribution of money flow strength, tends to indeed follow a Weibull or log-normal distribution [12]. Weibull and log-normal distributions are generated by similar mechanisms.

3.4 Random-walk centrality

Degree or strength determines the size of industries, which also tells us something about the importance of industries in the network. If high degree nodes are directly affected by a shock, many connected industries could potentially also suffer indirectly from this. The relation between cascade effects and degree has for example been researched by Acemoglu et al. [11]. However, only the size of industries might not suffice to describe the role of industries during economic shocks. Blöchl et al. [29], propose a metric called random-walk centrality, which can be interpreted as the "susceptibility during a supply shock". This metric is specifically designed for input-output matrices, which are highly connected networks with strong self-loops. The metric resembles closeness centrality, which describes the mean geodesic distance from all nodes to one. In other words, it describes how close other nodes are to you [5]. The intuition behind random-walk centrality is similar, but in this case the distance from other nodes to you is determined by random walks. Specifically, the Mean First Passage Time (MFPT) is used. The MFPT $H(s, t)$, from a source node s, to a target node t, equals the expected number of steps that a random walker needs to reach t for the first time, starting from s:

$H(s, t) = \sum_{r=1}^{\infty} r * P(s \to^{r} t)$

Here, r is the exact amount of steps it takes to reach t, and $P(s \to^{r} t)$, is the probability of this. The random walker decides which edges it follows, based on a probability distribution determined by edge weights. More specifically, if we represent our network as a matrix, we divide the entire matrix by its row sums. The row sums are equivalent to the out-strength v^{out} of industries. Now, each entry in the matrix represents the transition probability. Intuitively, a random walk reflects how money may flow through a network. More specifically, a random walker reflects the way in which a random euro traverses through the network. A random walker is more likely to progress over edges with a relatively large edge weight for a certain industry.

After computing $H(s, t)$ for all $s, t \in N$, where N is the set of nodes, the random-walk centrality is defined as "the inverse of the average MFPT to a given node":

$RW_C(i) = \frac{n}{\sum_{j \in N} H(j,i)}$

According Blöchl et al., nodes with a high random-walk centrality, are more sensitive to supply conditions in the economy, due to economic shocks. This is because many supplier industries are close to industries with a high random-walk centrality. In this research, we assume that the shock is already present in the network, caused by the intelligent lockdown. Thus, we expect to see cascade effects for nodes with a high random-walk centrality. We have concretely formulated this in research question 1.4: *To what extent are industries with a high random-walk centrality more susceptible to shocks?*. Here, we define the total sum of money flows to neighbouring industries as a performance indicator. Thus, industries need less input from other industries, when impacted by a shock.

3.5 Maximum flow

In this section we describe the maximum flow metric, which we use to answer research question 2, including its subquestions. First of all, we answer research question 2.1: *To what extent can we use maximum flow to describe cascading effects?*

We use the minimum cut-maximum flow theorem as proposed by Ford-Fulkerson to describe the impact of the change in consumer spending, during lockdown [30]. The intuition behind using this measure is that we analyze the maximum flow of value to industries, given that private individuals are the only external source of money. Maximum flow gives an upper-bound or carrying capacity on a flow network between a source and a target node. Essentially, we compute the maximum flow over different supply chains. The amount of value that can flow over the supply chains serves as a performance metric. A decline in maximum flow value for an industry, indicates overall disruptions on the supply chains from consumers to this particular industry.

For this theorem, we propose a directed network $G = (N, E)$, where N is the set of nodes, and E is the set of edges $E \subseteq \{(i, j) \in N \mid i \neq j\}$. This theorem assumes that flow runs from a *source* node $s \in N$ to a *sink* node $t \in N$. Edges have capacities $cap(i, j)$, indicating the maximum amount of flow that can pass this edge. A flow is a nonnegative function $flow(i, j)$, defined for all edges $i, j \in E$, for which the following applies:

- For every edge $i, j \in E$, $flow(i, j) \leq cap(i, j)$
- for every node $i \in N$, where $i \neq s$ and $u \neq t$: $\sum_j flow(i, j) - \sum_i flow(j, i) = 0$

Thus, flow can not exceed the capacity of an edge and nodes may not have more flow coming in than going out, except for the source and sink node. The goal is to maximize the flow between s and t, preserving the constraints. According to the minimum cut-maximum flow theorem, this is equal to the cut with the minimum capacity, or bottleneck capacity. For computing the maximum flow, we use the Edmonds-Karp implementation of the Ford-Fulkerson theorem, which uses BFS to find augmenting paths [34].

In reality, there are more external sources than private individuals, but by modelling it this way, we try to get an idea of 1) the direct effect and 2) the indirect effect of a change in consumer spending. In this research, we aim to discover such effects, which is also reflected in research question 2.2 and 2.3. The change in maximum flow describes a combination of direct and indirect effects: there can be a direct link from private individuals to a particular industry over which value flows, which comprises the direct component of maximum flow. The indirect component consists of all flows through other industries. Some industries might be more directly depending on private individuals, some are higher upstream on the supply chain, meaning that certain industries will be more likely to be affected directly by consumer spending behavior.

We compute the maximum flow values for every industry in the network. This means that for every industry, we construct a flow network with private individuals as source and the industry for which we compute the maximum flow, as target. By comparing the maximum flow values for the 2019 and the 2020 weights, we can see whether capacities in flow networks between consumers and industries have changed or not. Capacities of edges are equivalent to overall spending from one industry to another industry. A decline means that an industry spends less money on another particular industry. If maximum flow declines, this means that overall, the capacities have changed in such a way, that less value can flow over the supply chains. This may indicate cascade effects.

4 Results

This chapter contains the most important findings of our research. First, we describe the transaction data in more detail. After this, we show the results from our network analysis. This particular section is divided into two parts: general network description and the influence of consumer spending behavior.

4.1 Data description

In this section, we describe properties of the raw transaction data, before aggregation. We do this in order to see the effect of the different filtering choices and to get an overview of the transaction data. We will show distributions of the transactions, the transaction data over time, and specifically consider the 2019 with the 2020 period. As described in Section 3.1, there are two occasions during the data pre-processing and selection stage, when we remove transactions: (1) when filtering irrelevant transactions and (2) when removing transactions between external parties.

The distributions of the individual transaction values are depicted in Figure 8. Not taking the left-hand side of the distribution into account, we note a straight line on log-log scale, indicating a power-law distribution. This means that most transactions are "small" and there is only few large transactions. After filtering irrelevant transactions, many small transactions are removed. The median of the distribution also shifts to the right. After also removing external to external transactions, we especially notice that the largest transactions are removed. This is mostly due to the removal of public sector transactions.

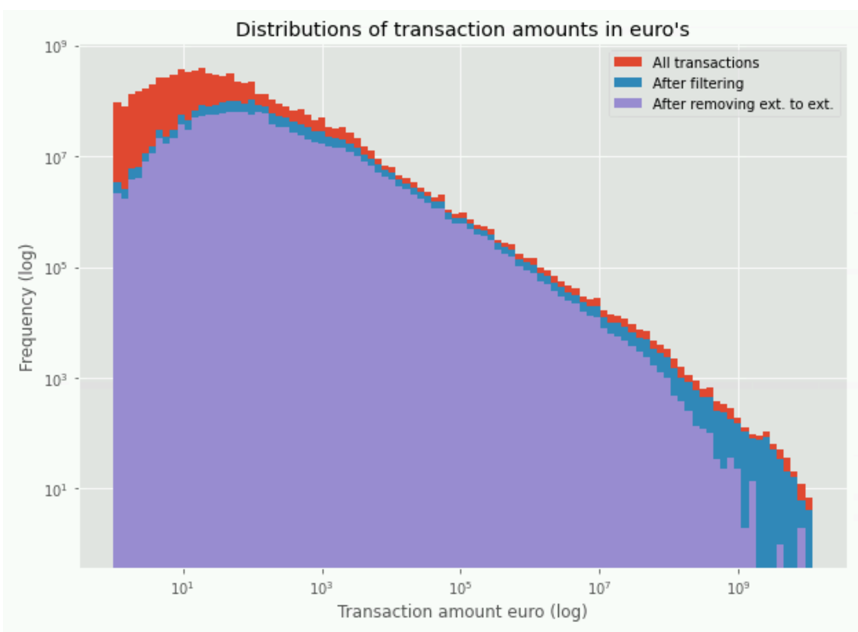

Figure 8: Distribution of the transaction amounts in EUR, for the different data selections

We can also see these effects in Table 2. Initially, the mean transaction value is 1820.73. After removing irrelevant transactions, this is 3477.40. We retain slightly more than 1/6th of the total number of transactions. If we look at the sum of the transaction values, this equals about 1/3th, again indicating that we remove a large quantity of smaller transactions. After removing irrelevant and external to external transactions, the average number of transactions per month equals 47.3 million. The average volume of these transactions is 164.4 billion. Finally, in the total dataset, there are 1,537,059 unique company identifiers. After removing transactions, this is 1,284,658.

	All transactions	**After filtering**	**After removing ext. to ext.**
Total sum (billions)	11,226.2	8243.4	3944.8
Count (millions)	6165.8	1708.6	1134.4
Sum per month (billions)	467.8	343.5	164.4
Count per month (millions	256.9	71.2	47.3
Mean value amount per transaction)	1820	4824	3477
Amount of companies	1,537,059	1,432,393	1,284,658

Table 2: Properties of different data selections for the period of October 2019 until September 2020

Considering Figure 9, we see a large difference between months. The 2019 and 2020 period are highlighted. Some properties of this period are also shown in Table 3. We see that overall, the difference of the sum and count of transactions between months remains largely the same before and after removing transactions. There is some regularity here, but it is hard to spot clear seasonality patterns. Considering the lockdown period, there especially is a drop in April, both in count and sum of transaction values. This is not visible in the 2019 period. For the month March of 2020, the sum of transactions is much higher, compared to March 2019.

Looking at Table 3, the total sum and count of transactions is lower in 2020, regardless of which data selection we take. The relative difference is also quite similar, for each data selection. Considering the monthly mean of the count and sum of transactions, we see that this is lower in the intelligent lockdown period, compared to the total averages of Table 2. For 2019 period, the averages are slightly higher compared to these total averages. Finally, we see that less companies make transactions in 2020 period, compared to the 2019 period.

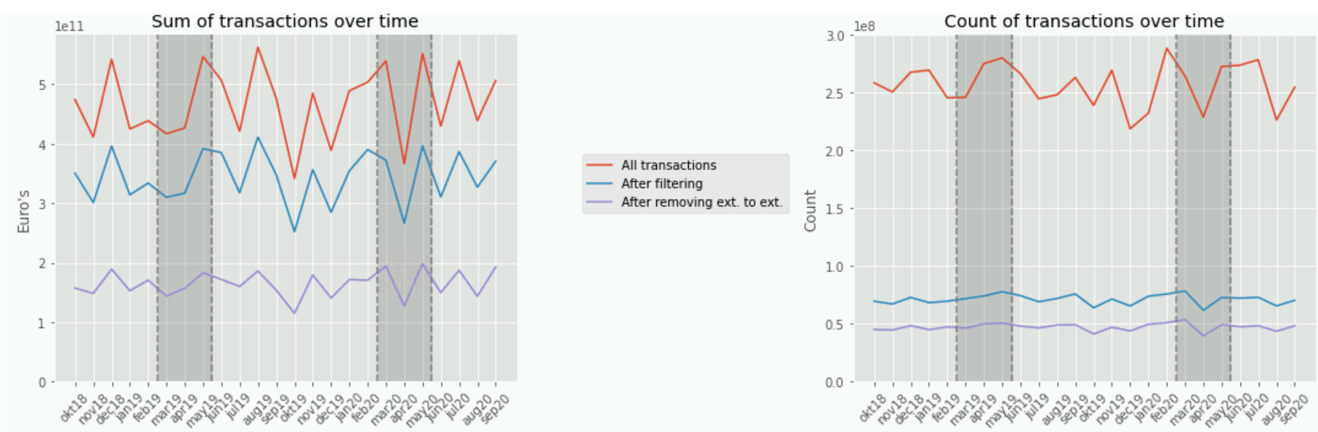

Figure 9: Total monthly sum and count of transactions of different data selections over time

	2019 period			2020 period		
	All transactions	**After filtering**	**After removing ext. to ext.**	**All transactions**	**After filtering**	**After removing ext. to ext.**
Total sum (billions €)	1470.1	1098.8	521.0	1295.0	949.1	456.4
Count (millions)	779.3	216.5	143.1	697.3	211.4	140.0
Sum per month (billions €)	490.0	366.3	173.7	431.7	316.4	152.1
Count per month (millions)	259.8	72.2	47.7	232.4	70.5	46.7
Mean amount per transaction (€)	1886	5075	3640	1857	4488	3260
Amount of companies	1,168,739	1,079,709	912,246	1,137,278	1,051,775	890,866

Table 3: Properties of different data selections where the 2020 period is compared to the 2019 period

4.2 Network description

In this section we analyze the network that we can create from the transaction data. For this we study the statistical properties and structure of the 2019 and 2020 network, using various network metrics.

