# Using Social Network Analysis for Fraud Detection

## Tracing the Path from Data to Value

**M.E. Clerx**

TU Delft

Nederlandse Voedsel- en Warenautoriteit
*Ministerie van Landbouw, Natuur en Voedselkwaliteit*

*This page was intentionally left blank*

# Using Social Network Analysis for Fraud Detection

**Tracing the Path from Data to Value**

Master thesis submitted to Delft University of Technology

in partial fulfilment of the requirements for the degree of

**MASTER OF SCIENCE**

in **Management of Technology**

Faculty of Technology, Policy and Management

In collaboration with the NVWA

By

Miriam Clerx

Student number: 4749596

August 2019

## Graduation Committee

| | |
|---|---|
| Chairperson & Second Supervisor | : Dr. S.W. Cunningham, section Policy Analysis |
| First Supervisor: | : Dr. H.G. van der Voort, section Organisation & Governance |
| First External Supervisor | : A. Suppers, NVWA |
| Second External Supervisor: | : T. Booijink, NVWA |

*This page was intentionally left blank*

# Executive Summary

Recent incidents in the food and beverage industry show that consumer goods remain vulnerable to tampering, imposing a risk on public well-being, animal welfare, and nature. Facing the age of digitalisation, inspection agencies are improving their way of working using a risk-oriented and data-driven approach to improve the enforcement efficiency and effectivity. Digitalisation leads to an era in which new data sources evolve, and new techniques for storing and analysing large data sets are enabling many new applications. One of these applications is Social Network Analysis (SNA). Social Network Analysis has proven to be a successful tool in predicting and understanding fraudulent behaviour. SNA emphasises the structural aspects of networks to detect and interpret patterns of social entities. A social network is commonly modelled by a graph consisting of nodes which are connected by patterns of contacts or interactions, called links. These networks can either be analysed graphically or mathematically. In the context of law enforcement, SNA has potentials for risk analysis and threat assessments, destabilising criminal networks, role identification, scenario building, support on the deployment of intelligence assets, and provide evidence for prosecution.

Although the benefits of using SNA in fraud detection seem to be clear cut, a more institutional view emphasises the different actors involved who shape the process from data to value creation. The assumption that just working data-driven leads to better deployment of enforcement assets may be too simplistic. In practice, the way in which business value is created from big data often remains unclear. The ability to create value from big data depends on having the right process in place to give meaning to the data, also known as the big data value chain. This requires collecting the right data, having access to data, obtaining trustworthy data, having the right skills in place for data analysis, and concrete actions to realise the potential of big data. There seems to be a gap between the promises of big data and its practical realisations, in particular in the public domain.

This research considered big data in the context of Social Network Analysis (SNA) from an institutional perspective. This means it is assumed that actors that shape the process from data as a raw material to the final deployment of inspection capacity based on the outcome of the network analysis. This research has underlined a decision-making perspective which states that the way the alternatives are framed impacts the alternative chosen by people and in turn the subsequent decision. What the institutional view exactly means in data-driven inspection required a deeper understanding in the activities involved in the big data process chain and the operational complexities and decisions that need to be made in turn. This led to the following main research question:

**How does the big data activity chain influence the potential value created by using Social Network Analysis for fraud detection?**

It was chosen to address this question by combining a literature study with two case studies executed at the Data Science Cluster of the Netherlands Food and Consumer Product Safety Authority (NVWA).

The first part of this research consisted of a literature study consisting of three main parts. To begin, existing literature has been reviewed to clarify the differences between expert-based and data-driven fraud detection. The traditional approach for fraud detection relies on human expert input, evaluation, and monitoring. Using an automated, data-driven system could lead to a more efficient and effective methodology for detecting fraud. However, machine learning models are no panacea in fraud detection. It might be very difficult (if not impossible) to explain how certain scores or decisions are obtained. Second, literature has been assessed to gain understanding in how SNA can be used to predict fraud; how is SNA usually applied and in what context, and how might it help to detect fraud. SNA helps to structure data important for making business decisions. Current literature

on SNA shows its application in a wide variety of areas, among which fraud detection. Fields that have used SNA for fraud detection include health care fraud, insurance fraud, mobile internet fraud, money laundering, mortgage fraud, online auction fraud, opinion fraud, security fraud tax fraud, social security fraud, insurance fraud, and telecom fraud. Yet, any scientific application in the consumer and food product domain is lacking. Finally, the last part of the literature review addressed the challenges that come along with applying SNA in the context of fraud. These challenges have different origins. Researchers on networks tend, with a few exceptions, to fall into two distinct categories, each operating under severe constraints. One the one hand, crime researchers have expertise in criminological theory and research but seem to lack analytical expertise and access to good data. On the other hand, data analysts may have strong analytical skills and access to various data sources but tend to lack knowledge on criminological issues such as fraud. In addition, the development of fraud detection methods is constrained by the scarcity of available data sets and the limited disclosure results in public. This imposes a severe limitation on the exchange of ideas in fraud detection. Not to mention that the majority of published methods are usually black-boxes where their workings are mysterious. Besides, both fraud and fraud detection methods are embedded in a specific context; a solution to fight credit card fraud cannot be applied in insurance companies.

The second part of this research empirically explored the use of SNA pertaining to two main purposes. First of all, the study has a functional purpose which aimed to explore the use of SNA in the fraud detection, more specifically fraud in the context of food and consumer products. Secondly, the research has an institutional purpose aimed to get insight into big data value creation in the public sector. To reach these goals, the researcher was actively involved as a participant in the implementation of SNA in two case studies. This research can therefore also be considered as action research, in which the researcher acted as a reflective practitioner. By applying reflective practice, taken-for-granted assumptions are questioned which supports the development of insights.

As followed from the analysis of two large real-world datasets, SNA has the potential to identify patterns in the network and in turn improve the inspection efficiency and effectivity. First of all, network visualisation offers a powerful solution to make information hidden in networks easy to interpret and understand. With one glance at the network one could identify who does business with whom, which entities act as bridges between two clusters, trace suspicious patterns, and gain insight into the overall structure of the networks (i.e. to what extent are entities connected). Secondly, applying network metrics may quickly result in findings and new insights into the large and complex networks that are not quickly graphically interpret. Four centrality metrics have been applied: Reach Centrality, Closeness Centrality, Stress Centrality, and Betweenness Centrality. These centrality metrics can help to identify the important players in the networks and can be used to evaluate or predict the possible consequences of removing specific actors from the networks to destabilise the networks. However, it requires close consideration to determine what can be regarded as important and whether the selected metrics have the potential to reach that goal, which turned out to be strongly case dependent. The former depends, among others, on the capacity of enforcement assets, and the latter depends on the structure of the network.

Although the emergence of big data opens great opportunities in the public domain, the analysis of big data is confronted with many challenges. From the case studies it followed that data-driven fraud detection within the context of the food and consumer product domain is still a long way to go. During the research, it became evident that various important assumptions and decisions have been which appeared to be fundamental for the outcome of the analysis. An argument that has not been cited previously in the context of fraud detection, is that even when a data-driven fraud detection system is capable of self-learning, it is still inevitable that people shape the process from data to value. Based on the CRISP-DM process, the various key activities related to SNA have been identified. The scope of the activity chain was confined to the process of data understanding to modelling, given that these

were the activities in which the researcher participated. Each of the activities is involved with operational complexities. The complexities are the result of prerequisites, dilemmas, trade-offs, or path dependencies leading to institutional or technological lock-ins. The CRISP-DM steps, activities, operational complexities, and key findings are highlighted below.

**Data Understanding**

- *Data source gathering*: Gathering the data that capture the problem;
- *Data source selection:* Evaluation data sources and selecting data sources;
- *Data acquiring:* Gaining access to data;
- *Data understanding:* Interpreting data and verifying data quality.

From the research it followed that the data understanding phase involves more than solely comprehending the meaning of the data. It requires an organisation to gather, select, acquire, interpret and verifying the data quality. Data understanding is characterised by an institutional lock-in. There is a tendency of organisations to become committed to develop in certain ways as a result of their organisational structure and practices or their beliefs and values. Especially in the context of fraud research, the value acquired from the utilisation of multiple datasets is often far higher than individual data sets, since many illicit activities occur below the radar. The act to think differently, unconventionally, or from a new perspective is likely to be constrained by ordinary cognitive and institutional processes.

**Data Preparation**

- *Data selection:* Specifying nodes (entity selection), specifying links (relationship selection), specifying geographical boundaries, specifying domain boundaries, specifying timeframe boundaries, and selecting risk factors;
- *Data preparation:* Preparing nodes, preparing inspection data, preparing risk factor data, and preparing GPS data;

Data preparation is a prerequisite for enriching the data and improving the accuracy of the outcome. More specifically, it determines to what extent the model (correctly) reflects reality. In general, data cleaning takes on many forms and is considered to be time-consuming. This makes data preparation an expensive, yet inevitably important, phase. The data preparation is characterised with a lot of discretionary freedom in which various trade-offs are to be made. As a data scientist, there is much room to act and decide in selecting the data tables and ranges, given the absence of effective control and ambiguous rules.

**Modelling**

- *Visualisation*: Specifying network size, specifying nodes (size colour, pictogram, label and underlying node details), specifying links (width, direction, colour, removal of self-referencing, and underlying link details);
- *Applying centrality metrics:* Specifying network size, selecting a level of analysis, selecting metrics, defining importance, specifying network rules, applying multiple metrics simultaneously;
- *Addition of risk factors:* Specifying inspection boundaries (visualisation), specifying risk factor boundaries (visualisation), selection of network size;
- *Plotting on a geographical map*: Specifying network size, selecting a coordinate system, selecting radius based selection method and radius;
- *Tracking changes over time:* Specifying time frame and specifying visualisation background.

In the modelling phase, the available alternatives for SNA were constrained either by the capabilities of the software or the way in which the data was prepared. This made data preparation and an iterative process. The restrictions turned out to be strongly case-dependent; centrality metrics that rely on shortest path analysis are less useful in a highly fragmented network. Software selection is a key example of path dependency leading to a technological lock-in situation. Due to the selection of the software, the design and user practices have become fixated to such an extent that applying alternatives would be unfeasible (or requires substantive investment in resources and capabilities) – even if the alternatives were more desirable. Besides, when the analysis relies on the graphical representation of a network, there seems to be a trade-off between the accuracy of the network and its interpretability.

The consequence of the complexities is that they prevent the creation of options or lead to options that are sub-optimal. Neglecting them would be at the detriment of any future SNA-ambition an organisation may hold. The decisions made could be fundamental for knowledge-sharing, commitment, representativeness, accuracy, validity, operational efficiency and the effectiveness and efficiency of the deployment of enforcement assets. The activities and complexities defined in this thesis are a valuable contribution to navigate organisations towards SNA as a method to detect fraud. Additional empirical studies are required to determine the relative impacts of how decisions on operational complexities shape the process from data to value creation. By developing a framework this research contributed both to the academical debate on big data value creation and the practical application of SNA for fraud detection. As the research into big data value creation is still in a very young state, the findings from this thesis form a starting block for future research to expand on.

# Preface

This thesis was written in fulfilment of the MSc Management of Technology at Delft University of Technology. It contains the results of a study that aims to identify how big data influences public decision-making in the context of fraud detection.

The research was conducted in collaboration with the NVWA. Herewith I would like to thank the NVWA's Intelligence & Research department for offering me the ability to conduct this research and for enabling me to use the NVWA's resources. In particular I would like to thank my supervisors, Anouk Suppers and Tom Booijink, who always managed to free up time from their busy schedules to provide me with day-to-day guidance. I would also like to thank my faculty supervisors, Haiko van der Voort and Scott Cunningham, for their expertise, feedback, and for enthusiastically supporting my thesis project. Finally, I would like to thank my parents and social environment for their open-ended support, motivational words, and understanding during my graduation period.

After all these words of gratefulness, the only thing left is to wish you a pleasant time of reading this thesis.

Delft, August 23rd 2019

M.E. Clerx

MSc Student Management of Technology
Delft University of Technology
Faculty of Technology, Policy, and Management

*This page was intentionally left blank*

# Table of Content

## List of Figures



**List of Tables**

**List of Abbreviations and Acronyms**

| | **Nederlands** | **English** |
|---|---|---|
| **BRS** | Bedrijf Registratie Systeem | Business Register System |
| **BuRO** | Bureau Risicoanalyse en Onderzoek | Office for Risk Assessment & Research |
| **CRISP-DM** | Cross-Industry Standard Process for Data Mining | Cross-Industry Standard Process for Data Mining |
| **DA** | Dierenambulance | Animal Rescue |
| **DIPO** | Dierenpolitie | Animal Control Service |
| **DSC** | Data Science Cluster | Data Science Cluster |
| **GDPR** | Algemene Verordening Gegevensbescherming | General Data Protection Regulation |
| **RF** | Risicofactor | Risk Factor |
| **FTE** | Voltijdequivalent | Full-time Equivalent |
| **GBD** | Gemeenschappelijk Veterinair Document van Binnenkomst | Common Veterinary Entry Document for Animals |
| **IOD** | Inlichtingen en Opsporingsdienst | Intelligence and Investigation Division |
| **I&R** | Identificatie en Registratiesysteem voor dieren | Livestock Identification and Registration |
| **K&O** | Kennis en Onderzoek | Intelligence & Research department |
| **KvK** | Kamer van Koophandel | Chamber of Commerce |
| **LED** | Gegevensbescherming bij Rechtshandhaving | Law Enforcement Directive |
| **LID** | Landelijke Inspectiedienst Dierenbescherming | The Dutch Society for the Protection of Animals |
| **LNV** | Landbouw, Natuur en Voedselveiligheid | Agriculture, Nature and Food Quality |
| **MOS** | Melding Ondersteuning Systeem | Notification Support System * |
| **NVWA** | Nederlandse Voedsel- en Waren Autoriteit | Netherlands Food and Consumer Product Safety Authority |
| **PLB** | Planbureau voor de Leefomgeving | PLB Netherlands Environmental Assessment Agency |
| **RVO** | Rijkdienst voor Ondernemend Nederland | The Netherlands Enterprise Agency |
| **SAS VI** | SAS Visual Investigator | SAS Visual Investigator |
| **SNA** | Sociale Netwerk Analyse | Social Network Analysis |
| **TRACES** | Trade Control and Expert System | Trade Control and Expert System |
| **VDM** | Vervoersbewijzen Dierlijke Mest | Animal Manure Consignment Note |
| **UBN** | Uniek Bedrijfsnummer | Unique Business Number |
| **UT** | Universiteit Twente | University of Twente |

*Translation

*This page was intentionally left blank*

# 1. Introduction

January 2013. The Food Safety Authority of Ireland revealed in an investigation that 10 out of 27 beef-labelled hamburger products contained horse DNA. A few weeks later, 11 out of 18 tested beef lasagne were also found to be contaminated with equine meat. Europe got shocked; it revealed a large-scale breakdown in the traceability of the food supply chain, and the risk that more harmful ingredients could have been included as well. Although the "Horse Meat Scandal" was not a direct health issue, the European food sector faced an enormous consumer trust crises. Several more scandals broke the news in the years that followed. A more recent well-known example is the fraud with fipronil-contaminated eggs. Hundreds of Dutch poultry farms were temporarily on lockdown after the toxic pesticide fipronil was found in their eggs. Thousands of hens were culled and countless eggs were destroyed. As these examples indicate, fraud with consumer products has an enormous impact on society. Research by the Consumentenbond (2016) showed that 21% of the investigated products contained undeclared deviations. These scandals suggest the need for compliance with regulations and detection of malpractitioners. In this chapter, the Dutch Food and Consumer Product Authority (NVWA) is introduced as an authority to safeguard public and environmental health, and the challenges and opportunities that come along with a rapidly evolving society. This results in the research problem statement, the research questions and the research objective. The introduction chapter ends with an outline of the thesis structure.

## 1.1 Societal Relevance and Ambition

As the first section indicates, safe food products and other consumer goods are crucial for society. This does not only mean that products must be safe for humans, but also for animals, and plants and flora. In addition, honesty and transparency with respect to consumers are considered to be of increased importance nowadays. These aspects can be summarised in four main public values that are being protected by the Netherlands Food and Consumer Product Safety Authority (NVWA): food safety, product safety, animal welfare, and plant- and animal health and environment. The NVWA is responsible for establishing, monitoring, and enforcing laws and regulations to that end.

Virtually every industry has been experiencing rapid, massive, and sometimes disrupting change over the last couple of years. As industries change, so need the agencies that control them. The same is true for the NVWA. The fast-changing society requires innovative ways to adapt and respond to these changes. The legislation is adapted continuously, requiring new ways of supervision in return. Besides, due to the trend of globalisation, the NVWA faces the challenge of increasingly complex supply chains crossing the national borders. Innovation enables a wider range of goods and services to be delivered to a worldwide market (Schilling, 2013; Van der Voort, Klievink, Arnaboldi, & Meijer, 2019). Tracing and tracking inputs and outputs throughout a complex supply chain and across multiple borders has resulted in much greater risk exposure (Schaadt, 2013). Next to the usual quality- and safety aspects, also fraud and other illegal activities must be detected and acted upon. A total of 8,376 administrative fines were imposed on offenders in The Netherlands in 2017 (NVWA, 2018). This is not to say that all offenders are aware of their misconduct; organisations might lack knowledge of the applicable legislation. Organisations committing deliberate fraud were charged for a total value of almost 75 million euros in 2017 (NVWA, 2018). However, this number only represents the financial burden on frauds that were actually detected. There is still considerable ambiguity about the exact size and impact of fraud. Studies provide estimations varying around several billion euros (Algemene Rekenkamer, 2014). Establishing the exact cost of fraud may be hard, if not impossible, because fraudsters actively conceal their activities. Even though the Algemene Rekenkamer (2014) provides rough estimates rather than exact measurements, the estimates do indicate the importance and potential impact of fraud, and therefore as well the need for governments to actively fight and prevent fraud with all means they have at their disposal.

## 1.2 Inspections in a Data-Driven World

One of the most striking changes of this decade is the rapid evolution of data science in various fields. The public sector has now become more and more aware of the fact that the data availability and new methods of its use may be utilised for the public benefit (Maciejewski, 2016). Also within the NVWA, there is a growing realisation that data is not simply a by-product of the organisations' primary processes, but that it is a valuable resource by itself. This means that the way in which the NVWA functions, the way in which is dealt with consumers and business, and the way she exercises her responsibility towards the federal government, politics, and society is changing. From that perspective, the NVWA continuously wants to develop itself, with the ultimate objective of being trustworthy, transparent, independent, and professional (NVWA, 2016). Massive volumes of data can be collected, modelled, and analysed e.g. to uncover the patterns of human behaviour and to help with predicting social trends. This changes the way we think about business, politics, education, and health, and due to the increasing amount of data that are collected, data science innovations will undoubtedly continue in the years to come (Tayebi, 2015). With regard to that end, the NVWA has formulated a 2020 mission. Following the NVWA's 2020 mission, the NVWA's driving pillars in safeguarding the public values are to work data-driven and risk-focused: acting in the most effective way to minimise the risks for public health. In addition, the NVWA wants to stimulate proactive rather than reactive supervision. In literature this can be linked to predictive policing: "any policing strategy or tactic that develops and uses information and advanced analysis to inform forward-thinking crime prevention" (Uchida, 2010, p. 1). This includes enhancing the preventative capacity in identifying offenders where none could previously be found under the standard inspection techniques (Teyebi & Glässer, 2016).

Due to the limited capacity of enforcement resources inspections have to be executed in an effective and efficient manner (Shimshack & Batten, 2014). Early identification and detection of offenders through predictive policing ensures that enforcement agencies like the NVWA can actively operate to fight fraud and "remove the threat on society" (Teyebi & Glässer, 2016). To summarise, may have two potential benefits, working data-driven and risk-focused has two potential benefits (NVWA, 2016):

(i) Better knowledge and information position: the better the analysis, the better the selection of the right firms with the highest risks, improving the NVWA's performance;

(ii) Flexibility: being more flexible and responsive to risks and societal developments.

The reasoning behind the use of prediction techniques is that most offenses are not random but happen in patterned ways (Teyebi & Glässer, 2016). Social Network Analysis (hereinafter SNA) has proven to be a successful tool in predictive policing and understanding criminal behaviours and extract criminal patterns. SNA emphasises the structural aspects of networks to detect and interpret patterns of social entities (Wasserman & Faust, 1994), in which the social network is commonly modelled by a graph which consists of groups, called nodes, connected by patterns of contacts or interactions, called edges or links (Sapountzi & Psannis, 2018). The unique feature of socially networked data is that it reveals new opportunities to understand individuals and their connections in a society (Lettieri, Altamura, Malandrino, & Punzo, 2017). Comparing networks, tracking changes in a network over time, indicating communities and important nodes, and determining the relative position of individuals and clusters within a network are some of its common procedures (Sapountzi & Psannis, 2018). SNA as a prediction method may foster proactive policing rather than reactive policing and is promising in improving intervention strategies by making more efficient use of limited resources (Teyebi & Glässer, 2016). These methods give law enforcement agencies like the NVWA a set of tools to do more with less and focus on the highest risks.

## 1.3 Research Problem

The current age of digitalisation leads to an era in which new data sources evolve, and new techniques for storing and analysing large data sets are enabling many new applications. The previous sections emphasise the opportunities that come along with the societal changes and the need of the NVWA to improve inspection effectivity and efficiency by working risk-oriented and data-driven. However, the use of a big data to "predict the future" (Goel, Hofman, Lahaie, Pennock, & Watts, 2010; Uchida, 2010) and "remove the threat on society" (Teyebi & Glässer, 2016) supposes that more data leads to a better information position, and more information leads to better supervision. Similarly, the NVWA's assumptions that data-driven and risk-oriented inspections lead to better to better governance might be too simplistic. In practice, the way in which business value is created from big data often remains unclear. The ability to create value from big data depends on having the right process in place to give meaning to the data, also known as the big data value chain. This requires collecting the right data, having access to data, obtaining trustworthy data, having the right skills in place for data analysis, and concrete actions to realise the potential of big data. There seems to be a gap between the promises of big data and its practical realisations. Section 2.7 further elaborates on the knowledge gaps in the existing literature.

This research views big data in the context of Social Network Analysis (SNA) from an institutional perspective. This means it is assumed that there are several actors that shape the process from data selection to the final deployment of inspection capacity based on the outcome of the network analysis. What the institutional view exactly means in data-driven inspection requires a deeper understanding of both the big data process chain and the actors that shape the process, as well as the opportunities and limitations of using SNA, in particular with respect to fraud detection.

## 1.4 Research Objective

This research has two main purposes. First of all, this thesis explores the use of SNA in the context of fraud detection, more specifically fraud in the context of food and consumer products. SNA has the potential to identify patterns in the fraud network and in turn improve inspection efficiency and effectivity. Secondly, this research applies an institutional view to get insight into big data value creation in the public sector. This means that through the evaluation of the big data chain process this research aims to get a better understanding in the factors influencing big data decision-making. According to Yin (2012), in the case of limited existing knowledge on a certain topic, a new empirical study will most likely be an exploratory study. Therefore, the study focuses on exploring the concepts rather than on developing general theoretical statements. To reach this goal, the next section presents the main research questions and the appurtenant sub-questions.

## 1.5 Research Questions

Considering the research problem and the research objective, the main research question is defined as follows:

**How does the big data activity chain influence the potential value created by using Social Network Analysis for fraud detection?**

In order to answer this question, it is necessary to define sub-questions on the different topics of the main research question. These will be introduced, after which the methodology per question will be elaborated on further in Section 1.6 .

*1: What are the differences between an expert-based fraud detection and data-driven fraud detection system?*

*2: What is Social Network Analysis, and how is it applied to detect fraud?*

*3: What challenges for using Social Network Analysis in fraud detection are mentioned in literature?*

*4: How does the process of data understanding to Social Network Analysis work?*

*5: What operational complexities are applicable in the process of data understanding to Social Network Analysis?*

The main objective of answering the research questions is to provide preliminary insights in how SNA can be a relevant concept in detecting fraud in the context of food and consumer products from an institutional perspective. The next section will briefly elaborate on the different research methods that will guide answering the research questions.

## 1.6 Research Methods

The intention was to approach the developed research questions with the use of multiple methods. The study can roughly be divided into two different steps: combining a literature study with two case studies executed within the Netherlands Food and Consumer Product Safety Authority (NVWA). In the upcoming sections, a short description of the different methods used in these steps will be presented. When applied, each method will be elaborated on further in the appurtenant chapter.

### 1.6.1 Literature review

The first step of the study consists of a literature review to get a better of fraud detection, SNA, and its opportunities and challenges.

Big data and analytics provide powerful tools that may improve an organisation's fraud detection system. In the first part a literature study has been performed to clarify the differences between expert-based and data-driven fraud detection. With regard to that end, the first sub-question is as follows:

*Sub-question 1: What are the differences between an expert-based fraud detection and data-driven fraud detection system?*

Second, literature has been reviewed to gain understanding in how SNA can be used to predict fraud; how is SNA usually applied and in what context, and how might it help to detect fraud. In other words, the opportunities of SNA must be clear if we want to be able to say something about to what extent SNA, or more general big data, reaches its potentials. This leads to the following sub-question:

*Sub-question 2: What is Social Network Analysis, and how is it applied to detect fraud?*

To answer this question, the question is split up into three parts. First of all, the concepts of fraud and more specifically fraud detection in the context of big data will be reviewed. The literature review starts from a broad perspective with the objective to get an understanding of the challenges that come along with detecting fraud.

In the second part, the literature study aims to get a better understanding of SNA and its mathematical foundations. A thorough study of the field will provide an overview, given the extensive amount of available literature on SNA yet unexhaustive, of the various metrics used in network analytics as well as the potentials of SNA for law enforcement, which will help answering the first sub-question.

Finally, the last part needed to answer this question is reviewing existing publications on applying SNA to evaluate the state-of-the-art in scientific journals. The review of SNA will be narrowed down to the context of fraud detection, as the base of literature on this field is very large. It is important to mention, though, that analysing SNA applications is explicitly focused on scientific publications,

which may not fully reflect reality given that the data sets are not made available and results are often not disclosed to the public (Yufeng, Chang-Tien, Sirwongwattana, & Yo-Ping, 2004). The exact review steps will be further elaborated in Section 2.5.2.

*Sub-question 3: What challenges for using Social Network Analysis in fraud detection are mentioned in literature?*

There is a vast amount of literature in which about the potential of analysing social networks to detect fraud and criminal activities. However, the identification of the challenges that come along with using SNA for fraud detection is useful for the practical application of SNA in this research.

### 1.6.2 Case Study & Action Research

In the second part, this research aims at empirically exploring how the big data activity chain impacts the value created by SNA. To do so, the researcher was actively involved as a participant in the implementation of SNA in two case studies. This research can therefore also be considered as action research, in which the researcher acted as reflective practitioner.

The fourth sub-question has been set-up to gain insight into the activities that are involved in the process of creating value from big data. To answer this question, the CRISP-DM model is used as a guideline to structure the activities executed in the two cases. The fourth sub-question is formulated as follows:

*Sub-question 4: How does the process of data understanding to Social Network Analysis work?*

Finally, the last sub-questions looks at the operational complexities involved in each activity. The operational complexities requires one to make decisions. Considering data value creating as a decision-making process, each of these decisions will have consequences for the decisions made in a subsequent phase.

*Sub-question 5: What operational complexities are applicable in the process of data understanding to Social Network Analysis?*

## 1.7 Relevance of the Research

To the best of the author's knowledge, this work is the first comprehensive attempt to explore the use of Social Network Analysis to detect fraud and suspicious activities in the context of the food and consumer products. From a scientific point of view, by applying SNA in fraud detection, this research will contribute to network analytics research in specific and more generally to research on big data value creation in the public sector. As will be elaborated on in the literature review in Section 2.5, the concept of SNA has barely been researched in the context of fraud detection, let alone included in the context of food and consumer products. Therefore, this thesis serves as a first exploration of the particular scientific subject. In addition, it offers a qualitative study meant to explore the field of big data and public decision-making and provides structure for further debate.

This makes that there are several major reasons that make this thesis a useful resource for readers with different backgrounds and goals: (i) The literature on SNA has been explored thoroughly to identify and understand applications of SNA to detect fraud; therefore, this work covers the fundamental applications in this context; (ii) The use of SNA is experimentally evaluated using two large real-world datasets producing high-quality results. The author is not aware not aware of any related work assessing performance using similar datasets; (iii) The institutional perspective provides solid understanding of complexities in big data value creation in the public sector; (iv) This multidisciplinary work is completed in close collaboration with law enforcement experts and data scientists and therefore offers a contribution to the knowledge exchange between them.

Ideally, the outcomes will help other researchers to shape the debate on the technical, organisational, ethical, and political shortcomings of big data. This is not only relevant in scientific terms, but also in social terms. Such research is vital for organisations like the NVWA to gain valid findings and conclusions when adopting data-driven systems. In addition, since the focus is on specifically on SNA, it offers practical insight into how SNA itself and how it could improve fraud detection and in turn inspection efficiency and effectivity.

## 1.8 Research Scope

The scope of this research is to identify the potential added value of the SNA for the detection of fraud and thereby improving inspection efficiency and effectivity in the context of consumer and food products. From a broader perspective, this research gives insight into big data value creation within the public sector. As indicated in the previous section, the research focuses on two cases executed by the data science team of the NVWA. The value created for data-driven inspection by the SNA depends both on the decisions made by the analyst as well as the way that the inspectors act upon the outcome of SNA. If the findings are not considered seriously by the inspectors, the promises of big data will not become reality. However, the latter will be considered out of scope. Nonetheless, this does not mean that the expertise and interests of the inspectors will not be taken into account when creating the SNA.

## 1.9 Thesis Outline

The thesis report is structured based on the steps as explained in Section 1.6. In Chapter 2, the (theoretical) background is introduced by means of literature studies. Chapter 3 outlines the case study in more detail. Chapter 4 is an empirical evaluation of data value creation within the NVWA, including the activities carried out for SNA. Chapter 5 and 6 present the results for both SNA cases. The decision analysis is detailed in Chapter 7. The last chapter, Chapter 8, takes on the conclusions, future research recommendations and a reflection.

# 2. Literature Review

The previous chapter introduced the need for the NVWA to work according to a data-driven and risk-oriented approach to safeguard the health and safety of humans, plants, and animals. This chapter examines existing literature on the use of big data as input to detect fraud through the analysis of social networks. The first part focuses on some key concepts and definitions related to big data. The second part introduces two different perspectives on big data. The third part introduces the 'Big Data Value Chain' as means to analyse big data value creation. Section 2.4 elaborates on fraud and fraud detection. After that, Section 2.5 introduces SNA as a method to detect fraud and suspicious activities, and reviews the key concepts and metrics related to SNA and the current directions in its use. Following Baesens, Vlasselaer, and Verbeke (2015) the social character of fraud is underlined. This means that it assumed that the probability of someone committing fraud depends on the people (s)he is connected to. Finally, the last part summarises the knowledge gaps in the literature that underline the theoretical significance of this research.

## 2.1 Big Data

An important driver for improvements with respect to fraud detection is the growing data availability (Baesens et al., 2015). The age of digitalisation leads to an abundance of available data and so-called big data. Big data refers to the large and complex data sets, which are hard to handle using traditional tools and techniques (Elgendy & Elragal, 2014). This is the result of specific big data features, commonly characterised with three at least three V's: Volume, Velocity, and Variety (McAfee & Brynjolfsson, 2012). As the big data field matured, other Vs have been added such as Variability and Veracity (data quality and uncertainty).

Big data is closely related to big data analytics, which refers to the ability to create value from the available data (Elgendy & Elragal, 2014). What benefits organisations perceive as "value" depends on their strategic goals for adopting and using big data (Ghoshal, Larson, Subramanyam, & Shaw, 2014). Maciejewski (2016) defines three different fields of applying big data in the public sector:

(i) Public supervision: Identifying irregularities (e.g. incompliance) and taking responsive action.;
(ii) Public regulation: Supporting policy development and execution by strengthening the information input and providing immediate feedback on policy and its impacts.;
(iii) Public service delivery: Improving certain public services or products (e.g. infrastructure such as roads).

It is important to note that the terms "data" and "information" have often been used interchangeably. In this research, data is distinguished from information based on the data–information–knowledge–wisdom hierarchy (Braganza, 2004): Data are raw materials, the ingredients of information, and information is the outcome of data analysis, used for a specific purpose.

## 2.2 The Big Data Paradigm

As big data is the primary input for meaningful network analysis in fraud detection, this section highlights two contrasting views concerning the use of big data within organisations: the rational view and the institutional view. This differentiation is useful for assessing how big data affects public decision-making (Van der Voort et al., 2019). The rational view represents a clear process in which big data can enhance the various activities in which information is required, whereas the institutional view represents a dynamic process, in which political or other aspirations partially determine when, where and how to use big data (Van der Voort et al., 2019). This section elaborates on these two views and argues for critical reflections on how organisations translate, as well as fail to translate, the potentials of such amount of data in actual value.

### 2.2.1 The Rational View of Using Big Data

The ongoing advancement of tools and technologies over recent years has created a new ecosystem with ample of opportunities for data-driven innovation. As described in the introduction, using big data to predict fraudulent activities has many potential benefits. The promise of big data to "predict the present" (Choi & Varian, 2012) and even the future (Goel et al., 2010; Uchida, 2010) gives big data the potential to be extremely helpful for decision-making in the public sector (Van der Voort et al., 2019). Big data can be used to extract trends that previously went undetected (Teyebi & Glässer, 2016; Van der Voort et al., 2019), and improving enforcement effectiveness and efficiency. In a similar vein, Baesens, Bapna, Marsden, Vanthienen, and Zhao (2014) describe big data as "the mother lode of disruptive change in a networked business environment". Van der Voort et al. (2019) frame this as the rational perspective of big data. The current emphasis on the potential merits of big data relies on the assumptions that big data leads to better information and therefore to better decisions.

### 2.2.2 The Institutional View of Using Big Data

Following from the previous section, big data in the public domain is commonly understood to contribute to better governance. Although the opportunities look promising, the assumption that big data and big data analysis result in better decisions might be too simplistic (Janssen, Van der Voort, & Wahyudi, 2017). Regardless of the hypes, extensive publicity, and high hopes that come along with big data, it does not guarantee the gaining of actual value. Instead, it may lead organisations to believe that they can gain more value from big data than they are actually capable of in practice (Ransbotham, Kiron, & Prentice, 2016; Ross, Beath, & Quaadgras, 2013). As the amount of available data rises to new heights, so too does the complexity (Kayser, Nehrke, & Zubovic, 2018). The institutional view implies that ability to create value from big data depends on having the right process in place to give meaning to the data. Big data applications need to manage the various V's that come with big data: Volume, Variety, Velocity, Veracity, and Variability. Organisations are challenged to create the right contexts, by acquiring the right data-sources, by shaping IT-structures and processes, and by asking the right questions that guide the data analysis (Kayser et al., 2018).

A core assumption in the institutional view is that the decisions made in the design phase are already significant for the final outcome. The quality of decisions made is not just solely dependent on the data, but also on the process in which the data are collected and the way data are processed (Janssen et al., 2017). For example, as data are the key input to any data analytics, the selection of data will have a deterministic impact on the analytical models that will be built in a next step (Baesens et al., 2015). One of the factors influencing the data selection is the often limiting accessibility to data. Distribution of data is usually restricted given the sensitive and personal nature and the ethical considerations related to their circulation (Kitchin, 2014). Another reason for limited accessibility is the asset attribute of data for the owner to create competitive advantage. Data might be limited in access so that the owner of the data can maximise the value or leverage income through the sale of data. In other cases, an organisation might not want to distribute data for the fear of what the data might reveal, with economic or political consequences as a result. Next to accessibility restrictions, data selection is concerned with data quality and veracity. Data quality refers to the cleanliness of data (error and gap free), the untaintedness (bias free), and consistency (few discrepancies). Veracity concerns the authenticity of the data and the extent to which it accurately and faithfully represents what it is meant to. (Kitchin, 2014)

As becomes apparent in the previous examples, there are multiple interrelated factors affecting the final quality of the decision based on big data (Janssen et al., 2017). These factors originate from various disciplines; from engineering to management, legal studies and public administration (Van der Voort et al., 2019). The current view that big data leads to better information provision and therefore better decisions neglects the several actors that shape the process from data selection, preparation, and generation to the final decisions taken (Sharma, Mithas, & Kankanhalli, 2014; Van der Voort et al., 2019). To summarise, the institutional perspective of using big data in a decision-making process challenges two main assumptions underlying the rational perspective (Van der Voort et al., 2019):

(i) Data revolution yields better information;

(ii) Better information leads to better decision-making.

## 2.3 The Big Data Value Chain

One way to get better insight into the institutional view on big data is analysing the big data value chain. Value chains have been used to model the series of activities that an organisation performs in order to deliver a valuable product or service to the market (Porter 1985). The value chain categorises the value-adding activities of an organisation allowing them to be understood and optimised (Curry, 2016). A value chain consists of activities, also referred to as steps or sub-systems, with each inputs, transformation processes, and outputs. As an analytical tool, the value chain can also be applied to understand the value creation of big data and big data analytics (Curry, 2016). In reality, there are many data sources, variations in flows and decisions involved to increase the quantity and quality of data over time (Janssen et al., 2017). A data value chain can be described as the series of steps needed to generate value and useful insights from data (Curry, 2016). Despite of its potentials, the data value chain is hardly taken as an analytical tool of looking at big data (Janssen et al., 2017).

In literature, several researchers have proposed different steps in literature that make up a big data value chain. For example, Bizer, Boncz, Brodie, and Erling (2012) propose five steps; problem definition, data searching, data transformation, data entity resolution, and answer the query or solve the problem. M. Chen and Liu (2014) define just three steps; data handling, data processing, and data moving. Oussous, Benjelloun, Ait Lahcen, and Belfkih (2018) identify six steps; data capturing, data storage, data searching, data sharing, data analysis, and data visualisation. According to H. Chen, Chiang, and Storey (2012) big data systems require the subsystems data generation, data acquisition, data transportation, data pre-processing, data storage, and data analytics. Åkerman et al. (2018) proposes data acquisition, data transfer, pre-processing and storage, data analysis, and feedback. Although various steps are identified, little attention is paid to who executes these steps, in what for context (e.g. public sector), and the effects of one step on the subsequent step(s).

## 2.4 Fraud and Fraud Detection: Concepts and Definitions

Following from the previous sections, analysing the big data value chain can be used to understand big data value creation from an institutional perspective. In this research, big data value creation is concerned with detecting fraudulent activities that impose a threat on society. Up until now, fraud, criminal activities, and illegal activities, have been used interchangeably. However, a better understanding of the scope of the research requires precise definitions and clear differentiation of the key concepts. This section introduces the key concepts and definitions related to fraud and fraud detection.

### 2.4.1 Defining Fraud

Understanding what counts as crime is important for everyone involved in crime research, as its definition determines several policy decisions concerning social control (Henry et al., 2001). Crime is an *"unlawful act punishable by a state or other authority"* (Oxford English Dictionary, 2nd Ed.). Nonetheless, the term 'crime' has no simple and universally accepted definition (Cane, Conaghan, & Walker, 2008). Whether activities are criminal depends on the applicable rules or legislation in a country (Cane et al., 2008).

Fraud can be considered as white-collar crime and is in the literature commonly defined as *"A criminal deception; the use of false representations to gain an unjust advantage"* (Concise Oxford Dictionary, 11th Ed.). However, as argued by Baesens et al. (2015) this definition does not clarify the nature and characteristics of fraud, and as such, does not provide a solid base for the requirements of a fraud detection system. Therefore, the following definition is proposed: "*Fraud is an uncommon, well-considered, imperceptibly concealed, time-evolving, and often carefully organized crime which appears in many types of forms"* (Van Vlasselaer, Eliassi-Rad, Akoglu, Snoeck, & Baesens, 2015). This

definition, instead, emphasises five key characteristics associated with fraud (Baesens et al., 2015). First of all, fraud is uncommon; only a minority of the involved cases typically concerns fraud. Second, fraud is imperceptibly concealed, meaning that fraud evidence is obfuscated to hide illegitimate activities. Fraudsters apply various hard to detect, yet rational, strategies to execute fraudulent activities. As a consequence, it is often well considered and planned on how to commit fraud. This gives fraud a purposive and intentional dimension. The third characteristic refers to the evolvement of fraud over time. To remain undetected, fraudsters adapt and refine, given time, their method frequently. Fourth, fraud is often carefully organised crime, meaning that fraud is often committed in a network of fraudsters. There are different forces underlying this fraud network, including interdependency, homophily, and differential association (see Section 2.5.2). Fraudsters do not operate fully isolated, but rather have allies and cooperate with other agents, i.e. they are not independent. Homophily is a theory in network science which states that entities that prefer to associate with people similar to them (McPherson, Smith-Lovin, & Cook, 2001). Differential association theory states that the likelihood to commit crime depends on the (anti)criminal norms of an actor's connections. Knowledge of how to commit camouflaged fraud is shared with other agents (Van Vlasselaer, Eliassi-Rad, et al., 2015). As such, using a social network analysis technique, which will be elaborated on in the next section, might reveal insights into the context of how and where fraud is committed (Baesens et al., 2015). Finally, fraud may occur in many different types of forms. These different forms originate from the wide range of methods used by fraudsters and the various contexts in which fraud occurs, such as tax, telecommunications, banking, medicine, e-commerce, and insurance (Šubelj, Furlan, & Bajec, 2011).

Motives or drivers to commit fraud can be explained by the fraud triangle that has been developed by the American sociologist Cressey (1953), who worked extensively in the fields of criminology and white-collar crime. According to Cressey (1953), the occurrence of fraud is conditioned by the joint existence of three elements: pressure, opportunity, and rationalisation (Figure 1).

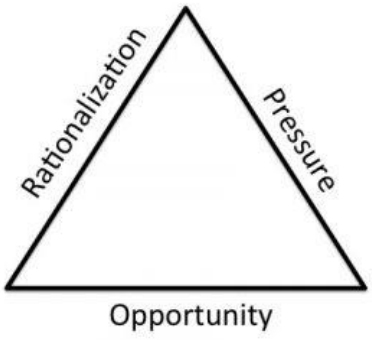



Figure 1 The three elements of the Fraud Triangle

- *Pressure:* fraud is committed because a problem or a need is experienced of financial, social, or any other nature, and it cannot be resolved or relieved in a legitimate manner. In other words, a certain pressure represents the reason to misrepresent results.;
- *Opportunity:* the precondition for fraud is the ability to commit fraud. Fraudulent activities can only be committed when the chance exists for the individual to solve the experienced pressure or problem in an illegitimate yet concealed manner.;
- *Rationalisation:* the cognitive mechanism that explains why fraudsters do not refrain from committing fraud and consider their behaviour as acceptable.

Although fraud, as defined in this research, has an explicit intentional dimension, there exist different images of fraud, or more general for organisational non-compliance. Kagan and Scholz (1980) differentiate between three different theories of why business firms violate regulations. In the first theory, organisations are framed as 'amoral calculators'. Motivated entirely by profit-seeking, firms

carefully assess opportunities and risks associated with non-compliance. Based on a cost-benefit analysis, the law is disobeyed when the anticipated fine and probability of being caught are small in relation to the profits that can be gained. In the second theory, organisations are seen as 'political citizens'; principally inclined to obey the law, partly because of belief in the rule of law, and partly as a matter of long-term self-interest. However, when regulations and enforcement officials treat them arbitrarily, impose unreasonable burdens, or when disagreement with regulations exists, this commitment becomes fragile. In the third theory, organisations are seen as 'organisational incompetent'. Violations of regulations are attributed to organisational failure in implementing an effective compliance strategy.

### 2.4.2 Fraud Detection and Prevention

To fight fraud, both fraud detection and fraud prevention are essential for an effective strategy. Fraud detection refers to the ability to recognise fraud patterns and discover fraudulent activities (ex-post approach), whereas fraud prevention refers to measures that can be taken to avoid fraud (ex-ante approach) (Baesens et al., 2015). Effective and efficient fraud reduction requires both methods to work in a complementary manner.

According to Baesens et al. (2015) successful data-driven fraud analytic models must meet the following five key characteristics:

- *Statistical accuracy*: the statistical significance, detection power and the correctness of the statistical model in labelling cases as being suspicious;
- *Interpretability:* is needed to when a deeper understanding of the detected fraud patterns is required. This could include validation of the model before putting it into use. Since interpretability may depend on the user's knowledge. Interpretability depends on the user's knowledge and therefore contains a certain degree of subjectivism;
- *Operational efficiency:* the time that is required to generate a certain output, i.e., the time required to evaluate whether a case is suspicious or not. Some cases need to be evaluated in real time (e.g. a money transaction), making operational efficiency crucial;
- *Economical cost:* developing and implementing a fraud-detection model involves a significant cost to an organisation. The total costs include the costs of selecting, gathering, and analysing data, and the costs to put the resulting analytical models into practice. Additional costs include human resources, computing power, and IT-structures. Possibly also external data has to be acquired to enrich the available internal data;
- *Regulatory compliance:* depending on the context, there may be specific external and/or internal regulations and legislation that apply to the development and application of a model.

The traditional approach for fraud detection is the expert-based fraud detection system, whereby the system relies on human expert input, evaluation, and monitoring (Baesens et al., 2015). Rather than searching for mathematical patterns in a data set, this is usually an intuitive and experienced based approach; expert opinions are collected on a number of decision criteria (Nisbet, Miner, & Yale, 2018). The problem with expert-based systems is that they rely on subjective inputs and tacit knowledge that may be contradictory to each other (Nisbet et al., 2018). In addition, as described by Dazeley (2006) this is a costly and time-consuming task and requires domain experts. In the age of big data, methods involving manual fraud detection are not only time-consuming, expensive, and inaccurate, but they are also impractical (West & Bhattacharya, 2016). It is impossible to detect all fraud by manual inspection over a large database (Y. Peng et al., 2006). Besides, small sets of well-written, heuristic rules may be transparent, relatively simply understandable, and can easily be translated into an automated system. Yet, since fraud usually evolves over time, the fixed rules need to be updated and new rules need to be added. For many problems, such as fraud detection, a rule-based system may become large and difficult to understand and to maintain (Ryman-Tubb, Krause, & Garn, 2018). In addition, expert-based systems are often restricted to identify and characterise

known patterns of fraudulent behaviour (McCue, 2015). As the size of databases increases, traditional fraud detection approaches may miss a great portion of fraud (Y. Peng et al., 2006). Large, data-driven systems, instead, can develop rules based on machine learning algorithms (Jurgovsky et al., 2018). The volume of data allows tools to extract insights, trends, and predictions that expert-based systems are not capable of. Predictive modelling uses mathematical and computational methods to predict an event or outcome. The system learns the fraudulent patterns in historical data to predict future fraud. Using an automated, data-driven system could lead to a more efficient and effective methodology for detecting fraud (Baesens et al., 2015). However, such approaches have also some disadvantages. Besides of needing big data in good quality, and much computational power and engineering competencies, machine learning models are generally black boxes (Holzinger, 2018). Data-driven approaches are becoming increasingly opaque, and even if the underlying mathematical principles of such models are understood, they still lack declarative knowledge (Holzinger, 2018). It might be very difficult (if not impossible) to explain to others how certain scores or decisions are obtained. Table 1 summarises the main differences between both systems.

Table 1 Fraud Detection Systems

| Characteristics | Expert-based system | Data-driven system |
|---|---|---|
| **Rules** | Expert rules | Data-driven rules |
| **Knowledge** | Tacit | Explicit |
| **Analytics** | Limited predictive capability | Both reactive and predictive |
| **Problems** | Difficult to maintain<br>Heuristic rules<br>Expert knowledge dependent<br>Capturing of tacit knowledge<br>Costly due to labour-intensity | "Black box" algorithms<br>Data-dependent<br>Data analysis competencies |

## 2.5 Social Network Analysis

In the recent years, the use of social media websites has gained an increasingly important role in many peoples' lives. People communicate through online social network sites like Facebook, Twitter, LinkedIn, Instagram, and so on, and share their experiences with friends, family, acquaintances and many more. By just one click you can update the rest of the world about all your whereabouts. And this is exactly where it becomes interesting. This whole interconnected network of people knowing each other seems to be an invaluable source of information (Baesens et al., 2015). As the name suggests, Social Network Analysis (SNA) is basically the analysis of social networks (Chauhan & Panda, 2015). However, the formation of social networks goes far beyond the use of online social media networks. In fact, networks are all around us. One could for example think of more traditional communities, transportation and mobility networks, epidemiological networks (the spread of a disease over the population), information networks, trade networks, biology networks, or utility networks (Kong, Shi, Yu, Liu, & Xia, 2019). In general, Social Networks (SNs) are "collections of individuals or organisations that are interrelated in a particular situation like collaboration and socialisation" (Kong et al., 2019). Analysis of these networks (Social Network Analysis) aims to understand the networks and participants and has two main focuses: the actors and their relationships in a specific social context (Cachia, 2008). Also the fraud detection domain might benefit from the analysis of social networks (Baesens et al., 2015). With regard to that end, the NVWA wants to use network analytics as a means to structure and analyse data and improve risk-oriented inspection.

Over the years, various approaches have been proposed in literature to counteract fraud based on big data, among which is SNA. SNA can be used for measuring and analysing the structure of relationships between actors (Steketee, Miyaoka, & Spiegelman, 2015). Relationships in a network are the basis on which actors in a network are connected. The relationships or connections

underlying the network can occur in a variety of forms, such as interpersonal relationships like friendship, advice seeking, or trust, which denote interactions between individuals; or inter-organisational relationships like knowledge and resource sharing, or trading goods, which denote the interaction between organisations as a whole (Steketee et al., 2015). One of the core assumptions of SNA is that the structure of these connections influences individual and organisational behaviour. For example, relationships between individuals or organisations might enable or restrain access to resources, exchange of information, or lead to exposure to social norms and culture (Steketee et al., 2015). SNA can be a powerful technique for researching social phenomena such as the flow of information through a network and the identification of key actors. Important for the flow of information is the intensity of the relationship (Baesens et al., 2015). To illustrate, the relationship between two best friends and the corresponding information exchange completely differs from the relationship between two distant acquaintances.

### 2.5.1 Network Components, Structures, Characteristics, and Metrics

Since the early 2000s, network science has begun to advance rapidly due to the integration of sociology, mathematics, and computer science (H. Chen et al., 2012). Various social network theories, mathematical models, network metrics, and topology, have been developed that help understanding network properties and relationships. Besides, there is a wide availability of software for network analysis nowadays, both commercial and freely available through the internet. In this section, a brief overview of the network components, structures, characteristics, and metrics applied in SNA is provided. Please note that the objective of this review is not to be exhaustive; merely exploratory and introductory.

**Network Components**

To begin, every social network includes nodes, relationships, and edges. In the social network literature, nodes have also been termed actors, vertices, units, or points, while edges have also been termed connections, links, or ties. All are accepted terminology referring to the same respective concepts (Steketee et al., 2015). A node represents the network members which can be, as mentioned in the previous section, individual persons or collective groups (e.g., boards of directors, social groups, clubs, or businesses). Relationships in a network are the basis on which nodes in a network are connected (e.g., friendship, collaboration, trading goods or services, or passing information). An edge represents a connection between two nodes on the basis of a given relationship (Steketee et al., 2015).

**Network Characteristics**

The edges depicting the relationships in a network can be characterised by various attributes. First of all, relations may be described either as directed (e.g., passing information or undirected (e.g., joint collaboration) (Steketee et al., 2015). The directed lines are often called arcs (Scott & Stokman, 2015). The direction assigned to the line represents the likely flow of information or resources between the two participants (Scott & Stokman, 2015). It is important to note that there may be multiple types of relationships existing between the same set of nodes (e.g., there may exist, at the same time, friendship, romantic, and collaboration connections among two co-workers in an organisation) (Steketee et al., 2015). This is similar to the multi-edge representation in Figure 3.

Secondly, edges can be distinguished by signs attached to them to indicate the type of relationship (Scott & Stokman, 2015). These signs can be positive or negative signs, e.g. the representation of positive and negative relations such as cooperation and conflict. The edges between the nodes may also be valued or weighted. The strength or intensity of a relation is then represented by a number, either an actual number (e.g. monetary value) or a scaled representation of strength (Scott & Stokman, 2015). Baesens et al. (2015) distinguish the following weights:

- *Binary weight:* The standard network representation. The edge weight is either 0 or 1 and reflects whether or not a link exists between two nodes. Binary weighted graphs can be extended by a negative weight (−1). Negative weights are then used to represent isolation, neutral weights represent an unconscious connection, and positive weights are used to represent friendships.;
- *Numeric weight:* A numeric edge weight expresses the affinity of a person to other persons (s)he is connected to. High values indicate a closer affiliation. A popular way is the "Common Neighbour" approach, i.e. the edge weight equals the total number of common activities or events both people attended.;
- *Normalised weight:* The normalised weight is a variant of the numeric weight where all the outgoing edges of a node sum up to 1.;
- *Jaccard weight:* This weight indicates the degree of similarity between both nodes (a and b), based on the events both nodes attend (A and B), using the following formula:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

Whereas edge weights represent the connectivity within a network, nodes are commonly characterised with labels. These labels can indicate demographic values, interests, beliefs or other characteristics of the node (Bhagat, Cormode, & Muthukrishnan, 2011). In general, a social network is multi-labelled: a node usually has various attributes (Soulie Fogelman, Mekki, & Sean, 2011). When analysing fraud networks, the fraud label can be integrated into the network (Figure 2); a node could be labelled fraudulent or legitimate (Baesens et al., 2015).

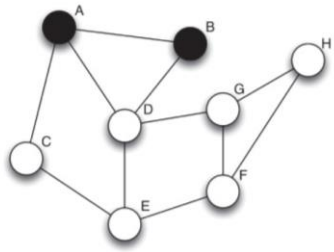



Figure 2 Example of a fraud network. Actors A and B are labelled as fraudulent (Baesens et al., 2015)

**Network Structures**

The simplest form of any social network is a dyad; a network containing two nodes with a single edge connecting them. A dyad is not necessarily very interesting but does represent the building block of larger networks (Steketee et al., 2015). When a third node is tied to both nodes in the group, a triad or clique is formed. From there, larger networks of expanding size and structural complexity are created (Steketee et al., 2015).

Although, in general, edges connect two nodes to each other, some special variants (Figure 3) are sometimes required to map reality more accurately (Baesens et al., 2015):

- *Self-edge:* A self-edge is a link from the node referencing to the node itself. For example, a person who transfers money from his/her account to another account (s)he owns.;
- *Multi-edge:* A multi-edge exists when two nodes are connected by more than one edge. For example, two people can be connected through a friendship as well as a business relationship;
- *Hyper-edge:* A hyper-edge links more than one node in the network. For example, three people who went to the same event, or three organisations share the same resource.

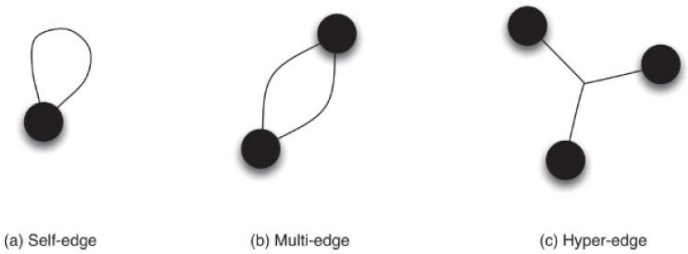



Figure 3 Edge Representation (Baesens et al., 2015)

When the nodes in a network consist of only node type, the networks are considered to be unipartite graphs (Figure 4). However, in many applications it might be useful to integrate a second node type in the network. Such networks are bipartite networks and represent the reason, also referred to as an event, why people connect to each other (Baesens et al., 2015). An event can, for example, be a shared resource (social security fraud) or insurance claim (insurance fraud) (Baesens et al., 2015). Whenever a network integrates more than two heterogeneous node types, a network becomes multipartite. A multipartite graph usually reflects the true reality more accurately, but simultaneously introduces a lot of complexity in the network. The network can quickly grow to immense sizes, requiring more computing power and imposing the need for scalable and efficient algorithms (Baesens et al., 2015).

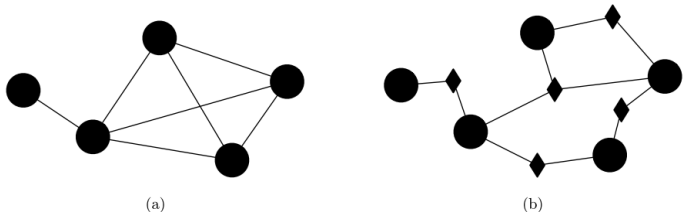



Figure 4 A unipartite graph (a) and a bipartite graph (b)

**Network Metrics**

In general, a network can be represented either graphically or mathematically. The graphical representation of a network, or sociogram, is the most intuitive and straightforward visualisation of a network (Baesens et al., 2015). Visualising graphical information in social networks enables experts to intuitively make conclusions about the social networks. Different methods of visualising the social network can be used to represent information, such as spatial position, colour, size, and shape (Teyebi & Glässer, 2016). Even though graphical network representations are appropriate for visualisation purposes, they cannot be used to compute useful statistics and extract meaningful characteristics of the network. Graph theory can be considered as the mathematical foundation for the analysis of networks (Chauhan & Panda, 2015). The adjacency matrix and the adjacency list are two ways to present the network in a mathematically interesting way (Baesens et al., 2015). This section focuses on the various metrics used in graph theory that measure the impact of the social environment on the nodes of interest. In general, two levels of analysis can be distinguished: node-level analysis (also known as ego-centric analysis) and network-level (socio-centred analysis). This section elaborates on those methods. In addition, attention is given to clustering, flow analysis, and collective inference. For more detailed information on the methods behind the SNA measures and metrics, see Wasserman and Faust (1994), Scott (2000), Carrington, Scott, and Wasserman (2005), Hanneman and Riddle (2005), Scott (2011), and Scott and Carrington (2011).

*Node-Level Analysis and Network-Level Analysis*

Node-level analysis is also known as the ego-centred approach and emphasises the individual actor and its immediate network neighbourhood (Boccaletti, Latora, Moreno, Chavez, & Hwang, 2006).

Node-level analysis thus focuses on the composition of local network structure. Entities may influence each another through their connections and adjust or select their connections based on the characteristics of their neighbourhood. The direct connections of an entity are usually most salient to its behaviour, but indirect connections such as their neighbours' neighbours (second degree connections) may be taken into account as well (Nooy, 2009). For example, if a fraudulent node is in the close neighbourhood, fraud might impact that node more intensively and "contaminate" the node, i.e., share knowledge on how to commit fraud (Baesens et al., 2015). In other words, an entity's local context or ego-network is likely to affect its behaviour (Nooy, 2009). Examples of node level analysis include degree centrality and (probabilistic) relational neighbour. Table 2 provides an overview of commonly used metrics (Baesens et al., 2015).

Instead of focusing on examining the attributes of individual nodes, network-level analysis uses the overall distribution of relations among nodes. Network-level metrics include centrality metrics such as closeness and betweenness. Centrality metrics can be utilised to quantify the importance of an actor in a social network (Boccaletti et al., 2006). Doing so can have important implications for future research, policy making, or planning (Steketee et al., 2015). It allows us to understand actors and their importance in a network. Centrality metrics can be useful in preventing the expansion of future fraudulent activities by identifying the central node(s), that is, nodes that might impact many other nodes (Baesens et al., 2015). If a network is very much centralised around a single node then the network can be easily fragmented by deactivating that node (Chauhan & Panda, 2015). Next to network centrality, also density is a commonly used network-level metric. Network density and centrality metrics are complementary to each other. Whereas density explains the general level of connectedness in a network, centralisation explains the extent to which the connectedness is focused around a particular node. Also the node-level metrics can be extracted based on the whole network structure, or on a subgraph of the network. Additionally, one could look at the network dynamics of how the structure of a network changes over time or how the network structures of two groups or

Table 2 Overview of Network-Level Metrics (Baesens et al., 2015)

| | | |
|---|---|---|
| **Triangles** | Number of fully connected subgraphs consisting of three nodes | |
| **Graph theoretic center** | The node with the smallest maximum distance to all other nodes in the network (lowest reach centrality) | |
| **Degree** | Number of connectons of a node (in- versus out-degree if the connections are directed) | $c_i = \sum_j a_{ij}$ |
| **Geodesic path** | Shortest path between two nodes in the network | $d(v_i, v_j)$ |
| **Closeness** | The average distance of a node to all other nodes in the network (reciprocal of farness) | $\left[\frac{\sum_{j=1\,(j\neq i)} d(v_i, v_j)}{n-1}\right]^{-1}$ |
| **Betweenness** | Counts the number of times a node or connection lies on the shortest path between any two nodes in the network | $\sum_{j<k} \frac{g_{jk}(v_i)}{g_{jk}}$ |
| **Network Constraint/ Structural holes** | Measures the value of network constraint based on structural holes (e.g. Consider three nodes A, B and C, where A is linked to B and C, but a link between B and C is missing. That missing link forms a "structural hole" in which A acts as a bridge between B and C. This gives advantage (e.g. information flow) to A, as B and C cannot interact directly. You can say that A is less "constrained" by its ego-network and could benefit from the structural hole existing between B and C.) | $c_{ij} = (p_{ij} + \sum_q p_{iq} p_{qj})^2 \quad , \qquad i \neq q \neq j$ |
| **Relational Neighbour** | Relative number of neighbours that belong to a class (e.g. to class fraud) | $P(c\|n) = \frac{1}{Z} \sum (n_j \epsilon Neighbourhoood_n \| class(n_j) = c)^{w(n,n_j)}$ |
| **Probabilistic Relational Neighbour** | Probability to belong to class c given the posterior class probabilities of the neighbours | $P(c\|n) = \frac{1}{Z} \sum (n_j \epsilon Neighbourhoood_n)^{w(n,n_j) P(c\|n_j)}$ |
| **Density** | The extent to which the nodes are connected to each other (reciprocal of farness) | $d = \frac{2M}{N(N-1)}$ |

organisations of similar size compare to each other (Steketee et al., 2015). Depending on the nature of the research, one could decide to apply network-level analysis as a starting point, and subsequently use node-level analysis to analyse the role of one node in the network (Silva & Saraiva, 2015).

*Clustering*

A commonly used analysis at network level is the identification and examination of network subgroups or substructures, sometimes referred to as clusters or clustering (Steketee et al., 2015). Clusters refer to interconnected subgroups in a network with a higher number and more intensive relationships among the members of the cluster than any other random subgraph in the network. Often clusters can be displayed visually as regions in a network with relatively high local density and relatively few links to other clusters (Hoppe and Reinelt, 2010). Uncovering such subgroups in a network can be highly useful in applied fields (Steketee et al., 2015). In a fraud context, communities can play an important role in influence dispersion. Groups often share, reinforce, and complement ideas and alternatives on how to commit fraud (Baesens et al., 2015). By using clustering methods one can consider whether people are more likely to commit fraud if they are influenced by a whole cluster instead of being influenced by only one fraudulent entity. The effect of one fraudulent cluster on a node is diminished if the node belongs to many legitimate or non-fraudulent groups. In many applications, the discovery of clusters can result in the detection of hidden fraudulent groups or existing fraudulent structures (Baesens et al., 2015).

Methods for clustering include graph partitioning, spectral clustering, and the Girvan-Newman algorithm (Baesens et al., 2015). Other advanced clustering measures exist such as small worlds, preferential attachment, core-periphery structures, and estimated random graph models (Steketee et al., 2015), but examining these measures falls out of the scope of this research.

*Flow Analysis*

Another type of analysis worth to mention is flow analysis. SNA can be used to model and explore how information, attitudes, or physical items spread through network connections (Steketee et al., 2015). For example, a high density corresponds to a high connectedness, which could be an indicator for nodes extensively influencing each other. This may be important in the analysis of fraud (Baesens et al., 2015).

Flow through a network is created by directional edges. However, flow additionally depends on the properties of actors, edges, and what goods or information are being transmitted through the network. Findings of the analysis may differ depending on the rules that govern flow (Borgatti, 2005). Examples of such considerations include (Steketee et al., 2015):

- Is the graph directional, and if so, can a directional path between A and B exist together with a path connecting B an A? E.g. A physical item can be returned, traveling from actor A to B and back to A, while a piece of information would be directional and not reverse its path.
- Could an actor be part of a path just once or more than once? E.g. if a disease leaves the host immune, the actor cannot be reached more than once, while a physical item may travel to the same actor multiple times.
- Could a path be travelled more than one time? E.g. physical items like a loaned book follows a path usually once, while others, such as money, can repeat a path multiple times.
- Could an actor be connect to multiple actors? E.g. a physical item cannot travel from actor A to both actors B and C at the same time, while a piece of information can duplicate itself and reach both B and C simultaneously.
- If duplication occurs, are copies identical or merely similar to the original item? E.g. Consider the difference between an email and a virus. A forwarded email is usually (merely) the same

replicated information. The spread of a virus instead reaches multiple recipients, but mutations and variations occur in the process.

Borgatti (2005) illustrates that measures such as closeness and betweenness may yield different results across different path restriction rules. As such the use of these measures should include consideration of whether closeness, for example, is truly capturing the distance or time by which goods or information flow through a network (Steketee et al., 2015).

Similarly, it is important to consider how to define "importance", which can be illustrated with the following example. As explained in the previous section, key players in a network can be identified using centrality metrics. Degree Centrality makes the assumption that important nodes have many connections, whereas Closeness Centrality assumes that nodes that are important are close to other nodes. This means that there are different ways to think about "importance", and depending on how centrality is defined, the outcome may differ.

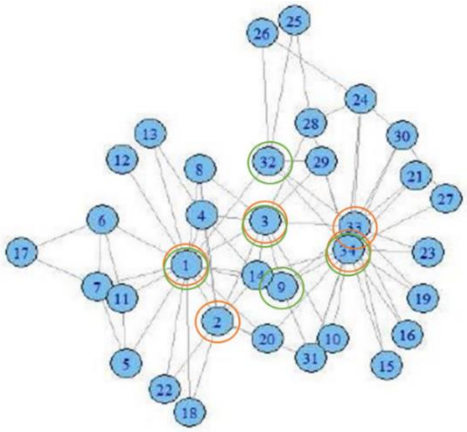



Figure 5 Network of a 34-person karate club (Zachary, 1977). The orange circles indicate the five most important players using Degree Centrality, the green circles indicate the 5 most important people using Closeness Centrality.

**Collective Interference Algorithms**

Collective interference algorithms can be used to make predictions about the unlabelled nodes in a network. Thus in a partially labelled network, where some actors are observed as (non)fraudulent but others are unobserved, a collective inference procedure infers a set of labels or probabilities for the unknown nodes (Baesens et al., 2015). Some common examples of collective inference procedures are PageRank (Page, Brin, Motwani, & Winograd, 1998), Gibbs sampling (Geman & Geman, 1984), Iterative Classification Algorithm (Lu & Getoor, 2003), Relaxation Labelling (Chakrabarti, Dom, & Indyk, 1998), Loopy Belief Propagation (Pearl, 1986). Again, such advanced concepts and measures deserve attention but fall outside the scope of this research.

### 2.5.2 Related Work: Applications of Social Network Analysis

The development and interest in Social Network Analysis have increased intensively over the last few decades. SNA collects the data important for making business decisions, market research, identification of influential users, and so on (Milovanović, Bogdanović, Labus, Barać, & Despotović-Zrakić, 2019). Current literature on SNA shows its application in a wide variety of domains. A search using the keyword "Social Network Analysis" retrieves a large number of articles (>6000). This section focuses on examples of applications rather than the methodological details. A more comprehensive collection of discussions and applications instead has been brought together by Scott (2011).

As the previous sections suggest, one might say, in general, that a network becomes a social network whenever the actors are people or groups of people like organisations (Baesens et al., 2015). Health

research, marketing research, and education research are three of many areas that have applied social network analysis (Steketee et al., 2015). SNA can, for example, be used to analyse a network of connected people to identify how a disease would spread across the population and which links need to be broken to prevent the whole network from getting infected (Chauhan & Panda, 2015). Marketing applications include the identification of market leaders and the exploitation of their position to influence other users in a network (Webster & Morrison, 2004). In scientific research, collaboration networks among researchers can be used to analyse and improve the distribution of research findings and products by researchers and research initiatives (Steketee et al., 2015).

In addition to the applications described in the previous paragraph, the study of crime may also benefit from the use of SNA (Scott & Carrington, 2011). Sparrow (1991) was pioneering in combining SNA and criminology and has summarised existing concepts of network analysis applied to crime analysis. Although the application of SNA in the crime domain is still very generic and a criminal act can encompass a wide range of activities (e.g. from bicycle theft terrorist attacks), fraud can be considered to be one of them. Fraud is, like other crime, often carefully organised crime and committed in a network of fraudsters (Baesens et al., 2015). Fraudsters are dependent on allies for cooperation and support, also known as interdependency. This underlines the social character of fraud.

This section focuses on the use of SNA in crime in general and in fraud in particular. As a reader you will be introduced to the underlying theories to apply network analytics in crime research from a sociology perspective, followed by the analytical purposes of SNA for law enforcement, and ends with the state-of-the art in practice.

**Supporting Theories**

One of the theories giving foundation to apply SNA in crime research is the differential association theory. This theory, first formulated by Sutherland (1939), states that criminal attitudes and behaviour are not innate, but are learned from "intimate personal groups." According to this theory, the propensity of a person to be delinquent is affected by the relative strength of criminal and anticriminal norms hold by his or her close connections (Scott & Carrington, 2011). In other words, an actor's connections influence the norms for committing a crime.

Next to the influence of the neighbourhood (relations) on the individual, the homophily theory emphasises the selection of other actors by the individual, that is, individuals prefer to associate with people who are similar to them (Scott & Carrington, 2011). Following Barone and Coscia (2018) in narrowing to a network perspective, homophily implies that nodes will likely connect if they share similar characteristics.

Based on these theories, one can assume that the probability of someone committing fraud depends on the people (s)he is connected to (Baesens et al., 2015). These are the so-called guilt-by-associations. "Guilt" distributed from guilty actors to their neighbours through relationships that constitute the network; if a neighbour connects to many guilty entities, it is likely to be guilty too. Whether this is the result of homophily or differential association is ambiguous. Nonetheless, there is a consensus that the two theories complement each other and that a correlation exists between an actor's delinquent behaviour and its surrounding network (Scott & Carrington, 2011). Studies on social security data sets show that fraudulent companies are significantly more connected to other fraudsters (Easley & Kleinberg, 2010; Park & Barabási, 2007; Van Vlasselaer, Eliassi-Rad, et al., 2015). Knowledge of how to commit fraud in a concealed manner is shared with other agents (Van Vlasselaer, Eliassi-Rad, et al., 2015). Fraudsters can be linked together as they seem to attend the same activities, are involved in the same crimes, use the same set of resources, or sometimes even are one and the same person (e.g. identity theft) (Baesens et al., 2015). As such, using a social network

analysis technique might reveal insights into the context of how and where fraud is committed (Baesens et al., 2015).

**Purposes of Social Network Analysis**

SNA may fulfil different analytical purposes depending on the context in which it is applied. Van der Hulst (2008) has summarised some of these purposes for law enforcement; each will be presented and discussed in turn:

- *Risk analysis and threat assessments*: SNA can be used for monitoring, evaluating and prediction of potential threats associated with the entities in a network. By doing this, one could obtain better understanding of the structure of crimes, how the involved actors and networks are positioned (e.g. criminal collaborations), and to what extent they pose a threat to society. The use of the SNA metrics can help identify potential risks that otherwise may be overlooked.
- *Destabilise networks:* SNA can help to identify the key players in a network and evaluate or predict the possible consequences of eliminating specific actors from a network to destabilise that network. Particular targets can be identified that would cause maximal disruption of the ongoing or planned illegal activities. Hence, knowing what determines the strengths and vulnerabilities of a network and the particular roles and positions associated with actors provides investigators with tactical options to demobilise (or reinforce) it.
- *Role identification:* SNA can also prove useful to identify roles in a network. These could be aliases (more nodes are one and the same actor due to similarity(ies) in the pattern of their social ties), identical roles (e.g. facilitators or brokers), or actors that serve as substitutes. Role identification would make sense for comparisons within a network as well as between networks. For example, potential successors of key players who are dismissed or eliminated from a network can be located, or roles and functions across networks can be compared in order to target important players (e.g., if actor A is the brain behind illegal activities in network X then actor B may be occupying an identical role in network Y if the social links and structures of both actors and both networks are similar to each other).
- *Scenario building:* SNA can be helpful in the reconstruction of scenarios based on speculation about what might have happened in order to "explain a current reality with a fixed outcome (e.g., a murder case)". Visualisation and analysis of networks could support creative thinking that may lead to the offender(s).
- *Support decisions on the deployment of intelligence assets:* Because of the ability to identify important key players and ties in a network, monitoring efficiency can be improved. SNA can support tactical decisions on the deployment of intelligence assets in the most optimal locations worth monitoring.
- *Evidence for prosecution:* SNA can become a powerful tool to support law enforcement by making offenders accountable for their role and involvement. However, this requires that roles and responsibilities can be proven based on social structures. Using SNA to serve as evidence for prosecution is yet far from realisation. This is mainly due to the imperfection of available data in law enforcement.