4.2.1 Density of the network

First, we describe the effect of the removal of edges on the density of the network. As mentioned, we removed insignificant edges from the network that did not reflect dependencies between industries, based on thresholds. The choice of threshold values, will not majorly affect our findings in terms of total flow, as we remove only edges with negligible weights. We remove 85% of all the edges, using our filtering approach. This is depicted in Table 4. The number of nodes and edges before removing noise was respectively **928** and **175,798**. The density of this network was **0.204**, meaning that the network contains rouglhy 1/5th of the total amount of possible edges. This is relatively dense considering the density of most real-world networks [5]. After removing insignificant edges, the network contains **788** nodes and **28,007** edges. The density is now **0.045**. We removed most edges, but considering the sum of edge values, we removed only 1.2%.

	Before removing edges	**After removing edges**
Number of nodes	928	788
Number of edges	175,798	28,007
Density	0.204	0.045

Table 4: Properties of the network before and after the removal of insignificant edges

4.2.2 Components

In Table 5, we see that nearly all 788 industries are part of the GWCC. This means that the majority of the economic activity is captured in the GWCC, which allows us to reliably use methods that operate on only one component. An example of such a method is for example the maximum flow algorithm. The GSCC contains 674 nodes of the in total 788 nodes. The GIN consists of 65 nodes, the GOUT of 45 nodes and there are 4 industries in the DC's.

Number of nodes	788
Number of nodes in GWCC	784
Number of nodes in GSCC	674
Number of nodes in GIN	65
Number of nodes in GOUT	45
Number of nodes in DC	4

Table 5: Amount of nodes in different components

The sum of the monthly transaction amount over all edges in the inter-industry network are respectively 18.220 billion in 2019 and 19.525 billion during the intelligent lockdown period. These numbers are without flows from external parties. This means that during the intelligent lockdown period, the total sum of inter-industry transactions was actually higher, compared to the same period in 2019.

4.2.3 Strength distributions

In this section we analyze the strength property of the different nodes in the network. We analyze the different strength distributions by considering the shape and the mean of these distributions. The shape is determined by determining which function is a best-fit for the distribution: power-law, Weibull or log-normal. For the strength distributions, we consider the networks without the removal of insignificant edges (mentioned in Section 3.2.2). This is not needed, as we look at an aggregate metric. Removing insignificant edges disrupts the left-hand side of our distributions, making it more difficult to fit a probability distribution. In terms of total flow in euro's, removing insignificant edges has virtually no impact. Figure 10 and 12 show the in-strength v^{in}, out-strength v^{out}, in-strength c^{in} and out-strength c^{out} distributions of 2019 and 2020. For all eight distributions, we also show the best fitting probability distribution functions. The data is plotted on a log-log scale, using logarithmic binning. This also means that industries with a strength of 0 are not displayed in the histograms. In the appendix we show all the strength distributions, including the different probability distribution functions that we fitted on the data (Figure 26, 27, 28 and 29).

For all distributions, it is clear that the power-law function does not fit well. Similar to [12], the distributions decay faster than a power-law. The Weibull and the log-normal functions provide better fits for the data. In Figure 10, we can see that the log-normal distribution fits the best for both distributions. However, the fits are quite different. In Figure 12, we see that for the 2019 data, a Weibull is the best fit. For the 2020 data, log-normal provides a better fit. In Table 6, we can see that for the out-strength v^{out}, both log-normal fits do have rather similar parameters.

Considering both figures, it is difficult to clearly differentiate between the 2019 and 2020 strength V distributions. On average industries have a higher industry output/input during the 2020 period. This can be explained by the largest industries. In Table 10 and 11 we depict the largest industries based on their in-strength v^{in} and out-strength v^{out}, for both the 2019 and 2020 weights. We can see that the largest hub, "Supermarkets and Other Grocery (except Convenience) Stores", especially had a spectacular increase in industry inflows and outflows.

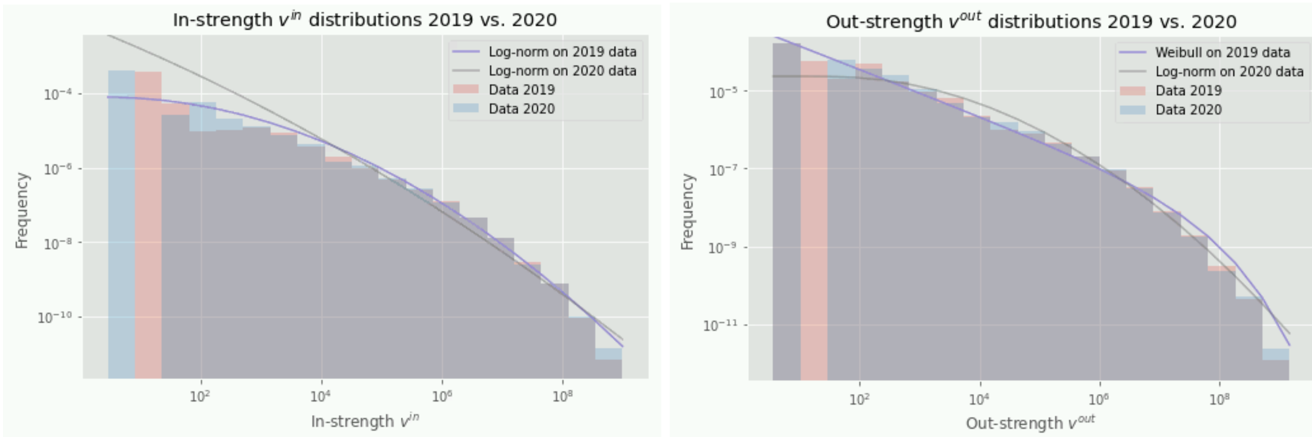

Figure 10: In-strength v^{in} and out-strength v^{out} distributions of 2019 and 2020 plotted on top of each other, including the best fitting probability density function

As noted in [35], there are two apparent reasons for the increase: first of all, consumers displayed hoarding behavior, stocking up on supplies which they normally would not have bought. Secondly, restaurants and many other food and drink related establishments were forced to close during the intelligent lockdown period. Supermarkets form a substitute for this demand, increasing their revenues. Next, the top 10 industry "Drugs and Druggists' Sundries Merchant Wholesalers" nearly doubled in both in-strength v^{in} and out-strength v^{out}. The reason for this is that consumers stock up on medical products, also for precautionary reasons. Other hubs, like "Third Party Administration of Insurance and Pension Funds" and "Miscellaneous Financial Investment Activities" also increased. Another example is "Toy and Hobby Goods and Supplies Merchant Wholesalers", which enters the top 10 of out-strength v^{out}. This may be explained by the fact that certain web shops fall under this category.

There is also some examples of industries declining in in-strength v^{in} and/or out-strength v^{out}, like "Direct Life Insurance Carriers" and "Security Guards and Patrol Services". For the life insurance carriers, this can be explained by a decline for regular healthcare demand, due to the Corona virus. It is also quite obvious why there is a decline for "Security Guards and Patrol Services", as stores and businesses had to close their doors and events were cancelled.

Most of the declines tend to be smaller. We can also see this in Figure 11, which depicts the difference in out-strength v^{out} between 2019 and 2020. The average in-strengths $\langle v^{in}\rangle$ of 2019 and 2020 respectively equal **15,470,327** and **16,459,709** ($\langle v^{in}\rangle=\langle v^{out}\rangle$). We can see that much more industries actually declined in their out-strength v^{out}. However, the increases tend to be larger.

The in-strength c^{in} and out-strength c^{out} distributions of 2019 and 2020 are depicted in Figure 12. Both the log-normal and the Weibull functions provide better fits for these four distributions, compared to the in-strength v^{in} and out-strength v^{out} distributions. The average in-strength $\langle c^{in}\rangle$ of 2019 and 2020 respectively equal **1784** and **1593**. We can clearly see that in 2019 much more transactions were made, compared to the lockdown period of 2020. This can also be seen in the picture, as the 2019 distributions cover a slightly bigger areas compared to the 2020 distributions.

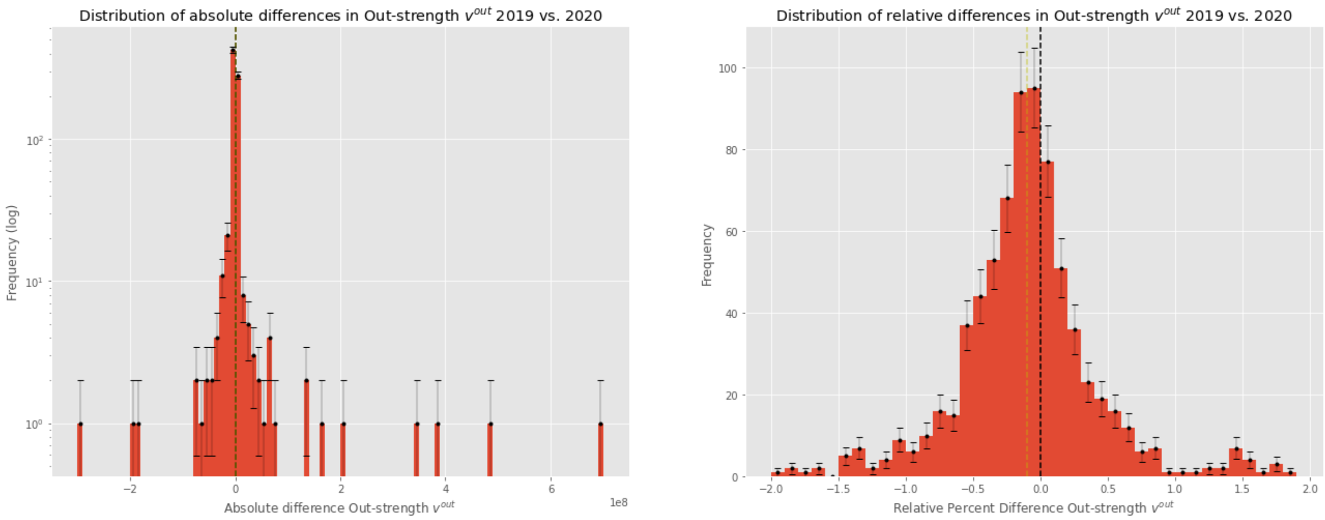

Figure 11: Distributions of absolute and relative differences for out-strength v^{out} of industries, 2019 compared to 2020

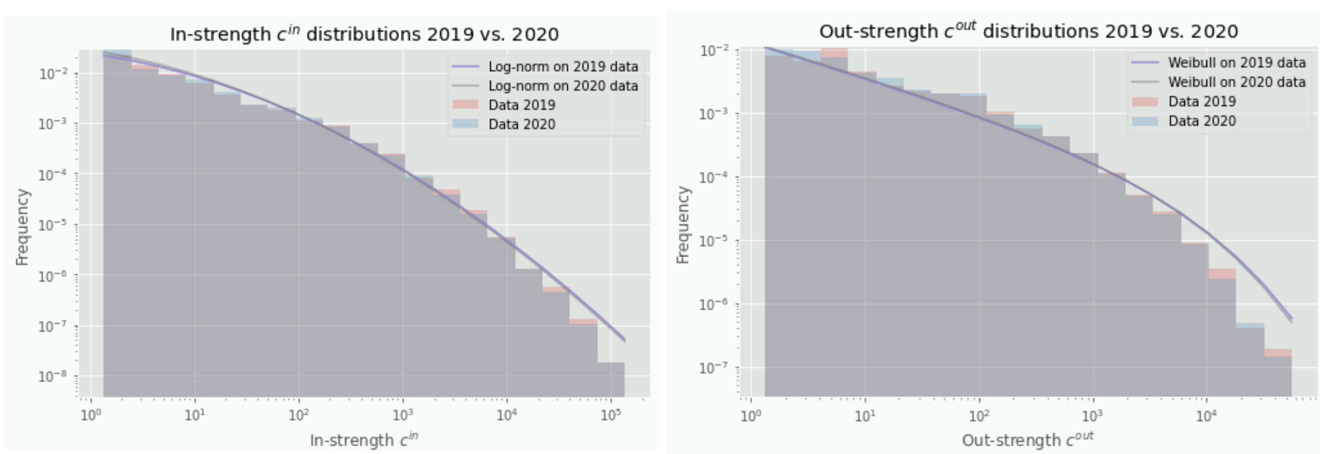

Figure 12: In-strength c^{in} and out-strength c^{out} distributions of 2019 and 2020 plotted on top of each other, including the best fitting probability density function

Overall, we can conclude that in terms of total value flow, the negative impact on the economy as a whole is difficult to detect. There are more industries that slightly decline in out-strength v^{out} and in-strength v^{in}, but there is also industries that profit from the intelligent lockdown effects. In terms of transaction counts, it is clear that the intelligent lockdown causes a decline here. However, the transactions in the 2020 period tend to be bigger. An explanation for this might be that consumers have less opportunity to visit stores meaning they make less transactions, but have to buy more when they do. Consumer payments are not represented within these distributions, however in general we could say that inter-industry payments higher on the supply chain often follow consumer demand patterns.