**From Theory to Reality**

Even though several authors report about the prospection of analysing criminal networks (Baesens et al., 2015; Scott & Carrington, 2011; Teyebi & Glässer, 2016), widespread applications of SNA in this domain are lacking. Scott and Carrington (2011) argue that the use of social network analysis in criminology is still in its infancy, since "the great majority of crime network studies consider only the characteristics of the members of the networks, and not of the structure of their relationships". Most analyses of network structures rely just on the examination of visual representations of networks, rather than computational analyses. Even the computational analyses tend to limit themselves to the

simplest network concepts, such as centrality and density (Scott & Carrington, 2011). As Van der Hulst (2008) has pointed out, researchers on criminal networks tend, with a few exceptions, to fall into two distinct categories, each operating under severe constraints. One the one hand, crime researchers have expertise in criminological theory and research but seem to lack analytical expertise and access to good data. Likewise, Steketee et al. (2015) argue that the potential for applying SNA is vast, but not without challenges. The collection of social network data can be problematic, especially when complete data on a network are needed to compute useful statistics. On the other hand, data analysts may have domain expertise and access to classified data but tend to lack knowledge on criminological issues – or may be prevented from publishing their work due to secrecy or privacy considerations. More generally, accurate and comprehensive data on "dark networks" are still rather difficult to obtain (Scott & Carrington, 2011).

To examine studies on the practical use of social network analysis explicitly in the context of fraud, a systematic literature review has been performed. The review consisted of search and analysis of journals with the aim of providing a descriptive overview of academic and practitioner-oriented studies related to fraud detection using SNA. More specifically, the review aimed to arrive at a set of papers that (i) focus on the adoption, implementation, or use of social network analytics to detect fraud and (ii) have specifically mentioned the terms "fraud" and "network analysis" or "link analysis[1]" in the title, abstract, key words, or within the article itself. For each publication, the fraud category, description of application, metrics used, and the size of data sets were obtained. Papers available since 2000 have been considered, given that this is when network science has begun to advance rapidly and large volumes of unstructured data gained momentum (H. Chen et al., 2012). ScienceDirect and Google Scholar were used to search on "Keywords" for journals. As such, the paper by Van Vlasselaer, Eliassi-Rad, et al. (2015) was retrieved which provides an overview containing published papers related to fraud detection using network analytics up until then. Given the scarcity of available publications, it was decided to add nine studies listed in the overview provided by Van Vlasselaer, Eliassi-Rad, et al. (2015) that meet the aforementioned conditions. These were papers mainly published before 2008. It should, however, be noted that papers published from the end of the 00's onwards better specify the network analysis terminology and the metrics used. In addition, it could be stated that the number and advancement of metrics used has, in general, increased considerably since then.

Overall, the studies varied from applications in social security fraud (Neville, Jensen, Komoroske, Palmer, & Goldberg, 2005; Van Vlasselaer, Eliassi-Rad, et al., 2015), credit card fraud (Soulie Fogelman et al., 2011; Van Vlasselaer, Bravo, et al., 2015) mortgage fraud (Nash, Bouchard, & Malm, 2013), insurance fraud (H. Chen et al., 2004; Galloway & Simoff, 2006; Šubelj et al., 2011), money laundering (Drężewski, Sepielak, & Filipkowski, 2015; Fronzetti Colladon & Remondi, 2017), opinion fraud (Akoglu, Chandy, & Faloutsos, 2013) telecom fraud (Chang, Lai, Chou, & Chen, 2017; Cortes, Pregibon, & Volinsky, 2001), security fraud (Fast et al., 2007; Neville et al., 2005), accounting fraud (McGloho Bay, Anderle, Steier, & Faloutsos, 2009), mobile internet fraud (Wei, Liu, & Liu, 2019), online auction fraud (Chau, Pandit, & Faloutsos, 2006; Chiu, Ku, Lie, & Chen, 2011; Pandit, Chau, Wang, & Faloutsos, 2007; Wang & Chiu, 2008; Yanchun, Wei, & Changhai, 2011) and health care fraud (Liu et al., 2015). An overview can be found in Appendix A.

According to this literature research, it should be pointed out that networks have not yet been used for detecting fraud concerned with food or consumer products. Besides, literature seems to have focused on the usage of SNA as fraud detection rather than the implementation and adoption of the technique in organisations. Despite their strong foundations and expressive power, the development of new fraud detection methods seems to be rather difficult. Several arguments are mentioned in

[1] Link analysis is accepted terminology for network analysis, mainly used in the early 2000s.

literature. Yufeng et al. (2004) argue that the development of fraud detection methods is constrained by the scarcity of available data sets and the limited disclosure results in public. This imposes a severe limitation on the exchange of ideas in fraud detection. Ryman-Tubb et al. (2018) argue that an important factor limiting the impact of fraud research is that the majority of published methods are black-boxes where their workings are mysterious; "the inputs and its decision on fraud can be observed but how one becomes the other is opaque". Given the little similarities between aforementioned cases (e.g. between the metrics used and the size of data sets) is consistent with an argument made earlier by Almeida (2009) that "each of the application areas has particular characteristics, so a solution to fight credit cards fraud cannot be applied in insurance companies".

The previous paragraphs indicate the fragmented nature and the absence of a comprehensive methodology to apply SNA in fraud detection. In short, there is a promising number of publications that describe for what purposes SNA can be used to detect offenders, but actual research designs that explain in depth what works, why, and when are lacking. This confirms previous literature findings on the early stages of development of SNA in the criminal domain.

## 2.6 The Institutional View of Big Data in Fraud Detection

The institutional perspective, as introduced in Section 2.2, challenges the NVWA's assumptions of the benefits of a data-driven approach (Section 1.2). Within the context of inspection, one additionally has to deal with the complex nature of fraud. The fraud definition as introduced in Section 2.4 already reveals much of this complexity. Fraud is usually a rare event and can take and an unlimited variety of different forms (R. J. Bolton & Hand, 2002), making its identification very difficult. Because of its uncommon and stealthy nature, the majority of records is usually legitimate (Nisbet et al., 2018). Intentional concealment of fraud results in a diffused fraud signal, which's detection requires rigorous methods. One often has to deal with an unbalanced dataset, meaning that datasets have very few fraudulent records, compared to the non-fraudulent. Consider a problem where 99% of the data are non-fraudulent and only 1% is fraudulent: classifying every record is non-fraudulent would have an accuracy of 99% (or an error rate of only 1%) (Pierre, 2018). This looks like an excellent model but seems to be useless in fraud detection (Pierre, 2018). Not to mention that fraud detection usually requires much computing power and systems capable of analysing enormous volumes of data. Fast and efficient algorithms must be developed to process all these data in time to act on any information related to fraud (Nisbet et al., 2018). Besides, many potentially predictive variables require external sources, imposing accessibility, trust, and privacy restrictions (S. Peng, Yu, & Mueller, 2018). On top of that, fraud researchers do not want to disclose their practices, because this might provide fraudsters means to defeat the detection system (Nisbet et al., 2018). In addition to the data complexity, fraud detection and prevention is a dynamic process. Fraudsters are adaptive and, over time, will usually find novel and increasingly subtle ways to circumvent detection measures (R. J. Bolton & Hand, 2002).

Even though the institutional view of using big data in decision-making challenges the assumption that big data leads to better public decision-making, empirical evidence in literature has been limited. Several authors express that studies have barely touched upon how organisations translate its potential into actual value. Günther, Rezazade Mehrizi, Huysman, and Feldberg (2017) argue that future research needs to empirically examine how different actors within organisations work with big data in practice, how organisational models are developed, and how organisations deal with different stakeholder interests to realise value from big data. Also Sharma et al. (2014) argue for further research needed to identify the process and conditions under which insights generated from big data lead to better quality decisions. Likewise, Kitchin (2014) addressed the quest for research focussed on the technical, organisational, ethical, and political shortcomings of big data. This research needs to be conducted across types of sectors and types of data to establish the issues and solutions pertaining to different data and systems. Depending on the context, different types of

challenges may arise in an organisational setting, ranging from very specific analytical capabilities to principal big data issues (e.g., no own data infrastructure, expert-based approaches) (Kayser et al., 2018). Especially research of unlocking the full potential of big data in the public sector is lacking (Maciejewski, 2016). The public sector may require a different approach to that of the private sector, which means adapting approaches to public interests, tasks, and policies. Also Van der Voort et al. (2019) and Janssen et al. (2017) emphasise the need for further research on how big data influences governance in the public sector. Such research is vital for valid findings and conclusions on adding value for organisations adopting data-driven systems (Kitchin, 2014).

The previous sections give reason to apply the institutional view on using big data for decision-making the context of inspection. The complexity of using big data in fraud detection requires data-analyst to make decisions. Since there are many different ways in which fraud can occur, means that there are several different ways of computing risks models (R. J. Bolton & Hand, 2002). In that sense, the use of big-data in creating fraud detection models can be seen as a decision-making process as theorised by Simon (1960), in which the way the alternatives are framed impacts the alternative chosen by people and in turn the subsequent decision. Complex organisational decision-making processes are often involved in creating options, evaluating them, and committing to a particular option (Sharma et al., 2014). Often, these are likely to be sub-optimal due to the complex circumstances, limited time, and inadequate mental computational power which impact the quality of decisions (Bok, Kankanhalli, Raman, & Sambamurthy, 2012). When applying the institutional view it is questioned if the promises of using a data-driven approach rather than an experience-based approach (Baesens et al., 2015) underestimate the human interferences within the data-driven process that shape the final outcome. This means a shift in the context of public decision-making from general public decision-making, like policymaking and agenda setting, to more operational decision-making in the public sector, i.e. decision-making done in the context of inspection.

From an institutional perspective, there is an opportunity of the different actors to introduce bias in the fraud detection and inspection process. On the one hand, data analysts can, wittingly or not, because inspectors hardly ever fully understand the complex algorithms that give meaning to the data (Van der Voort et al., 2019). In Social Network Analysis, for instance, the analyst is responsible for picking and choosing links from the database, asking the computer to visualise them, and computing meaningful statistics (Sparrow, 1991). Besides, deciding on the set(s) of nodes that lie within a network is a difficult problem for network studies (Carrington et al., 2005). On the other hand, inspectors will not in general be programmers, and are likely to express their expertise in terms that cannot immediately be translated into a program (Norvig, 1992). In addition, at the end of the process, inspectors may not absorb information from data analysts and adapt their operations, even if it is evidence-based (Van der Voort et al., 2019). Instead, they have the flexibility to act on their "gut feeling", regardless of the evidence provided by the data analysts.

To summarise, the ability to create from big data value depends on a whole chain of activities in which various actors play a role (Anderson, 2015). This requires collecting the right data, trustworthy data, having the right skills in place for data analysis, and concrete actions to realise the potential of big data. As introduced in Section 2.3, this chain of activities can be referred to as the big data chain (Janssen et al., 2017). This chain perspective might reveal insights in how the different actors and their actions shape the value of big data and the final quality of the decision-making. Within the context of inspection, this first requires better identification of the actors and their contribution to the big data chain. Second, further exploration needs to be done to establish the different decisions made by data scientist. These decisions may impact the potential of using big data for inspection.

## 2.7 Knowledge Gaps in Existing Literature

The literature review in this chapter suggest the necessity for an in-depth understanding of the context to understand what the institutional view on big data means for using SNA in fraud detection, and how the big data chain influences the value as such. As becomes clear in the previous literature review, several knowledge gaps can be identified on different levels:

(i) Application of SNA to detect fraud and suspicious activities related to food and consumer products (Section 2.5.2). Unlike other research, this research does not only apply SNA, but also emphasises the organisational complexities faced when to extract value from SNA.

(ii) Application of the big data chain perspective on data-driven inspection (Section 2.3). Considering the scope of the project, the big data chain will from now onwards be referred to as the 'big data activity chain'.

(iii) Application of the institutional view of big data in the context of public supervision and operational decision-making (i.e. decision making on the deployment of inspection capacity) (Section 2.6) ;

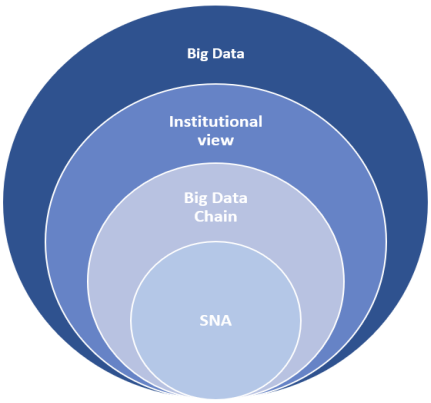



Figure 6 Knowledge gaps in data-driven fraud detection from an institutional perspective

# 3. Research Methods

As in the literature review is explained, the current age of digitalisation leads to an era in which new data sources evolve, and new techniques for storing and analysing large data sets are enabling many new applications. Literature is characterised by a strong focus on the opportunities that big data provides for organisations. However, the exact business value of any big data application is often unclear. The institutional view implies that there is a wide range of factors influencing the value of big-data decision-making. What this means for data-driven fraud detection requires a better understanding of the context as well as the big data chain (Janssen et al., 2017). From that perspective, this research aimed to apply Social Network Analysis as a data-driven means to detect fraud and suspicious activities with the ultimate aim to improve inspection efficiency and effectivity.

This chapter starts with a brief introduction to the research area, after which is elaborated on the research method and an introduction to the case studies.

## 3.1 Research Area

A total of ten large inspectorates exist in The Netherlands (Rijksoverheid, 2019), of which the NVWA is the largest agency with about 2600 employees (van Lint, 2018). All agencies are responsible for the enforcement and the regulation of laws, but differ in size, inspection domain, and area of expertise. As a result, the organisational structure and the way inspections are executed varies among the agencies. The NVWA is not only the largest inspection agency, but is also characterised by its broad range of inspection domains. Whereas most agencies have a very specific domain, the NVWA is responsible for safeguarding the safety of food and consumer products, the health of plants and animals, the wellbeing of animals, and the enforcement of the nature laws (NVWA, 2018).These are reflected in 22 different domains, of which the ten most important domains and their relative deployment of inspection capacity are represented in Figure 7.

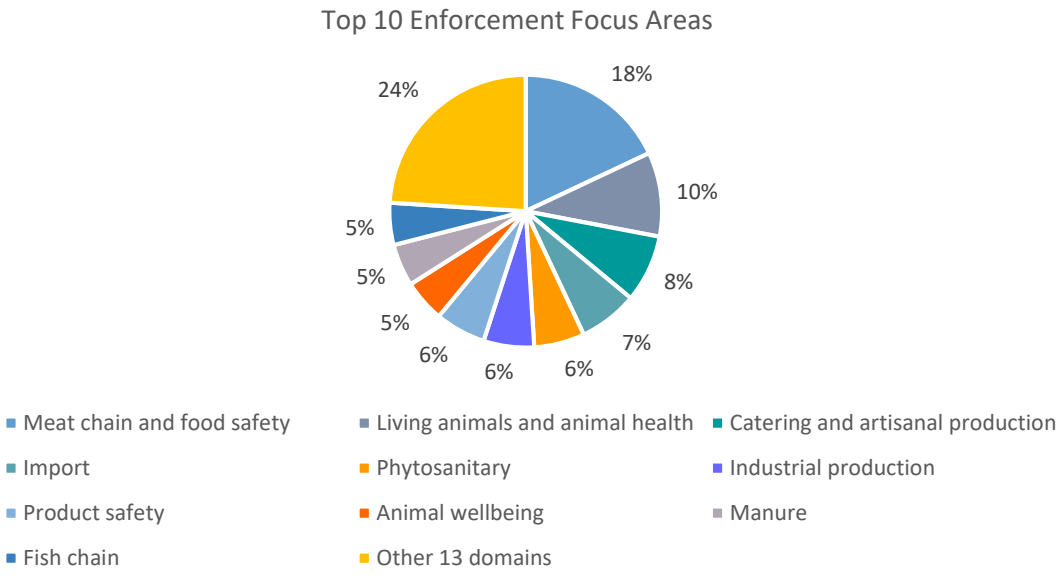



Figure 7 NVWA Enforcement focus areas

The NVWA is overseen by the ministry of Agriculture, Nature and Food Quality and spends about one million hours on inspection each year (NVWA, 2017a). The NVWA organises her enforcement activities based on three intertwined levels: strategic level, tactical level, and operational level. On strategic level is decided what is to be done (i.e., focussing on the highest risks), on tactical level how it needs to be executed (i.e., what instrument is used) and the operational level is in charge of the actual execution of enforcement activities (NVWA, 2016).

In the age of digitalisation, all inspectorates face the same challenge: using data to improve inspection efficiency and effectivity. The current economic climate has forced public supervision to reduce

spending and get by on fewer resources. The inefficient manual, expert-based, processes currently used to look for fraudulent and suspicious activities waste valuable time and resources – something inspectorates like the NVWA today simply cannot afford. Therefore, the organisation has formed a Data Science Cluster (DSC) for finding novel and innovative ways to extract value from big data with regard to that end. The Data Science Cluster is part of the Intelligence & Research department (K&O), which is located under the Office for Risk Assessment & Research (BuRO), an independent department of the NVWA which provides both solicited and unsolicited advices related to the different domains which are part of the organisation's responsibility. The Office for Risk Assessment & Research is in turn part of the organisation's Strategy division.

## 3.2 Research Method: Case Study

In this research it is hypothesised that the value created by SNA as a fraud detection method is influenced by various factors. Since the overall objective is to explore how the big data activity chain influences the added value of a data-driven approach to improve fraud detection and inspection, a case study approach is selected. A case-study approach is useful to get a close and in-depth understanding of a small number of cases, set in their real-world context (Yin, 2012). The motivation behind this research is to facilitate organisational learning in data-driven inspection and fraud detection. This organisational learning is based on knowledge of how the decision-making on big data eventually impacts the final outcome.

### 3.2.1 Case Selection

According to the case study design matrix presented by Yin (2012), a holistic multiple case study approach has been used to analyse what decisions are to be made in creating an SNA. The rationale for using more than one case is that evidence from multiple cases is often stronger, and the overall study can, therefore, be regarded as being more robust (Herriott & Firestone, 1983). In addition, it allows for comparison and understanding differences and similarities. On the downside, a multiple-case design is usually more difficult to implement compared to a single-case design and may require more resources (Yin, 2012). The number of cases that can reveal considerations influencing the quality of data-driven inspection is, however, limited. For that reason, the accessibility was the main criteria for case selection and identifying the kind of empirical data that is necessary to make convincing arguments. Nevertheless, the case selection has to be justified and considered carefully (Blatter, 2014).

This research considered two specific cases in which SNA has been applied. Given the amount of time and resources of both the researcher and the organisation, increasing the number of cases would not have been feasible within the current timeframe. Adding more cases might have reduced the possibilities and probabilities to find convincing answers for each individual case (Blatter, 2014). Besides the pragmatic considerations of time, access, and expertise, there are also methodological justifications for preferring the two selected cases. In general, when selecting cases one desires (i) a representative sample that allows to develop rigorous and detailed explanations of how the cases relate to the others in a broader universe and (ii) useful variation on the dimensions of theoretical interest (Seawright & Gerring, 2008).

In both cases, the unit of analysis is the big data activity chain process. The research aims to improve the understanding on how big data analysis, or in particular SNA, can help decision-makers, or in particular inspectors, and to put this understanding to use by applying network analytics to detect fraud and suspicious activities.

### 3.2.1 Case Introduction

To answer the research question, an in-depth case study has been performed at Data Science Cluster of the NVWA. The NVWA is currently changing its way of working towards a more data-driven and risk-oriented approach. Part of this plan is to apply Social Network Analysis as a means to detect fraud. SAS® Visual Investigator has been introduced in the organisation as a tool to detect fraud. SAS Visual Investigator is designed for banks and financial institutions looking for fraud and money

laundering, national security and law enforcement organisations looking for terrorism and criminal activities, legal firms conducting discovery, and hospitals and public health organizations guarding against disease outbreaks. It combines network detection and visualisation algorithms with the ability to mine massive amounts of data. Although there are many ways in which misconduct happens, the NVWA's primary intention of using SNA is to get insight into the actors pertaining a role in deliberate fraud.

There were two potential cases of interest within the NVWA: manure fraud and illegal dog trade. The reasons for choosing these specific cases were twofold: their representativeness and their diversity. Both cases present a clear potential for fraud with the purpose of unjust financial gain. The next section will elaborate on the differences. Manure is a project that focuses on organisations that commit fraud by fictitious release of manure. The manure production in the Netherlands is too large so that it becomes detrimental for the environment. About 80% of farms in the Netherlands produce more manure than they can legally use on their farm (Rijksoverheid, 2016). As a result, farmers dump their manure illegally, and the country breaks EU regulations on phosphates designed to prevent groundwater contamination, and the high levels of ammonia emissions effect the air quality. Due to the high costs that come along with manure transport and disposal, there is a strong incentive for the actors involved in the entire manure chain to commit fraud and put individual interest above societal interest. Illegal dog trade is a project that aims to get insight into the fraud network of illegal dog traders. Puppies are often imported from eastern European countries, countries with the prevalence of rabies. Rabies is a deadly disease that is transmitted when an infected animal scratches or bites an animal or human. Also saliva from an infected animal can spread rabies if the saliva comes into contact with eyes, mouth or nose.

### 3.2.3 Case Comparison

Since both cases are inherently different an extensive case comparison has been carried out. Based on the fraud definition as introduced in Section 2.4.1 the five fraud characteristics of each case are described. In addition, both cases are differentiated based on political attention, geographical concentration, stage of data-driven inspection, and the fraud triangle. The information that makes up the comparison results from reports and documentation and is written in consultation with domain experts. Table 3 summarises these key case characteristics. An elaborate case reflection can be found in Appendix B. One should note, however, that this reflection is done to create a generic idea of the context in which Social Network Analysis will be applied and is therefore not mutually exclusive. In practice, inspectors make case-by-case judgments about the reason why and how a specific violation occurred and about the motives and capabilities of the firms. In addition, the organisation's Intelligence and Investigation Division (IOD) has made 'intelligence images' which could provide additional information on the nature of fraud, but publishing this information is restricted due to its sensitive nature.

Table 3 Case Characteristics

| | **Manure** | **Illegal Dog Trade** |
|---|---|---|
| **Political attention** | Top-down | Bottom-up and top-down |
| **Size of fraud and the fraud triangle** | An estimated 25-40% | An estimated 25% |
| *Pressure* | Financial (survival) | Financial (making money) |
| *Opportunity* | - Low inspection rate<br>- Low perception of getting caught<br>- Open-ended lawsuits | - Low inspection rate<br>- Low perception of getting caught<br>- Complex organisation of enforcement activities |
| *Rationalism* | - Industry standard<br>- Cost-benefit analysis | - Little empathy for animal suffering |
| **Types and variations of fraud** | E.g. cheating with the weights and volumes of transported and stored manure, the nitrate and phosphate content, the VDMs (transport licenses) whether or not combined with the AGR/GPS systems, or the declaration of "non-existing" pieces of land | E.g. cheating with the trade recognition, health certificate, vaccinations, the age of the puppies, identification document, the chip, the registration and the professional skills of the trader, and registration of dogs |
| **Uncommon** | No | No |
| **Intentional** | Mostly; amoral calculator, political citizen, and organisational incompetence. | Yes, amoral calculator |
| **Concealed** | Yes | Yes |
| **Time-evolving** | Yes | Yes |
| **Individual vs. Organised** | Organised | Organised |
| **Stage of data-driven inspection** | Notification-driven and outlier detection | Mainly notification driven |
| **Goals of SNA** | Identification of different roles | Identification of the highest risks |
| **Geographical concentration** | Oost-Brabant / Noord-Limburg | Entire country (import of dogs) |
| **Stakeholders** | Farmers, licensing authority, agricultural advisor, lawyers, accountants, equipment supplier, government, Wageningen University and Research (WUR), social connections (friends and families), veterinarians, landlords, garages, independent samplers, labs, intermediaries (transport companies), truck manufacturers, contractors, advisors, inspection agencies, grant providers, financial institutions, equipment supplier | Breeder, intermediary ("tussenhandelaar"), transporter, dog trader, passport producer, veterinarians, chipper, chip producer, buyer |

## 3.2 Action Research

The objective of this study was to develop an overall understanding of the decisions that influence data-based value creation in the public domain. This research did not only aim at exploring how the big activity data chain impacts the value created by the SNA, but the researcher was also actively involved as a participant in the implementation of SNA. The application of SNA was for a large extent new to the organisation. Together with the project team, the researcher was engaged in exploring the possibilities of SNA. For that reason, the research can be considered as action research. Action research is "an orientation to knowledge creation that arises in a context of practice and requires researchers to work with practitioners" (Bradbury-Huang, 2010). This means that the action researcher is not an independent observer, but becomes a participant (Benbasat, Goldstein, & Mead,

1987). Thus, the researcher has two objectives: taking action to solve a problem and contributing to science in the development of concepts (Benbasat et al., 1987). This specific research method "is unique in the way it associates research and practice, so research informs practice and practice informs research synergistically" (Avison et al., 1999). As a participant, the researcher fulfilled two major roles in this research: executing SNA and applying reflective practice. Reflective practice is the ability to reflect on actions so as to engage in a process of continuous learning (Schön, 1983). It involves "paying critical attention to the practical values and theories which inform everyday actions and decisions, by examining practice reflectively and reflexively. This leads to developmental insight"(G. Bolton, 2010). A key rationale for reflective practice is that experience alone does not necessarily lead to learning; deliberate reflection on experience is essential (Loughran, 2002).

The research objective was achievable using action research for two reasons. First of all, action research involves introducing organisational change (Shani & Pasmore, 1982), which was the main goal of the study (i.e., using data to fight fraud and improve inspection efficiency and effectivity). Secondly, action research aims to develop a holistic understanding (Coughlan & Coghlan, 2002), which was achieved by reflective practice.

In practice, studies that involve action research are embedded in a specific context (C. Lim et al., 2018). Like any other research methods, this has its strengths and weaknesses. The strength is the in-depth understanding the researcher obtains. In contrast, a weakness comes from the potential lack of objectivity as a result of the researcher's stake in effecting a successful outcome for the organisation. In addition, it must be remembered that data collection tools in action research are themselves interventions that generate data. Interventions may evoke feelings of like anxiety, suspicion, and empathy or create particular expectations in the organisation of interest (Coughlan & Coghlan, 2002). This may be critical information to the success of a project.

While the results of action research may contribute to the available scientific knowledge, they tend to be specific and are generally used for the improvement of the target organisation (A. Lim & Chai, 2015). This makes generalisations to other situations where intervention is applied by people less knowledgeable than the researcher difficult (Benbasat et al., 1987). Therefore, clearly specifying the problem setting, like done in the Section 3.2, is crucial in action research. Indicating the contextual aspects of a study is necessary to discuss the extent to which the research findings can be transferred (Mathiassen, Chiasson, & Germonprez, 2012).

# 4. The Big Data Value Chain

In literature, several researchers have proposed different steps that make up a big data value chain. In this chapter, the data chain perspective is applied to understand how the NVWA wants to extract value from data-driven inspection (Section 4.1). Section 4.2 elaborates on relation between the NVWA and SNA. Finally, section 4.3 zooms into the R&D process of SNA, in which each activity is described in detail.

## 4.1 Data-driven Value Creation at the NVWA

The Data Science Cluster (DSC) was formed in July 2017 as a response to societal developments and the organisation's mission to work risk-oriented and data-driven. The cluster is primarily responsible for finding novel ways to work with data with the ultimate goal of improving inspection efficiency and effectivity. The DSC aims to use an agile way of working, in which ideas and organisational problems flow through the so-called 'data science funnel' (Figure 8).

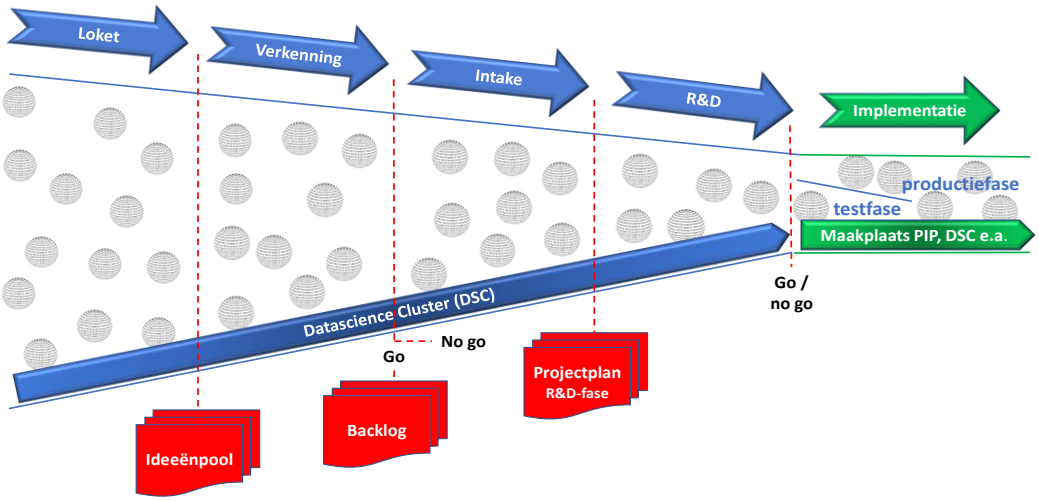



Figure 8 Data Science Funnel

**Phase 1: The counter – idea pool**

The data science funnel starts with phase 1: the counter. This is the phase of receiving, gathering, and prioritising of incoming data-related ideas and requests. These ideas can enter the counter in three possible ways:

(i) Submission via the department's counter (K&O);
(ii) Submission via the organisation-wide counter (NVWA);
(iii) Submission of ideas by the Data Science Cluster members (as a result of consultation with the business or own ideas).

**Phase 2: Exploration**

The incoming request are qualified according to three criteria: innovativeness, complexity (solving the problem requires advanced data analytics), and urgency. This results in three possible outcomes:

(i) The request is *definitely not* part of the cluster's responsibility. This is fed back to the requester and if possible the right department or person is referred to.;
(ii) The request *belongs* to the cluster. The project is transferred to the department's backlog.;
(iii) The request *potentially belongs* to the cluster. This means that there is not enough information available yet to make a decision. In cooperation with the requester the proposal is explored, evaluated, and assigned to one of the two previous categories.

**Phase 3: Intake – Backlog Prioritisation and Project Plan**

The third phase is called 'intake' and consists of prioritising the backlog and making a project plan.

The backlog is made up of projects that have substantial size and meet the DSC project criteria as mentioned in the first phase. The demand for data science expertise exceeds the available capacity, which makes prioritising necessary. Prioritising occurs monthly, once the employees' occupation is evaluated. After that, project plans are created, starting with the projects given the highest priority. This occurs in agreement with the requester, which is also commonly named as 'the client' or 'the business'. After the content check, the project plan 'version 0.95' is submitted to the head of K&O department and project plan 'version 0.99' is submitted to the client for signature.

**Phase 4: R&D**

The R&D phase is characterised by active exploration of new data tools and methods. In other words, the DSC experiments with what is possible, what data are needed, and how it works. Skills and competences may already be available in the DSC or be acquired through training or external expertise. During the R&D phase, the project is structured according to the CRISP-DM (Cross-Industry Process for Data Mining) model. The CRISP-DM model provides an approach to guide datamining projects and is made up of six different steps (Figure 9).

*1. Business understanding:* Before the model is constructed, the underlying problem that is to be solved should be identified and analysed.

*2. Data understanding:* Data understanding requires the DSC to think about what data is available, what the data mean and to consider whether it is possible to solve the problem with the data.

*3. Data preparation:* Data preparation concerns all the tasks to make the data ready for modelling. This means that the data need to be selected and the quality of the data need to be analysed. Often the data is not directly ready to be used for modelling, but need to be cleaned, integrated, formatted, and missing data needs to be accounted for. Data preparation is usually a time-consuming process.

*4. Modelling:* Modelling involves the actual datamining such as outlier detection, clustering, classification, regression-analysis, or as in the case of SNA the creation and analysis of networks.

*5. Evaluation:* Evaluation includes assessment of datamining results and validation of the model. One reflects on the outcomes based on the business understanding. This could lead to a new CRISP-DM cycle in which the current model is improved, to the development of a new model, or to a combination of both.

*6. Deployment:* This step is about the translation of results in concrete actions. Successful deployment often requires a clear strategy.

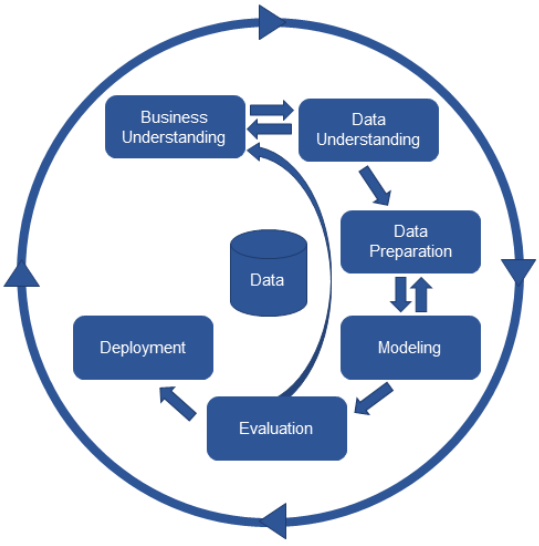



Figure 9 CRISP-DM Model (Wirth & Hipp, 2000)

## 4.2 Introduction of SNA at the NVWA

The use of Social Network Analysis as a means to detect fraud and suspicious activities and thereby improving inspection efficiency and effectivity is a subject that entered the DSC through consultation with other businesses. The Dutch Tax Authority successfully deployed SNA to uncover illegitimate activities in the context of carousel fraud and shared their experiences with the NVWA. At that moment, the organisation experienced severe difficulties with understanding the trade network of dogs and the size of fraud which could potentially be addressed with network analytics.

After a first experimentation project with SNA in the illegal dog trade project, the DSC expected that SNA could also be of added value in other chains. At that time, there was a cross-department IT project consisting of various "sub-projects" focused on improving data-driven inspection in the manure chain. For that reason, the organisation decided to start an experimental project for SNA in this chain as well. As both cases are inherently different considering the type and volume of (available) data that could construe the network, SNA in the manure chain required some experimenting first.

In the case of illegal dog trade, instead, some exploratory experimenting was already done using SAS® Enterprise Guide. In the meanwhile, the DSC decided to upgrade SAS® Enterprise Guide to SAS Viya. SAS® Viya extends the previous SAS platform by providing more advanced analytics tools and realising innovative results faster. Using network analytics is not unique to the SAS software; other programs also provide methods to analyse and link data. SAS's advantage is, however, its capability to extract data from distant or complex data sources. This data is often not held in one table, let alone one database. SAS can easily collect data from disparate data sources and integrate this data quickly (SAS, 2019). Since the data preparation and data modelling in SAS® Enterprise Guide were far from the final product that was to be realised, it was decided to redo the analysis while using the new software environment.

## 4.3 The SNA R&D Projects

As described in Section 4.1, the DSC aims to structure R&D processes according to the CRISP-DM process model. In this section, each of the process steps from business understanding to modelling will be outlined in the context of SNA, since these are the steps in which the researcher participated. The last two steps of the CRISP-DM model fall outside the scope of this study.

Within the manure domain, several IT-projects are currently under development. However, this was the first time that the DSC was deployed to work on a manure project. For the dog trade project, instead, some preliminary experimenting with SNA had already been done using the SAS® Fraud Framework. For the new project, it was decided to use the same data sources, though, redo the analysis, yet more extensive, using the SAS VA 8.3.1 environment in the modelling phase (see Section 4.2). In general, the use of SNA was fairly new for the DSC; the current DSC had neither worked with network analytics nor the software before. The preliminary SNA experiments were carried out by a single person who already left the DSC and the work was not accessible anymore due to an expired software license. In addition, the project team was not familiar yet with the data.

### 4.3.1 Business Understanding

The first two steps of the CRISP-DM model (Business Understanding and Data Understanding) were executed by means of two workshops. These workshops were mainly organised with the aim of knowledge-sharing, positioning the DSC within the organisation, and gaining commitment from the work field. This commitment is assumed to be crucial for the allocation of resources and the successful implementation of data-driven inspections in a later stage. It is important to note that these two steps were executed independently from the SNA project. In fact, their main purpose was to collect useful data sources for a diversity of data-driven projects that could contribute to effective and efficient inspection detection of fraud and suspicious activities.

*Manure Case*
The first workshop (March 2019) was organised to evaluate the project goals and to review risk indicators based on the barrier model. The barrier model (see Appendix E) visualises the main steps in the manure chain and the related stakeholders and violations that occur, and the indicators, facilitators, opportunities, and barriers to commit or prevent fraud. The group ranked all risk indicators on scale from 1-3 based on their importance. The researcher attended the workshop together with experts from various fields (8 persons): two data scientists, one inspector, one programmer, two domain experts (Expertise Dier) and two external researchers from the University of Twente (UT) who had expertise in external open data sources.

*Dog Trade Case*
For the dog trade project, the risk indicators and related data sources had already been identified in a workshop held in May 2016. The data sources have been evaluated and also the accessible documentation on the experimental project has been reviewed and discussed with a member of the project team. Besides, various additional meetings took place between the domain experts and the data scientists. Executing social network analysis was not the only ongoing project. In fact, it was part of a larger process to improve data-driven inspection within the domain.

### 4.3.2 Data Understanding

*Manure Case*
The second workshop was centralised around collecting data sources based on the risk indicators that were identified in the first workshop. The number of attendees was significantly larger compared to the first workshop. In total 24 people were present, representing the DSC, Expertise Dier, Programming, UT, RVO, and the ministry of LNV. During the workshop the group was divided into three different teams, consisting of people with various backgrounds to ensure cross-disciplinary input. Each of the groups was asked to think about different data sources that could be relevant for a particular risk indicator that was identified in the first workshop. After that, the groups were asked to select the ten most interesting data sources based on five different criteria: relevance, usefulness, availability, trustworthiness, and quality. Subsequently, these ten data sources were categorised in a relevance versus feasibility matrix. Relevance is then defined as the usefulness the extent to which it may reflect misconduct, and feasibility as the ability to get access to the data within a reasonable time-frame.

As a result of the many different potentially interesting data sources that followed from the workshop, the DSC decided to focus on the internal data first. The DSC considers internal data to be the data that internally available. This means data that is collected through own inspections as well as data that is readily available in the datamart from inspection partners such as RVO or agreements that are already in place with parties such as the KvK. The NVWA and RVO are both responsible for inspection of the manure chain. Whereas RVO focuses on administrative supervision, the NVWA executes physical inspections. An overview of the selected sources for SNA can be found in Table 4. The NVWA governs an in-house data mart in which various data sources can be found. Note that not all RVO data sources are internally available.

Table 4 Overview Data Manure Case

| Data Source | Datamart | Description | Owner | Network Component |
|---|---|---|---|---|
| **VDM** | DM_VDM_ VERVOERSBEWIJZEN | Manure transportation licenses | RVO | Link |
| **BRS** | CVU_RELATIES_99 | Business register number details | RVO | Node |
| **-** | DM_BAG | Coordinates | NVWA | Node (descriptive information) |
| **SPIN** | DM_VTE_ACTIVITEIT | Inspection data | NVWA | Node (descriptive information) |
| **Feed** | External | Annual statement on the amount of animal | RVO | Node (descriptive information) |
| **Parcel** | External | Land usage | RVO | Node (descriptive information) |
| **Administration** | External | Administration data by farmers and intermediaries related | RVO | Node (descriptive information) |

However, some of the data that has been received from RVO lacks contextual information: i.e., within the DSC, no information was available on how the data are collected, processed, and aggregated. The DSC has tabled questions to be answered by RVO, but these remained unanswered during the course of the project. Therefore, only the VDM, BRS, and SPIN data were prepared for modelling.

*Dog Trade Case*

Questions related to the data were asked to the organisation's Expertise department. During the data understanding phase, the project members also had to team-up with various domain experts. The data that was planned to be used SNA is presented in Table 5. A more elaborate description of the data sources can be found in Appendix G.

Table 5 Overview of Data Illegal Dog Trade

| Data Source | Datamart / Data file | Description | Owner | Network Component |
|---|---|---|---|---|
| **TRACES** | DM_TCS_CERTIFIATEN | Register for imported dogs | NVWA | Link |
| **I&R Dog** | DM_CCD_MELDING | Register for transportation of dogs in the Netherlands | RVO | Node |
| **UBN** | 'Lenie_pdl_huisdieren_22-11-2018.xlsx' | Unique business number for commercial animal holder (location) | RVO | Node (descriptive information) |
| **SPIN** | DM_VTE_ACTIVITEITEN | Inspection data | NVWA | Node (descriptive information) |
| **MOS** | DM_MOS_MELDINGEN | Notification system | NVWA | Node (descriptive information) |
| **Trade register** | 'Handelserkenningen overige diersoorten 20 november 2018.xlsx' | Certificate register for traders | NVWA | Node (descriptive information) |
| **KVK** | DM_NHR_ONDERNEMING | Business register | KVK | Node (descriptive information) |

Similar to the manure project, it was decided to focus on the internal data first. Although the LID possesses inspection data as well, that data was considered to be inaccessible due to the organisation's hesitance to share data with other parties. More data sources of interest were considered during the workshop, such as DIPO information, the EU dog registration, advertisements on social media, foreign number plates, and municipal licenses. However, obtaining that data also

seemed to be unfeasible due to legal and accessibility constraints and was considered to be out of scope for the first models of network analysis.

### 4.3.3 Data Preparation

During the data preparation phase, the data sets were selected and prepared for modelling. Sometimes already cleansed data was available, whereas other time raw data was provided that had to be cleaned and reformatted before it could be used for modelling and analysis. Data had to be prepared for five modelling activities that were expected to be feasible with the selected data: network visualisation, application of network metrics, addition of risk factors to the network, plotting the network on a geographical map, and tracking changes over time. Each will be elaborated on in the modelling step. All data used for network analysis was prepared with SAS Enterprise Guide 7.1. The data preparation can be found in Appendix I and Appendix II (separate files).

Two additional comments need to be made. First of all, the data only shows the registered transactions, which does not have to be consistent with the actual transactions (i.e. some transactions occur 'below the radar'). Secondly, it should be pointed out that the projects are still in an experimental phase. This means that the data quality requirements are not as strict as in an implementation phase and that data is not cleaned according to fixed standards. Within the R&D phase, the DSC itself is in charge of preparing the data and improving the data quality towards a dataset that is suitable for analysis. Only when projects are actually implemented within the organisation, a separate data governance department will be involved to clean data according regulations and legislation.

**Manure Case**

Before the data was prepared for analysis, the nodes and links of the network had to be specified. In the manure case, it was decided to use VDM data as a basis to construe the network. The rationale for selecting VMD data as a basis for the network rests on the fact that it captures all (registered) actors involved in manure transportation, who represent a large part of the stakeholders involved in the manure problematics. The selection of VDM data will be discussed first, followed by the actual preparation of the nodes.

*(i) Selection of VDM data:*

Each manure transportation has a unique transport number that is registered in the VDM (Animal Manure Consignment Note). Besides, a VDM contains details on the location of loading (supplier) and unloading (customer) and the transporter. All of these actors have a unique BRS (Business Register System) number. For analysis, only the VDMs in which none of these numbers are missing were selected. As stated in the organisation's project plan, the focus of the network analysis is on the provinces of Noord-Brabant and Limburg (De Peel), as this is considered to be the most critical region in The Netherlands. The data for 2018 was selected and the geographical selection was made based on loading zip codes. Only the transportations with a loading and unloading site within The Netherlands have been selected. In addition, a selection was made on VDMs containing hug manure.

*(i) Node Preparation:*

The data input for modelling in SAS VA version 8.3.1 requires the specification of "source" and "target" values. For ungrouped networks, SAS VA 8.3.1 displays a node for each value of a source data item, then linking another node corresponding to the target value. One way to think of the structure of ungrouped data set is to consider records as conceptual objects that have parent associations (Overton & Zenick, n.d.). All possible nodes should exist as objects in the source column, even parent nodes that do not have any additional connections. These are referred to as terminal objects because they have no target or parent node to reference (Overton & Zenick, n.d.). Figure 10 shows an example

of ungrouped data. Records with no parent values are terminal records. To represent terminal in an ungrouped network analysis, one must add rows to the data where the terminal value is the value for the source data item and the target data item is missing (e.g. Source item 1,2E+08). These are necessary to complete an ungrouped network diagram.

| Source | Target | Frequentie | Type | Inspectie | AANTAL_INSPECTIES | AANTAL_AKKOORD | AANTAL_NIET_AKKOORD | LATITUDE_ADRES | LONGITUDE_ADRES |
|---|---|---|---|---|---|---|---|---|---|
| 2,04E+08 | 2,02E+08 | 2 | Afnemer | Geen inspectie | 0 | 0 | 0 | 51,56851 | 4,144278 |
| 2,04E+08 | 2,02E+08 | 21 | Afnemer | Akkoord | 2 | 2 | 0 | 51,26935 | 5,968617 |
| 90086791 | 2,02E+08 | 6 | Afnemer | Niet akkoord | 3 | 0 | 3 | 52,02819 | 4,199017 |
| 2,02E+08 | 2,02E+08 | 1 | Afnemer | Geen inspectie | 0 | 0 | 0 | 51,75632 | 5,728461 |
| 1,2E+08 | | 6 | Afnemer | Geen inspectie | 0 | 0 | 0 | 51,21054 | 5,821936 |

Figure 10 Ungrouped Data Structure

Each node in the network was meant to represent a BRS that is registered in a VDM. A node could either be a supplier, transporter, customer, or a combination of them (e.g. a farmer who transports is own manure). According to the data structure as displayed above, each BRS number is thus defined as source value. Once all unique BRS numbers were defined as source values, supplementary data was added to the BRS numbers. This included company details, inspection data (SPIN), the scoring on risk factors, and the company's GPS coordinates.

- *Company details:* The company details added to the nodes included the address of a BRS which came from the CVU_RELATIES_99 data mart. Redundancy and duplicates had to be removed from the data. Besides, the month of operation was added to be able to track the network dynamics.;
- *Inspection data:* In the SPIN data, only the inspections related to manure were selected and filtered on 2016-2018. For each BRS, the number of approved and/or disapproved inspections were linked to the source values.;
- *Risk factor score:* Three risk factors were selected (National Holidays, Time of Transportation, and VDM Modifications) which can all be extracted from the VDM data. The factor of National Holidays involves all transportations as established in a VDM on a Dutch national holiday in 2018. Similarly, time of transportation involves all VDMs that include transportations between 11 p.m. and 4 a.m. Lastly, a separate list was available for all VDM adjustments which could be used for the factor of VDM Modifications.
- *GPS data:* Latitude and longitude values are required to plot a node on the map. The DM_BAG datamart contains the coordinates for address. However, only the BRS numbers with a KvK number have a specified street address including number. The most accurate address for all BRS numbers is a six digit zip code (1234AB), making zip code the key for linking the coordinates. Since a zip code covers a range of coordinates the maximum coordinates were chosen for each single zip.

**Dog Trade Case**

Like all real-world data, the selected data contained errors and noise from various sources. Several data issues were encountered during data preparation, making extensive pre-processing necessary to find and remove mistakes and to create an understanding of the data. Two important issues need to be mentioned. First of all, it is only since 2014 possible to link chip numbers to intra-trade certificates. This means that only dogs imported from 2014 onwards can be used for modelling. A second issue was the data sources. There is a lot of noise in the address data and many codes lack a

specific house number. At least 70% of the data would be lost when linking this data based on addresses. Besides, about 5% of the intra-trade certificates is not linked to a chip number or contains a range of passport numbers. This is a consequence of the fact that passport numbers and chip numbers are self-reported by de dog owner. Dog owners sometimes mix up passport and chip numbers, resulting in discrepancies between them. Due to the many issues involved with linking the data, it was decided to do a first SNA on TRACES data that construes the network.

*(i) Selection of TRACES data:*
All dog imports have a unique transport number that is registered in the TRACES certificate database. In addition, the TRACES data contains details of the exporter and importer, such as name, address, postal codes, city, and country, and the type of certificate, data of import, and the species. For analysis, only the Intra-Trade certificates of 2018 were used with a valid status and the 'Carnivora' species.

*(i) Node preparation:*
As the modelling phase was also executed in SAS VA 8.3.1, a similar source-target structure was required as explained in the previous section. Each node in the network was meant to represent either an exporter or an importer based on name. Once all importers and exporters were defined as source values, supplementary node data was added. This included company details as registered in TRACES, the risk of rabies, and the entities' GPS coordinates.

- *Company details:* The company details added to the nodes included the address (street, postal code, city, and country) of the sender and receiver. The house number was not specified in many cases.;
- *Risk of rabies:* Based on the import details that are specified in TRACES, the risk of rabies could easily be linked. Latvia, Poland, and Romania are classified as risk country. All imports from risk countries and non-risk countries are assigned a 1 and 0 respectively.;
- *GPS data:* To add coordinates to the nodes, the average latitudes and longitudes for a zip code have been selected, since complete addresses were lacking.

### 4.3.4 Modelling
The modelling phase consisted of five main activities: network visualisation, application of network metrics, addition of risk factors to the network, plotting the network on a geographical map, and tracking changes over time. These activities were determined based on the researcher's expertise on SNA, the project goals, and the expected software capabilities. The methodology described below is one approach to SNA, but also other designs could be used. A description of the modelling and analysis process is given for each activity. All modelling activities were executed in SAS Viya 4.3 with VA 8.3.1

**Activity 1: Network Visualisation**
The first activity was to create a network visualisation – the most straightforward and intuitive form of network analysis. Within the network visualisation, the node size, colour, pictogram, label, and underlying node details were determined. Similarly, various properties of the links between the nodes were adjusted: the link width, direction, colour, and underlying link details. These variables are especially useful for highlighting key node characteristics and relationships between nodes using discrete or numerical values. Depending on the purpose, the specification for each of the variables may differ. The results of the network visualisation for the manure case and the dog trade case are presented in Chapter 5 and 6 respectively.

**Activity 2: Application of Network Metrics**
The second activity was to apply different metrics on the network. Disconnected Network ID was

used to find each group of connected nodes. All nodes that are connected have the same value for disconnected network ID. The following centrality metrics were available for SAS VA 8.3.1:

*Reach Centrality* ($C_R$)*:* A metric that indicates how many links away the farthest connected link is. Let $L(v, w)$ be the length of the shortest path from node $v$ to node $\boldsymbol{w}$, when $\boldsymbol{w}$ is reachable from $\boldsymbol{v}$. Reach Centrality for a node $v$ is the greatest $L(v, w)$ for any reachable $\boldsymbol{w}$. Nodes that are in the middle of the network have a smaller reach value.

$$C_R = L(v, w)_{max} \tag{1}$$

*Closeness Centrality* ($C_C$)*:* A metric that indicates how close a node is to all of its connected node. A high score indicates the actor is close to all other nodes. Let $S(\boldsymbol{v})$ be equal to the sum of $L(\boldsymbol{v}, \boldsymbol{w})$, that is, the sum of the lengths of the shortest paths from node $\boldsymbol{v}$ to all other reachable nodes. Let $Smax$ = the greatest $S(\boldsymbol{v})$.

$$C_C = \frac{S_v}{S_{max}} \tag{2}$$

*Stress Centrality:* A metric that indicates how frequently a node would be crossed when taking the shortest paths between nodes. For a node $v$ , let $Nv$ be equal to the number of times $v$ is crossed on shortest paths, even those that are multiple optimal. A high score indicates a node is a likely path for information flows.

$$C_S = \frac{N_v}{N_{max}} \tag{3}$$

*Betweenness Centrality:* A metric that indicates how often a given node is part of the shortest paths between nodes as a fraction of all of the shortest paths between each pair of nodes. Betweenness Centrality accounts for cases when there is more than one shortest path between two nodes (multiple optima). A high score indicates that is a likely path for information flows. Let $T\boldsymbol{x}, \boldsymbol{y}$ be equal to the total number of shortest paths from a node $\boldsymbol{x}$ to reachable node $\boldsymbol{y}$. Let $T\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{v})$ be the number of those paths that cross node $\boldsymbol{v}$. Therefore, the fraction of shortest paths that cross $v = T\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{v}) \,/\, T\boldsymbol{x}, \boldsymbol{y}$. Let $B\boldsymbol{v}$ = the sum of $T\boldsymbol{x}, \boldsymbol{y}(\boldsymbol{v}) \,/\, T\boldsymbol{x}, \boldsymbol{y}$ for all pairs of reachable nodes $x$ and $y$. If $Bmax$ is equal to the greatest $B\boldsymbol{v}$, then the Betweenness Centrality for node $\boldsymbol{v} = B\boldsymbol{v} \,/\, Bmax$.

$$C_B = \frac{B_v}{B_{max}} \tag{4}$$

For the manure case, all centralities were calculated for the largest cluster (group disconnected IDs). For the dog trade case, the metrics have only been applied for all network structures consisting of more than 3 nodes, given the fragmented nature of the network and the fact that all centrality metrics (Reach Centrality, Closeness Centrality, Stress Centrality, and Betweenness Centrality) rely on shortest path analysis. After that, the correlations between the metrics were determined to see to what extent the metrics would yield similar central nodes.

**Activity 3: Addition of Risk Factors**

The risk factors that had been prepared during the data preparation phase were incorporated into the network, both graphically and mathematically. For the manure case, also inspection data was included.

The graphical integration of risk factors and inspection data was done by colouring a node based on its scoring on inspections and risk factors. To be able to do so, boundaries for risky and non-risky behaviour had to be set. In the manure case, separate visualisations were made for each risk factor (National Holidays, Time of Transportation, and VDM Modifications) and inspection data. In the dog trade case, the only risk factor prepared for analysis was the risk of rabies. Three different scenarios have been applied, in which each subsequent scenario adds additional accuracy:

- *Black and white scenario:* Entities that have violated the law, or operate on suspicious days, times, or have modified VDMs are given a particular colour and nodes which do not meet these criteria are given a different colour.;
- *Traffic light scenario:* Since network entities have shown different behaviour over time, a traffic light scenario adds one additional category to the previous scenario: firms which have shown both compliant and incompliant behaviour.;
- *Compliance interval scenario:* The last scenario colours a node based on a risk factor score within a certain interval.

The main reason for differentiating between inspection data and the risk factors (RF) is that inspection data is only available for a limited number of entities. The selected risk factors, instead, are based on data that is captured in VDMs or TRACES certificates(see data preparation phase) and are thus available for every transaction (link) which in turn can be related to the nodes of the network.

To gain additional insight into how the three risk factors relate to the (central) actors in the network, the risk variables have been linked to the centrality values. All nodes are ranked on an ordinal scale based on their centrality value. In the context of efficient and effective inspection, it would be rational to focus on the central nodes first as they have the biggest influence on the network. In the manure domain, 305 unique firms (based on BRS number) have been inspected (see Appendix F) between 2016 and 2018, which corresponds to ±100 unique firms per year. For that reason, the top 100 central nodes were compared based on their centrality values and the average scoring on a variable. Since it is not realistic to carry out all inspections based on network analysis, also the results of the top 50 and top 20 most central nodes are presented. The same numbers were used for analysing the results of the dog trade case, since inspection data was not prepared for analysis. Note, however, that dog trade is a much smaller domain and a smaller selection is likely to be required.

Next to comparing the scores of the most central nodes to the full data set, the correlation between the variables and the centrality metrics was tested.

**Activity 4: Plotting on a Geographical Map**

In an attempt to plot the network on a geographical map, SAS VA 8.3.1 includes several different coordinate system configurations. As the data available in the organisation did not match one of VA's predefined geography types, the option of Custom Coordinates was used. Also when using Custom Coordinates, the coordinate space must be considered. A coordinate space is simply a grid designed to cover a specific area of the earth. It is critical that the geography variable and the dataset use the same coordinate space. This tells VA how to align the grid used by the data with the grid used by the underlying map. In order to display the data correctly on a map, these grids must align. If they do not match, the data will appear in the wrong location or may not be displayed at all. By default, VA uses the World Geodetic System (WGS84) as coordinate space.

Both cases required the use of Custom Coordinates. Within the VA Data panel, the dataset was selected and the source values (Figure 10) were used for mapping. The classification of the source values was changed from 'Category' to 'Geography' in the classification drop-down. The 'Edit Geography Item' window appeared. Custom coordinates in the 'Geography data type' dropdown was

selected. Three new dropdown lists appeared that are specific to the Custom Coordinates data type: 'Latitude (y)', 'Longitude (x)' and 'Coordinate Space'. To tell VA where to find the spatial data in the dataset, the Latitude (y) and Longitude (x) dropdown lists were used to add the source value coordinates. The 'Coordinate Space' dropdown defaults to World Geodetic System (WGS84) and is suitable for both cases. This procedure was repeated for the target values.

Radius-based selection on a geographical map was used to determine distances from a location pin. To create a radius-based selection one must click on a location pin on the map or select a location from the search results and then select geographic selection. In this case, the latter option was used. The type of selection is a circular selection based on distances in miles or kilometres. Once the type is selected, the radius could be specified and the selection can be drawn on the map.

**Activity 5: Tracking Changes over Time**

A serious drawback of tracking changes over time is that there is no pre-defined method on how to investigate the network dynamics that can be used within the SAS VA 8.3.1 software. With the data as-is, it was explored how one could get more insight into network fluctuations. In the manure case, the only variable that changed over time was the shape of the network; who does business with whom. The inspection compliance and the risk factors were kept as fixed variables, resulting from average data from 2016-2018 and 2018 respectively. Similarly, the only variable that changed over time in the dog trade network was the shape of the network; who does business with whom. The risk of rabies and company details were taken as a fixed variable as of the date of data preparation.

First of all, visualisations of the monthly networks have been made. For each month, the number of clusters (based on disconnected IDs) and the network size (the number of nodes and links of the network) have been determined. Second, the centralities of the nodes that operate in each month have been calculated. It is interesting to see whether the central nodes that pop up in a particular month are also important nodes in any other month or on a yearly basis. The centralities of the nodes in the network have been compared on a monthly basis. Depending on the month, each node may have a different centrality. For each month, all nodes are ranked based on their centrality. Again the top 100, top 50, and top 20 were used as a basis for analysis. All top 100, top 50, and top 20 nodes were aggregated to determine how many unique nodes can be differentiated. Finally, the number of central nodes that end up in the top 100 in a particular month as well as in the top 100 most central nodes on a yearly basis have been determined for each month.

# 5. Social Network Analysis Results: Manure Case

Following from the research gaps identified in the literature review, the concept of SNA has barely been researched in the context of fraud detection, let alone included in the context of food and consumer products. This chapter presents the results based on the network analysis of the manure domain executed in SAS VA version 8.3.1, within the SAS Viya 4.3 environment. For each activities presented in the previous chapter, a small synthetic dataset is used to explain the and interpret the results. After that, the results of the entire network are presented. Looking at a very simple network makes it easier to understand the network and the centrality metrics. Supporting tables and information can be found in Appendix H.

The small sample dataset consists of nodes that have been subtracted from the full network. One node was randomly selected and a selection of its first-degree and second-degree connections were also obtained randomly. The nodes still have their original firm type, but all other attributes have been made up to facilitate explanation of the results. The large dataset consists of all manure transportations that are established in VDMs in 2018, including inspection data that has been gathered during 2016-2018 from both RVO and the NVWA.

To begin, the network visualisation will be presented. After that the results of four centrality metrics (Reach, Closeness, Stress, and Betweenness) will be presented and the interpretation of those values will be discussed. The subsequent section integrates various risk factors in the network. Again, there will be elaborated on the visualisations and the metrics. The last part of this chapter presents the visualisation of the network on a geographical map and focuses on the changes of the network over time.

## 5.1 Activity 1: Network Visualisation

This section first explains the results by using a small data set, after which the results of the full data set are presented.

### 5.1.1 Sample Network Diagram

The network in its simplest form consists of nodes (BRS; Unique Business Number) and the relationships between them (VDM). In general, manure is transported from a supplier via an intermediary (transporter) to a customer. When connecting the nodes and links in a network, one could see the structure of the network, but adding labels and other properties could help to understand the relationships better. The following properties of the network were adjusted to customise the look and feel of the visualisation:

- *Node size:* The node size indicates the centrality of a particular node. Nodes with a higher centrality have a larger size. This network uses Closeness Centrality as a basis for node size.;
- *Node colour:* The node colour visualises differences in numeric or character-based values that represent intensity or call attention to specific nodes. This network uses firm type as a character-based value. As will be shown later, the node colour could also represent other node attributes.;
- *Node pictogram:* The node pictogram is a character-based property of the node that, just like the node colour, indicates the type of firm: suppliers, transporters, and customers are represented in the network.
- *Node label:* The current node label indicates the firm ID. For simplicity, the this network uses alphabetic letters to specify the node. However, whenever the network becomes too complex, one could decide not to add any node attributes.;
- *Link size*: Similar to node size, the link size is useful for visualising numeric values that represent the strength of a relationship. This network uses transportation frequency between the actors of the network as a basis for link width.;

- *Link direction*: The link direction indicates the direction of manure transportation (i.e. from supplier to customer via a transporter). This is not to say that ideas on how to commit fraud also follow a direction. Instead, this could be transferred both ways. However, for visualisation purposes it makes sense to add a direction to the graph to understand in which direction manure is transported.
- *Link colour*: No specific link property has been assigned to the link colour in this network.

### 5.1.2 Full Dataset

Figure 11 shows the entire network of suppliers, transporters (intermediaries), and customers of manure in 2018; 5.928 nodes connected by 27.332 edges. Note, however, that all self-referencing links (self-edges) have been removed from the network (total item count is 155). These are farmers who dispose their manure on their own land. Within this level of abstraction there are no multi-edges present in the network; the only way the entities are connected is through a VDM and the link between two nodes which is a frequency count of the transactions between them. Multi-edges would only be present if either all VDMs were seen as separate links or if entities were connected to each other by other means than VDMs (e.g. the same advisory firm or laboratory). The former would not be feasible considering the size and complexity of the network (frequency values of VDMs rise up to 4.306), and the latter requires additional data to supplement the network which was not available yet at the time of modelling.

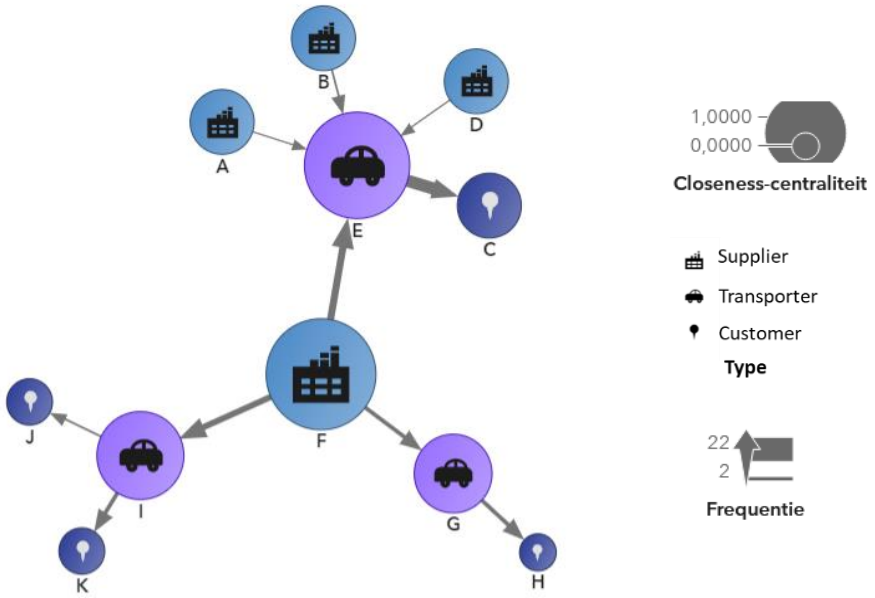



Figure 11 Visualisation of Sample Network Diagram Manure

Similar properties as for the small dataset described in Section 5.1.1 were used to adjust the visualisation. Node labels are removed from the network since this would not serve visualisation purposes. Another main difference is the grey node that pops up in the network. Whereas in the sample network dataset all nodes had a specific role (either supplier, transporter, or customer), this network shows that some nodes fulfil multiple roles at the same time, adding additional complexity to the network. This means that a node could simultaneously be a supplier and transporter, or a transporter and customer, as well as a supplier, transporter, and customer. One could for example interpret this as a farmer who transports its own manure. A node may fulfil different roles depending on the transaction that has been specified in the VDM.

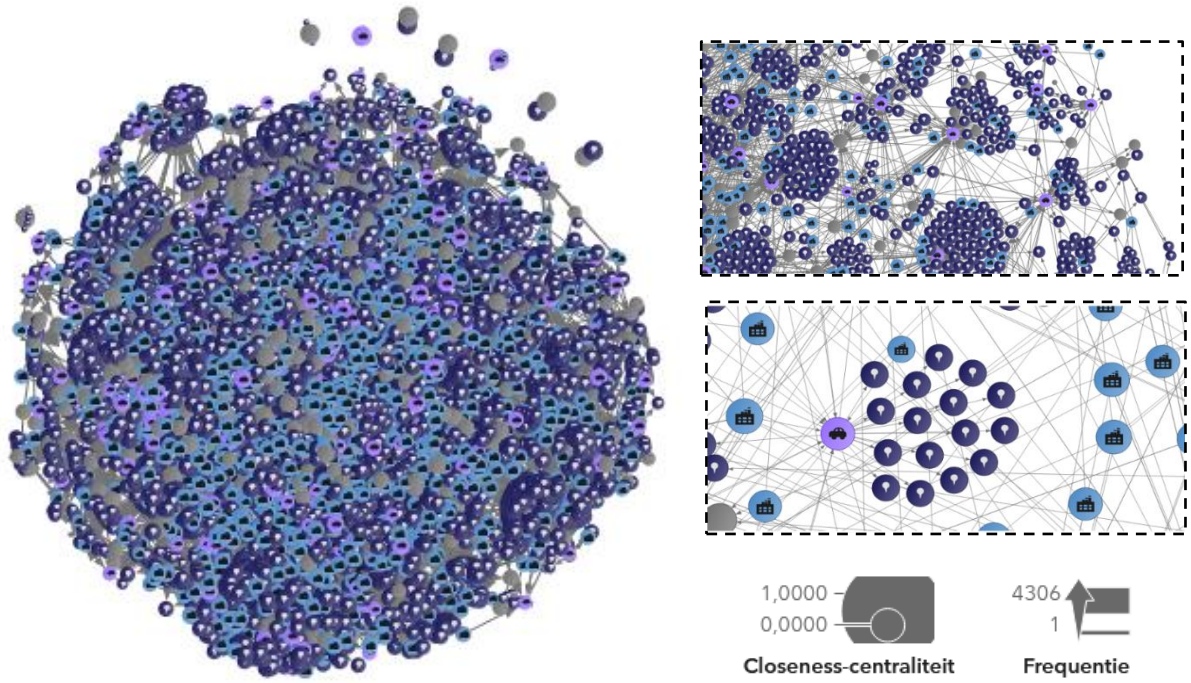



Figure 12 Visualisation of Full Manure Network

The visualisation of the entire network (Figure 12) reveals that most nodes cluster together in one network. Only 26 nodes (which can be assigned to an additional 9 disconnected clusters) were not part of this 'big network'. The size of these clusters varies from 2 to 7 nodes (in which a 2-node network consists of nodes which fulfil multiple roles at the same time as explained in the previous paragraph). The visualisation of clusters based on disconnected networks can be found in Appendix H8.

The network is quite complex and thus requires drilling into particular nodes of interest to perform further analysis. Figure 13 provides zoomed snapshots of parts of the network. When moving your cursor over the network, additional characteristics of the selected link(s) or node(s) (see Figure 13) are shown. This becomes particular interesting when more node and link attributes are added to see how these characteristics are distributed over the network. This will be elaborated on in Section 5.3. Right clicking on a node (or multiple nodes) or using the "Squared Selection" option allows one to filter the visualisation by selecting "Include Only Selection". The selected nodes can be included or excluded from the network. Once node values are filtered, the visualisation can be used to dig deeper into the selected node values.

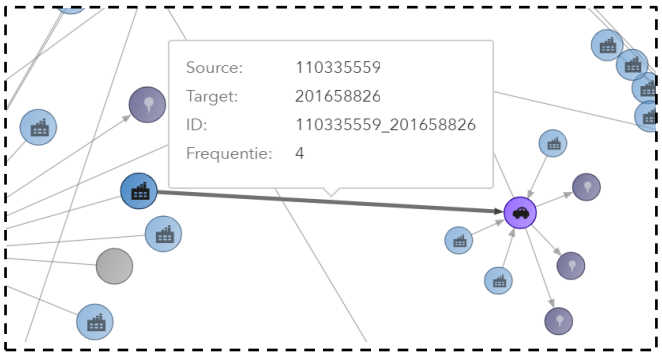



Figure 13 Detailed visualisation of the Manure Network

Network visualisation offers a powerful solution to make information hidden in networks easy to interpret and understand. Inspecting the visual representation of a network can be part of an ex-ante

approach in which the inspector is familiarised with the data which can often quickly result in some first findings and insights. Besides, the network is a useful representation to verify mathematically obtained results (i.e. network metrics). In an ex post approach, the network can be used by inspectors or intelligence offices to trace back how an event has happened and which other entities are potentially at risk.

## 5.1 Activity 2: Application of Different Network Metrics

Next to the visual representation of networks, networks can also be analysed mathematically. In this section the results of the different network metrics that have been explored are presented. This section focuses on the results and interpretation of the values. Each of the centrality metrics that has been applied (Reach, Closeness, Betweenness, and Stress) is based on the analysis of the shortest path between nodes. For the latter three algorithms, at least one node in the network has a value of one. In the model it was assumed that links are weighted all the same. Table 6 presents the values for each metric respectively based on the network constructed in Section 5.1. The meaning of each metric will be discussed briefly in the next section.

### 5.2.1 Sample Network Centrality Values

Based on the network constructed in Figure 12, the corresponding centralities of the nodes have been calculated (Table 6). Although the transportation frequency is used to indicate the link width, frequency is no variable for calculating the centralities; it is only used for visualisation purposes. The same is true for the link direction. Even though manure transportation is directed, the spread of fraudulent activities is assumed to be multi-directed.

Table 6 Centrality Values Sample Network Manure

| **Node Specification** | | | **Centralities** | | | |
|---|---|---|---|---|---|---|
| ***Item Type*** | ***Label*** | ***Firm type*** | ***Reach*** | ***Closeness*** | ***Stress*** | ***Between-ness*** |
| **Node** | A | Supplier | 4,0000 | 0,3750 | 0,0000 | 0,0000 |
| **Node** | B | Supplier | 4,0000 | 0,3750 | 0,0000 | 0,0000 |
| **Node** | C | Customer | 4,0000 | 0,3750 | 0,0000 | 0,0000 |
| **Node** | D | Supplier | 4,0000 | 0,3750 | 0,0000 | 0,0000 |
| **Node** | E | Transporter | 3,0000 | 0,9375 | 0,9677 | 0,9677 |
| **Node** | F | Supplier | 2,0000 | 1,0000 | 1,0000 | 1,0000 |
| **Node** | G | Transporter | 3,0000 | 0,5625 | 0,2903 | 0,2903 |
| **Node** | H | Customer | 4,0000 | 0,0000 | 0,0000 | 0,0000 |
| **Node** | I | Transporter | 3,0000 | 0,6875 | 0,5484 | 0,5484 |
| **Node** | J | Customer | 4,0000 | 0,1250 | 0,0000 | 0,0000 |
| **Node** | K | Customer | 4,0000 | 0,1250 | 0,0000 | 0,0000 |

- *Reach Centrality:* Reach Centrality is probably the easiest to understand; it indicates how many connections the farthest connected link is. This is not the longest possible path, instead, it longest of all shortest paths for a given node. Nodes that are in the middle of the network (i.e. higher centrality) have a smaller value. One could see in Table 6 that node F has a reach of two, and nodes E, G, and I have a reach of three. To interpret this; it takes at most two and three nodes respectively to go from those particular nodes to any other nodes in the network (see Figure 11). A reach value of one is only possible when the network consists of only two nodes.
- *Closeness Centrality:* Closeness Centrality can be explained as the mean distance of one node in a network to all other nodes in a connected network; it thus indicates how 'close' a node

is to all of its connected nodes. Based on the algorithms the smallest possible value is 0, and the maximum value is 1. Within this example, the node F has the highest Closeness Centrality.

- *Stress Centrality:* Stress Centrality indicates how many times a certain node is on the shortest path between two other nodes. When analysing the data it shows that node F has a Stress Centrality of 1,000 and node E, I, and G have a Stress Centrality of 0,9677, 0,2903 and 0,5484 respectively. In this network, all other nodes are not intermediate nodes on any shortest paths, so they have a stress value of 0. One way to interpret this network is that node F is the most frequently crossed node, or the heaviest traffic node in the network. Indeed, if a critical problem occurs with this node, then connections between other nodes would fail. Node I and G are also heavily trafficked, but not as frequently as node E. Losing one of the nodes with a stress value of 0 does not cause any loss of connectivity between other nodes.
- *Betweenness Centrality:* Betweenness is like Stress Centrality, in that it identifies the nodes that are crossed most frequently using shortest path. However, betweenness also accounts for multiple shortest paths between two nodes. With betweenness, one focuses less on the node itself, and more on the global dispersion of path options. When multiple shortest paths exists in a network, the value assigned to a crossed node is the fractional part of crossings based on the number of multiple shortest paths between two nodes. For example, there might be three ways to go from node X to Z, of which one of them is via Y. Then the fractional part of crossings of Y is 1/3. The highest betweenness scores represent nodes that are critical to lots of source-target pairs, but are not necessarily stressed the most. For this node network, there are no multiple shortest paths between nodes; this means that the values for betweenness match the values for stress. A high betweenness score indicates that nodes act as bridges between nodes or individuals, who facilitate the flow of information around the system.