Distribution	Period	Weibull		Log-normal		Log-likelihood	Best fit
		k	λ	μ	σ		
In-strength v^{in} (value in)	**2019**	0.333	23,493,156	13.760	3.741	-285.103	Log-norm
	2020	0.305	49,294,457	13.634	7.121	-276.636	Log-norm
Out-strength v^{out} (value out)	**2019**	0.930	19,652,489	13.992	3.265	-292.259	Weibull
	2020	0.374	25,275,083	13.852	3.314	-287.960	Log-norm
In-strength c^{in} (transaction count in)	**2019**	0.470	1352.680	5.223	2.720	-167.148	Log-norm
	2020	0.574	1384.707	5.039	2.748	-167.933	Log-norm
Out-strength c^{out} (transaction count out)	**2019**	0.453	2772.828	5.728	2.341	-157.457	Weibull
	2020	0.453	2470.017	5.615	2.328	-156.238	Weibull

Table 6: Overview of the Weibull and log-normal fits on the different strength distributions

4.3 Random-walk centrality

In this section, we analyze the random-walk centrality RW_C of industries in the network. We describe the most important industries based on the random-walk centrality, as well as the question to what extent industries with a high random-walk centrality are more susceptible to shocks (research question 1.3 and 1.4). The top ten industries for both 2019 and 2020 are depicted in Table 7.

Industry	RW_C 2019	**Industry**	RW_C 2020
Miscellaneous Financial Investment Activities	0.0448	Miscellaneous Financial Investment Activities	0.0441
Offices of Notaries	0.0160	Offices of Other Holding Companies	0.0276
Offices of Other Holding Companies	0.0157	Offices of Notaries	0.0147
General Freight Trucking, Local	0.0145	General Freight Trucking, Local	0.0135
Commercial and Institutional Building Construction	0.0140	Administrative Management and General Management Consulting Services	0.0123
Employment Placement Agencies	0.0133	Employment Placement Agencies	0.0119
Direct Life Insurance Carriers	0.0127	Direct Life Insurance Carriers	0.0113
Commercial Banking	0.0125	Offices of Real Estate Agents and Brokers	0.0111
All Other Legal Services	0.0083	Insurance Agencies and Brokerages	0.0110
Other Heavy and Civil Engineering Construction	0.0081	Commercial and Institutional Building Construction	0.0010

Table 7: Top 10 industries based on random-walk centrality RW_C in 2019 and 2020

Most industries in the top 10 are the same for 2019 and 2020. The industries that score high on RW_C are generally industries that provide services for a wide range of other industries. An example of this is the largest industry in both 2019 and 2020 is "Miscellaneous Financial Investment Activities". Other highly ranked industries are administrative industries, such as: "Offices of Notaries" and "Offices of Other Holding Companies". A decline or increase in RW_C would not necessarily mean a decline in incoming money flows for a certain industry. It could for example also mean that neighbouring industries pay relatively more or less to other industries, compared to the 2019 period. In other words, the industry is less central in the network.

According Blöchl et al., industries with a high RW_C are most immediately impacted by a shock. By relating the absolute change in out-strength v^{out} to the random-walk centrality of industries in 2019, we try to get an idea of this effect. We consider the absolute value as there are industries that profit from the intelligent lockdown and industries that do not. In this way, we consider the change in out-strength v^{out}, either negatively or positively. We take the RW_C of 2019, as we consider this to be the ground-truth value for industries, in a "normal situation". We expect that industries that are central in the normal situation, are more susceptible for shocks.

Figure 13 shows the random-walk centrality plotted against the absolute change in out-strength v^{out}, including two zoomed-in plots. In the most left plot, there is a correlation of 0.5236, which indicates a weak correlation. This could be highly influenced by the outliers. For the highest industry regarding RW_C, “Miscellaneous Financial Investment Activities”, we also see a high absolute change in out-strength v^{out}. In this case, we notice an increase. As many other industries are close to this industry, this increase might be an aggregation over different shocks. When removing few of the outliers, we obtain a slightly lower correlation value, as is depicted in “Zoom1”. A clear pattern is still difficult to detect here, and might also be affected to much by bigger industries. Finally, when zooming in even more, we find a correlation value of 0.5765, which is shown in the most right plot. We see that some industries with a high random-walk centrality indeed have a high absolute change in out-strength v^{out}. However, we also can spot many cases for which this is not true. When zooming even further in, we do not obtain clearer patterns. These plots are given in the Appendix.

Correlation between RWC 2019 value and absolute difference in out-strength v^{out}

Zoom1 Zoom2

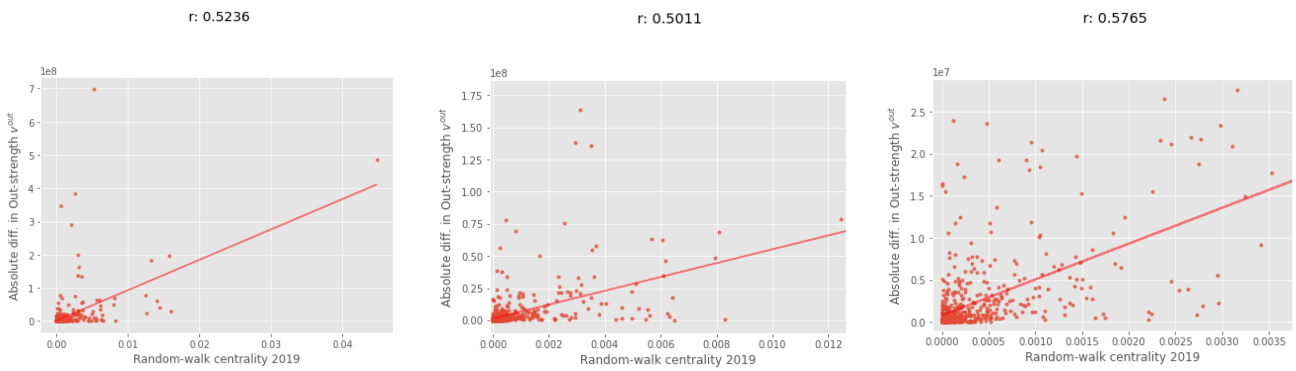

Figure 13

To conclude, we can see that the expectation of Blöchl et al. does not hold up, when empirically testing it using a real-world example. Although there is a slight correlation between the change in out-strength v^{out} and the random-walk centrality, the evidence that centrality indicates a higher susceptibility for shocks is too weak. This means that it is not the case that industries with a high random-walk centrality are more likely to be affected during the shock of the intelligent lockdown. An explanation for this might be that we can not speak of “one clear shock”. The impact of the Corona crisis depends on a range of different factors, affecting the economy at multiple places at the same time. This also leads to a set of different positive and negative shocks.

In the Appendix we include related findings, where we also tried to detect a relation between shock susceptibility and random-walk centrality. We were however not able to find such a relation in these plots. In Figure 32, we depict the relative change in out-strength v^{out} plotted against the random-walk centrality. Next to this, we consider the “reduced” random-walk centrality, where the transition probabilities are altered in such a way, that self-loops and node pairs receive less weight. Finally, in Figure 33, we plotted the relation between average MFPT and relative change in out-strength v^{ut}, where we colored industries based on the sector they belong to. We did this to see whether the relation between MFPT and susceptibility for a shock is influenced by sector or not.

4.4 Impact of the consumer

In this section, we specifically analyze consumer spending behavior. During the intelligent lockdown, consumer spending declined. Some industries were impacted directly by this decline, some industries more indirectly. We want to quantify these effects. First, we consider total consumer spendings over time, where we distinguish between card payments and cashless payments. This is depicted in Figure 14. We can see that in total, consumers spend more on cashless transactions compared to card payments. As already mentioned, the cashless category comprises all transactions that can be related to private individuals and are not associated with card payments. This category consists mostly of online payments. Considering the entire course of the plot, we first notice two peaks in the December months, both for cards and cashless payments. This effect is caused by the holiday period in December. When specifically comparing the 2020 period to the 2019 period, we see that for cashless payments, the consumer spendings were slightly lower in March and May. However, we can see an increase in April 2020, compared to 2019. Over the sum of all relevant transactions (Figure 3), we could see a clear decline in April. This discrepancy may be explained by the growth of the online payment channel, during the lockdown period, causing the sum of cashless payments to increase.

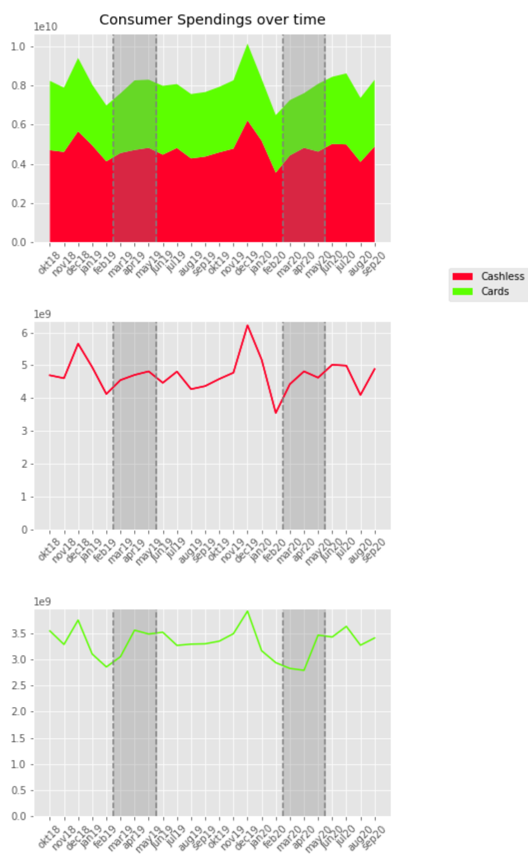

Figure 14: Sum of payment values for cashless and card payments of consumers over time. The top chart is the stacked chart of the other two graphs. The highlighted areas depict the 2019 and 2020 period

For card payments however, we do notice a clear decline, especially for the month April. The fact that physical stores had to close, would be the most logical explanation for this. The reason we do not see a clear decline in March, might be caused by the hoarding behavior of consumers in the first weeks of March 2020. The cashless payment channel appears to have become a substitute of the card payments channel in some cases, although this does not completely make up for the decline in card payments. This is especially the case for companies in the Wholesale Trade sector, as we can also see in Figure 15 and Figure 16. These figures give a broad idea which parts of the economy were affected the most by changes in consumer spendings during lockdown.

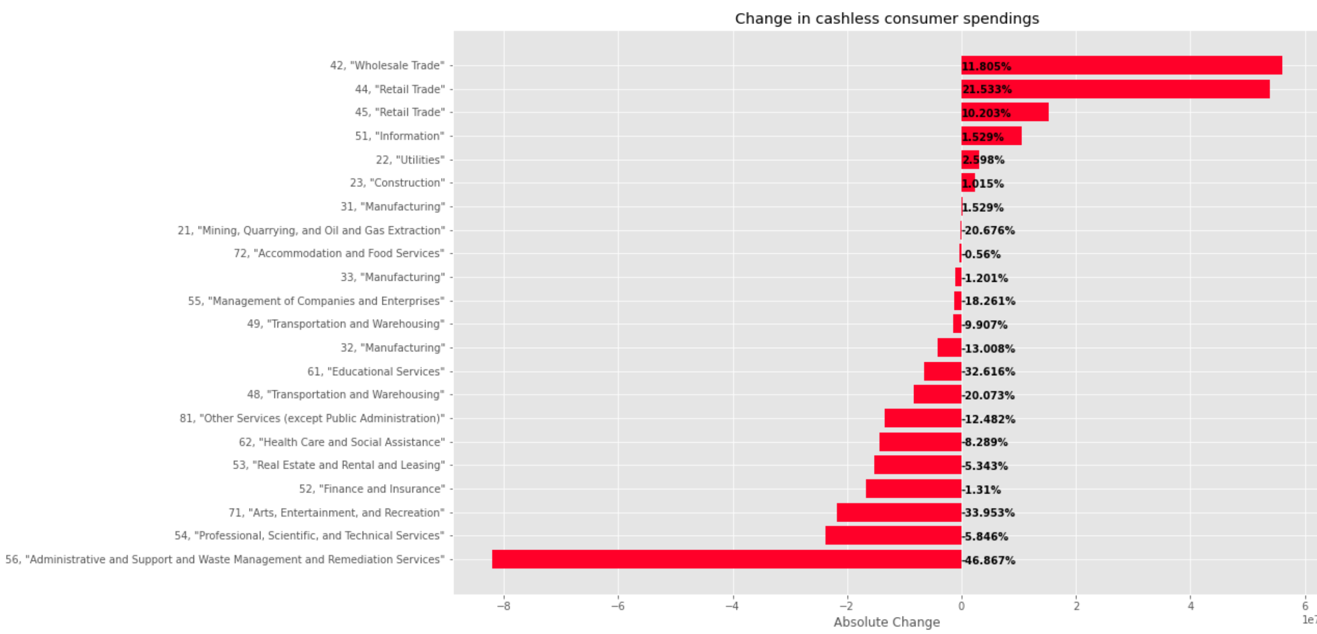

Figure 15: Sectors sorted on absolute change in total cashless consumer income, over the intelligent lockdown period compared to the same period in 2019 (the text displays the relative changes)

We must note that we look at an highly aggregated level, meaning there is a plethora of factors at play. In the next section, we will analyze more closely what industries cause these effects. For cashless payments, we see the biggest increase for “Wholesale Trade” and “Retail Trade”. This is largely influenced by the rise of supermarkets and web shops. Further, the increase in “Information” sector may be explained by a higher usage of wireless telecommunication and other information services. The biggest decline we can see is for sector “Administrative and Support and Waste Management and Remediation Services”, which amongst others consists of travel agency organizations. Further, we see a decrease for “Arts, Entertainment, and Recreation”, which forms the “Cultural sector”. The change in cards expenditures for individual sectors is depicted in Figure 16. Here, it is clear that most sectors declined. As expected, we notice big decreases for the “Accommodation and Food Services” and “Arts, Entertainment and Recreation” sector. Next, we see sectors decline associated with transportation and traveling. In the next section we will take a closer look at these results.