As becomes clear in this example, node F has the highest centrality score for each metric. However, when datasets and complexity increase, the ranking of central nodes may vary depending on the metric used. This will be elaborated on in the next section.

### 5.2.2 Full Dataset

Similar to the small dataset, the aforementioned centralities have been calculated for all the nodes in the network. The total network consisted of 5.928 unique BRS numbers which all represent a particular node in the network of manure transportation in 2018. As explained in Section 5.1, the visualisation of the entire network (Figure 12) revealed that most nodes cluster together in one network. Next to the 'big network', 8 additional clusters can be differentiated. Although the centrality values for the nodes in these separated clusters are relatively high (the networks are small, so all other nodes in the cluster can easily be reached), the nodes in these clusters are not considered for further analysis (total item count is 28). Based on the number of nodes and the number of manure transportations their influence on the big network is assumed to be negligible. The network consists for 24% (1.432 nodes) of suppliers, for 3% (163 nodes) of transporters, for 67% (3.926 nodes) of customers and 6% (379 nodes) of entities which fulfil multiple roles at the same time.

All nodes are ranked on an ordinal scale based on their centrality value. The top 100 central nodes were compared based on their centrality values. Since it is not realistic to carry out all inspections based on network analysis, also the results of the top 50 and top 20 most central nodes are presented. Depending on the metric used, the ranking of centrality for a particular node differs. The 100 nodes with the highest centralities for each metric can be found in Appendices H1-H4. In total 192 different BRS (unique nodes) numbers can be found in any top 100, 82 in any top 50, and 38 for any top 20. Note, however, that the ordinal scale does not give any indication of the magnitude of the differences between the ranks.

- *Reach Centrality:* The Reach Centrality value in the network varies from 5 to 10. Within the top 100 nodes with the lowest Reach Centrality value (i.e. higher centrality in the network), there is only one node in the network that is just 5 steps away from any other node; all other nodes have a value of 6 (see Figure 14). This makes ranking the nodes based on Reach Centrality problematic; in total 418 nodes have a Reach Centrality of 6.
- *Closeness Centrality:* When ranking the nodes based on Closeness Centrality (and normalise the values between 0 and 1), it becomes clear that decimals become significant in assigning a position. The node with the highest closeness value also has the highest betweenness value. The centrality values in the top 100 range from 0,845 to 1,000 (in which 1,000 indicates the highest centrality).
- *Stress Centrality:* The range of centrality values in the top 100 is considerably larger compared to Closeness Centrality when assigning the importance of a node in a network based on the Stress Centrality. The values for Stress Centrality vary from 0,061 to 1,000 (in which 1,000 indicates the highest centrality) among the 100 most central nodes. Stress Centrality reaches 0,000 from rank 1.559.
- *Betweenness Centrality:* The values for Betweenness Centrality vary from 0,046 to 1,000 (in which 1,000 indicates the highest centrality). The increased complexity of the network due to the larger number of nodes and connections result in multiple shortest paths between nodes. For that reason, the values of stress do not match the values for betweenness anymore. However, when comparing the top 100 of most central nodes, only 11 nodes are unique to each ranking (i.e. they are only present in the in the top 100 of one of these metrics), in which the lowest ranked node for has a position of 211. In other words, the metrics of betweenness and stress still yield comparable results.

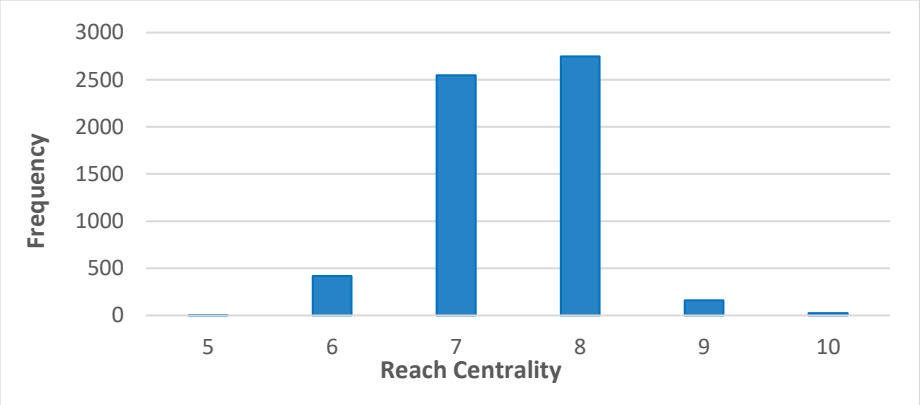



Figure 14 Frequency Distribution Reach Centrality Manure Network

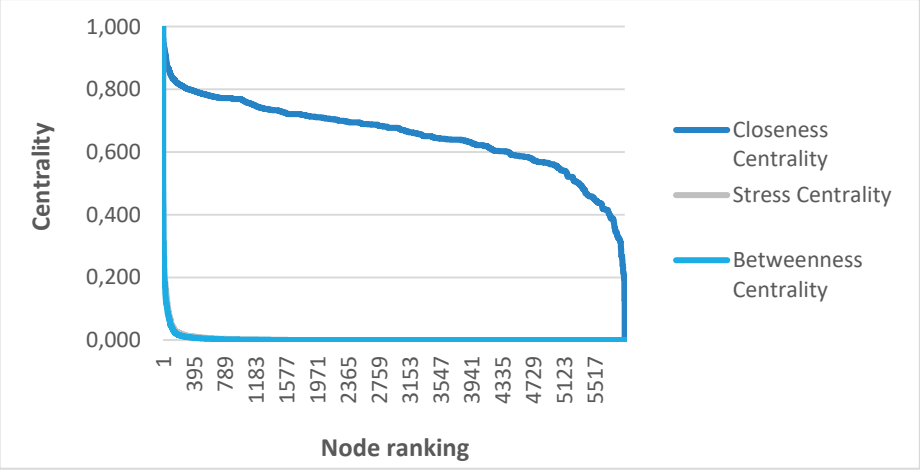



Figure 15 Centrality Distribution Manure Network. Note that Stress and Betweenness Centrality follow a similar pattern

The composition of actor types among the central nodes is presented in Table 7. The overall dispersion of node types related to the centrality ranking can be found in Appendix H.5. Although the full network consists for the largest extent of customers (67%), the nodes that fulfil a central role in the network are mainly actors who fulfil multiple roles (which have only 6% share of the full network). The composition of central nodes is clearly different from the composition of the full network.

Table 7 Composition of Node Types Manure Network

| **Number of nodes** | ***Type*** | **Reach Centrality** | **Closeness Centrality** | **Betweenness Centrality** | **Stress Centrality** |
|---|---|---|---|---|---|
| **Top 100** | *Supplier* | 42% | 13% | 4% | 4% |
| | *Transporter* | 6% | 21% | 30% | 31% |
| | *Customer* | 7% | 0% | 0% | 0% |
| | *Multiple* | 45% | 66% | 66% | 65% |
| **Top 50** | *Supplier* | 14% | 2% | 0% | 0% |
| | *Transporter* | 10% | 14% | 2% | 28% |
| | *Customer* | 2% | 0% | 0% | 0% |
| | *Multiple* | 74% | 84% | 98% | 72% |
| **Top 20** | *Supplier* | 0% | 0% | 0% | 0% |
| | *Transporter* | 15% | 15% | 5% | 35% |
| | *Customer* | 0% | 0% | 0% | 0% |
| | *Multiple* | 85% | 85% | 95% | 65% |

The centrality results suggest that if inspections would be based on centrality values, it matters which centrality metric is applied. Another way to look at this is calculating the correlation between the different metrics. Figure 16 presents an overview of how the different metrics are correlated.

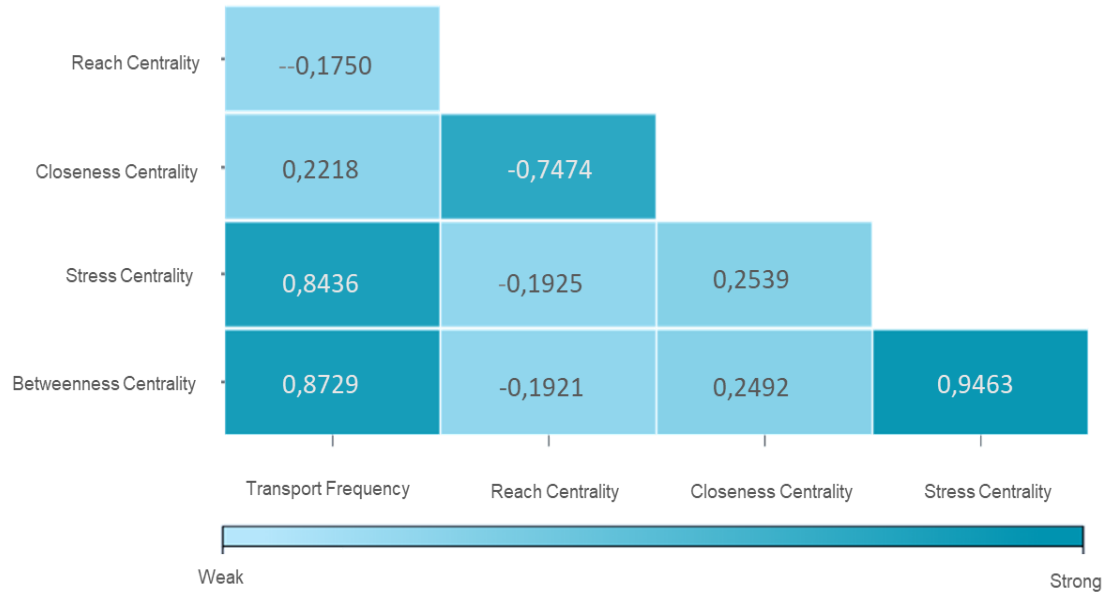



Figure 16 Correlation between Centrality Metrics Manure Network

There seems to be a strong correlation between Stress and Betweenness Centrality. This negative correlation indicates that there are many multiple shortest paths between the nodes, making destabilisation of the network difficult. In other words, a node that is important based on Stress Centrality is not considered to be important based on Betweenness Centrality; the node accounts

only for a fraction of all of the shortest paths between each pair of nodes. The moderate negative correlation between Closeness Centrality and Reach Centrality suggests that both metrics yield to a great extent similar results; a high Closeness Centrality value and a low Reach Centrality value both imply a high centrality. Please notice that centrality is not the same as transportation frequency. Consider the following example. A node which has 100 transportations to 1 single node has a lower centrality than a node with just 5 transportations to five different nodes. Yet, depending on the metric, frequency and centrality are correlated (see Figure 16).

## 5.3 Activity 3: Addition of Risk Indicators

Although the network as visualised in Section 5.1 provides insight into who does business with whom, it does not give any insight into fraud or suspicious activities. Therefore, three risk indicators have been added to see they are distributed over the network: VDM Modifications, National Holidays, and Time of Transportation. In addition, inspection data has been added to get an idea on how entities in the network are performing. The results of the inspection will be presented first and the results of the risk indicators thereafter. The analysis covers both visualisation and the application of centrality metrics in relation to the inspection data and risk indicators.

### 5.3.1 Sample Network Diagram

To begin, inspection status is used a character-based value (approved, disapproved, and no inspection), which is indicated with the node colour using a black and white scenario. All other characteristics are similar as described in Section 5.1. As both node I and J never had an approved inspection, the node is coloured red. Due to the limited inspection capacity, not all entities have been inspected (indicated in grey). One could argue that node K (not inspected) has a high propensity to be delinquent as its business partners proved to be delinquent too. When analysing this network based on visualisation, risk-oriented inspection would suggest to inspect node K.

In a similar way, inspection data can be added using a traffic light scenario or a compliance interval scenario, each adding additional accuracy to the model. The results of the sample network diagram are visualised in Appendix H.7. Instead of adding inspection data to the network, also the three risk indicators for suspicious activities can be added: VDM Modifications, National Holidays, and Time of Transportation. Since the underlying principle is the same, no further explanation is given for the sample network diagram; the results of the full dataset will directly be presented in the next section.

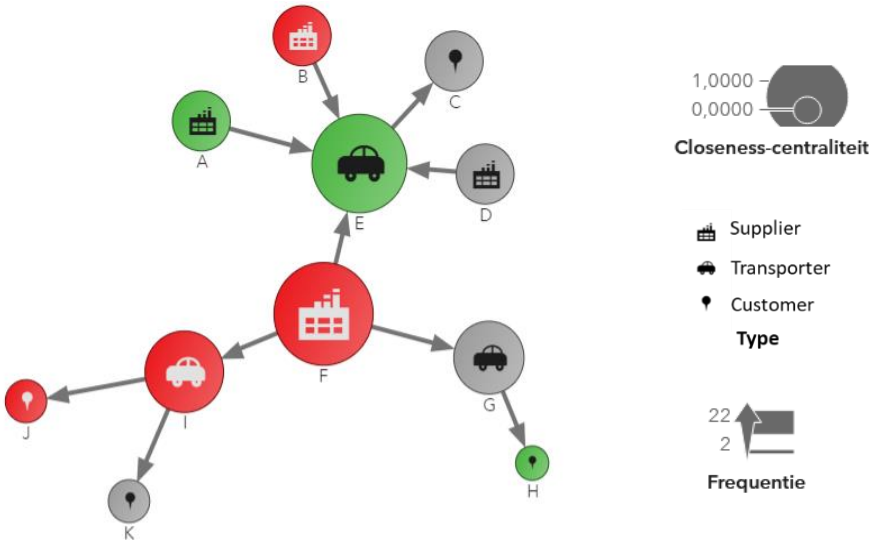



Figure 17 Sample Network Manure with Integrated Inspection Data

### 5.3.2 Full Dataset

This section presents the results of the entire network of 2018 with integrated inspection data and risk indicators, both visually and mathematically. To be able to do so, the term 'risk' had to be further specified; what does and what does not count as risk. There were neither pre-determined values known, nor did plotting compliance rate versus risk factor reveal any insights (Appendix H.9). For that reason, some assumptions had to be made; each will be elaborated on in turn.

#### Visualisation

*(i) Inspection Data*

Adding inspection data to the full dataset clearly visualises that many entities have not been inspected (Figure 18). Also clusters in the network in which every node is inspected are lacking. Figure 18 presents the network using a black and white scenario: nodes that have approved inspections are coloured green and nodes that have disapproved inspections are coloured red.

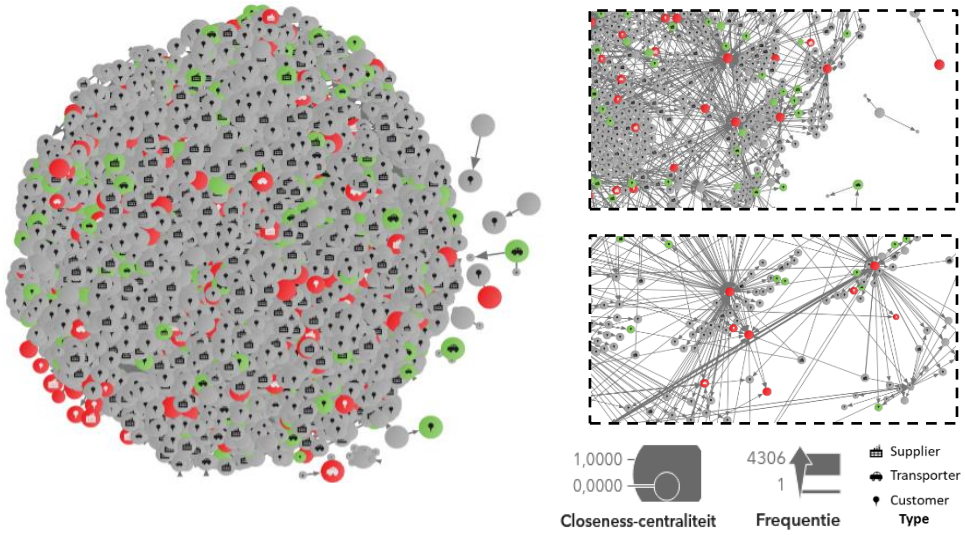



Figure 18 Network Manure with Integrated Inspection Data using a Black and White Scenario

In Figure 19, the traffic light scenario is applied. In this case, the nodes that have both approved and disapproved inspections are coloured orange. The scarcity of inspection data makes it hard to extract patterns from the network based on inspection data. Therefore, some filters on the network were applied to see what additional insights inspection data in the network could provide.

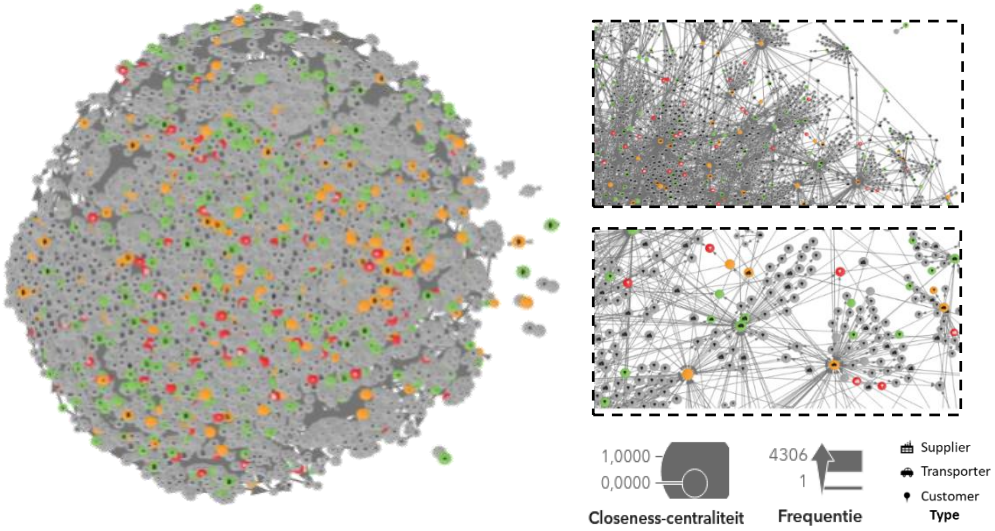



Figure 19 Full Network Manure with Integrated Inspection Data using a Traffic Light Scenario

For example, only the entities of which incompliant behaviour is known (the orange and red nodes) could be selected. This results in the following visualisation consisting of 298 nodes and 551 links distributed over 32 clusters (Figure 20). One could see that most of the nodes that have shown incompliant behaviour are connected to each other based on a shared VDM. Based on this network, new central nodes could be determined.

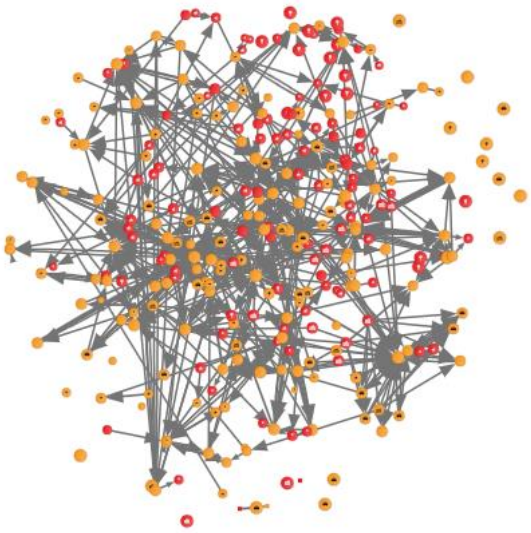


Figure 20 Network Manure of nodes with incompliant behaviour

However, there is a problem with visualising the network as-is is; many nodes fall within the same range. For example, the character of 'approved and disapproved' involves all inspections have had least one (dis)approved inspection. Regardless of the compliance rate (the number of approved inspections over the total number of inspections), all nodes that have at least one approved and disapproved inspection are coloured orange. It does not matter whether this entails a firm which has 4 out of 5 (compliance of 20%) inspections disapproved or just 1 out of 50 (compliance of 98%). Therefore, it is suggested to assign the node colour based on a certain compliance interval (e.g. compliance rate between 70-80%). Figure 21 illustrates this compliance interval scenario. This figure reveals, for instance, a triad in which the three nodes all have a low compliance rate. Further

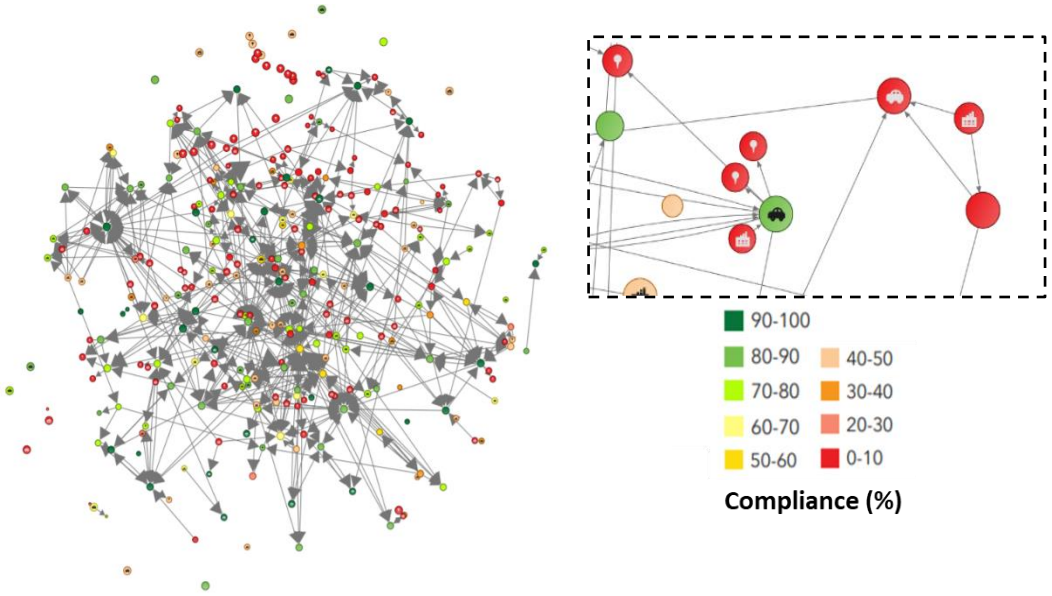



Figure 21 Network Manure with Integrated Inspection Data using Compliance Intervals

investigation should point out whether there is any similarity in the type of violations and which nodes that have not been inspected yet are likely to show identical behaviour.

*(ii) National Holidays*

Transporting manure on national holidays can be considered as a suspicious activity. Farmers are likely to be aware of the fact that no to limited inspections take place during national holidays; this could be an incentive to violate the regulation with a low chance of getting caught. For that reason, transportation of manure on national holidays can be considered as a risk indicator. All Dutch national holidays in 2018 are presented in Appendix H6. A simple calculation shows that 3% of the days in a year are national holidays. Any firm that transports more than 3% of its manure on holidays, could be considered as risky; they transport relatively more manure on holidays than any other day. In total 5,38% of the firms (319 unique BRS) meet this criterium and are visualised in Figure 22. More details on the distribution of the percentage of transportations on national holidays can be found in Appendix H.10 [2]. The network generated consists of 321 links and 156 nodes distributed over 90 disconnected entities (clusters).

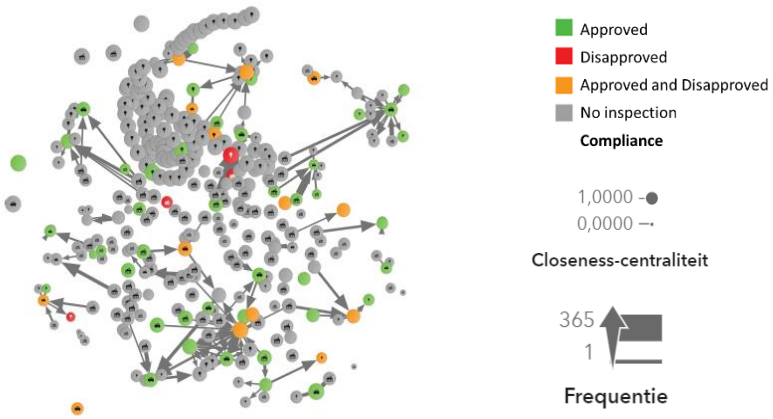



Figure 22 Manure Network National Holidays

*(iii) Time of Transportation*

Time of transportation is considered to be a risk factor based on the same rationale as transportation on holidays; there is a low perception of getting caught. Suspicious times of transportations involve all transportations between 11p.m. and 4a.m.. The risk factor is expressed as a percentage of the number of transportations that take place within this time slot over the total number of

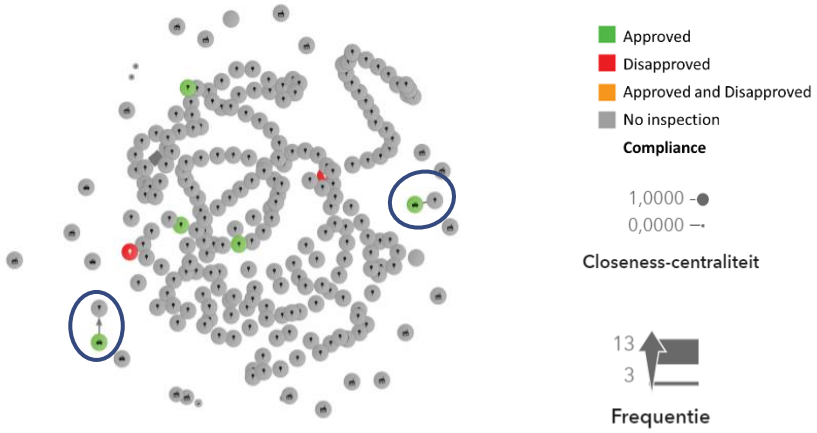



Figure 23 Manure Network Time of Transportation

[2] This table also includes data on self-referencing nodes which are not visualised in the network

transportations. Any firm that scores higher than 21% transports relatively more manure on suspicious times than other times. The frequency distribution can be found in Appendix H.11 [3]. This criterium is met by 103 firms (1,74%) and the corresponding firms are visualised in Figure 23. There seems to be little network formation based on shared VDMs; only 4 clusters can be identified.

*(iv) VDM Modifications*

The last risk factor that has been considered is VDM modifications. A VMD modification is any adjustment to the transport license. As VDMs are the basis on which the nodes are connected, any modification is attributed to both the nodes that are connected by the VDM-link. In other words, the VDM modification does not refer to the firm who modified a specific VDM, but to both entities that are established in that particular VDM.

Differentiating between what can be considered as risk and what not turned out to be particularly difficult for VDM modifications as there is no reference for normal or deviating behaviour. The frequency table in Appendix H.12 [4] shows that in total 235 firms have modified all their VDMs. Plotting all nodes reveals that there are only two pairs of connected nodes (dyads) that have adjusted 100% of their VDMs. Further inquiry shows that from a modification percentage of 78% or lower additional clusters and more complex structures arise in the network (264 nodes). Put differently, network formation between companies disappears when 78% or more VDMs are modified. Further research is required to see what the VDM adjustments exactly entail and if similar (suspicious) adjustments are being found among the connected nodes.

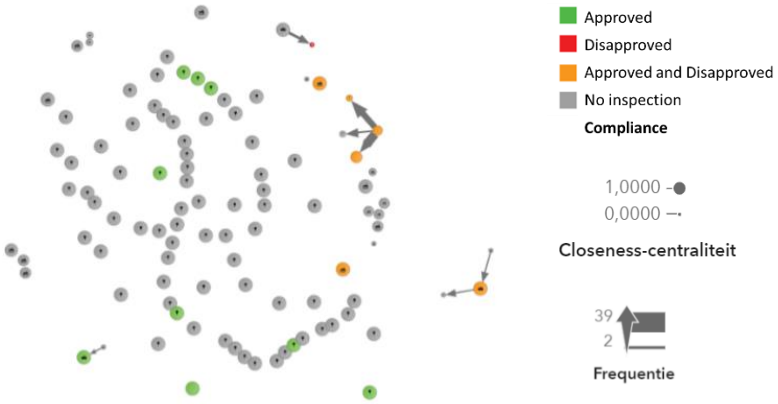



Figure 24 Manure Network VDM Modifications

**Centrality Metrics**

To gain additional insight into how the three risk factors (National Holidays, Transportation Time, and VDM modifications) and inspection data relate to the (central) actors in the network, the variables have been linked to the centrality values. Figure 25 presents how the different risk factors are correlated to each other and to the different centrality metrics. As the results show, no strong correlation can be found between the different variables. This implies that there is no clear relation observed between the centrality of a node in the network and the scoring of a node on the risk factors. As described in Section 5.2 the value of centrality and the ranking accordingly depend on the metric that is applied. Therefore, the value of each indicator is compared for each metric. Again, as the most central nodes are considered to be the most interesting in, values are compared for the top 20, 50, and 100 central nodes

[3] This table also includes data on self-referencing nodes which are not visualised in the network
[4] This table also includes data on self-referencing nodes which are not visualised in the network


Figure 25 Correlations Centrality Metrics and Risk Factors Manure

*(i) Inspection Data*

First of all, inspection data has been analysed. Appendix F provides additional information on the number of companies that have been inspected. As shown in Table 8, the number of inspections executed, the number of firms inspected, and the average (weighted) compliance differs depending on the applied metric. For example, when comparing the 100 most central nodes based on the centrality metric applied, Betweenness Centrality gives the most inspections among the central nodes. This is evenly true for the top 20 nodes with the highest centrality.

Probably more interesting is the number of firms with or without inspection. Based on Closeness Centrality, 15 out of the 100 most central nodes has no inspection history. When narrowing down to the top 50 and top 20, this number reduces to 5 and 1 respectively. Similarly, looking at Reach Centrality gives 43 firms which lack inspection data. However, as mentioned in Section 5.2, this could be explained by the fact that node 2 to 418 all have a reach value of 6. On average the inspection rate among all BRS is only 5,07%, so one could argue that the inspection rate among central nodes is significantly higher.

Looking at the average compliance of the central nodes reveals that, in general, central nodes have a higher compliance rate. This might be attributed to number of inspections that have been executed. The number of inspections executed is larger for firms which have many manure transportations (there is a moderate correlation of 0,576 between the frequency of transportations and the number of inspections) as they form a higher risk for environmental pollution. In general, these largely seem to correspond to the central nodes in the network (see previous paragraph). The more inspections have been executed, the more realistic the percentage of compliance. Several companies with lower centralities have been inspected just once. When this inspection was labelled as 'disapproved', the compliance percentage of such company is 0%. This illustrated by the increasing standard deviation when the numbers of ranked items increase.

Table 8 Inspection Analysis Manure

| | *N.o. Ranks* | **Reach** | **Closeness** | **Stress** | **Betweenness** |
|---|---|---|---|---|---|
| **No. Inspections** | *100* | 1132 | 1667 | 1681 | 1712 |
| | *50* | 1108 | 1292 | 1262 | 1283 |
| | *20* | 704 | 616 | 609 | 899 |
| **No. Inspected firms** | *100* | 57 | 85 | 93 | 92 |
| | *50* | 43 | 45 | 46 | 46 |
| | *20* | 19 | 19 | 18 | 20 |
| **No. Firms without inspection data** | *100* | 43 | 15 | 7 | 8 |
| | *50* | 7 | 5 | 4 | 4 |
| | *20* | 2 | 1 | 2 | 0 |
| **Average compliance (if inspected)** | *100* | 84,46% | 84,06% | 86,27% | 85,18% |
| | *50* | 85,99% | 84,40% | 87,48% | 88,26% |
| | *20* | 82,51% | 89,20% | 88,98% | 88,87% |
| *Standard Deviation* | *100* | 25,05% | 20,83% | 20,18% | 22,16% |
| | *50* | 20,03% | 19,64% | 19,00% | 17,58% |
| | *20* | 24,40% | 12,53% | 12,58% | 11,98% |
| **Weighted average compliance (if inspected)** | *100* | 88,43% | 88,42% | 89,35% | 89,54% |
| | *50* | 88,54% | 88,93% | 91,20% | 90,88% |
| | *20* | 90,34% | 90,10% | 91,13% | 91,66% |

*(ii) National Holidays*

The first risk factor that has been considered is National Holidays. Table 9 presents the results for the average percentage of transportations carried out during national holidays. All Dutch national holidays in 2018 are presented in Appendix H.6. On average less transportations take place during the national holidays for the selected most central nodes compared to the full network (1,28%). Similarly, there is more variation with regard to transporting manure on holidays when considering the whole data set rather than just the central nodes.

Table 9 National Holidays Analysis Manure

| | *N.o. Ranks* | **Reach** | **Closeness** | **Stress** | **Betweenness** |
|---|---|---|---|---|---|
| **Average % Transportations during Holidays** | *100* | 1,13% | 0,88% | 1,26% | 1,22% |
| | *50* | 0,96% | 0,77% | 1,09% | 0,99% |
| | *20* | 0,97% | 0,88% | 0,90% | 1,04% |
| *Standard Deviation* | *100* | 2,02% | 1,69% | 1,69% | 1,63% |
| | *50* | 1,18% | 0,99% | 1,39% | 1,10% |
| | *20* | 0,93% | 1,09% | 0,72% | 1,13% |
| *Minimum* | *100* | 0,00% | 0,00% | 0,00% | 0,00% |
| | *50* | 0,00% | 0,00% | 0,00% | 0,00% |
| | *20* | 0,00% | 0,00% | 0,00% | 0,05% |
| *Maximum* | *100* | 11,67% | 11,11% | 9,01% | 9,01% |
| | *50* | 4,66% | 4,66% | 7,05% | 4,66% |
| | *20* | 3,95% | 4,66% | 2,40% | 4,66% |

*(iii) Time of Transportation*

The second risk factor is the average percentage of transportations that has been carried out on an unusual time (see Table 10). Although the percentage of transportations on unusual times among the 50 most central nodes is on average quite similar for all metrics, this value varies considerably among the top 20 and top 100 rankings. In addition, there is no clear trend between the number of ranked nodes and the corresponding percentage of suspicious transportation times. Only when the most central nodes are compared to the full dataset (Average Time of Transportation equals 1,27%), it seems that central nodes have a higher percentage of transportations on unusual times: only the 20 most central nodes based on Stress Centrality have a lower average percentage.

Table 10 Time of Transportation Analysis Manure

| | *N.o. Ranks* | **Reach** | **Closeness** | **Stress** | **Betweenness** |
|---|---|---|---|---|---|
| **Average % of Suspicious Times** | *100* | 1,91% | 2,41% | 1,57% | 1,64% |
| | *50* | 2,23% | 2,25% | 2,10% | 2,26% |
| | *20* | 2,50% | 1,50% | 1,23% | 2,58% |
| *Standard Deviation* | *100* | 3,04% | 4,44% | 2,71% | 2,71% |
| | *50* | 3,29% | 3,26% | 3,32% | 3,33% |
| | *20* | 4,11% | 1,86% | 1,16% | 4,33% |
| *Minimum* | *100* | 0,00% | 0,00% | 0,00% | 0,00% |
| | *50* | 0,00% | 0,00% | 0,00% | 0,00% |
| | *20* | 0,00% | 0,00% | 0,00% | 0,00% |
| *Maximum* | *100* | 18,82% | 31,58% | 18,82% | 18,82% |
| | *50* | 18,82% | 18,82% | 18,82% | 18,82% |
| | *20* | 18,82% | 7,07% | 3,96% | 18,82% |

*(iv) VDM Modifications*

The last risk factor that has been considered is the modifications in VDMs. The average percentage of VMD Modifications for all nodes equals 6,97%. Based on Betweenness, Stress, and Closeness Centrality all central nodes have a lower percentage of VDM modifications. Only the top 100 central nodes based on Reach Centrality show a higher percentage of VDM modifications. However, as mentioned before, ranking based on Reach Centrality is problematic and it may therefore be more useful to compare the risk score for a particular reach rather than using rankings.