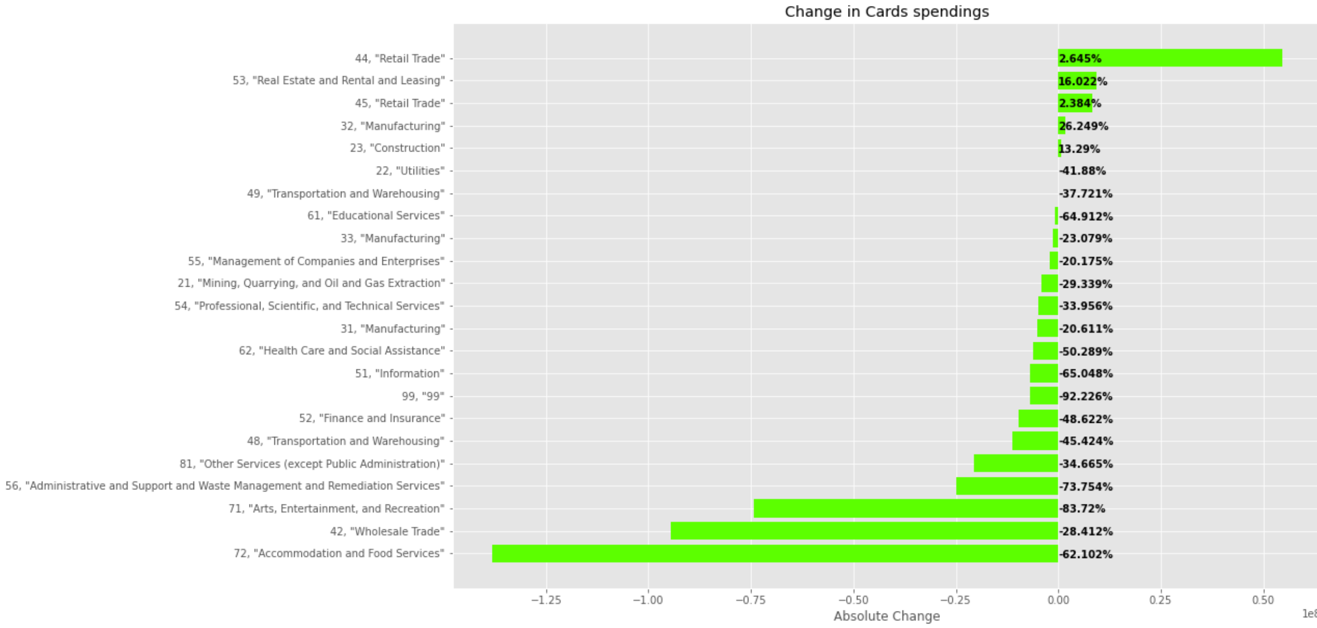

Figure 16: Sectors sorted on absolute change in total card consumer income, over the intelligent lockdown period compared to the same period in 2019 (the text displays the relative changes)

4.5 Maximum flow analysis

In this section, we analyze the maximum flow results, in order to get an idea of the effects of the changed consumer spending behavior. In particular, we try to get an idea of the indirect effect of this, using the maximum flow value as mentioned in Section 3.5.

4.5.1 Maximum flow results

The maximum flow values for industries in 2020 opposed to 2019 are depicted in Figure 17. We also provide zoomed-in versions to see changes in maximum flow for smaller industries. The average maximum flow value in 2020 is **25,760,655** is slightly higher compared to 2019, where it is **25,257,358**. This is not according to expectation, however, we also saw this for the average in- and out-strength v_{in} and v_{out}: the hubs that increase, influence the averages. In the plots in Figure 17, we can also see that the industries with the biggest flow tend to be on the left side of the dashed line. When zooming in, we can however see that for most industries there tends to be a decline in maximum flow, especially for smaller industries. We can also see this in Figure 19.

The industries for which the maximum flow declined and increased the most, are depicted in Figure 18. We could say that these industries are impacted the most by the changing consumer spending behavior, as well directly as indirectly. In Figure 19, we can see that most industries declined in maximum flow. The highest increases tend to be bigger in absolute terms, compared to declines in maximum flow.

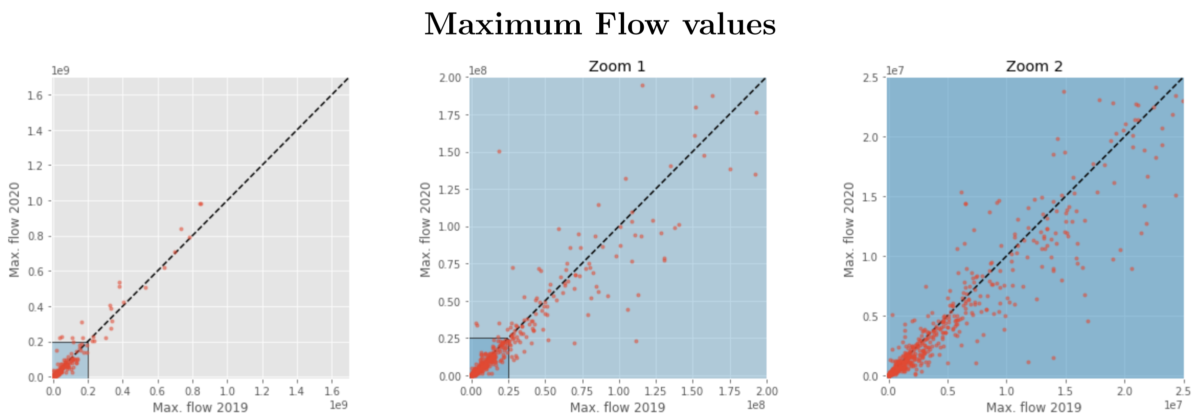

Figure 17: Maximum flow values 2019 vs. 2020, with two zoomed-in plots

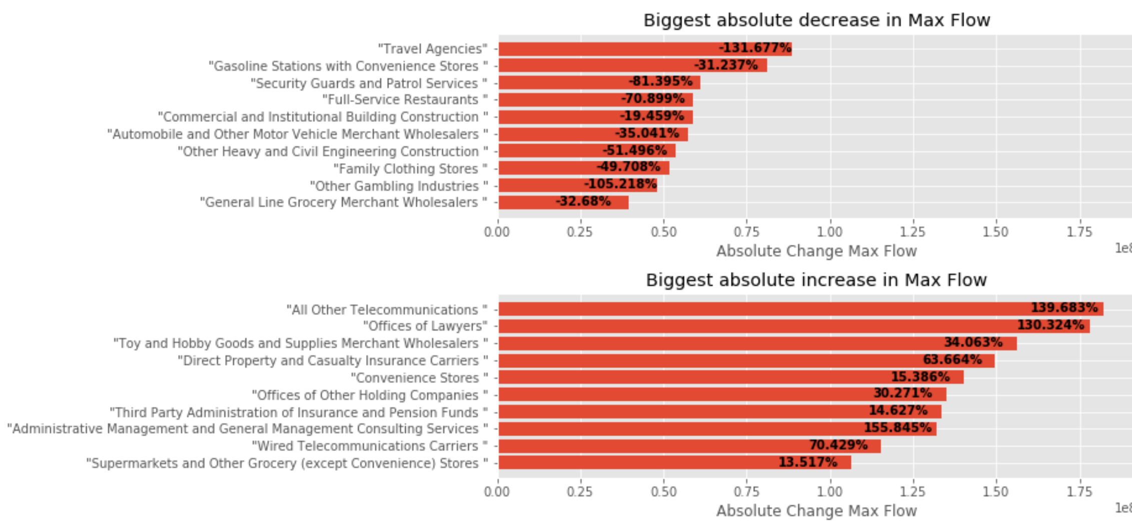

Figure 18: Top ten biggest increases and decreases in Max flow

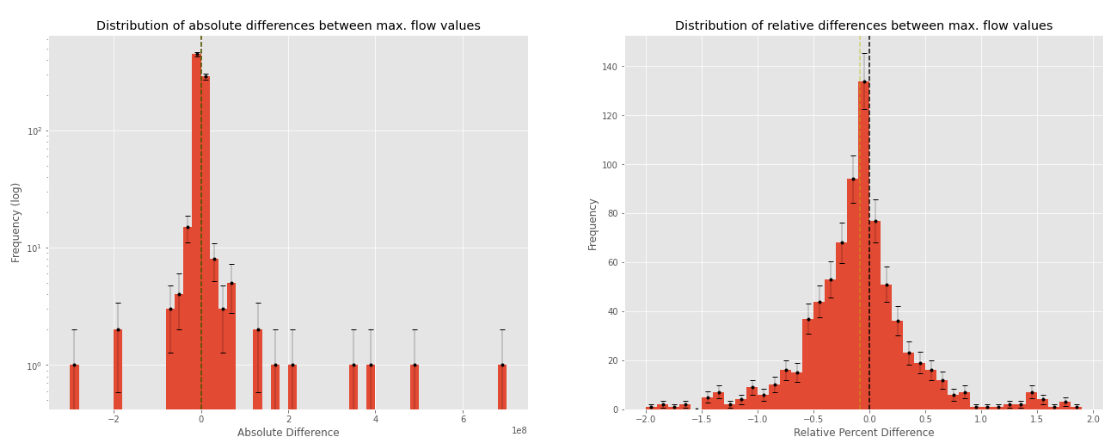

Figure 19: Distributions of absolute and relative differences for Max. flow of industries 2019 compared to 2020

The maximum flow gives a combined metric for direct and indirect effects of changes in consumer spending behavior. Some industries rely mostly on direct flow, some industries mostly on indirect flow and for other industries the maximum flow value is an equal mix of both components. On average, over all industries, roughly 43% of the maximum flow can be contributed to direct flow. It is interesting to consider the direct and indirect effects separately and also which of these effects contributes most to the maximum flow value. In the following sections we analyze these effects.

4.5.2 Direct effect: To what extend did consumer spending change during lockdown?

First we analyze how the direct consumer spending changed in the intelligent lockdown. 747 out of 784 industries received direct flow from consumers, either in 2019 and/or in 2020 (given our removal of insignificant edges). However, some of these industries depend more on consumer inflow than others. For the direct effect, we only consider the industries that depend most on consumer spending. Here, we introduce a threshold: if an industry receives more than 10% of its total inflow from consumers, either in 2019 or 2020, then we state that this industry depends on direct inflow. This applies for 296 out of 784 industries. The average of direct consumer payments for these industries declined with **3.6%**. For 2019, 10,462,364 card payments and 10,924,893 cashless payments were made on average over all industries. In 2020 this was 9,636,099 and 10,984,241 respectively. So we see on average a decline of **7.9%** in card payments, and a small increase of **0.5%** in cashless payments. As already mentioned, the decline in card payments may be explained by the decline in "physical payments". In Figure 20, the direct flow of the 2020 period versus the 2019 period is depicted. It is clear that most industries have declined, as we can also see in the histograms of the change in direct flow in Figure 22. The top ten biggest increasing and decreasing industries regarding direct flow are shown in Figure 21.

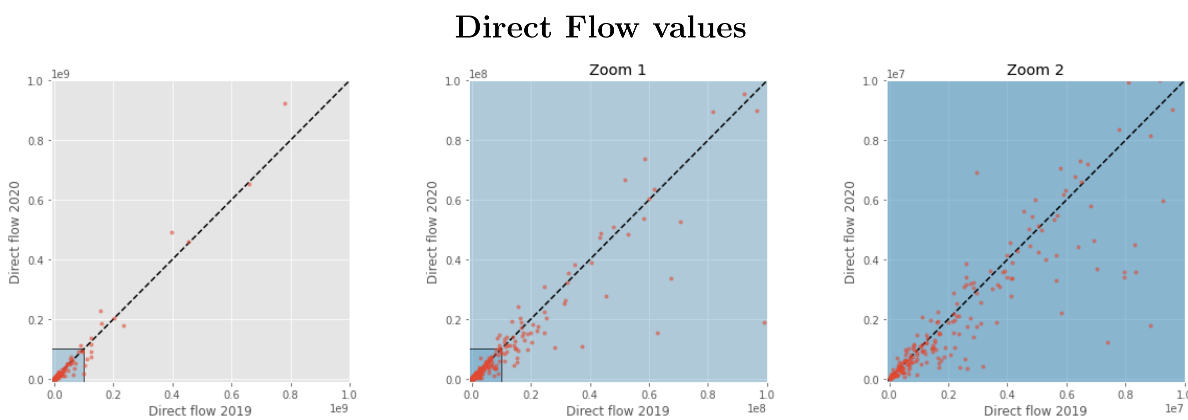

Figure 20: Direct flow values 2019 vs. 2020, with two zoomed-in plots

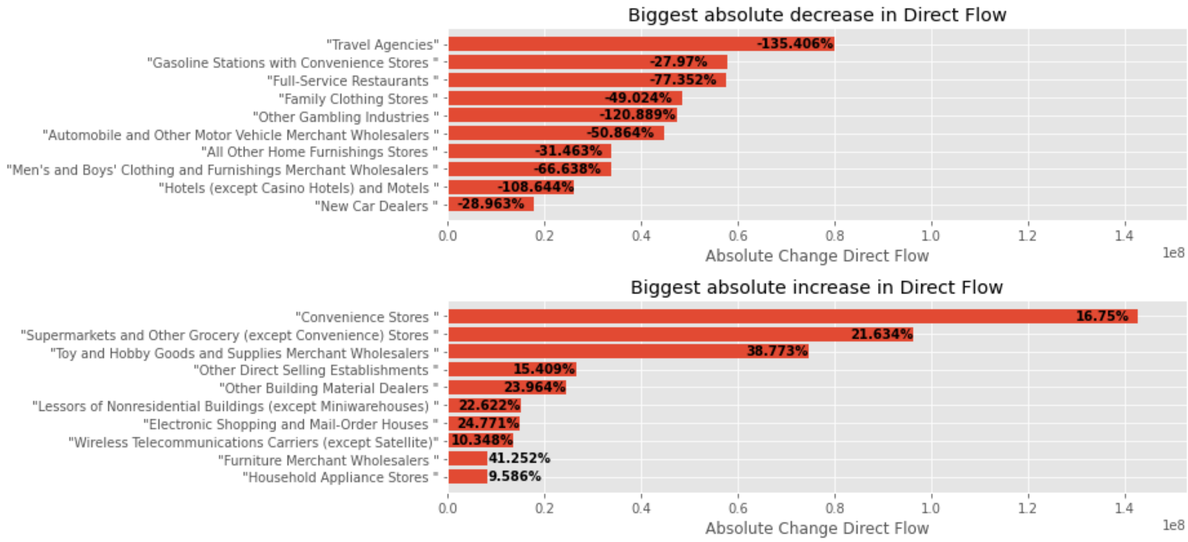

Figure 21: Top ten biggest increases and decreases in Direct flow

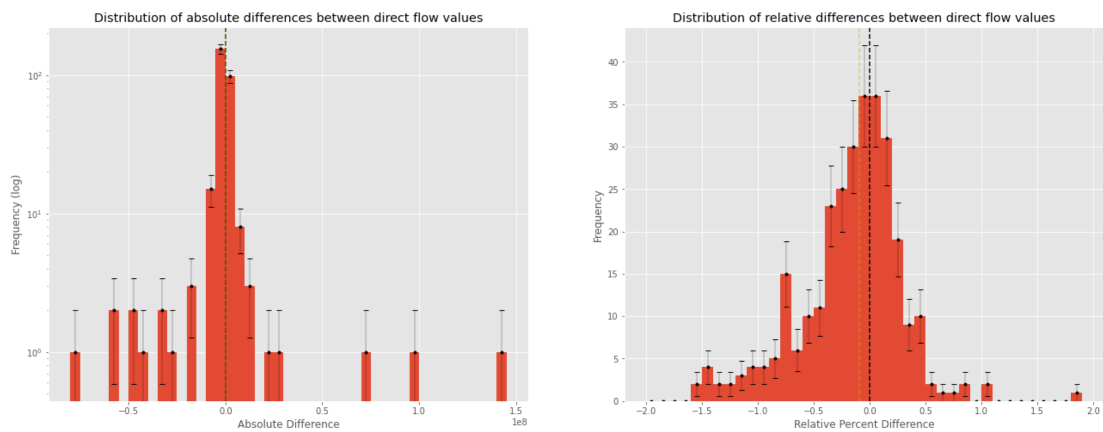

Figure 22: Distributions of absolute and relative differences of direct flow for industries in 2019 and 2020

4.5.3 Indirect effect: What are indirect effects of the change in consumer spending?

We have seen the change in direct consumer spendings. Indirectly, industries might also be impacted by a change in consumer expenditures. For this we consider the industries that do not receive more than 10% of their total income from consumers, both in 2019 and 2020. This is true for 488 industries out of the 784. We could say that these industries are higher on the supply chains starting from consumers. For the indirect component, we subtract the direct consumer spending from the maximum flow value. In Figure 23, we see plot the change for the indirect flow value for these 488 industries. The average in 2020 is higher than in 2019, namely **22,107,872** versus **21,175,247**. We can see some major increases, as Figure 24 and Figure 25 also indicate. This is not the case for decreasing industries. Overall, we see that more industries decline for indirect flow. In total, 288 industries decline and 160 increase. Figure 24 depicts the top ten biggest increases and decreases regarding the indirect component of the maximum flow.

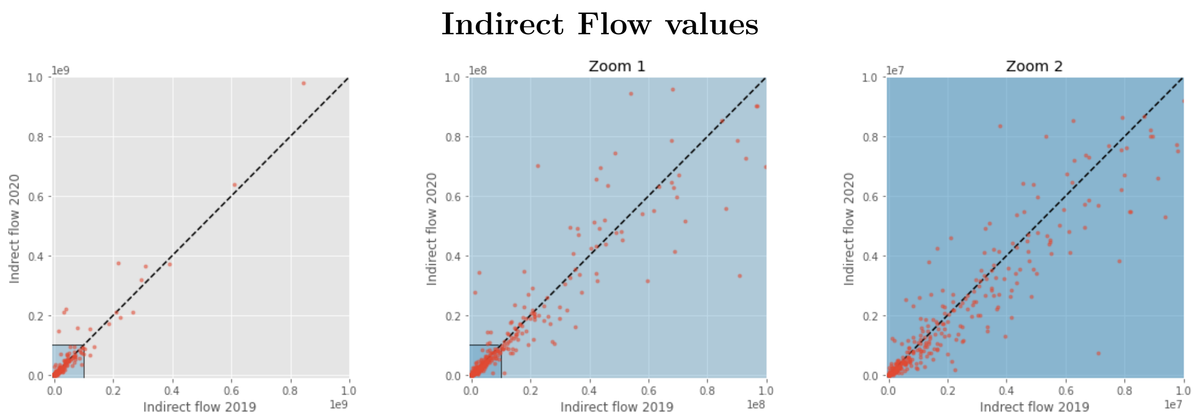

Figure 23: Indirect flow values 2019 vs. 2020, the middle and right plot are zoomed-in versions of the left plot

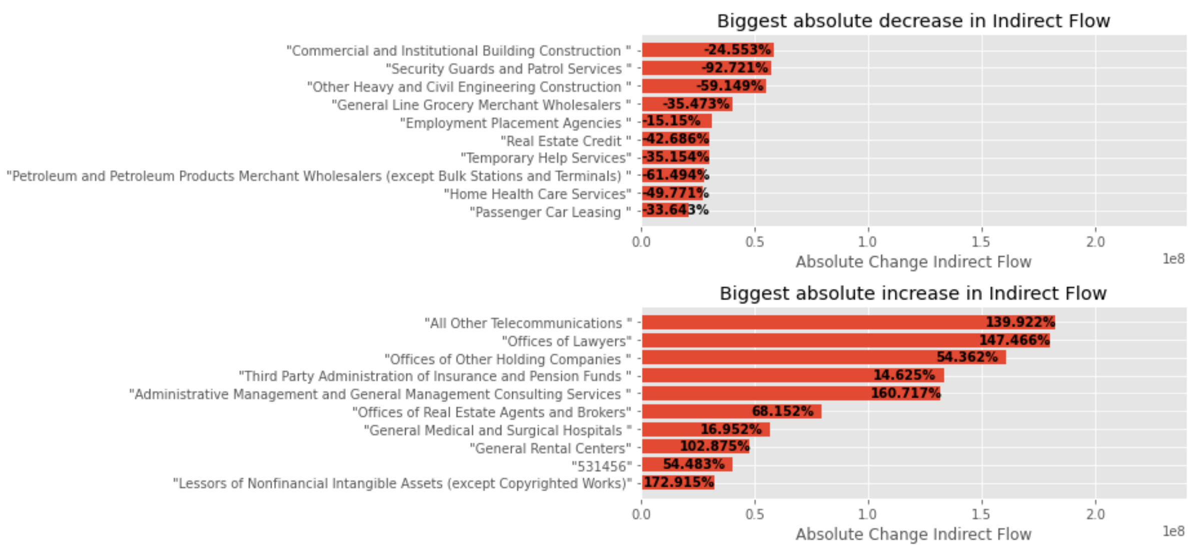

Figure 24: Top ten biggest increases and decreases in indirect flow for industries in 2019 and 2020. The "531456" industry refers to some real estate industry

We see similar results when separating the maximum flow measure into its direct and indirect components: more industries decrease than increase and considering the biggest differences, the increases tend to be larger than the decreases, which influences the averages of the metrics. Considering Figure 18, some of the biggest differences are caused by a change in direct flow, as we can also see these industries back in Figure 21. Examples are "Travel Agencies" and "Supermarkets and Other Grocery (except Convenience) Stores". Some of these biggest differences are caused primarily by the indirect effect, as we can see these industries back in Figure 24. Examples are "Security Guards and Patrol Services" and "All Other Telecommunications". The direct flow component gives a representation of industries that are initially impacted by the external shock. This gives intuitive results, as for most industries in Figure 21, there is an explanation for the difference in consumer spending, caused by the intelligent lockdown.

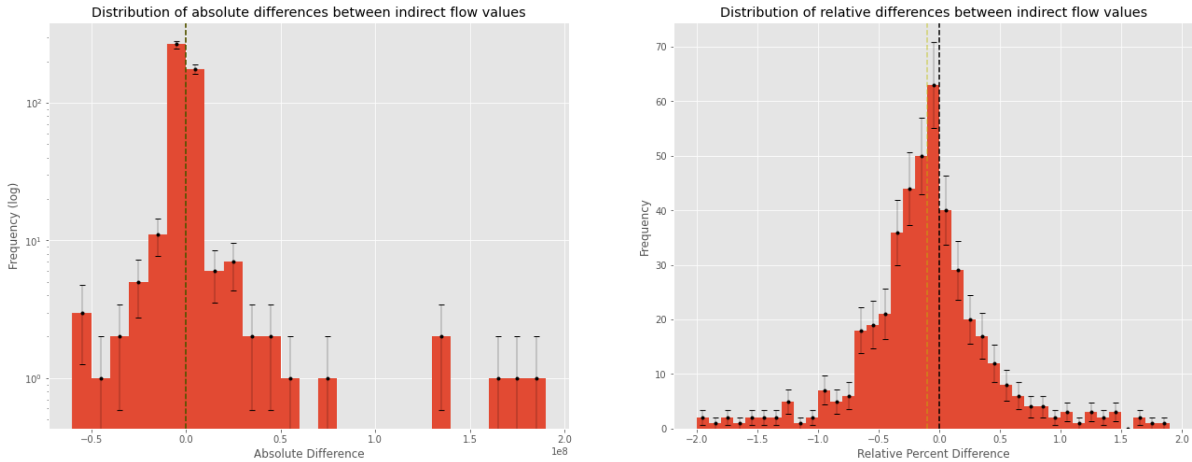

Figure 25: Distributions of absolute and relative differences of Indirect flow

The indirect component tries to capture network effects. Considering the top ten increasing and decreasing industries in Figure 24, there are indeed examples for which we can attribute the decrease in indirect flow to a decrease in consumer expenditures in a downstream industry. The decline for "General Line Grocery Merchant Wholesalers" might be explained by the fact that food and leisure organizations do not need to buy supplies anymore. The decline for " Petroleum and Petroleum Products Merchant Wholesalers" is a logical consequence of the lower expenditures at gasoline stations. For other examples, it is hard to determine whether these effects are actually caused by the decline in consumer spendings for most of them. In general, we could say that these are industries that are affected most by Covid-19 measures. We already saw the decline for "Security Guards and Patrol Services", which is partly caused by the fact that consumers do not visit physical establishments anymore. Furthermore, we notice declines for construction industries. We can also notice effects of the "reduce in movements", regarding the decline for petroleum wholesalers and "Passenger Car Leasing". For the top 10 increases, we again notice the increase for "All Other Telecommunications", which might be an indirect effect of the fact that consumers need to be able to communicate remotely. Furthermore, we notice many increases for administrative services. For most industries it is difficult to claim that the increase is due to a rise in consumer demand. The indirect component seems to give an indication of industries' supply chains that are in general mostly affected by the intelligent lockdown and not necessarily by consumer spending behavior.