Table 11 VDM Modification Analysis Manure

| | ***N.o. Ranks*** | **Reach** | **Closeness** | **Stress** | **Betweenness** |
|---|---|---|---|---|---|
| **Average VDM** | *100* | 7,19% | 4,88% | 5,93% | 5,98% |
| | *50* | 5,18% | 5,53% | 5,33% | 5,31% |
| | *20* | 5,40% | 4,27% | 5,06% | 5,03% |
| *Standard Deviation* | *100* | 14,43% | 5,09% | 5,95% | 5,93% |
| | *50* | 4,24% | 4,20% | 4,31% | 4,26% |
| | *20* | 4,41% | 2,45% | 2,86% | 4,20% |
| *Minimum* | *100* | 0,00% | 0,00% | 0,00% | 0,00% |
| | *50* | 0,00% | 0,22% | 0,38% | 0,38% |
| | *20* | 1,66% | 0,89% | 1,66% | 1,66% |
| *Maximum* | *100* | 100,00% | 37,12% | 31,56% | 31,56% |
| | *50* | 20,58% | 21,04% | 21,04% | 21,04% |
| | *20* | 20,58% | 10,02% | 20,58% | 11,36% |

For example, in total 418 nodes have a Reach Centrality of 6 and 1 node has a Reach Centrality of 5 (the lowest reach in the network). The average percentage of VDM modifications is 6,83%, which is slightly lower than the overall network average. As can be seen in Table 11 there is a central node based on Reach Centrality (ranking 58) who has modified all of its VDMs. This node similarly gets a relatively high position based on Closeness (563), Stress (138), and Betweenness (155). Further inquiry is required to establish what exactly has been modified in those VDMs.

## 5.4 Activity 4: Plotting on a Geographic Map

It is intuitive to say that most farmers use a transporter that is most nearest to them; the further the manure is transported, the higher the costs involved. Although a network analysis is useful to see the connection between the farmers, transporters, and customers, it does not give any insight about distances; plotting the network on a geographic map helps to show the distances over which manure is transported. However, plotting the network that includes suppliers, transporters, and customers in its current format will not represent the actual distances over which manure is transported. Manure is transported from supplier to customer in which the transporter only acts as facilitator. So the actual location of the transporter does not necessarily play a role; in fact it is likely to misrepresent the actual distance over which manure is transported. This is illustrated with the sample network visualisation.


Figure 26 Farmer A supplies his manure to customer C facilitated by transporter B. It is unlikely that the manure actually passes point X in the network

Even though it does not represent the exact distance over which manure is transported, selecting a transporter which is located far away from the supplier can be considered as a suspicious. Similarly, this is likely to involve higher costs. Please note, however, that this assumption is based on domain expertise and not verified yet with hard numbers. One could see a clearer picture when using radius-based selection. From a node in the network, a geographical selection can be created. For example, if one wants the intermediary to be maximum 25 km from the supplier (Figure 27) a radius can be plotted on the map. Nodes within assigned radius can be (de)selected for further analysis. Unfortunately, there is again no convincing reference value that indicates what distances are considered to be suspicious.

The current software license is limited to creating a circular selection based on the distance in miles or kilometres. Given the size and complexity of the network and the narrow geographical scope consisting of the provinces of Noord-Brabant and Limburg results in a densely connected network. Plotting the full network on a geographical map results in too much detail on the map. Therefore, a smaller data selection is required to make sense of the data. One could for example think of a monthly transportation network, which is presented in the next section.

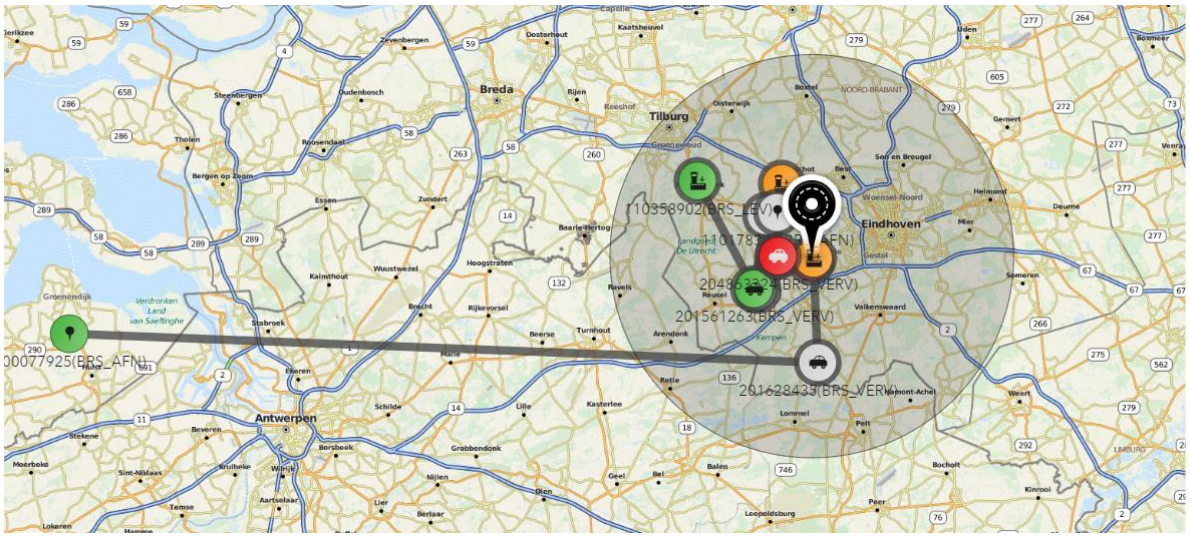



Figure 27 Radius Based Selection Manure

## 5.5 Activity 5: Tracking Changes over Time

So far, only the central nodes in 2018 have been defined. However, the registration of date and time of transportation in the VDMs provides the opportunity to track how the network changes over time. Since only data for 2018 was prepared for analysis, the analysis of changes of the network over time is based on monthly intervals of 2018.

First of all, visualisations of the monthly manure transportation networks have been made which can be found in Appendix H.14. This allows one to understand who does business with whom on a specific time or over a particular time period. Given the size of the network, the network has been plotted on a geographic map. Without this map, it would be unfeasible to see which nodes appear and disappear in the network. The number of clusters (based on disconnected IDs) and the network size (number of nodes and links of the network) of each month are presented in Table 12. The number of disconnected IDs varies from 14 clusters in March and November to 29 clusters June.

Table 12 Network Size Manure (2018)

| | **Jan** | **Feb** | **Mar** | **April** | **May** | **June** | **July** | **Aug** | **Sept** | **Oct** | **Nov** | **Dec** | **Std. Dev.** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **N.o. Clusters** | 27 | 16 | 14 | 16 | 24 | 29 | 26 | 15 | 34 | 18 | 14 | 17 | 7 |
| **Network Size** | | | | | | | | | | | | | |
| *N.o. Nodes* | 1041 | 1338 | 2056 | 2647 | 2269 | 1640 | 1551 | 2095 | 1513 | 877 | 978 | 975 | 554 |
| *N.o. Links* | 1848 | 1611 | 2511 | 3445 | 3349 | 1962 | 2140 | 2890 | 1817 | 1041 | 1196 | 1154 | 785 |

The diagram in Figure 28 illustrates that the number of nodes and links follow a similar trend; in other words, for each node added or replaced in the network one additional link is added or replaced. Whenever the line of links is steeper than the line of nodes means that links are added or removed that act as bridges between already existing nodes in the network. The number of clusters in a network on a yearly basis (9) is significantly smaller compared to clusters monthly basis. Therefore, one could argue that many nodes are not connected on monthly basis but are connected on yearly basis.

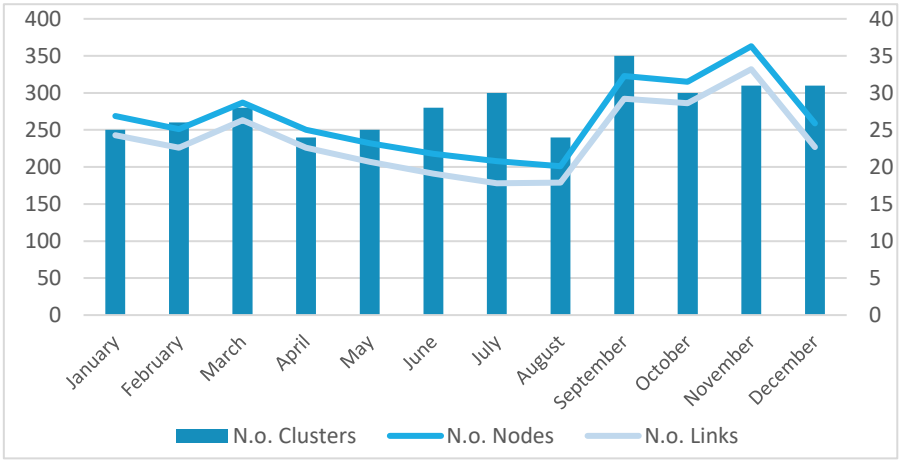



Figure 28 Graphical Representation Network Size per Month Manure (2018)

Second, the centralities of the nodes that operate in each month have been calculated. It is interesting to see whether the central nodes that pop up in a particular month are also important nodes in any other month or on yearly basis. This provides more insight into what role the selection of timeframe means for assigning centrality to a node and what this means for inspection in turn.

All centralities of the nodes in the network have been compared on monthly basis. For each month, all nodes are ranked based on their centrality. Using the same rationale as explained in Section 5.2, the top 100, top 50 and top 20 central nodes are used as a basis for analysis. For example, comparing the monthly 100 most central nodes based on Betweenness Centrality reveals 420 unique central companies on yearly basis; 246 unique central companies for the top 50 most central nodes; and 133 unique central companies for the top 20 central nodes. There are no companies which are considered to be important in each single month of 2018 based on Betweenness Centrality; 6, 3, 0 companies respectively for the top central nodes based on Closeness Centrality.

Table 13 Important Nodes per Month on Yearly Basis Manure (2018)

| | ***N.o. Ranks*** | **Closeness** | **Betweenness** |
|---|---|---|---|
| **N.o unique firms (BRS) in with high* centrality *any* month** | *100* | 472 | 420 |
| | *50* | 253 | 246 |
| | *20* | 138 | 133 |
| **N.o. unique firms (BRS) with high* centrality in *each* month** | *100* | 6 | 0 |
| | *50* | 3 | 0 |
| | *20* | 0 | 0 |

** High is defined as top 100, top 50, or top 20 respectively*

Table 14 provides further insights into the correspondence between the monthly central nodes and the yearly central nodes. It shows the number of central nodes that end up in the top 100 in a particular month as well as in the top 100 most central nodes on yearly basis. For example, in January 2018 there were 13 central nodes based on Closeness Centrality that also were considered as important over the whole year of 2018 based on Closeness Centrality.

Table 14 Number of Nodes Important in Month and Year Manure (2018)

| | **Jan** | **Feb** | **Mar** | **April** | **May** | **June** | **July** | **Aug** | **Sept** | **Oct** | **Nov** | **Dec** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Closeness Centrality** | 13 | 10 | 11 | 11 | 8 | 12 | 9 | 21 | 4 | 8 | 20 | 15 |
| **Betweenness Centrality** | 30 | 28 | 23 | 1 | 20 | 21 | 24 | 31 | 15 | 21 | 25 | 26 |

# 6. Social Network Analysis Results: Illegal Dog Trade Case

The second case that has been analysed is the Illegal Dog Trade Case. This chapter presents the results based on the network analysis executed in SAS VA version 8.3.1, within the SAS Viya 4.3 environment. The data consists of all registered imported dogs (13.925) from any country in Europe to the Netherlands in 2018. The total import from each country can be found in Appendix I.1. However, there is still a blind spot consisting of dogs which neither have a registered birth date in the Netherlands nor a registered trade number (import number). Just like in the manure case the results of the network visualisation, centrality metrics, risk indicators, geographic map and changes over time are presented. Since the metrics applied in both cases are in principle the same, no further explanation will be given on the interpretation of the values based on a small dataset. Instead, the results of the full dataset will be presented directly.

## 6.1 Activity 1: Network Visualisation

The network in its simplest form consists of nodes (sender and receiver or exporter and importer) and the connection between them based on a TRACES certificate. The labels and other properties added help to understand the relationships better. The following properties of the network were adjusted to customise the look and feel of the visualisation:

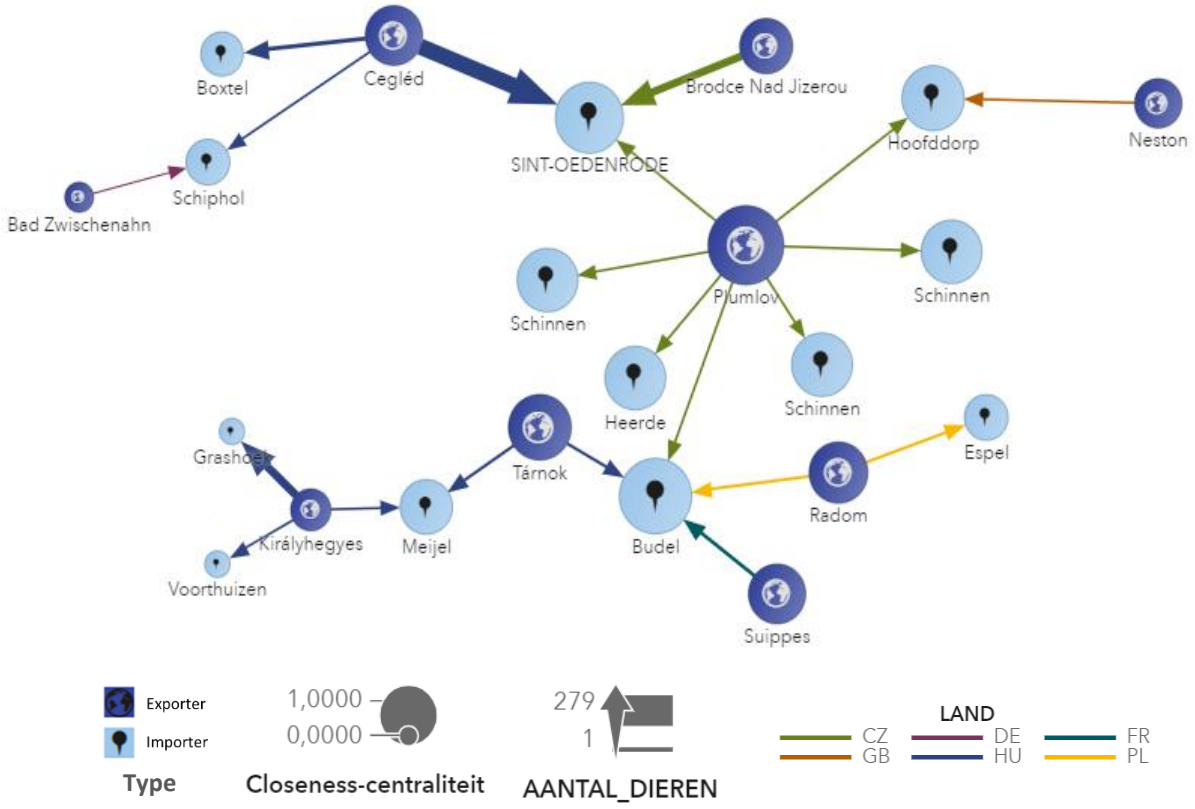



Figure 29 Sample Network Dog Trade

- *Node size:* The node size indicates the centrality of a particular node. Nodes with a higher centrality have a larger size. This network uses Closeness Centrality as a basis for node size.;
- *Node colour:* This network uses origin and destination as a character-based value.;
- *Node pictogram:* The node pictogram is a character-based property of the node that, just like the node colour indicates origin or destination.;
- *Node label:* The node label indicates the city name of origin and destination. Basically any other node attribute can be added as node description.;

- *Link size*: Similar to node size, the link size is useful for visualising numeric values that represent the strength of a relationship, in this casus the number of dogs that are imported from origin to destination.;
- *Link direction*: The link direction indicates the direction of trading transportation (i.e. from origin to destination). The link direction mainly serves visualisations proposes in understanding the trade flows of dogs.
- *Link colour*: Since the there is only a limited number of countries involved in the network, link colour can be used to assign the country of origin.

The network visualisation helps with understanding who imports dogs from whom, through which importers the various exporters in the European countries are connected, and through which exporters the various importers are connected. For example, node Tárnok exports to Meijel and Budel and acts as bridge between the two of them. Whenever any incident is detected in Meijel, Buldel might also be at risk since they both import dogs from Tárnok, which is just one of the two suppliers of dogs to Meijel. Similarly, when Tárnok leaves the network, one could expect that Meijel will import its dogs from Királyhegyes. All of this can be detected with just one glance at the network.

The full network representation of Figure 29, consisting of 2.032 nodes connected by 2.321 edges can be found in Appendix I.2. Figure 30 shows a similar network, but in this representation colour is used to indicate the different clusters (disconnected IDs). The whole network of all registered imported dogs in 2018 can be divided into 109 clusters. The nodes can be differentiated in 122 places of origin (exporters) and 2010 destinations (importers). Within this level of abstraction there are no multi-edges present in the network; the nodes are in no other way connected than via a TRACES registration link, which is a frequency count of the dogs registered in the transactions between them.

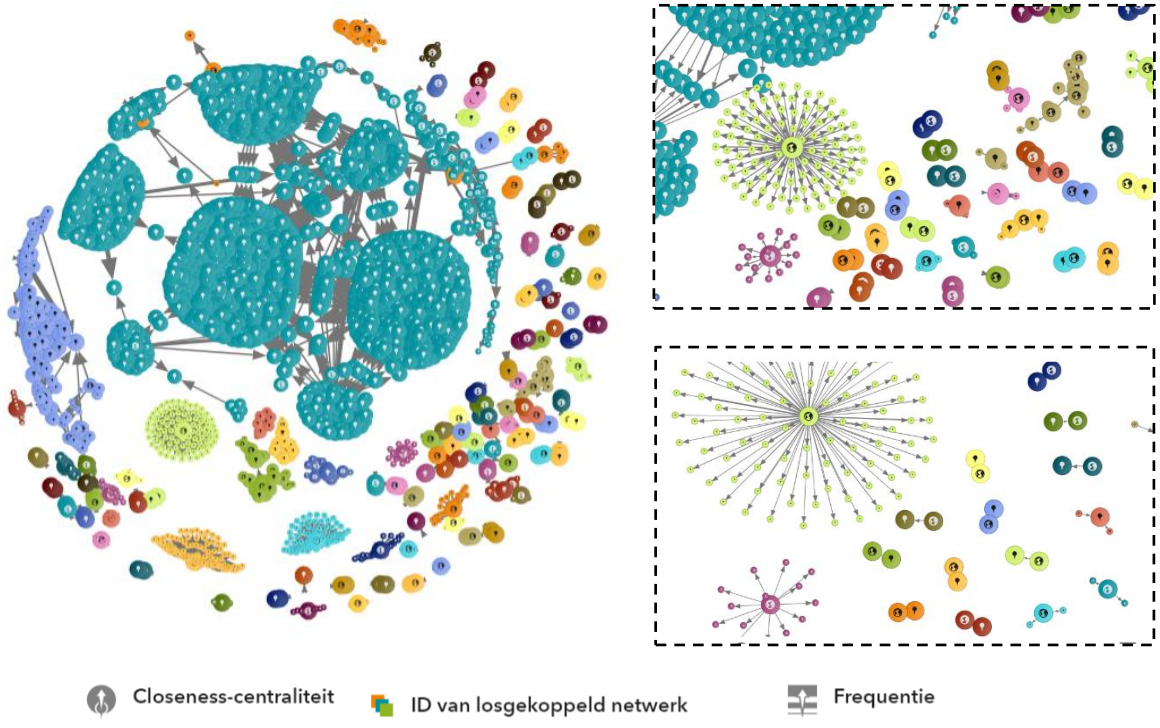



Figure 30 Full Network Visualisation Dog Trade

Given the size and complexity of the network, the node labels have been removed from the network since this would not serve visualisation purposes. Instead, the city of origin and destination is a node attribute that is visualised when clicking on a particular node. Another difference is that

representation in Figure 30 assigns the node colour not based on origin or destination, but on cluster. From Figure 30 it easily follows that the network is rather fragmented; there is one relative big cluster, several medium-size clusters and many small clusters. These small clusters are dyad or star-shaped topologies in which just one firm supplies dogs to respectively one or multiple destinations. Since the transactions visualised in the network are all import registrations from a European country to the Netherlands, a node is either a supplier or a receiver; a node cannot fulfil multiple roles at the same time like in the manure case (see Section 5.1).

## 6.2 Activity 2: Application of Different Network Metrics

The principles of the various metrics are explained in Section 5.2 based on a small sample data set. As the meaning of the metrics is not inherently different for the dog trade case, the results of the full network are presented directly. For more information on the interpretation of the values, please check Section 5.2.

The fragmented nature of the network makes it difficult to detect the relevant central nodes. In fact, since all metrics rely on shortest path analysis, each separate cluster will have a central node, regardless of the size of the cluster. Ranking all the nodes in the 2018 network based on centrality would not make sense as all the dyad structures would end up high in the list, which are assumed to have only limited influence on the other entities. For that reason, all two- or three-node clusters were filtered out of the network, and the analysis of centralities is based on the residual nodes. This boundary is, however, arbitrary.

The residual network consists of 24 different clusters, of which 6% is exporter and 94% is importer. The clusters vary in size from 4 to 1448 nodes. For further analysis, the residual nodes have been ranked based on their centrality value; a higher ranking indicates a higher centrality. Since there are only two types of entities in the network (exporters and importers), the centrality values have been compared for both of them (see Appendix I.3)

- *Reach Centrality:* The Reach Centrality value in the network varies from 1 to 14. A 10-link path is with a frequency count of 886 the most common maximum shortest path between the nodes, followed by 385 nodes which have a Reach Centrality of 2. The frequency distribution of Reach Centrality values is shown in Figure 31. The average Reach Centrality of exporters and importers is 6,943 and 8,269 respectively. There are 12 nodes in the network with a Reach Centrality of 1. This means that there are 12 nodes in a network (consisting of more than 3 nodes) who export dogs to various customers and these customers do import dogs from other exporters (i.e. disconnected IDs).
- *Closeness Centrality:* From the 2.032 nodes in the network, 1.739 nodes have a Closeness Centrality greater than 0,000. From Figure 32 follows that Closeness Centrality covers the greatest centrality dispersion over the nodes. Among the 100 most central nodes 54% is exporter and 46% is importer.
- *Stress Centrality:* Stress Centrality approaches 0,000 from node 190 onwards. Out of those 190 nodes 72 nodes (38%) are exporter. The composition of exporters and importers related to a centrality rank is visualised in Appendix I.3. For example, among the 100 nodes with the highest Stress Centrality, 54% is exporter and 46% is importer.
- *Betweenness Centrality:* As both Betweenness and Stress Centrality rely on how many times a node is crossed using shortest path, metrics approach 0,000 from node 190. This is illustrated in Figure 32; the lines cross 0,000 at the same node ranking. This means that not only the numbers of nodes with a centrality higher than 0,000 is similar, but also the composition of exporters and importers among the 190 most central nodes; 38% and 62% respectively. When looking at the top 100 most central nodes 56% is exporter, and 44% importer.

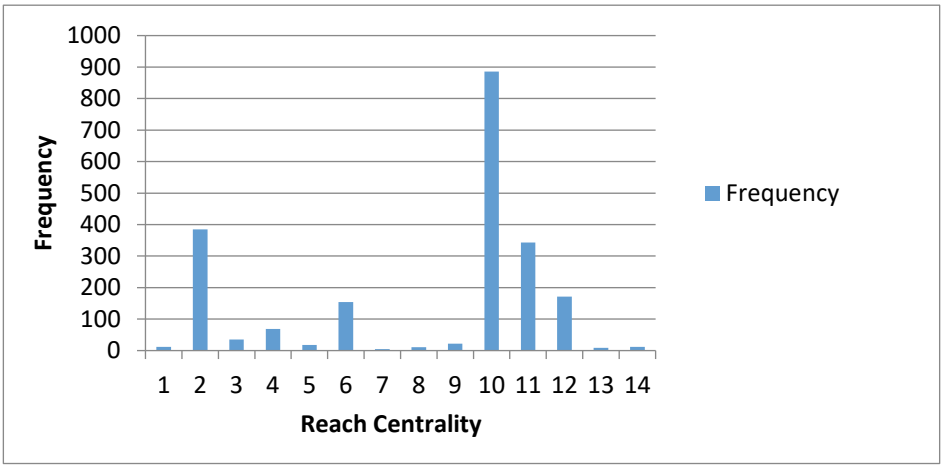



Figure 32 Frequency Distribution Reach Centrality Dog Trade

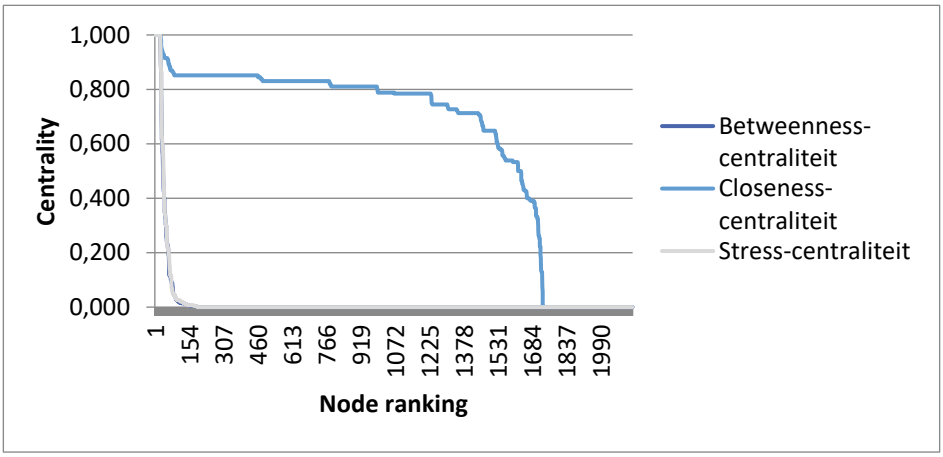



Figure 31 Centrality Distribution Dog Trade. Note that Stress and Betweenness Centrality follow a similar pattern

The strong correlation between Betweenness Centrality and Stress Centrality is also indicated in Figure 33. This means that there are not many shortest paths between the nodes, making destabilisation of the network easy. Furthermore, there seems to be a strong correlation between the values of Closeness and Reach Centrality. However, the latter one refers in fact to variation of central or important nodes based on the metrics; the nodes ranked high based on Closeness Centrality have a low Reach Centrality ranking. If Closeness Centrality and Reach Centrality would yield similar results, one would expect a strong negative correlation (see Section 5.2). This is because Closeness Centrality and Reach Centrality are interpret differently. A node with the highest Reach Centrality ranking in this dataset has a value of 1. The greater the Reach centrality value, the lower the actual centrality in the network. Closeness Centrality, instead, varies between 0,000 and 1,000 in which 1,000 indicates the highest centrality in the network. In other words, the higher the Closeness Centrality of a particular node, the more central that particular node is. The positive correlation between Reach Centrality and Closeness Centrality thus refers to a difference in assigning importance to a node in the network.

Note that centrality is not the same as trade frequency. Consider the following example. A node that trades 100 dogs with one importer has a lower centrality than a node that trades 5 dogs with five different importers. The correlations between the centrality and the trade frequency are illustrated in Figure 33.

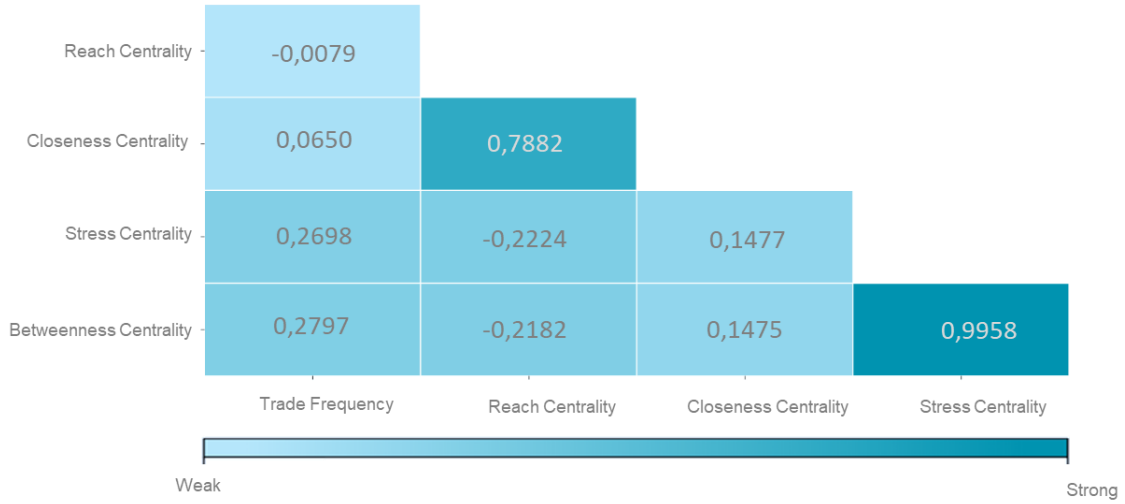



Figure 33 Correlations Centrality Metrics Dog Trade

## 6.3 Activity 3: Addition of Risk Factors

Now that the network has been visualised and the important nodes have been calculated based on centrality metrics, risk factors can be added to see and understand how they are distributed over the network. First the visualisation results will be presented, after which the risk factor is related to the centrality metrics.

### Visualisation

One of the primary risks involved with illegal dog trade is the risk of rabies. The presence or absence of rabies in domestic and wild animals is known for each country. Figure 34 illustrates how the risk of rabies is distributed over the network. The risk of rabies of a node is indicated with a particular colour: present (1), not present (0), or partly present (0 < risk value < 1). A risk value >0 means for an exporter that the country of export is a risk-country. For an importer this means that at least one TRADE certificate contains information on the import of dogs from a high risk country. A particular

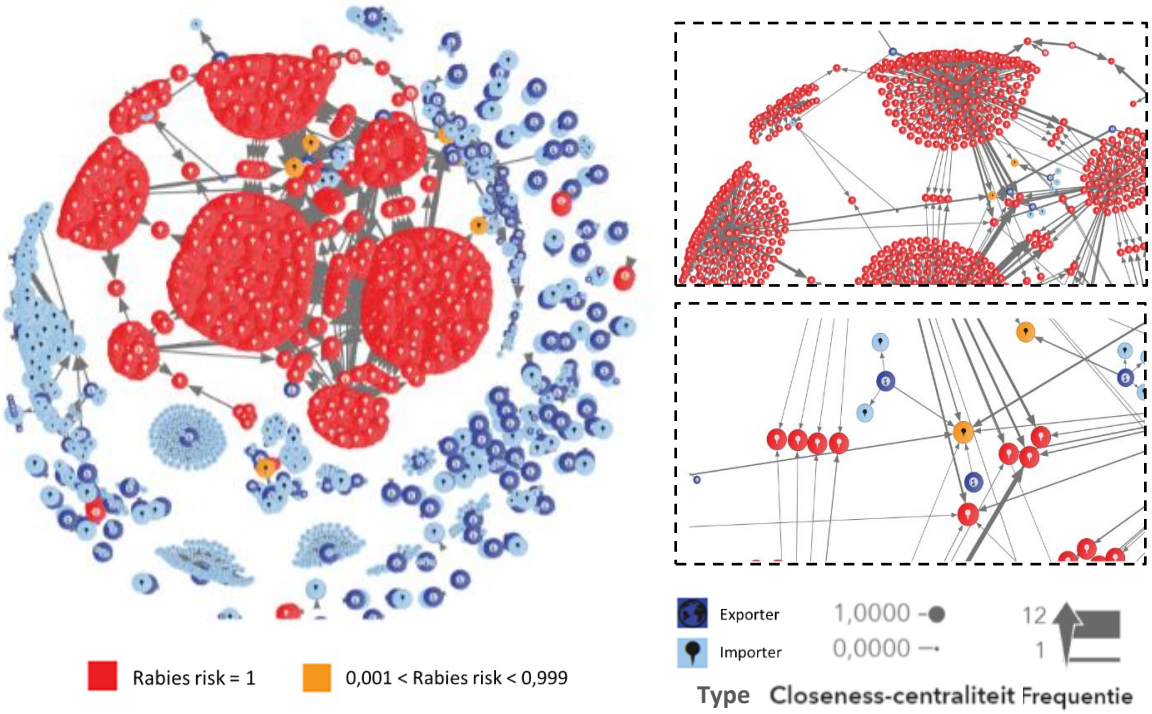



Figure 34 Risk of Rabies Dog Trade Network

node is coloured red in Figure 34 whenever all transactions involve a risk country (1). Whenever a node trades with both risk and non-risk countries the node is coloured orange (0 < risk value < 1). The risk value is then the weighted average of its risk and non-risk trades. For example, an importer who imports 30 dogs from a risk country and 10 dogs from a non-risk country has a risk value of 0,75. The exact risk value can be seen when clicking on a particular node. From Figure 31 it follows that many of the risk exporters and importers are connected to each other; there are only 6 orange nodes in the network which connect risk and non-risk countries.

The network of nodes with risk of rabies is shown in Figure 35. There are 5 different clusters based on disconnected IDs. The node colour is now used to indicate country. One could clearly see that almost all Romanian exporters are connected to each other via various Dutch importers.

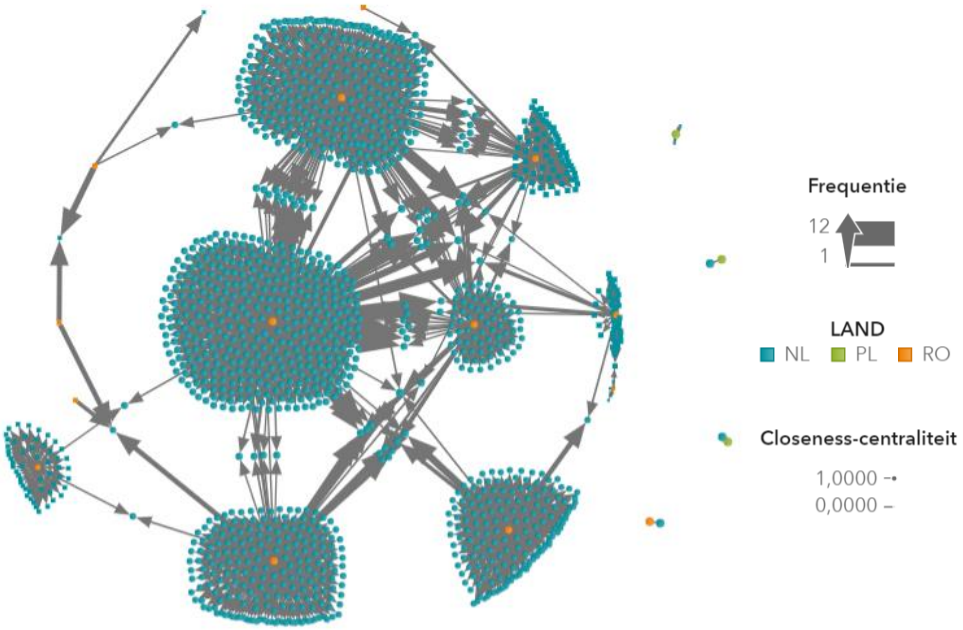



Figure 35 Network Risk Countries Risk of Rabies Dog Trade

**Centrality Metrics**

When looking how the centrality relates to the risk of rabies, the focus is on the same network as presented in Section 6.2 (i.e. excluding all 2 and 3 node network structures). Table 15 shows the number of nodes in a particular centrality ranking that have a risk of rabies. The overall distribution of the risk versus non-risk nodes over the total network is visualised in the diagram in Appendix I.4.

Remarkably is that based on Reach Centrality none of the central nodes has a risk of rabies. This is in contrast to the 70 out of the 100 central nodes with rabies risk based on Closeness Centrality. The lack of risky central nodes based on Reach Centrality may be due to the fact that ranking based on Reach Centrality is problematic. Within the top 100 central nodes, there are 12 nodes with a Reach Centrality of 1 and 88 nodes with a Reach Centrality of 2. However, the total number of nodes in the network with a Reach Centrality is 385 (see frequency diagram in Figure 31).

Table 15 Centrality Analysis Risk of Rabies

| | *N.o. Ranks* | **Reach** | **Closeness** | **Stress** | **Betweenness** |
|---|---|---|---|---|---|
| **Number of Nodes with rabies risk** | *100* | 0 | 70 | 29 | 27 |
| | *50* | 0 | 20 | 5 | 6 |
| | *20* | 0 | 0 | 0 | 1 |

A similar observation is made when looking at the correlations between the centrality metrics and the risk of rabies. The correlations are based on a centrality value and the corresponding risk of rabies, measured over the full dataset. There is a strong correlation between Closeness Centrality and the risk of rabies and Reach Centrality and the risk of rabies (Figure 36). However, the positive correlation between Reach Centrality value actually implies that the nodes with a low risk of rabies have a relatively central position in this network. The positive correlation between Closeness Centrality and the risk of rabies, instead, implies that the nodes with a high risk of rabies have a central position in this network.