5 Conclusion

In this thesis, we empirically described the effect of the intelligent lockdown on the Dutch inter-industry transaction network. Using transaction data is an appropriate method to describe changes in an economy. However, explaining why these changes occur is difficult, as it is hard to attribute them to one or a few factors. Constructing a sensible aggregated network from raw company transaction data is not trivial. It requires a careful process of data pre-processing and selection, as well as aggregation and filtering. However, creating a network from the data allowed us to go beyond counting transactions and to describe network effects.

The impact of the intelligent lockdown was measured by comparing two networks with each other, constructed from different snapshots of the data. These networks were first of all compared by their strength distributions. Doing this, we were able to learn about the structure of both networks. Even though we can see a decline in the raw transaction data over the 2020 period, both for the transaction count and transaction value, the structure of both networks remained quite similar. The shape of the strength distributions can be estimated by either a log-normal or Weibull function. Considering the strength in terms of the value of transactions, most industries declined. The average strength actually increased, primarily caused by hubs. For the strength in terms of the transaction count, we could both see a decline for most industries as well as for average values.

Next to strength, we also considered the random-walk centrality of nodes. This metric learned us which industries are most central in an economy, in terms of how close industries are to you, based on money flows. The intuition behind random-walk centrality is, that industries with a high value are more susceptible for shocks. Findings showed that the top 10 industries for this metric are generally industries that offer services to a wide range of other industries, such as such as investments, administration and transport services. It was difficult to test whether these industries were most immediately affected by the intelligent lockdown shock, as we can not speak of "one shock": during the intelligent lockdown, various macro-economic factors either positively or negatively affected industries in the network.

By analyzing consumption, we focused on one of these factors. Here, we could also see a mixture of positive and negative effects. Especially industries in the food and accommodation, and cultural sector, were directly impacted by a decline in consumer spendings. Supermarkets and webshops were positively affected. The type of payment method that consumers use, either cashless or by card, also influences the effect. To take also the indirect effect of the change in consumer spending into account, we used the maximum flow metric. Considering the difference in maximum flow over the 2019 and 2020 network, could indicate network effects. This is because it depicts a change in how much value can flow through edges over the network, also for industries higher on the supply chain. In some cases, we can clearly see that a decline in consumer spending to a downstream industry, can affect upstream industries. However, in most cases, the change in maximum flow seems to give a general indication of the impact of the lockdown on industries' supply chains. It is hard to attribute findings in the network to only one external factor, as in reality, a variety of factors are at play.

Future Work

This work can be improved in several ways. First of all, the used methods can be extended. We can for example use the mean first passage time, used to compute the random-walk centrality, to determine which industries are closest to private individuals and relate this to the change in the output or input of industries. Regarding the maximum flow measure, we can consider other external parties to be the source of the money. An example is foreign organizations. A consequence of the intelligent lockdown is an impediment on foreign trades. This could also be the cause of a shock, influencing certain industries directly and indirectly.

Regarding both the random-walk centrality and maximum flow measure, it is difficult to find actual network effects and especially to attribute them to certain factors. An alternative way to quantify network effects, is describing the cascade in the network. This can be done using an input-output model, as described in Section 1.2. Some of these models include macroeconomic factors in order to make the prediction more realistic. There are also variations on input-output models, that incorporate the fact that a cascade is caused by either demand or supply shocks, as described in [36]. A drawback to using input-output models, is the assumption that a cascade is a linear function.

Another possibility would be to use other flow-based metrics, apart from the ones used in this work. An example is using the Hemholtz-Hodge decomposition or the map equation method [15]. The Heholtz-Hodge decomposition could be especially useful to analyze the hierarchical structure of the network, learning us more about the position of industries on the supply chain. This could be related to the impact of shocks. The map equation is a flow-based community detection algorithm. The community structure might also be related to the impact of the intelligent lockdown: industries within communities might be more susceptible for cascade effects, as they interact more intensively with each other.

References

[1] Rijksoverheid, "Coronavirus tijdlijn." https://www.rijksoverheid.nl/onderwerpen/coronavirus-tijdlijn. Accessed at January 2021.

[2] CBS, "Economic contraction of 8.5 percent in q2 2020." https://www.cbs.nl/en-gb/news/2020/39/economic-contraction-of-8-5-percent-in-q2-2020. Accessed at January 2021.

[3] CBS, "Household spending almost 13 percent down in may." https://www.cbs.nl/en-gb/news/2020/30/household-spending-almost-13-percent-down-in-may. Accessed at July 2020.

[4] CBS, "Economic impact of covid-19." https://www.cbs.nl/en-gb/dossier/coronavirus-crisis-cbs-figures/economic-impact-of-covid-19. Accessed at December 2020.

[5] A.-L. Barabási *et al.*, *Network science*. Cambridge University Press, 2016.

[6] D. A. Smith and D. R. White, "Structure and dynamics of the global economy: network analysis of international trade 1965–1980," *Social forces*, vol. 70, no. 4, pp. 857–893, 1992.

[7] D. Garlaschelli and M. I. Loffredo, "Structure and evolution of the world trade network," *Physica A: Statistical Mechanics and its Applications*, vol. 355, no. 1, pp. 138–144, 2005.

[8] K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna, "The international trade network: weighted network analysis and modelling," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 02, p. P02002, 2008.

[9] A. Kharrazi, E. Rovenskaya, B. D. Fath, M. Yarime, and S. Kraines, "Quantifying the sustainability of economic resource networks: An ecological information-based approach," *Ecological Economics*, vol. 90, pp. 177–186, 2013.

[10] W. Leontief, *Input-output economics*. Oxford University Press, 1986.

[11] D. Acemoglu, V. M. Carvalho, A. Ozdaglar, and A. Tahbaz-Salehi, "The network origins of aggregate fluctuations," *Econometrica*, vol. 80, no. 5, pp. 1977–2016, 2012.

[12] J. McNerney, B. D. Fath, and G. Silverberg, "Network structure of inter-industry flows," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 24, pp. 6427–6441, 2013.

[13] G. Iosifidis, Y. Charette, E. M. Airoldi, G. Littera, L. Tassiulas, and N. A. Christakis, "Cyclic motifs in the sardex monetary network," *Nature Human Behaviour*, vol. 2, no. 11, pp. 822–829, 2018.

[14] S. R. de la Torre, J. Kalda, R. Kitt, and J. Engelbrecht, "On the topologic structure of economic complex networks: empirical evidence from large scale payment network of estonia," *Chaos, Solitons & Fractals*, vol. 90, pp. 18–27, 2016.

[15] K. Yuichi, I. Takashi, I. Hiroshi, and I. Hiroyasu, "Hierarchical and Circular Flow Structure of the Interfirm Transaction Network in Japan," Discussion papers 19063, Research Institute of Economy, Trade and Industry (RIETI), Aug. 2019.

[16] V. M. Carvalho, M. Nirei, Y. Saito, and A. Tahbaz-Salehi, "Supply chain disruptions: Evidence from the great east japan earthquake," *Columbia Business School Research Paper*, no. 17-5, 2016.

[17] H. Inoue and Y. Todo, "Propagation of negative shocks across nation-wide firm networks," *PloS one*, vol. 14, no. 3, p. e0213648, 2019.

[18] H. Inoue and Y. Todo, "The propagation of economic impacts through supply chains: The case of a mega-city lockdown to prevent the spread of covid-19," *PloS one*, vol. 15, no. 9, p. e0239251, 2020.

[19] F. Kyriakopoulos, S. Thurner, C. Puhr, and S. W. Schmitz, "Network and eigenvalue analysis of financial transaction networks," *The European Physical Journal B*, vol. 71, no. 4, p. 523, 2009.

[20] 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, no. 1, pp. 317–333, 2007.

[21] M. Elliott, B. Golub, and M. O. Jackson, "Financial networks and contagion," *American Economic Review*, vol. 104, no. 10, pp. 3115–53, 2014.

[22] D. Acemoglu, A. Ozdaglar, and A. Tahbaz-Salehi, "Systemic risk and stability in financial networks," *American Economic Review*, vol. 105, no. 2, pp. 564–608, 2015.

[23] M. Boss, M. Summer, and S. Thurner, "Contagion flow through banking networks," in *International Conference on Computational Science*, pp. 1070–1077, Springer, 2004.

[24] J. Garcia-Bernardo, J. Fichtner, F. W. Takes, and E. M. Heemskerk, "Uncovering offshore financial centers: Conduits and sinks in the global corporate ownership network," *Scientific Reports*, vol. 7, no. 1, pp. 1–10, 2017.

[25] D. E. Van Kuppevelt, F. W. Takes, and E. M. Heemskerk, "Understanding evolving communities in transnational board interlock networks," in *2018 IEEE 14th International Conference on e-Science (e-Science)*, pp. 312–313, IEEE, 2018.

[26] C. Mattsson, "Networks of monetary flow at native resolution," *arXiv preprint arXiv:1910.05596*, 2019.

[27] M. Zanin, D. Papo, M. Romance, R. Criado, and S. Moral, "The topology of card transaction money flows," *Physica A: Statistical Mechanics and its Applications*, vol. 462, pp. 134–140, 2016.

[28] D. Kondor, M. Pósfai, I. Csabai, and G. Vattay, "Do the rich get richer? an empirical analysis of the bitcoin transaction network," *PloS one*, vol. 9, no. 2, 2014.

[29] F. Blöchl, F. J. Theis, F. Vega-Redondo, and E. O. Fisher, "Vertex centralities in input-output networks reveal the structure of modern economies," *Physical Review E*, vol. 83, no. 4, p. 046127, 2011.

[30] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian journal of Mathematics*, vol. 8, pp. 399–404, 1956.

[31] O. of Management and B. U. States, *North American industry classification system*. Bernan Press, 1998.

[32] Economic Census, "Naics codes & understanding industry classification systems." https://www.census.gov/programs-surveys/economic-census.html. Accessed at December 2020.

[33] A. D. Broido and A. Clauset, "Scale-free networks are rare," *Nature communications*, vol. 10, no. 1, pp. 1–10, 2019.

[34] J. Edmonds and R. M. Karp, "Theoretical improvements in algorithmic efficiency for network flow problems," *Journal of the ACM (JACM)*, vol. 19, no. 2, pp. 248–264, 1972.

[35] ING, "Nowcast: Corona en het effect op de economie." https://www.ing.nl/zakelijk/kennis-over-de-economie/onze-economie/de-nederlandse-economie/publicaties/nowcast-impact-van-coronavirus-op-de-economie.html. Accessed at January 2021.

[36] S. Kelly, "Estimating economic loss from cascading infrastructure failure: a perspective on modelling interdependency," *Infrastructure Complexity*, vol. 2, no. 1, pp. 1–13, 2015.

6 Appendix

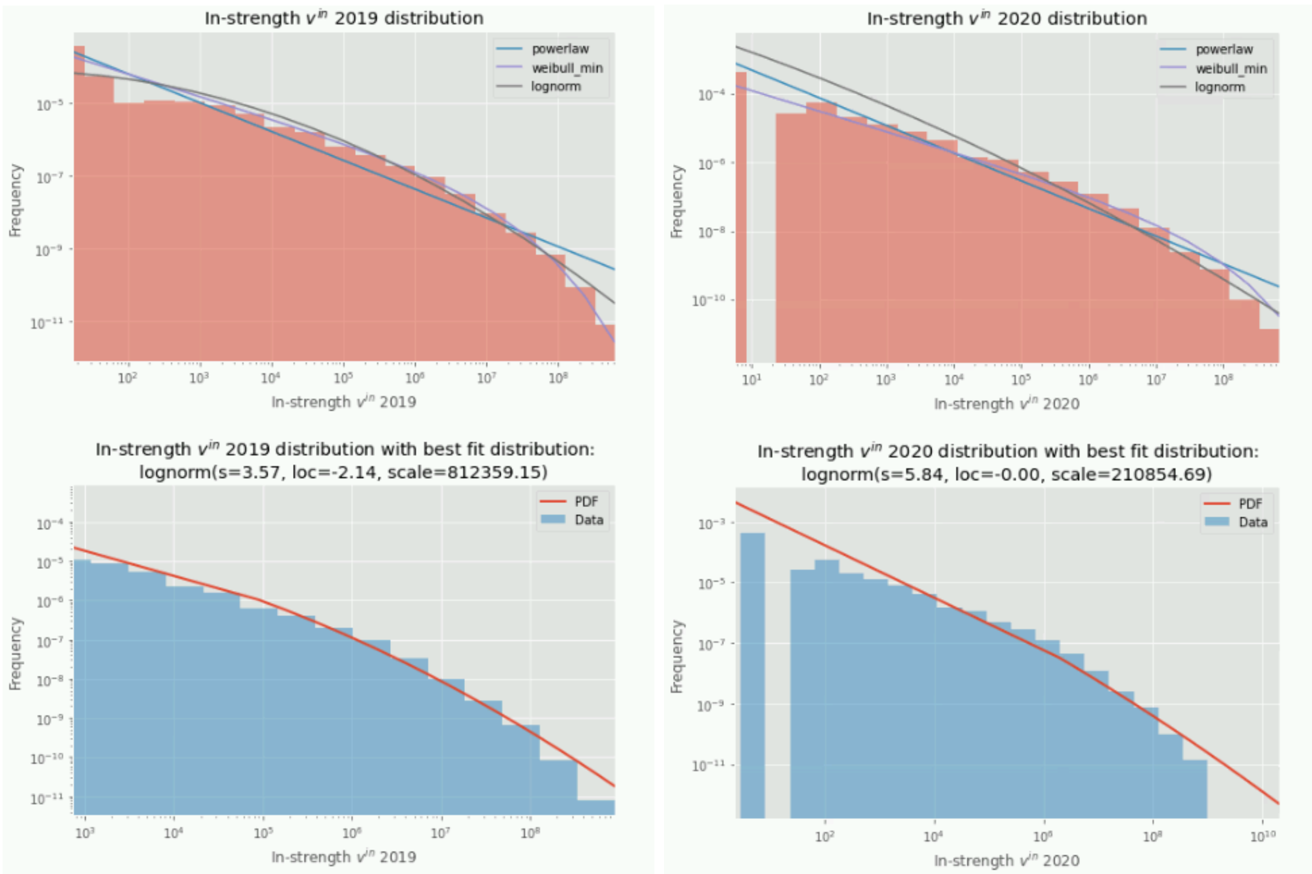

Figure 26: Different probability density functions, fitted on the in-strength v_{in} distributions of 2019 and 2020. The bottom plots show the best fitting distributions, including the distribution parameters.

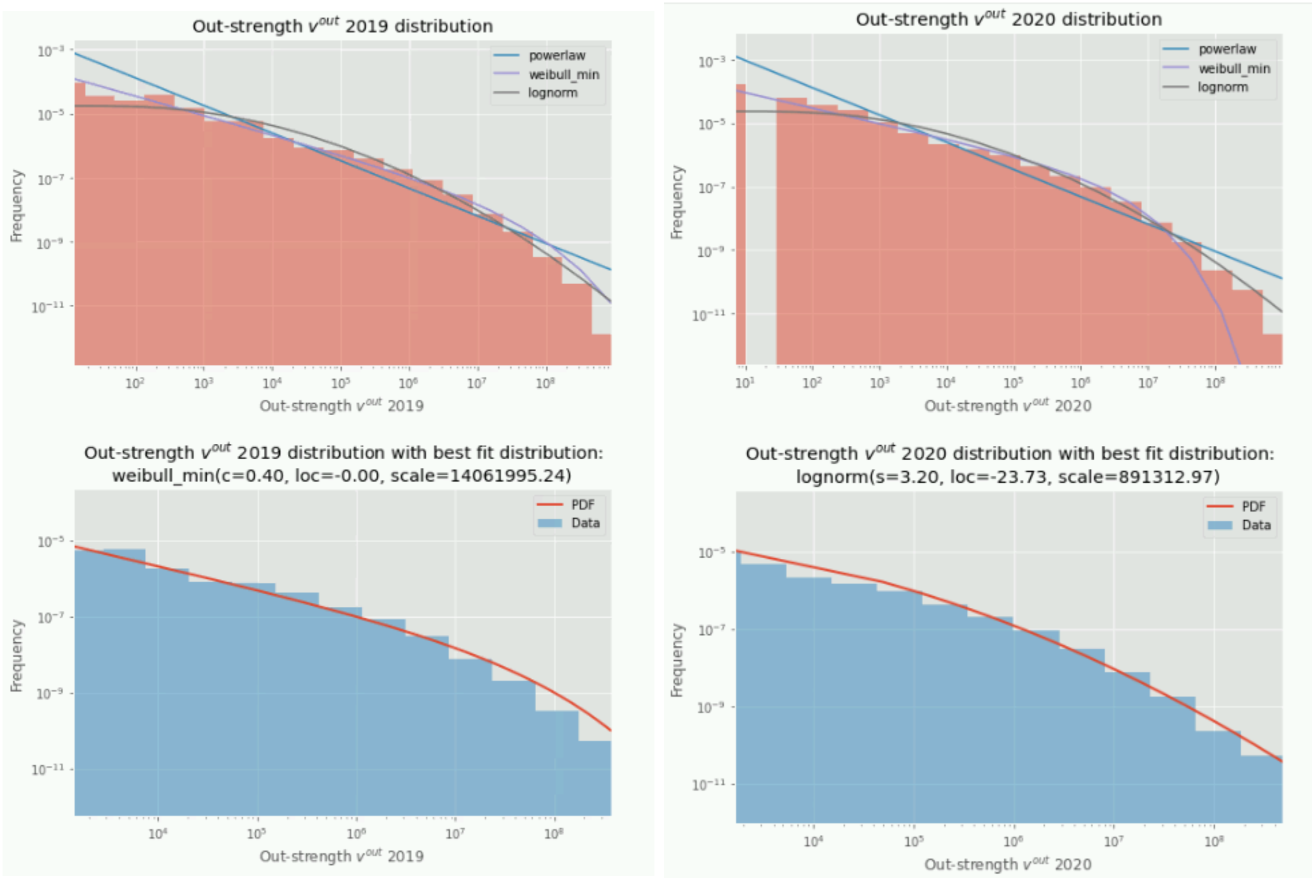

Figure 27: Different probability density functions, fitted on the out-strength v^{out} distributions of 2019 and 2020. The bottom plots show the best fitting distributions, including the distribution parameters.

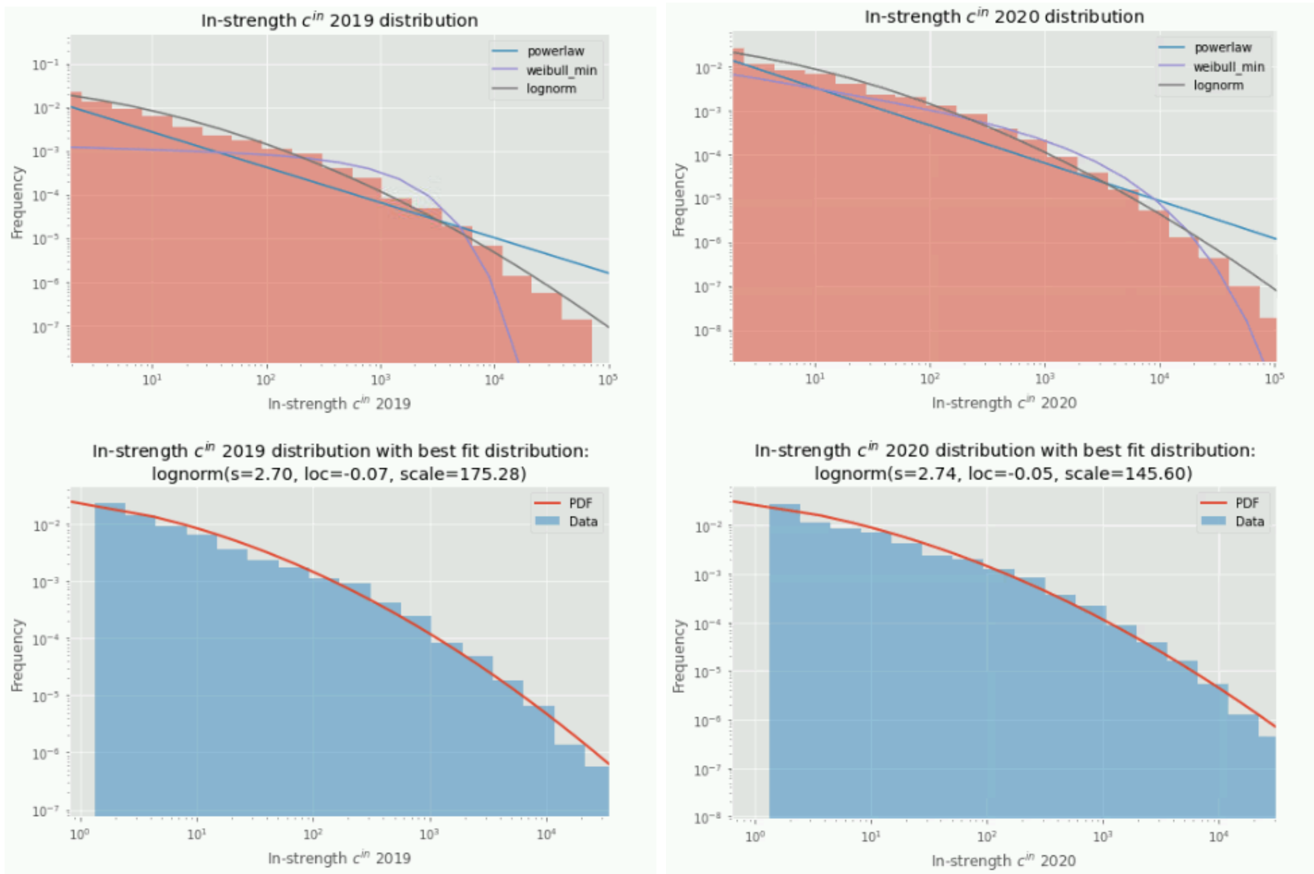

Figure 28: Different probability density functions, fitted on the in-strength c^{in} distributions of 2019 and 2020. The bottom plots show the best fitting distributions, including the distribution parameters.

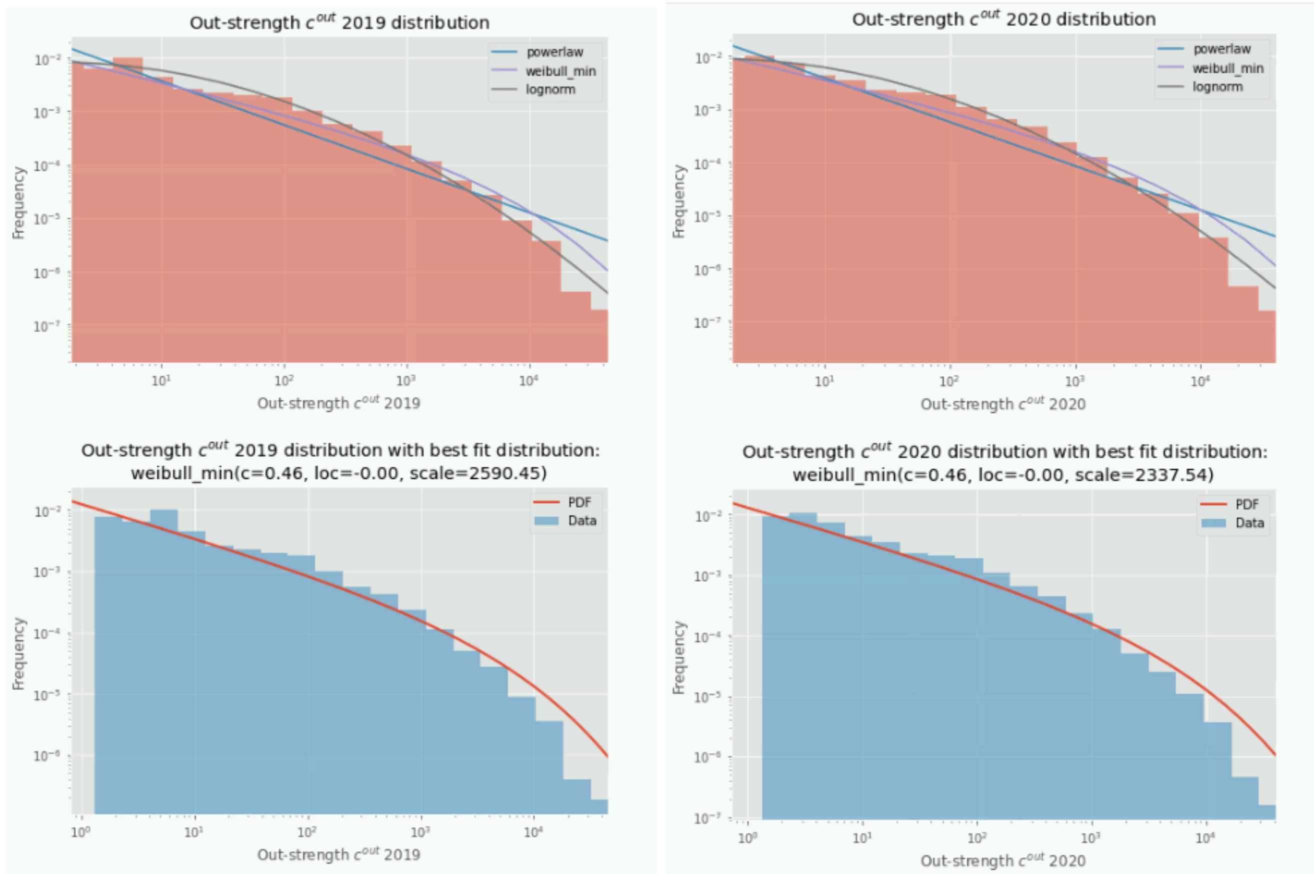

Figure 29: Different probability density functions, fitted on the out-strength c^{out} distributions of 2019 and 2020. The bottom plots show the best fitting distributions, including the distribution parameters.