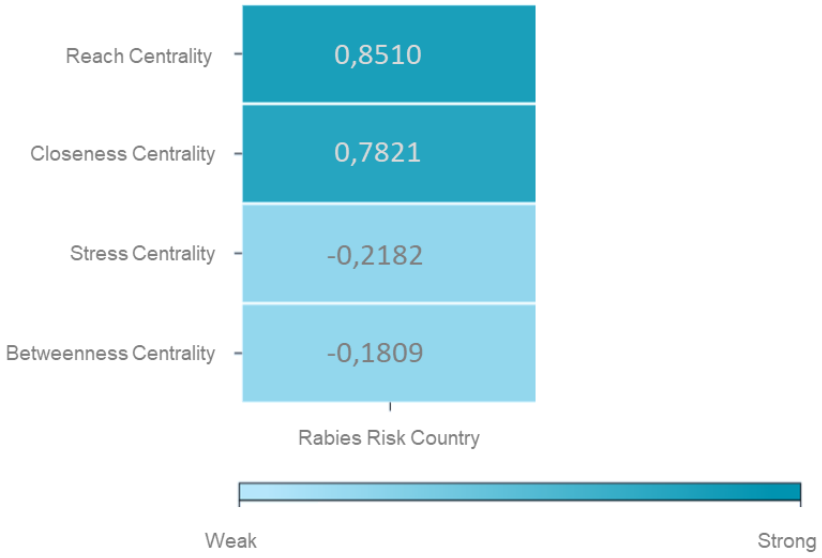



Figure 36 Correlation between Centrality and Rabies

## 6.4 Activity 4: Plotting on a Geographical Map

Plotting the network on a geographical map could provide more insight into where (registered) dog trade takes place. The address and zip codes of both the exporter and the importer are registered in the TRACES certificate which, in theory, allows one to plot the network on a geographical map. However, when looking at the actual data one could see that only the four numeric zip code digits are registered. To plot the network on a geo map, the software requires specific latitudes and longitudes of the particular location. Since a specific zip code covers a range of coordinates, the average latitude and longitude of the zip code has been used to plot the node on the map. Figure 37 shows all registered import locations in The Netherlands in January 2018. A red node indicates that the node imports all dogs from countries with a risk of rabies, an orange node indicates that the node imports some dogs from countries with a risk of rabies. Figure 34 shows, for instance, that there were no registered imports in the province of Zeeland in January 2018. Radius-based selection could be used to see how many firms in a circular selection import dogs and from which (risky) countries.

It is known that fraudsters use different names to import dogs, while they are actually just one and the same person. Radius-based selection could be used to see whether (suspicious) firms appear and disappear frequently within the same neighbourhood. Just like in the manure case, the software is limited to creating a circular selection based on the distance in miles or kilometres rather than travel distance (irregular selection based on travel distance using roads) or travel time (irregular selection based on the distance that can be travelled in the specified amount of time).

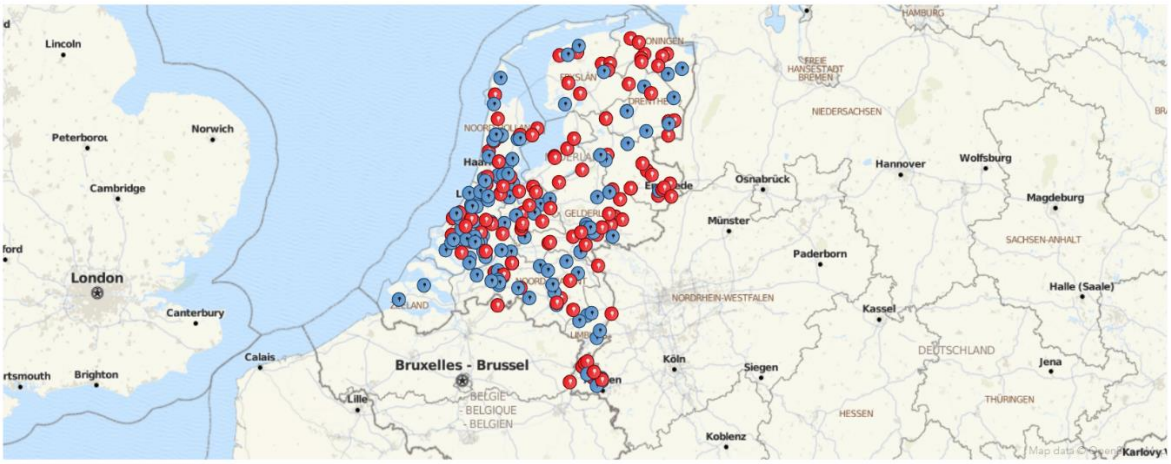



Figure 37 Registered Dog Trade Import Locations in The Netherlands in January 2018

Although the visualisation in Figure 37 is useful in quickly scanning potential risky areas, no network formation is shown yet; this requires plotting the exporters on the geographical map as well. Figure 38 shows all TRACES transactions in January 2018 between The Netherlands and Romania. The link colour in Figure 38 indicates the place of export. At one glance one could see that many dogs are imported from the same region. However, there are several issues with plotting the network in this way and the figure should be interpret with caution. First of all, adding foreign locations turned out to be particularly difficult; lacking zip codes, cities written in foreign script, and registered cities that turned out to be provinces. The poor quality and lack of foreign geographical understanding make linking node addresses to coordinates painstaking work that is prone to error. And again, the aggregated average latitude and longitude values of a particular city or region that is registered in TRACES have been used to plot the exporter on the map.

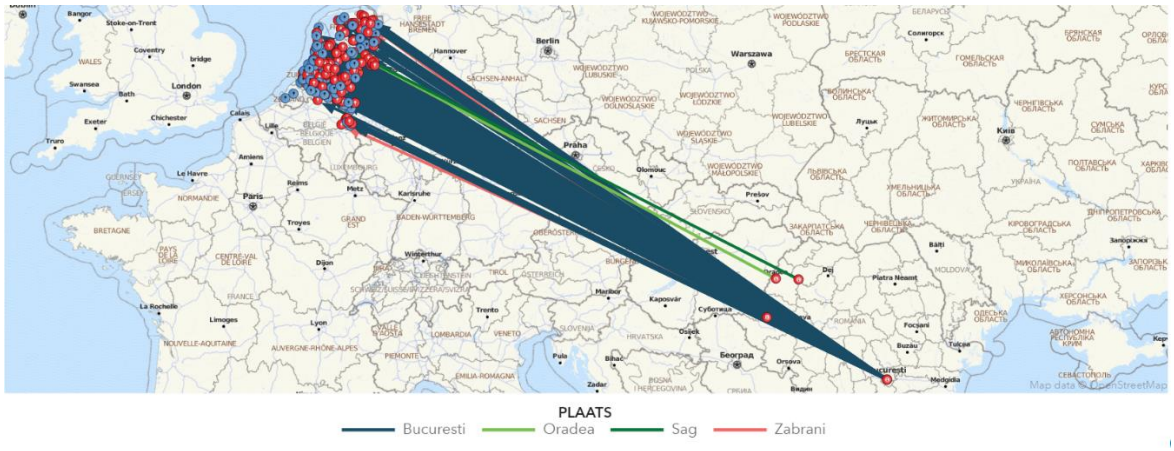



Figure 38 Doge Trade Network between Romania and The Netherlands in January 2018

Secondly, one could already see that a visualisation of just a tiny part of the whole dataset results in much detail on a map which is difficult to interpret. In Figure 38, the link colour indicates the place or origin. However, the concentrated network means that the geographical visualisation needs to be interpreted with caution.

The aforementioned issues could be solved by restructuring the network in which the link between two importers is based on a common exporter. However, this requires restructuring of the data and is considered to be out of scope.

## 6.5 Activity 5: Tracking Changes over Time

So far, only the central nodes in 2018 have been defined. However, dog trading is a dynamic process and the network is likely to vary over time. Since the import of dogs is related to a specific time which is registered in TRACES allows one to track how the network changes over time. Since only data for 2018 was prepared for analysis, the analysis of changes of the network over time is based on monthly intervals of 2018.

First of all, visualisations of the monthly manure transportation networks have been made which can be found in Appendix I.9. This allows one to understand who imports dogs from whom on a particular time or over a specific time period. The number of clusters (based on disconnected IDs) and the network size (number of nodes and links of the network) are presented in Table 16. The number of disconnected IDs varies from 24 clusters in April and August to 35 clusters September.

Table 16 Network Size Illegal Dog Trade (2018)

| | Jan | Feb | Mar | Apr | May | June | July | Aug | Sept | Oct | Nov | Dec | S.D. |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **N.o. Clusters** | 25 | 26 | 28 | 24 | 25 | 28 | 30 | 24 | 35 | 30 | 31 | 31 | 3 |
| **Network** | | | | | | | | | | | | | |
| *N.o. Nodes* | 269 | 251 | 287 | 250 | 232 | 218 | 208 | 201 | 323 | 315 | 363 | 259 | 47 |
| *N.o. Links* | 243 | 226 | 263 | 226 | 207 | 191 | 178 | 179 | 292 | 286 | 332 | 227 | 46 |

When plotting this data in a diagram Figure 39 one could see that the number of nodes and links follow a similar trend; in other words, for each node added or replaced in the network one additional link is added or replaced. If the of the number of links would have shown a steeper line than the line of nodes, this could have referred to a new exporter entering the network from which many importers, who are already present in the network, start to import their dogs. Similarly, this could refer to a new importer entering the network who imports dogs from several exporters (i.e. just one node is added linked to several existing exporters).

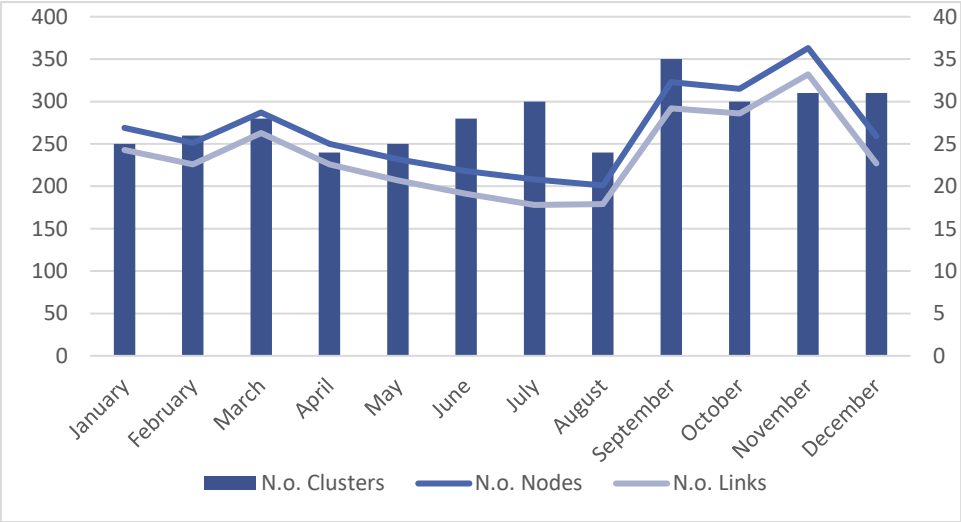



Figure 39 Graphical Representation Network Size per Month Dog Trade (2018)

The number of clusters in a network on yearly basis (24) is not significantly smaller compared to clusters monthly basis. To see whether the central nodes in each month are similar to the yearly central nodes, the centralities of the nodes that operate in each month have been calculated. This

provides more insight into what role the selection of timeframe means for assigning centrality to a node in the network.

All centralities of the nodes in the network have been compared on monthly basis. For each month, all nodes are ranked based on their centrality. Using the same rationale as explained in Section 6.2, the top 100, top 50 and top 20 central nodes are used as a basis for analysis. For example, comparing the monthly 100 most central nodes based on Betweenness Centrality reveals 701 unique central entities on yearly basis; 285 unique central entities for the top 50 most central nodes; and 81 unique central entities for the top 20 central nodes. There are 5 entities which are considered to be important (top 100), 2 (top 50), and 1 (top 20) in each single month of 2018 based on Betweenness Centrality; 2, 2, 0 entities respectively for the top central nodes based on Closeness Centrality. The 2 important entities identified in each month based on Closeness Centrality (top 100) also appear in the 5 identified important entities based on Betweenness Centrality (top 100).

Table 17 Important Nodes per Month on Yearly Basis Dog Trade (2018)

| | ***N.o. Ranks*** | **Closeness** | **Betweenness** |
|---|---|---|---|
| **N.o nodes in with high* centrality *any* month** | *100* | 652 | 701 |
| | *50* | 211 | 285 |
| | *20* | 79 | 81 |
| **N.o. nodes with high* centrality in *each* month** | *100* | 2 | 5 |
| | *50* | 2 | 2 |
| | *20* | 0 | 1 |

** High is defined as top 100, top 50, or top 20 respectively*

Table 18 provides further insights into the overlap between the monthly central nodes and the yearly central nodes. It shows the number of central nodes that end up in the top 100 in a particular month as well as in the top 100 most central nodes on yearly basis. For example, in January 2018 there were 20 central nodes based on Closeness Centrality that also were considered as important over the whole year of 2018.

Table 18 Number of Nodes Important in Month and Year Dog Trade (2018)

| | **Jan** | **Feb** | **Mar** | **April** | **May** | **June** | **July** | **Aug** | **Sept** | **Oct** | **Nov** | **Dec** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Closeness Centrality** | 20 | 19 | 18 | 12 | 16 | 14 | 13 | 17 | 13 | 12 | 13 | 18 |
| **Betweenness Centrality** | 27 | 29 | 27 | 25 | 30 | 23 | 31 | 23 | 32 | 31 | 32 | 33 |

# 7. Functional and Institutional Reflection

This research had two main purposes. First of all, this thesis aimed to explore the use of Social Network Analysis in the context of fraud detection, more specifically fraud in the context of food and consumer products. As follows from the previous chapters, SNA has the potential to identify patterns in the fraud network and in turn improve inspection efficiency and effectivity. Secondly, this research aimed to apply an institutional view to get insight into big data value creation in the public sector. This means that through the evaluation of the big data chain process this research aims to get better understanding factors that influence big data decision-making; there seems to be a gap between the promises of big data and its practical realisations. This chapter is centralised around reflection. To begin, a functional reflection will be carried out in which the various SNA activities are evaluated and differences and similarities between both cases are emphasized. The second part of this chapter focuses on a reflection from an institutional perspective. A key rationale for reflective practice is that experience alone does not necessarily lead to insight; deliberate reflection on experience is essential (Loughran, 2002). From an institutional perspective, it is assumed that actors that shape the process from data, the raw material, to decisions on the final deployment of inspection capacity based on the outcome of the network analysis. When reflecting on the previous chapters, it appears that various important assumptions and decisions have been made. Sometimes, these decisions turned out to be fundamental for the final outcome. By applying reflective practice, taken-for-granted assumptions are questioned which supports the development of insights.

## 7.1 Functional Reflection

During the research, five different activities of Social Network Analysis have been explored: network visualisation, the application of different metrics, the addition of risk factors, plotting the network on a geographical map, and tracking changes over time. The results are presented in Chapter 5 and 6. Please remember that there is no pre-defined methodology to apply SNA in the context of fraud detection, in particular not in the domain of food and consumer products. There are several ways in which SNA can be conducted. This section reflects on the steps that have been taken during this research, which was primarily exploratory in nature. Both the results and usefulness of each activity are compared for the two cases that have been analysed during this study.

### 7.1.1 Network Visualisation

Network visualisation offers a powerful solution to make information hidden in networks easy to interpret and understand. With one glance at the network one could identify who does business with whom, which entities act as bridges between two clusters, and gain insight into the overall structure of the networks (i.e. to what extent are entities connected). When looking at the visual network representations of both cases, one could already see some major differences. First of all, whereas in the manure network one big cluster is identified in which almost all nodes are connected, the dog trade network is way more fragmented. Up until know, there was no insight into the structure of both networks. Second, given the fragmented nature of the dog trade network, one could easily identify nodes that act as bridges between the clusters and which might be crucial for information flow. When one of these nodes leaves the network, predictions could be made to which nodes they are likely to connect. Besides, the network could be actively destabilised by deactivating that node. In the context of fraud detection and inspection this means that particular targets can be identified that would cause maximal disruption of the ongoing or planned illegal activities. Third, it is important to note that SAS VA 8.3.1 does not visualise self-referencing nodes. For the manure network, this means that a certain number of transactions were eliminated from the network. Although this number was considered to be insignificant in relation to the full network size, it may have more crucial consequences in other cases. This problem did not arise in the dog trade network. Finally, nodes in the dog trade network could either be categorised as exporter or importer. In the manure network, it becomes clear that a node fulfils different roles at the same time.

In general, network visualisation may serve two main goals. First of all, inspecting the visual representation of a network can be part of an ex-ante approach in which the inspector is familiarised with the data which can often quickly result in some first findings and insights. Besides, the network is a useful representation to verify mathematically obtained results (i.e. network metrics). In an ex post approach, the network can be used by inspectors or intelligence offices to trace back how an event has happened and which other entities are potentially at risk. In either case, the network visualisation can be adjusted to serve analysis purposes. Although the entire network looks quite complex, one could drill into specific nodes or make a node selection for further inquiry.

### 7.1.2 Application of Different Metrics

Next to the visual representation of networks, networks can also be analysed mathematically. The mathematical analysis in this research was restricted to four metrics: Reach Centrality, Closeness Centrality, Stress Centrality, and Betweenness Centrality. All of these metrics rely on shortest path analysis. Depending on the network, the similarity in central nodes that follow from the network may differ. In the dog trade case, there is a strong correlation between Stress and Betweenness Centrality, indicating that there are not many shortest paths between the nodes. This makes destabilisation of the network easy. In contrast, in the manure case, there is a strong negative correlation between these two metrics. In other words, a node that is important based on Stress Centrality is not considered to be important based on Betweenness Centrality; the node accounts only for a fraction of all of the shortest paths between each pair of nodes. This negative correlation indicates that there are many multiple shortest paths between the nodes, making destabilisation of the network more difficult. When comparing Reach and Closeness Centrality, there is a positive correlation in the dog trade case, but a negative correlation in the manure case. A negative correlation between these two metrics implies that the metrics yield similar results. When solely looking at Reach Centrality, one could see that there is much more variation in the reach of the nodes in the dog trade network compared to the manure network.

It is important to note that centrality is not the same as frequency. However, depending on the network structure, both variables may be correlated with each other. In the manure network, there is a strong correlation between Betweenness Centrality and the frequency of transportation and Stress Centrality and the frequency of transportation. This underlines that nodes which are positioned at any shortest path between two nodes, often have a high frequency of transportations. No correlation was found between any of the centrality metrics and the trade frequency in the dog trade case.

Based on the centrality networks that have been applied, it turned out that in both cases the composition of node types among central nodes is different from the composition of node types among the whole network. Whereas the manure network consists for the largest extent of customers, the most central nodes are mainly nodes which fulfil multiple roles (supplier, transporter, customer) at the same time. In a similar vein, the most central nodes in the dog trade network are the exporters, though they represent only 6% of the full network. Although the latter one is not surprisingly in itself (i.e. an exporter trades with multiple customers), it may have consequences to fight illegal dog trade and the risk of rabies; inspection of exporters is out of control of the organisation.

However, before applying centrality metrics, it is important to consider on what assumptions a certain metric is based and why and for what purpose a metric can be useful. Take the dog trade network. Whenever the network is fragmented, such as in the dog trade network, assigning centrality based on shortest path analysis becomes more complicated – depending on the purpose. If the purpose is to deploy inspection capacity on the most central nodes of the network, mainly the small

network structures will end up with a high centrality. Due to technical limitations of the software, it is not possible to take into account the edge weights of the nodes in determining the centrality of the node. As a result, even nodes with just one transaction score high on centrality. Any metric that relies on other assumptions than shortest path, such as Eigenvector Centrality or Degree Centrality would be more useful in such case.

The problem with fragmented and small network structures is also dependent on the network metric that is applied. For example, for Closeness Centrality the value is normalised between 0 and 1 over a cluster of disconnected IDs. So in a star-shaped network, only one node will be considered as central. When applying Reach Centrality to a star-shaped network, all nodes out of that network will pop up as important, since the maximum shortest distance is only two steps. This example also confirms an argument made in the previous section: network visualisation may support the mathematical analysis of networks.

It is not only important to consider on what assumptions a metric relies, but one must also consider what can be considered as important. In other words, in a 5.928-node network (manure case) or 2032-node network (dog trade case), it must be considered how many central nodes one wish to identify. This determines also the suitability of a metric. In the manure case, 418 nodes have a Reach Centrality of 6. But if only 20 nodes are considered to be important, using this metric would not make sense. In addition, it must be taken into account that there is a neat difference between risk oriented inspection and effective and efficient inspection.

Nonetheless, when these issues are carefully considered, applying network metrics may quickly result in findings and new insights. This becomes especially relevant when networks become large and complex that are not easily interpret. Centrality metrics can help to identify the key players in a network and evaluate or predict the possible consequences of removing specific actors from a network to destabilise that network. Because of the ability to identify important key players and links in a network, monitoring efficiency can be improved. SNA can support tactical decisions on the deployment of enforcement assets in the most optimal locations worth inspecting.

### 7.1.3 Addition of Risk Factors

Although network visualisation and the application of network metrics provide insight into who does business with whom and the central nodes of the network, it does not give any insight into fraud or suspicious activities. Therefore, risk indicators and inspection data have been added to see they are distributed over the network.

A network visualisation with integrated risk factors allows to extracting patterns of suspicious behaviour. It could be investigated whether nodes that are connected show similar behaviour on the risk factors. The other way round, one could investigate whether a certain risk score shows connection between the nodes. In the dog trade network, for example, one could see that the most Romanian traders were connected via Dutch importers. In the manure case, for example, one could see that nodes with disapproved inspections were connected to each other. However, these insights should be taken with caution; the visualisation does not account for randomness.

For the dog trade case, the risk factor was predefined. However, differentiating between what can be considered as risk and what not turned out to be difficult in the manure case. The current risk boundaries are arbitrary. The evidence from this study suggests that there are no correlations between the risk factors and the compliance rate. This means that either the usefulness of the risk indicators need to be reconsidered or the boundaries need to be finetuned with domain experts. Making accurate predictions requires better specification and more data to extract meaningful correlations out of the data. But even when the risk boundaries are determined, one should carefully

consider how the risks visualised. There seems to be a trade-off between accuracy and interpretability, as will be elaborated on in the next section.

The mathematical analysis determines how central nodes score on a risk factor. For example, in the manure case, it is shown that the average compliance of the 100 most central nodes is higher than the overall network, and are on average also more frequently inspected. In addition, fewer transportations take place on average during the national holidays for the selected most central nodes compared to the full network. In contrast, central nodes seem to transport relatively more on suspicious times. The percentage of VDM modifications is lower for the central nodes. Yet, the differences between the central nodes and the overall network differ depending on the metric that is applied. When applying Closeness Centrality in the dog trade network, there is a positive correlation between the centrality of a node and the risk of rabies, and when applying Reach Centrality, there is a negative correlation between the centrality and the risk. No correlation was established between Stress and Betweenness Centrality and the risk of rabies.

### 7.1.4 Plotting on a Geographical Map

Using a network diagram and placing this data on a map gave a good picture of the geographical distances over which entities are connected. Although the manure case focused on a geographical selection based on loading zip code, one could clearly see that the connections are spread over the entire country. In the dog trade case, the visualisation of the network on a geographical map was more problematic. The current data structure does not seem to be the optimal way to represent the entire network on a map. However, specific nodes can be selected that require further inquiry. Also by looking at the network on a map and colouring the node based on the risk of rabies, one could quickly identify in which areas have a higher risk. With the distance analysis capability based on radius-based selection one could quickly determine distances around a specified point on the map and begin the discovery and analysis that could lead to new insights.

### 7.1.5 Tracking Changes over Time

When tracking the network dynamics relies on the visual representation of a network, a small network size is required. When the network is too large and complex, it is infeasible to discover which nodes are active over time. Plotting the network on a geographical map helps to interpret the dynamics. The network dynamics can also be analysed mathematically. In this research, it was examined whether the central nodes per month are also central nodes on a yearly basis. For example, in the manure network on average 12 nodes that end up in the monthly top 100 also end up in the yearly top 100 of central nodes. In the dog trade network, this number is 15. Although it is not the purpose to compare the nature of these two networks, one could argue that the manure network is more dynamic. Rather than comparing both networks with each other, it would be more valuable for its functional purpose to see how this value changes over the years within each case.

## 7.2 Institutional Reflection

Whereas up until now the research was mainly focused on the practical application of SNA to detect fraud in a food and consumer product context, this chapter focuses on the institutional implications. In practice, the way in which organisations create value from big data often remains unclear. The ability to create value from big data depends on having the right process in place to give meaning to the data. This requires collecting the right data, having access to data, obtaining trustworthy data, having the right skills in place for data analysis, and concrete actions to realise the potential of big data. There seems to be a gap between the promises of big data and its practical realisations. In this section, SNA is considered from an institutional perspective. This means it is assumed that there are actors that shape the process from data as a raw material to the final deployment of inspection capacity based on the outcome of the network analysis. When reflecting on the previous chapters, it appears that various important assumptions and decisions have been made. Sometimes, these decisions appeared to be fundamental for the final outcome. Although decision-making in general is

about creating options, evaluating them, and committing to the (most favourable) option, decisions turned out to be sub-optimal due to various circumstances.

In this sub-chapter, the key activities based on the steps in the CRISP-DM process chain are identified (sub-question 4). For each activity, the complexities that arise are discussed in detail and the decisions made are reconstructed (sub-question 5). These complexities can take a variety of forms, including prerequisites, dilemmas, trade-offs, or path dependency and lock-ins. A prerequisite is an act that is required as a prior condition for something else to happen, even though this could be time consuming or very complex. In dilemmas, the decision-maker, in this research the data scientist, is faced with conflicting extremes in an "either.. or.." situation, though the costs and benefits seem to be equally weighted in each course of action (Vandrevala, Hampson, Daly, Arber, & Thomas, 2006). A trade-off is a problem situation in which there are many possible solutions, each striking a different balance between competing pressures (Jucevicius & Juceviciene, 2015). It means that the choice is not so much "either.. or..", but rather 'how much.." or "to what extent.." the factors in each available option are considered (Jucevicius, 2014). Most organisational decisions fall within this category (Jucevicius & Juceviciene, 2015). The analysis takes a perspective as theorised by Simon (1960), in which the way the alternatives are framed impacts the alternative chosen by people and in turn the subsequent decision. This may lead to path dependency. Path dependency is a term in economic and social sciences which denotes that once a certain design choice is taken or once a certain user practice has become prevalent, it becomes almost impossible to change. This may eventually lead to a lock-in situation. Lock-ins in a process can discourage or even prevent the creation of options for decision-making or lead to options that are sub-optimal.

The analysis in this chapter covers three steps of the CRISP-DM model: data understanding, data preparation, and modelling (Section 7.2.1 to 7.2.3). The reason for focusing on these steps is that the researcher has been actively involved in these steps and therefore could act as a reflective practitioner. Reflective practice involves questioning taken-for-granted assumptions which lead to the development of insights (see Chapter 3). A key rationale for reflective practice is that experience alone does not necessarily lead to insight; deliberate reflection on experience is essential (Loughran, 2002). During the reflection, a link to the potential impact of the decisions is made for each of the identified activities. These impacts are crucial in understanding how the big data chain affects the value creation. To conclude this chapter, Section 7.2.4 summarises the overall analysis in a framework.

### 7.2.1 Data Understanding

Data understanding requires the DSC to think about what data is available, what the data mean and to consider whether it is possible to solve the problem with the data. It can be divided into four main activities: data source gathering, data source selection, data acquiring, and data understanding. Depending on the project, some activities may or may not be applicable. In the dog trade case, some exploratory experimenting had already been carried out and various data sources were present. These were considered to be suitable for a second, yet more comprehensive, analysis. Therefore, this phase in the dog trade case was centralised solely about data understanding. Unlike the manure project, there were no additional requests for RVO or other parties to acquire data that could supplement the internally available data.

#### (i) Data Source Gathering

First of all, the data sources had to be selected that capture in some way the problem that is faced. As this was the first data-related project in the manure domain for the DSC, data understanding and gaining insight into the possible sources was of particular importance. The gathering of data sources was organised around workshops. It was chosen to involve participants from various backgrounds

and departments to ensure cross-disciplinary input. The main purpose was to think 'out of the box' when gathering potential data sources.

*Impact:* Although there may no critical decisions faced when 'the sky is the limit', there are some natural constraints for organisations to go off the beaten paths. There is a tendency of organisations to become committed to develop in certain ways as a result of their organisational structure and practices or their beliefs and values. The act to think differently, unconventionally, or from a new perspective is likely to be constrained by ordinary cognitive and institutional processes. Nonetheless, it is important to mention that the purpose of data source gathering was more than just obtaining data sources. It is a knowledge-sharing process which is considered to be crucial for future commitment and successful organisational change towards data-driven inspection.

**(ii) Data Source Selection**

Once the data sources were gathered, the most suitable data sources had to be selected based. Each of the data sources resulting from the data gathering phase has been ranked based on a trade-off between five criteria: relevance, availability, quality, trustworthiness, and usefulness. Although these criteria were pre-defined in NVWA policy, they are no more than the starting point for the selection of data sources. Only internal available data was selected; it turned out that accessibility was the main rationale for selecting the data sources.

*Impact:* Not all potential data sources that have been identified were used for modelling. This was mainly due to the lack of direct access to the data. However, data source selection may affect the representativeness of the model that is supposed to address the problem. Therefore, the inclusion and exclusion of data sources requires good consideration. The data selection is important for how the network will be constructed and the scoping of the network and the underlying problem. Any source that is not selected, is not likely to be used as a basis to construct the network; there is a the tendency of organisations to become committed to develop along the traditional paths. Once the foundation of a network is made, additional data sources are only likely to be used to supplement the network, since returning this path may bring significant investments in time and other resources.

**(iii) Data Acquiring**

Once the data is selected it must also be confirmed that the data that is aimed for can actually be acquired. In the manure case, for instance, one source of RVO was requested but not provided during the course of this research. When the desired data is unavailable, there is a trade-off between various alternatives: adjust the scope, substituting with available data sources, gather alternative data, or a combination of them. In the manure case, it was decided to continue the project based on convenience considerations; the inaccessible data was only considered to be supplementary. Again, this underlines a course of path dependency; it is often easier and more cost-effective to continue along an already set path than to create an entirely new one.

*Impact:* Similar arguments for the impact of unacquired data can be made as for data that is not gathered or selected. One might miss out on important data that could potentially construe the network, especially since there was a certain rationale to select the data that was not accessible in the end. During this research, only supplementary data could not be acquired. Although it was considered as supplementary, it might have served greater purposes when actual insight into the data was obtained when the data was explored in a subsequent stage.

**(iv) Data Understanding**

- *Interpreting data:* Finding out the meaning of the data and the implications for the analysis proved to be challenging, though is a prerequisite for any further activity. Questions related to the data were asked to the organisation's Expertise department. During the data

exploration phase, the project members had to team-up in both cases with various domain experts. Collaboration and knowledge sharing among DSC and domain experts were found crucial for building the model. In the dog trade case, for example, a TRACES certificate is only obligatory from five dogs onwards. This means that the network will be composed of transfers of more than five dogs unless smaller numbers are voluntary and wittingly registered. This kind of data understanding is key to the final interpretation of the network.

- *Verifying data quality:* Also the verification of the data quality was problematic. This is the result of the fact that no source description is available and meta data per column is missing. In some cases the contextual information was lacking and no information was available on how the data are collected, processed, and aggregated. Also a quality analysis of the data source was not available and it often remained unclear whether errors had their origin in the source or the datamart. Nonetheless, even without a thorough quality analysis it already appeared that the quality of the data in the dog trade case was very poor. Yet, it was considered to be sufficient to move forward given the lack of alternatives. This is an example of a lock-in situation; the use of data has not yet been deeply embedded in the organisation. The IT systems in place have been designed for the sake of capturing data rather than data analysis. As the organisation's history has shown, successfully implementing an IT-system to overcome this lock-in turned out to be immensely difficult (Rijskoverheid, 2019).

*Impact:* Since it was difficult to assess the quality of the data sources, this has consequences for the accuracy of the final model and correct interpretation of the results. Besides, good data understanding allows one to make better decisions on the data selection and may reduce the time for data preparation. Following from the cases, data that was not understood properly was even a rationale for not selecting it for modelling.

### 7.2.2 Data Preparation

In the data selection phase, the actual data sets had to be selected. Although accessibility was the key rationale for selecting data sources, when preparing the data it turned out that many of the data lacked contextual information. For that reason, it was decided to focus on internal data first. From this point onwards, the activities were targeted at network analysis. Data is differentiated into two main activities: data selection and data preparation.

#### (i) Data Selection

Data selection is about specifying the sources, tables, and ranges that are relevant for SNA. Although the selection of nodes, links, and boundaries is a prerequisite for SNA in itself, it involves a clear scoping dilemma; selecting a small or broad scope. A smaller scope is associated with operational efficiency, whereas a broader scope may result in a more complete picture of the network which might be needed to extract meaningful patterns from the network, yet requires additional resources.

- *Node selection:* Node selection is all about selecting the important stakeholders. In both cases to nodes in the network are extracted from just one data source, indicating a small network scope. In the manure case, each single node in the network represents a unique BRS number based on the BRS numbers established in the VDM licenses. However, the nodes could as well have presented the location of loading or unloading manure that is established in the VDM licenses. Then, in turn, this requires specifying the location (e.g. using zip codes or addresses). Since much additional data is based on BRS-numbers (that could be easily linked to the nodes), it was decided to take the BRS-numbers as a basis for the nodes in the network rather than the postal codes. In addition, this seemed an easy way to integrate to transporters as a type of facilitators into the network (instead of solely suppliers and customers). Transporters can be regarded as intermediaries since they are responsible for everything around transportation; registration, weighting the freight, manure sampling, and analysing the manure. Manure transporters are considered to be important players in the

trade network of manure, and integrating them in the network could provide better insight into how the actors interact. In the dog trade case, each node represents either an exporter (sender) or an importer (receiver) of dogs based on a shared TRACES certificate: a cross-boundary transaction. However, as will be elaborated on in the node preparation section, also the translocation of dogs within The Netherlands was initially planned to be integrated in the network (I&R data). This was, however, not feasible yet as linking the data of international transactions to national transactions was problematic due to the poor data quality. In addition, it was decided not to add intercontinental TRACES certificates to the network, as no information was available yet how this could be linked to I&R data. The current network which is based on just European exporters and importers seems to be far from complete to assess the risk of rabies within the country.

- *Link selection:* In the construction of the networks, two nodes are connected based on a shared VDM (transport license link) or a TRACES certificate. Yet, with the growing data availability these days, people are connected in many other ways. Therefore, the current representation might only be a fraction of how the entities in a network are related. In addition, this network only represents the transactions that are actually registered. In the dog trade case for example, it is known that an estimated 25% (Appendix B) of the dogs are traded illegally. To get a better grasp of the trade network, one should also think about a way to connect entities that is not so self-evident. Social media data or company ownership structures (KVK data) could, for instance, be used to link actors in a network. However, one should take in mind that integrating different types of links introduces an additional complexity: specifying the relative strength of the links. The small scope used in both cases (i.e. just one type of link) minimises this complication.
- *Selecting geographical boundaries:* In the manure case, a geographical selection was made based on the location of manure loading. The loading zip codes have been used as selection criterium. However, that does not mean that the companies are also located within that region; a farmer may have multiple places where manure is loaded without being accommodated there. Evenly true, this geographical selection could also be made based on the unloading location. Within the dog trade case, the geographical boundaries were set at Europe. The reason for this is that the European transactions and inter-continental transactions are registered in a different manner (Intra-Trade versus GDB, see Appendix G). This had two consequences. First of all, inter-continental transaction data (GDB) could not directly be linked to the Dutch I&R data. Secondly, within the GDB datamart, it is only possible to filter on 'Carnivora' which also includes ferrets and cats. The Intra-Trade data, instead, offers to possibility to select 'Canis Familiaris' which specifically relates to dogs.
- *Selecting domain boundaries:* Domain boundaries resulted from the project scope. In the manure case, the decision was made to focus on hog manure. However, the manure problematics involve more than just pigs. This is a decision criterium that was already set when identifying the project goals. In the data selection phase, all VDMs for hog manure could be selected. For the dog trade case, the decision was made to focus on dogs. As explained, this decision made it difficult to include GDB data (inter-continental transactions), since it was not known yet how to differentiate between ferret, cat, and dog transactions.
- *Selecting timeframe boundaries:* In both cases, data of 2018 was selected to construe the network. Nonetheless, this could also have been a monthly selection or even a multiple year network selection. However, as the network data for just 2018 already rose to enormous volumes of data is was considered to be a good starting point for constructing the network. In addition, building a dog trade network based on data that dates back from earlier than 2016 would not have been possible, as new registration requirements have been implemented only since then.