Industry	In-strength c^{in} 2019	Industry	In-strength c^{in} 2020
Wireless Telecommunications Carriers (except Satellite)	121644.00	Wireless Telecommunications Carriers (except Satellite)	126272.33
Offices of Other Holding Companies	52738.33	Insurance Agencies and Brokerages	47256.33
Insurance Agencies and Brokerages	47059.67	Offices of Other Holding Companies	45508.00
General Freight Trucking, Local	37527.33	General Freight Trucking, Local	34599.00
Miscellaneous Financial Investment Activities	33891.67	Electric Power Distribution	32908.00
Water Supply and Irrigation Systems	33264.33	Water Supply and Irrigation Systems	31611.33
Wired Telecommunications Carriers	31852.00	Miscellaneous Financial Investment Activities	31256.00
Electric Power Distribution	31201.67	Wired Telecommunications Carriers	28215.67
Custom Computer Programming Services	30967.67	Pharmacies and Drug Stores	23156.33
Other Grocery and Related Products Merchant Wholesalers	27711.33	Direct Property and Casualty Insurance Carriers	22715.00

Table 8: The ten biggest industries based on in-strength c^{in}

Industry	Out-strength c^{out} 2019	Industry	Out-strength c^{out} 2020
Direct Health and Medical Insurance Carriers	55422.33	Direct Health and Medical Insurance Carriers	50189.00
Insurance Agencies and Brokerages	41713.67	Insurance Agencies and Brokerages	41339.33
Third Party Administration of Insurance and Pension Funds	35730.00	Miscellaneous Financial Investment Activities	34316.33
Miscellaneous Financial Investment Activities	34968.67	Third Party Administration of Insurance and Pension Funds	30961.67
Full-Service Restaurants	32848.33	General Freight Trucking, Local	24967.00
General Freight Trucking, Local	26553.00	Full-Service Restaurants	24267.67
Custom Computer Programming Services	26189.33	Offices of Other Holding Companies	23073.67
Offices of Other Holding Companies	23766.33	Security Guards and Patrol Services	19007.67
Security Guards and Patrol Services	23494.33	Commercial and Institutional Building Construction	18638.67
Collection Agencies	20776.33	Offices of Real Estate Agents and Brokers	16764.67

Table 9: The ten biggest industries based on out-strength c^{out}

Industry	In-strength v^{in} 2019	Industry	In-strength v^{in} 2020
Supermarkets and Other Grocery (except Convenience) Stores	971.3	Supermarkets and Other Grocery (except Convenience) Stores	1667.8
Third Party Administration of Insurance and Pension Funds	885.2	Miscellaneous Financial Investment Activities	1033.6
Direct Life Insurance Carriers	835.0	Third Party Administration of Insurance and Pension Funds	979.1
Miscellaneous Financial Investment Activities	819.6	Drugs and Druggists' Sundries Merchant Wholesalers	861.0
Offices of Other Holding Companies	561.7	Direct Life Insurance Carriers	678.0
Drugs and Druggists' Sundries Merchant Wholesalers	489.9	Offices of Other Holding Companies	465.1
Employment Placement Agencies	483.4	All Other Telecommunications	409.0
Security Guards and Patrol Services	480.5	General Freight Trucking, Local	398.0
Offices of Notaries	401.0	Offices of Notaries	381.3
Commercial and Institutional Building Construction	374.5	Wired Telecommunications Carriers	377.2

Table 10: The ten biggest industries based on in-strength v^{in} (millions)

Industry	Out-strength v^{out} 2019	Industry	Out-strength v^{out} 2020
Direct Health and Medical Insurance Carriers	1341.3	Supermarkets and Other Grocery (except Convenience) Stores	1634.9
Supermarkets and Other Grocery (except Convenience) Stores	936.1	Direct Health and Medical Insurance Carriers	1505.1
Direct Life Insurance Carriers	831.8	Miscellaneous Financial Investment Activities	1085.2
Miscellaneous Financial Investment Activities	560.0	Drugs and Druggists' Sundries Merchant Wholesalers	855.0
Offices of Other Holding Companies	585.8	Direct Life Insurance Carriers	807.4
Insurance Agencies and Brokerages	570.4	Insurance Agencies and Brokerages	619.1
Security Guards and Patrol Services	475.8	Toy and Hobby Goods and Supplies Merchant Wholesalers	397.3
Drugs and Druggists' Sundries Merchant Wholesalers	470.1	Offices of Other Holding Companies	388.3
Third Party Administration of Insurance and Pension Funds	352.5	All Other Telecommunications	371.7
Commercial Banking	340.5	Third Party Administration of Insurance and Pension Funds	370.0

Table 11: The ten biggest industries based on out-strength v^{out} (millions)

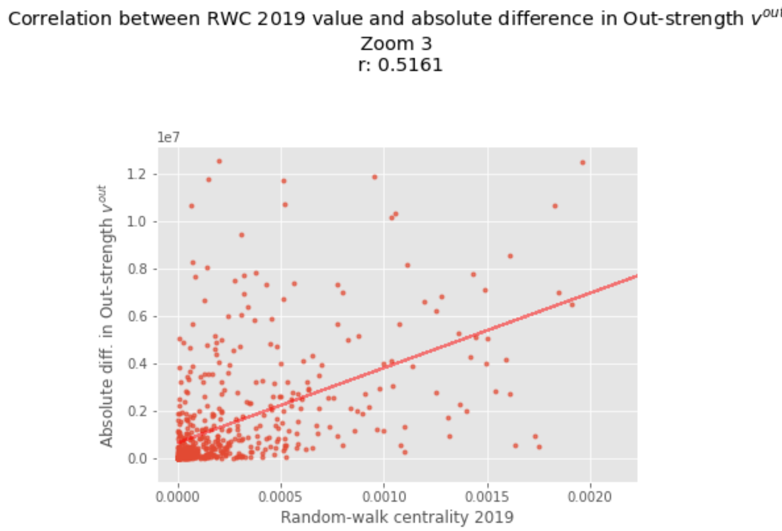

Figure 30: Zoom3

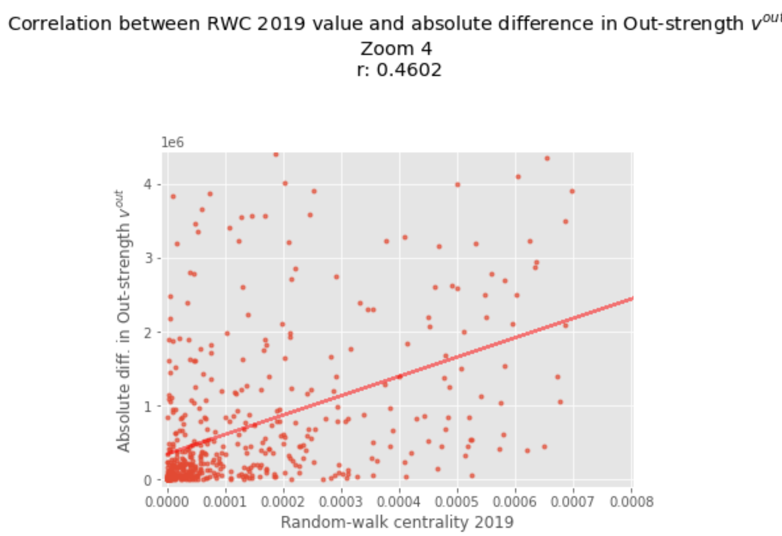

Figure 31: Zoom4

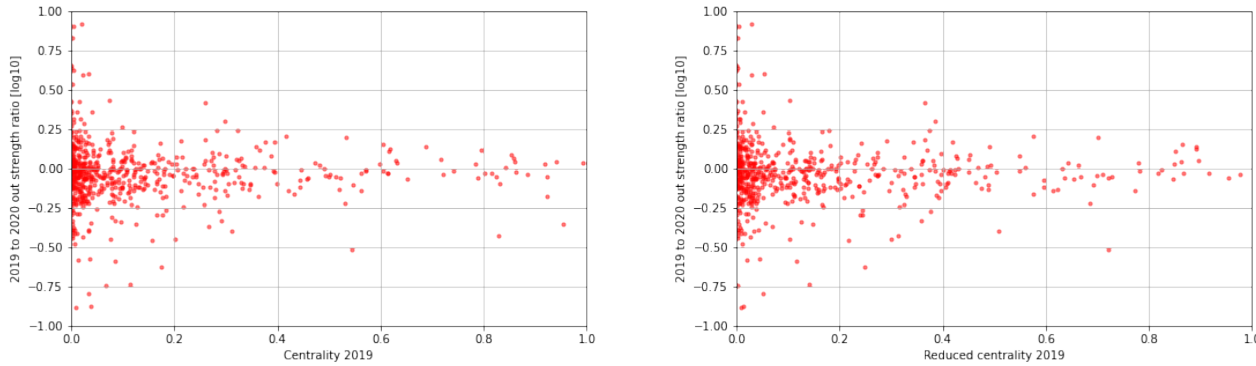

Figure 32: The left figure shows the relative change in out-strength v^{out} plotted against the random-walk centrality. The right plot shows the same, but with a "reduced" random-walk centrality. Computing this metric, works the same as for the standard random-walk centrality. We only change the transition probabilities in the matrix M. Specifically, we alter the matrix M in such a way, that self-loops and node-pairs receive less weight. The matrix M was obtained simply by dividing edge weights in the matrix by its row sums. To get the reduced probabilities, we multiply M by $1 - M^{-1}$, and then again divide the transition probabilities by its row sums. In this way, if much value immediately flows back through self-loops or neighbours, the transition probability will be lower.

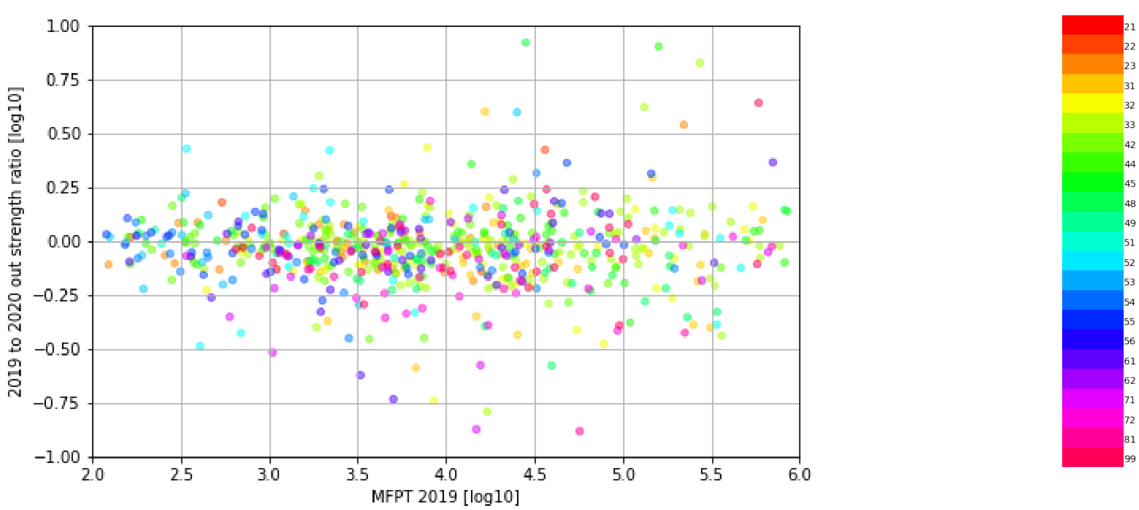

Figure 33: Relation between out-strength ratio v^{out} and average MFPT, where industries are colored according the sector they belong to.