- Selecting risk factors: In the manure case, each of the selected risk factors is based on data that is captured in the VDMs. This has both its pros and cons. The network, including risk factors, is based on only one data source. On the one hand, this means that the risk factors could easily be linked to the nodes in the network. On the other hand, the risk analysis relies on one source, meaning that the network is not considered from different perspectives. This may impose a severe limitation on the detection of suspicious patterns. In the dog trade case, the only risk factor that was directly available was the risk of rabies of a country. So again, accessibility turned out to be a key rationale for selecting risk factors.

*Impact:* The phase of the in detail data selection is mainly about scoping and determining the network construction based on the data sources that have been selected in the previous stages. This involves specifying the nodes and links of the network and setting geographical boundaries, domain boundaries, and timeframe boundaries, which in these cases can be characterised with a lot of discretionary freedom. As a data scientist, there is a lot of room to act and decide in selecting the data tables and ranges, given the absence of effective control and ambiguous rules. Yet, the decisions made during data preparation are of vital importance; it determines the actual foundation and building blocks of the network and in turn its level of accuracy and abstraction. This requires of course that the data correctly interpret. In both cases, only one dataset was used for the foundation of the network. However, the value acquired from the utilisation of multiple datasets is often far higher than the sum value of individual data sets. Even though in the dog trade case an attempt was done to combine datasets, this failed due to the lack of understanding of how to aggregate the datasets. Another important note is that this was the first stage which was specifically targeted at SNA, meaning that validity and representativeness issues arise. The first refers to whether the nodes and links in the network well-founded and likely to correspond accurately to how and with whom the entities interact and spread ideas in the real world. The latter refers to what extent the selected nodes and links are a complete representation of the actual trade network. In other words, if many transactions are unregistered (occur below the radar), the network will not be representative.

**(ii) Data Preparation**

Data preparation is about the pre-processing of the raw data into a form that can readily and accurately be used for modelling and analysis. Data preparation is a prerequisite for obtaining a data format suitable for SNA and is required to obtain meaningful data, though considered to be a time-consuming process. Next to some prerequisites, the process of data preparation involved various dilemmas and trade-offs.

- *Preparing nodes:* The main complexity for preparing the nodes was understanding the required data structure for the software. Nobody in the current DSC had worked with the either SNA or the software before. In addition, SAS VA 8.3.1 is a new software release, so explanatory documentation was lacking. This made data preparation, besides the usual issues that come along with data preparation, very time demanding. To build and analyse social networks, the software requires an ungrouped or hierarchical data format. It is important to note that the network structure as presented in Chapter 4 is applicable for an ungrouped network. A hierarchical network requires a different data structure. The ungrouped network structure was the first structure that was understood via various iterations, which turned out to be the reason for building an ungrouped network in both cases. This denotes a trade-off between the efficiency and effectivity of using SNA.
- *Preparing inspection data:* In the dog trade case, inspection data was not prepared for modelling yet due to the lack of human resources. In the manure case, selecting and preparing inspection data involved various complexities. First of all, it had to be decided which inspections were selected. A company that operates in the network is not just solely a trader of manure. Instead, it may have a range of business operations of which one or more are being inspected. The question that arises is whether or not, or to what extent include

other inspection data as well. In other words, one needs to consider whether someone wants to know how the organisation is performing in general or more specific related to the manure problematics. In the manure case, the latter was considered. Second, one needs to decide on the timeframe of inspections. In the manure case, it was decided to use three years of inspection history. This gives an indication of how the firm is operating not just recently, but also provides more information about the firm's performance over the years. In addition, the scarcity of inspection data compared to the network size seems to suggest to use more years of inspection data. However, this means that there is a trade-off between the relevance of inspection history and the coverage of inspection data in the network. A third issue that arose is the level of detail of inspections. In the manure case, it was decided to differentiate solely between approved and disapproved inspections. However, one could also argue to relate the nature of the violation(s) to a particular node. Finally, there is more data than just inspection data to gain an understanding of how a firm is performing. There might be investigations or other reports that provide more details on suspicious or fraudulent activities. These are all trade-offs which require close consideration when preparing your data for network analysis.

- *Preparing risk factor data:* In the manure case, each of the selected risk factors is based on data that is captured in the VDMs. During the data preparation for each factor, some important boundaries were set. The National Holiday risk factor seems to be clear cut; all Dutch national holidays are known. However, one could argue about whether or not transportations on Good Friday are as suspicious as transportations on Christmas Day. Similarly, there seems to be a discrepancy between national holidays and other festives, such as New Year's Eve for example. This is not an official national holiday but could be regarded as suspicious. The same is true for the Time of Transportation factor. The boundary for suspicious activity is currently set between 11p.m. and 4a.m. This boundary is, however, arbitrary. Lastly, also specifying the boundary for VMD brings some complexity. The network in its current design only emphasizes whether or not a VDM has been modified. However, much more detail could be added about what exactly has been modified. VDM modifications include a broad range of possible adjustments. In the dog trade case, the only risk factor that was directly available within the internal data was the risk of rabies of a country. Although this is just one risk factor, the advantage is that clear boundaries for (non)-risks have already been specified. Since the country of export is registered in the TRACES data, the risk of rabies could easily be linked to the nodes.
- *Preparing GPS data:* When one starts working with maps and spatial data, having a fundamental understanding of coordinate systems and map projections becomes necessary. GPS data of both the source and target values are needed to plot a network on a geographical map. Therefore, latitude and longitude values had to be specified. The DM_BAG datamart contains the coordinates for an address. In the manure case, the problem arose that only the BRS numbers with a KvK number have a specified street address including number. The most detailed address that is available for all BRS numbers is a six-digit zip code (1234AB), making zip code the key for linking the coordinates. Since a zip code covers a range of coordinates the maximum coordinates were chosen for each single zip code. However, there are, of course, also other alternatives to solve the coordinate problem, indicating the nature of the dilemma. For instance, in the dog trade case, the aggregated average latitudes and longitudes of a city were used. Both the GPS data for the importer (target) and the exporter (source) had to be specified. Adding foreign coordinates turned out to be particularly difficult; lacking zip codes, cities written in foreign script, and registered cities that turned out to be provinces. The poor data quality and lack of foreign geographical understanding make linking node addresses to coordinates painstaking work that is prone to error. In this research, the city was used as a basis to specify the coordinates of the exporters. Since a city

or region covers a range of coordinates, the aggregated average latitude and longitude values of a particular city or region that is registered in TRACES have been used to determine the coordinates of a node.

*Impact:* Data preparation is important in enriching the data and improving the accuracy of the final outcome. More specifically, it determines to what extent the model (correctly) reflects reality. Given the "garbage in, garbage out" principle, dirty data will not be able to provide data inspectors with correct information. There appeared to be various causes for dirty data: duplicate records, missing data, entry mistakes, spelling variations, unit differences, outdated data. This was especially true for company details (addresses) that were stated in TRACES. In general, data cleaning takes on many forms and is considered to be time-consuming. The DSC states that data preparation takes up to 80% of the time. This makes data preparation an expensive, yet inevitably important, process.

### 7.2.3 Modelling

All modelling activities were executed in the SAS Viya 4.3 platform. Two options for network analysis are available in the SAS Viya platform; one in the Visual Analytics (VA) environment and one in the Visual Investigator (VI) environment. On the one hand, the Visual Analytics environment could be used to create a full network overview but is limited in its displaying options, which becomes in particular a restriction when using large and complex data sets and multiple node attributes. On the other hand, the Visual Investigator tool is developed with a specific purpose to identify, investigate and act on suspicious activities and events of interest quickly, including capabilities that govern the complete life cycle of an investigation. It is described as an easy-to-use network viewer backed by powerful intelligence analytics with the ability to visualise and interactively explore entire social networks and their structure (SAS, 2019). However, this tool required additional expertise which could not be provided immediately. Besides, the VI tool is only capable of mapping a node and its first- and second-degree connections. Considering the large size of the network it was decided to start the modelling process in the SAS VA environment.

In the modelling phase it is noticed that the available alternatives for Social Network Analysis are constrained by the capabilities of the software. The software selection is a clear example of path dependency leading to a lock-in situation. The decision to acquire the software was made independently from the SNA projects. Yet, due to the selection of the software, the design and user practices have become fixated to such an extent that applying alternatives would be unfeasible (or requires substantive investment in resources and capabilities) – even if these alternatives were more desirable. The software selection, therefore, seems to be a crucial step in extracting value from big data. This section reflects on each of the activities in the modelling phase.

**(i) Visualisation**

- *Selecting network size:* In the visualisations presented in Chapter 5 and Chapter 6, the entire networks of 2018 are constructed (path dependency). In other words, all data that was prepared for analysis. However, when visualising the network there is a plain trade-off; one could use smaller network sizes to improve the interpretability, though facing the risk of missing out of crucial links in the network.
- *Nodes specification:* The node pictogram, label, colour, size, and underlying details can be decided on. As follows from the result analyses in Chapter 5 and 6, these variables may serve different goals. The type of variables are predefined by the software; an option to adjust the node shape was not available, thus restricting the possibility to differentiate on an additional node attribute. Although this may not be a major concern, it illustrates the lock-in effect of the software; displacing the software with any alternative requires a significant investment in resources and capabilities. Besides, there is a dilemma to be faced within the visualisation phase; it must be decided on which node attributes are being displayed with one of the

aforementioned variables, and which attributes are only visualised when selecting a particular node.

- *Links specification:* Following from the data selection and preparation phases, the basis on which the nodes are connected is a shared VDM or TRACES certificate. In the visualisation phase, one needs to consider how this connection is displayed. The link width, direction, colour, and underlying link attributes have to be determined. In the manure case, due to absence of specification alternatives, self-referencing links were removed from the network. The total item count was only 155 which seems to be negligible given the magnitude of manure transportations. However, depending on the case, the removal of self-referencing links might be more or less critical. In the dog trade network, the problem of self-referencing links did not appear. Within this network, the first degree relations are always disparate to the node itself: an exporter is always directly connected to importers and importers solely directly connected to exporters.

*Impact:* Network visualisation offers a powerful solution to make information hidden in networks easy to interpret and understand. In building the network, various variables can be adjusted to customise the look and feel of the network. This is important for the operational efficiency and the interpretability of the model to be able to quickly and easily analyse if and to what extent entities are related. Any data scientist should be conscious of the importance of choices made in the node and link specification. The network visualisation is, in fact, an information package that is delivered to the inspectors and the basis on which further inquiry is executed and conclusions are drawn. Nonetheless, visualisation options appeared to be restricted either by the software or the way in which the data is prepared. This makes modelling and data-preparation an iterative process which may become time-consuming. Software limitations are strongly case-dependent. For example, the restriction to visualise self-referencing nodes only turned out to be an issue in the manure case.

**(ii) Metrics**

- *Selecting network size:* Similar as the visual presentation of the network, the centrality analysis in Chapter 5 (manure case) is based on 2018. Selecting a smaller network size based on the time variable yields different centrality results. Depending on the purpose of analysis, smaller selections can be made to drill into particular clusters. This was done in the dog trade case. The centrality analysis is based on the data of 2018 minus all 2 and 3 node network structures. The reason to eliminate these structures is based on the available network metrics within the software. This will be elaborated on next.
- *Selecting level of analysis:* Another decision to make regards which network level is used for analysis; network-level analysis, node-level analysis, or a combination of both. In this research, centrality metrics have been used to identify the important nodes of the large and complex networks. Only little attention has been paid to local network structures and the identification of roles and gaining understanding of how nodes influence one another. Inevitably, this requires additional domain expertise.
- *Selecting metrics:* Within the network analysis, four different metrics have been applied: Reach Centrality, Closeness Centrality, Stress Centrality, and Betweenness Centrality. Several problems arise when selecting centrality metrics. First of all, as is outlined in Chapter 5 and Chapter 6, each of these metrics yields different outcomes. Thus, it requires close consideration which of the metrics is applied to analyse social networks. Second, as becomes clear in the literature review, there are many more metrics that could be applied when analysing social networks. In fact, these four metrics are only a small part of the available metrics, and, in addition, all rely on shortest path analysis. By using the SAS VA version 8.3.1 software and determining central nodes, one implicitly assumes that shortest path is the way to assign centrality and in turn importance to a node in a network. In the dog trade case, an

additional complexity arises. As explained in section 7.1, the fragmented nature of the network makes it difficult to apply centrality metrics that rely on shortest path analysis. The magnitude of two- and three node-network structures results in many central nodes. Even when just one transaction exists between an exporter and importer (dyad network structure), both nodes end up with a high centrality value. A metric such as Degree Centrality or Eigenvector Centrality, that relies on the number first-degree respectively second-degree connections of a node, would be more suitable in this case. This option is, however, precluded by using the SAS VA 8.3.1 software and thus a clear example of a lock-in situation.

- *Defining importance:* Although it might be partly overlapping the point made above, it is important to define 'importance'. Whereas the previous point refers to defining importance upfront (i.e. selecting the metrics), this point refers to the interpretation of the results. In both cases, the results of the top 20, top 50, and top 100 central nodes were compared. However, the inspection compacity for the dog trade is considerably smaller than for the manure domain. It is therefore slightly irrational to select the same number of important nodes. For that reason, one should carefully consider how importance is related to the size of the network and the priority of inspecting a domain. This may, of course, also depend on the metric that is applied. In any case, a prerequisite for defining importance is domain expertise.
- *Specifying network rules:* Following Borgatti (2005), the findings of the analysis may differ depending on the rules that govern the network. By default, two important assumptions are made by the software, which again, can be considered as a lock-in situation. First of all, the link direction can only be used for visualisation purposes. The centrality values for a directed and undirected network yield exactly the same results. In other words, direction does not seem to matter, which is unlikely given the size and complexity of the network. Although this is not a primary concern within both cases (the exchange of ideas on how to commit fraud is in principle undirected), it is worth mentioning this implicit assumption. Secondly, this network assumes that links are weighted all the same. However, it could be possible that shortest path centralities actually do consider edge weights that might, for example, be distances in geographical locations that the edges represent (e.g. distance of manure transportation) or the strength of a connection between two nodes (e.g. frequency of transactions). These implicit assumptions could become more relevant for the dog trade network when also I&R data is added to the network (i.e. domestic dog trade).
- *Applying multiple metrics simultaneously:* The software is limited to apply only two centrality metrics at the same time. So, when one wishes to apply more than two metrics simultaneously to compare the results and see how a node scores based on different metrics, the data needs to be exported and linked in a new environment. This is a cumbersome process and may lead an analyst to stick to just two metrics. Using four metrics instead, each relying on different assumptions, might result in better insight into the central nodes of the network.

*Impact:* Considering the size and the complexity of the network, using centrality metrics is useful in identifying the important nodes of a network. However, decisions made in selecting metrics, defining importance, specifying network rules, and combining centrality metrics are crucial in the sense that they become the rationale for the deployment of enforcement assets. Selecting the most appropriate metrics to the context emerged to be a challenge, hence, this step should receive special attention. Again, the availability of alternatives is restricted by the selection of the software. It therefore requires close consideration whether the selected metrics have a potential to reach the goal of the network analysis – which is in turn case dependent. Centrality metrics relying on shortest path are likely to be suitable for identifying central nodes in the manure case but seem to be less favourable in the dog trade case.

**(iii) Addition of Risk Factors and Inspection Data**

- *Specifying inspection boundaries (visualisation):* There are multiple ways in which inspection data can be visualised in the network. One could, for instance, choose for a black and white scenario; firms either have violated the regulations or not which is then indicated with a particular colour. A second alternative is to apply a traffic light system in which green indicates solely approved inspections, red solely disapproved selection, and orange a combination of both. However, the reality is much more complex. A company which scores negative on just one inspection out of 50 would fall in the same range as a company which scores negative on 49 out of 50 inspections. Therefore, an alternative solution is to use the compliance percentage to visualise the performance of a firm, which in turn is divided into different intervals. In total ten intervals could be differentiated, which all have a specific colour. Nonetheless, even when using compliance intervals there seems to be a trade-off between the number of colours used that represent the intervals and the ability to interpret what it actually means. Although inspection data was not yet added to the network in the dog trade case, it is expected that similar issues will arise.
- *Specifying risk factor boundaries (visualisation):* Similar as for the inspection data, also the risk factors need to further specified. In the manure case, there was no correlation between the percentage of compliance and the scoring on risk factors. This means that either the usefulness of the risk indicators need to be reconsidered or the boundaries need to be finetuned. Subsequently, what can be considered as risk might be either in specified in intervals or specific values, but requires in any case careful consideration. In the dog trade case, visualising the risk of rabies for exporters was simple: dogs are either exported from a classified risk country or not. However, integrating the risk of rabies for importers was more complicated. The same issue arises as when inspection data is integrated into the network. An entity could import just one dog from a risky country versus 49 dogs out a non-risky country. A black and white scenario would assign the same colour to a node who imports all dogs from a risky country. Similarly, a traffic light scenario would assign the same colour to a node who imports 99% of the dogs from risk countries or just 1%. Thus, what can be considered as risk requires close consideration and further input from domain experts.
- *Selecting network size:* By integrating risk factors and inspection data, new variables are added which can be used to adjust the size of the network. One could filter the network based on a selected risk score to see whether there are any connections between the nodes or apply centrality metrics to determine the central nodes. Besides, one has to decide on which and how many risk factors are visualised simultaneously.

Please note that the complexities that arise when applying centrality metrics are also valid when applying network metrics on a risk-incorporated network.

*Impact:* The arguments made for visualisation and centrality metrics are also valid when integrating risk factors and inspection data in the network. In fact, the analysis of risk relies on the network visualisation and the metrics that are applied. However, a simple network representation only provides insight into how the entities are related but does not say anything about suspicious (possibly fraudulent) activities. Given that the organisation wants to work risk-oriented, makes decisions on the network visualisation and metrics even more important; a risk factor integrated network will be the basis on which inspections are executed. As mentioned in the section above, when visualising networks there seems to be a trade-off between the number of colours used that represent the risk intervals and the ability to interpret what it actually means. When applying centrality metrics, having a carefully considered definition of 'important' is paramount.

**(iv) Geographical Map**

- *Selecting network size:* The primary concern when plotting the network on a geographical map is the network size. Considering the manure network of 2018, that consists of 5.928 nodes and 27.332 edges, simply results in too much detail on a map which is concentrated on the provinces of Brabant and Limburg. Therefore, smaller network selections have to be made. However, when looking at how the network changes over time, there are fluctuations in the central nodes that appear on monthly basis and on a yearly basis (see Table 13 and Table 17). There seems to be a trade-off between the accuracy of the network (i.e. the extent to which it represents reality) and the interpretability. Although the dog trade network is considerably smaller than the manure network, visualising a yearly network on a geographical background still results in a rather condensed map. Even when just one country is selected it is difficult to interpret how the nodes are related. Selecting a smaller network size could mean that one misses out on important entities or links in the network. Selecting the network size thus requires careful consideration. As mentioned before, one could also decide to restructure the data and make the commonality in exporter the bases on which two importers are connected. In that case, neither the foreign coordinates need to be specified nor the network needs to be projected over entire Europe. This denotes the iterative process of modelling and data preparation.
- *Selecting coordinate system:* A prerequisite for plotting the network on a geographical map is the selection of a coordinate system. SAS VA 8.3.1 includes several different coordinate system configurations. They are selected from a drop-down list during the geography variable setup. A coordinate space is simply a grid designed to cover a specific area of the earth. The four coordinate spaces included with VA are World Geodetic System (WGS84), Web Mercator, British National Grid (OSGB36), Singapore Transverse Mercator (SVY21). However, the data available in the organisation did not match one of VA's predefined geography types. Therefore, the option of Custom Coordinates was used in both cases.
- *Radius-based selection:* When applying radius-based selection, two important points of consideration arise. First of all, the method of radius-based selection needs to be chosen. The SAS VA version 8.3.1 software license is limited to creating a circular selection based on the distance in miles or kilometres. Travel distance (irregular selection based on travel distance using roads) or travel time (irregular selection based on the distance that can be travelled in the specified amount of time) are only available if premium settings are enabled in the SAS Visual Analytics settings. Especially the latter would be an interesting alternative in the manure case. Secondly, the radius size needs to be selected. This may be less critical if it is only used to zoom into a particular part of the network. However, more caution must be exercised when using radius-based selection to extract patterns of suspicious behaviour. Thus, the size of the radius may depend on the investigation purposes and is not a fixed design requirement. The exact added value of applying radius-based selection needs to be discussed with domain experts.

Please note that the complexities that arise when visualising the network are also valid when plotting a network on a geographical map.

*Impact:* When plotting the network on a geographical map, the selection of the network size turned out to be important for both cases. Plotting a network on a geographical map is meant for graphical analysis. Therefore, the arguments made for network visualisation can also be made for geographical mapping. Although a yearly network might give better insight into the important entities of the network, it results in too much detail on the map which is difficult to interpret. Selecting a network of smaller sizes may be easier to interpret, but involves introducing bias in the system by assigning importance to a node which only turns out to be a minor player in the long run. Therefore, the choices

that are made during geographical may have significant consequences for the entities that are given closer attention.

**(v) Tracking Changes Over Time**

- *Time frame selection:* When looking at the network dynamics, a certain time frame needs to be chosen. In both cases, a monthly comparison was used to see how the structure of a network changes over time. One should not only choose a certain time frame, but one should also think about what the time dimension means for the (suspicious) patterns that are to be extracted from the network. The other way round, one should think about what the (suspicious) can be extracted from a dynamic network. These considerations will, however, require additional domain expertise. The rationale for selecting a certain timeframe also depends on the size and complexity of the network and whether its purposes lie within the graphical and/or mathematical analysis of the network. When the network becomes too large and complex, the ability to track changes over time based on the visual representation might be limited.
- *Specifying visualisation background:* If the tracking of changes of the network over time relies on the graphical representation of the network, networks likely need to be plotted on a map. If a network consists of too many nodes it would be unfeasible to interpret the network dynamics on a plain background. A geographical map gives a fixed position to the nodes which helps to understand how the structure of a network changes. This is particular important when one wants to investigate which nodes appear and disappear in the network. However, this first requires to plot the map on the map in a conventional manner (see point iv). Even then, there is a trade-off between the interpretability and the relevance of the network size.

Please note that the complexities that arise when visualising the network and applying centrality metrics are also valid when tracking changes over time.

*Impact:* The selection of time and the network size are related to each other. A larger timeframe selection results in a larger network. Therefore, the arguments made about network size are also applicable for the time frame selection, the basis on which the network dynamics are measured. Tracking changes over time is important to give meaning to the centrality value that has been assigned to a node. When the a network is highly dynamic, the centrality of a node may become less relevant over time. In other words, if "important" nodes change over a small period of time, centrality metrics may not be the right way for deploying enforcement assets. In contrast, in a static network there is less variation in the central nodes, giving more reason to apply centrality metrics as a basis for inspection.

### 7.2.4 Towards a Framework

The overall analysis is summarised in Table 19. The first column represents the CRISP-DM process step. The second column represents the key activities that are involved based on these steps (sub-question 4). The subsequent column represents the organisational complexities involved each of these steps (sub-question 5). The final column summarises why each of the decisions made in the activities matter. In other words, what are the impacts of the decisions on the value that is created. By developing a framework this research contributed both to the academical debate on big data value creation and the practical application of SNA for fraud detection. As the research into understanding big data value creation is still in a very young state, the findings from this thesis form a starting block for future research to expand on. In its practical form, the table can be used as a framework for structuring SNA processes. It reminds the user of the fact that the value created by SNA is not solely dependent on the data, but also on the process in which the data is collected and the way that data is processed for final analysis.

Table 19 Activity, Operational Complexity and Impact Framework

| CRISP-DM step | Activity | Operational complexity | Impact |
|---|---|---|---|
| Data Understanding | Data source gathering | Gathering data that captures the problem | Knowledge-Sharing, Commitment, Representativeness |
| | Data source selection | Evaluating data sources | Representativeness |
| | | Selecting data sources | |
| | Acquiring Data | Gaining access to data sources | Representativeness |
| | Data understanding | Interpreting data | Accuracy, Operational Efficiency |
| | | Verifying data quality | |
| Data Preparation | Data Selection | Specifying nodes: entity selection | Representativeness, Validity |
| | | Specifying links: relationship selection | |
| | | Specifying geographical boundaries | |
| | | Specifying domain boundaries | |
| | | Specifying timeframe boundaries | |
| | | Selecting risk factors | |
| | Data Preparation | Preparing nodes | Accuracy, Representativeness, Validity |
| | | Preparing inspection data * | |
| | | Preparing risk factor data | |
| | | Preparing GPS data | |
| Modelling | Network Visualisation | Specifying network size | Effectiveness, Efficiency** |
| | | Specifying nodes: size colour, pictogram, label and underlying node details | |
| | | Specifying links: Width, direction, colour, removal of self-referencing, and underlying link details | |
| | Applying Metrics | Specifying network size | Effectiveness, Efficiency** |
| | | Selecting level of analysis | |
| | | Selecting metrics | |
| | | Defining importance | |
| | | Specifying network rules | |
| | | Applying multiple metrics simultaneously | |
| | Adding Risk Factors and Inspection Data | Specifying inspection boundaries (visualisation) * | Effectiveness, Efficiency** |
| | | Specifying risk factor boundaries (visualisation) | |
| | | Selecting network size | |
| | Plotting on Geomap | Specifying network size | Effectiveness, Efficiency** |
| | | Selecting a coordinate system | |
| | | Selecting radius based selection method and radius | |
| | Tracking Changes over Time | Specifying time frame | Effectiveness, Efficiency** |
| | | Specifying visualisation background | |

* Only applicable for the manure case

** Of the deployment of intelligence assets

# 8. Conclusion and Recommendations

In this chapter, the main findings of the research are presented. The chapter starts with answering the five sub-questions. Based on the answers to the sub-questions and the institutional and functional reflections, both functional and institutional conclusions are provided in Section 8.2 Finally, Section 8.3 presents a reflection on the quality of the research and recommendations for future research.

## 8.1 Conclusion on the Sub-Questions

In this paragraph, each of the five sub-questions is answered. Answers to these questions will logically lead to answering the main research question.

*Sub-question 1: What are the differences between an expert-based fraud detection and data-driven fraud detection system?*

Fraud is an adaptive crime, so it needs special methods of intelligent data analysis to detect and prevent it. The traditional approach for fraud detection is the expert-based fraud detection system, whereby the system relies on human expert input, evaluation, and monitoring. Using an automated, data-driven system could lead to a more efficient and effective methodology for detecting fraud. However, machine learning models are no panacea in fraud detection. It might be very difficult (if not impossible) to explain to others how certain scores or decisions are obtained. Table 20 summarises the main differences between both systems.

Table 20 Fraud Detection Systems

| **Characteristics** | **Expert-based system** | **Data-driven system** |
|---|---|---|
| **Rules** | Expert rules | Data-driven rules |
| **Knowledge** | Tacit | Explicit |
| **Analytics** | Limited predictive capability | Both reactive and predictive |
| **Problems** | Difficult to maintain<br>Heuristic rules<br>Expert knowledge dependent<br>Capturing of tacit knowledge<br>Costly due to labour-intensity | "Black box" algorithms<br>Data-dependent<br>Data analysis competencies |

*Sub-question 2: What is Social Network Analysis, and how is it applied to detect fraud?*

Over the years, various approaches have been proposed in literature to counteract fraud based on big data, among which is Social Network Analysis. SNA can be used for measuring and analysing the structure of relationships between actors. Relationships in a network are the basis on which actors in a network are connected. The relationships or connections underlying the network can occur in a variety of forms, such as interpersonal relationships like friendship, advice seeking, or trust, which denote interactions between individuals; or inter-organisational relationships like knowledge and resource sharing, or trading goods, which denote the interaction between organisations as a whole. One of the core assumptions of SNA is that the structure of these connections influences individual and organisational behaviour. The relationships between individuals or organisations might enable or restrain access to resources, exchange of information, or lead to exposure to social norms and culture. SNA can be a powerful technique for researching social phenomena such as the flow of information through a network and the identification of key actors.

The development and interest in SNA have increased intensively over the years. SNA helps to structure data important for making business decisions. Current literature on SNA shows its application in a wide variety of areas, among which fraud detection. Fields that have used SNA for fraud detection include health care fraud, insurance fraud, mobile internet fraud, money laundering, mortgage fraud, online auction fraud, opinion fraud, security fraud tax fraud, social security fraud, insurance fraud, and telecom fraud. Yet, any scientific application in the consumer and food product domain is lacking.

To detect fraud, social networks can be analysed both graphically and mathematically. Whereas graphical analysis is the most straightforward and intuitive form analysis, it cannot be used to compute useful statistics and extract meaningful characteristics of the network. This becomes especially relevant when networks become large and complex. The graphical approach rests on two perspectives: node-level analysis and network-level analysis. While the node-level perspective looks at the structural characteristics of individual nodes, the network-level perspective considers the overall network structure. Analysis of social networks may potentially lead to risk analysis and threat assessments, destabilisation of networks, role identification, scenario building, support decisions on the deployment of intelligence assets, and evidence for prosecution.

*Sub-question 3: What challenges for using Social Network Analysis in fraud detection are mentioned in literature?*

There is a vast amount of literature which speaks about the potential of analysing social networks to detect fraud and criminal activities. Nonetheless, widespread applications of SNA in this domain are lacking. The use of social network analysis in criminology is still in its infancy, since "the great majority of crime network studies consider only the characteristics of the members of the networks, and not of the structure of their relationships". Most analyses of network structures rely just on the examination of visual representations of networks, rather than computational analyses. Even the computational analyses tend to limit themselves to the simplest network concepts. Besides, literature seems to have focused on the usage of SNA as fraud detection rather than the implementation and adoption of the technique in organisations. Despite their strong foundations and expressive power, the development of new fraud detection methods seems to be rather difficult.

Several arguments are mentioned in literature for the difficult development of SNA in fraud detection. Researchers on criminal networks tend, with a few exceptions, to fall into two distinct categories, each operating under severe constraints. One the one hand, crime researchers have expertise in criminological theory and research but seem to lack analytical expertise and access to good data. The collection of social network data can be problematic, especially when complete data on a network are needed to compute useful statistics. On the other hand, data analysts may have strong analytical skills and access to various data sources, but tend to lack knowledge on criminological issues such as fraud – or may be prevented from publishing their work due to secrecy or privacy considerations. The development of fraud detection methods is constrained by the scarcity of available data sets and the limited disclosure results in public. This imposes a severe limitation on the exchange of ideas in fraud detection. Not to mention that majority of published methods are considered as black-boxes where their workings are mysterious. Besides, both fraud and fraud detection methods are embedded in a specific context; a solution to fight credit cards fraud cannot be applied in insurance companies.

*Sub-question 4: How does the process of data understanding to Social Network Analysis work?*

Following from the research gaps identified in the literature review, the concept of SNA has barely been researched in the context of fraud detection, let alone included in the context of food and

consumer products. To answer the fourth sub-question, the process of data understanding to SNA has been analysed empirically at the DSC of the NVWA. The DSC uses the CRISP-DM model as a guideline to structure their activities. Figure 40 summarises the key activities that were carried out during the two cases in which SNA has been applied. In both cases, the data understanding phase was executed independently from the SNA project. During the process, many activities were executed in a linear order. Only between data preparation and modelling several iterations took place. The data preparation phase of cleaning and transforming the raw data prior to SNA was both time and labour intensive. The basis for the modelling phase consists of network visualisation and applying network metrics. Depending on the purpose, the network can be plotted on a geographical map, risk factors can be added or the network dynamics can be analysed both graphically or mathematically. Please remember that this is a reconstruction based on this research. The activities and the iterations involved in the chain may be strongly context dependent.

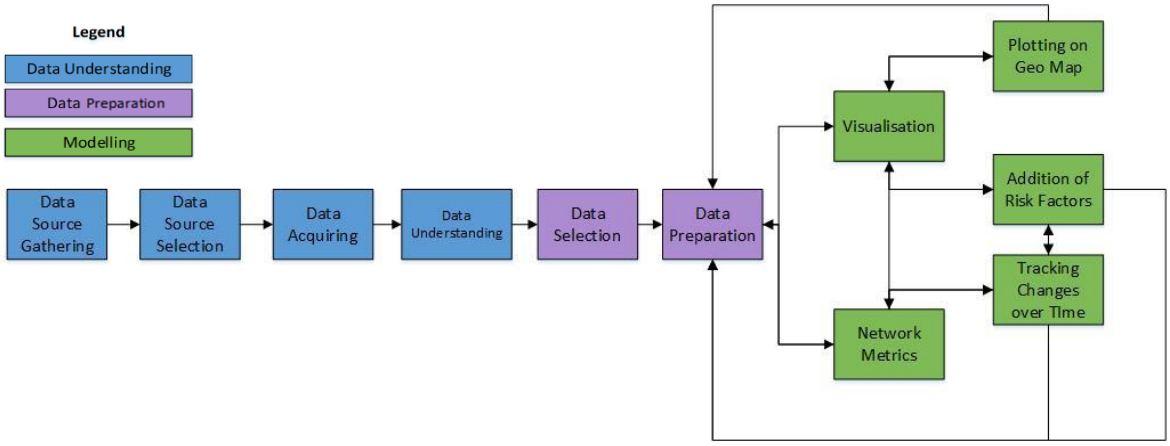



Figure 40 Activity Chain from Data Understanding to Modelling based on CRISP-DM

*Sub-question 5: What operational complexities are applicable in the process of data understanding to Social Network Analysis?*

Although there are already many successful big data applications, executing SNA is involved with several practical problems have to be solved for a meaningful analysis. Each of the activities identified in Figure 40 is concerned with operational complexities. These complexities are the result of prerequisites, dilemmas, trade-offs, or path dependencies and lock-ins during the process of data understanding to SNA. This results in decisions and assumptions that need to be made for which there is no optimal solution, or the optimal solution is unknown.

**Data Understanding**

- *Data source gathering:* Gathering the data that captures the problem;
- *Data source selection:* Evaluation data sources and selecting data sources;
- *Data acquiring:* Gaining access to data;
- *Data understanding:* Interpreting data and verifying data quality.

**Data Preparation**

- *Data selection:* Specifying nodes (entity selection), specifying links (relationship selection), specifying geographical boundaries, specifying domain boundaries, specifying timeframe boundaries, and selecting risk factors;

- *Data preparation:* Preparing nodes, preparing inspection data, preparing risk factor data, and preparing GPS data;

**Modelling**

- *Visualisation:* Specifying network size, specifying nodes (size colour, pictogram, label and underlying node details), specifying links (width, direction, colour, removal of self-referencing, and underlying link details);
- *Applying centrality metrics:* Specifying network size, selecting level of analysis, selecting metrics, defining importance, specifying network rules, applying multiple metrics simultaneously;
- *Addition of risk factors:* Specifying inspection boundaries (visualisation), specifying risk factor boundaries (visualisation), selection of network size;
- *Plotting on a geographical map:* Specifying network size, selecting a coordinate system, selecting radius based selection method and radius;
- *Tracking changes over time:* Specifying time frame and specifying visualisation background.

## 8.2 Conclusion on the Study

This research had two main purposes. First of all, this thesis aimed to explore the use of Social Network Analysis in the context of fraud detection, more specifically fraud in the context of food and consumer products. Secondly, this research aimed to apply an institutional view to get insight into big data value creation in the public sector, using two case studies of SNA. This section provides some key messages for each purpose.

### 8.2.1 Functional Conclusions

As follows from the previous chapters, SNA has the potential to identify patterns in the fraud network and in turn improve inspection efficiency and effectivity. First of all, network visualisation offers a powerful solution to make information hidden in networks easy to interpret and understand. With one glance at the network one could identify who does business with whom, which entities act as bridges between two clusters, and gain insight into the overall structure of the networks (i.e. to what extent are entities connected). However, when visualising networks, there seems to be a trade-off between the interpretability and the accuracy of the network. Secondly, applying network metrics may quickly result in findings and new insights into the large and complex networks that are not easily visually interpret. Centrality metrics can help to identify the important players in the networks and evaluate or predict the possible consequences of removing specific actors from the networks to destabilise the networks. Nonetheless, it requires close consideration what can be considered as important and whether the selected metrics have the potential to reach that goal, which turned out to be strongly case dependent. The former depends, among others, on the capacity of enforcement assets, and the latter depends on the structure of the network. Any metric that relies on shortest path analysis seemed to be effective in the manure case, given its interconnected structure, but less effective in the dog trade case, given its fragmented network structure.

When the graphical or mathematical foundation of the network is set, other options are available to extend the network analysis. Three options were considered in this research. First of all, a network visualisation with integrated risk factors allows extracting patterns of suspicious behaviour. On the one hand, it could be investigated whether nodes that are connected show similar behaviour on the risk factors. On the other hand, one could investigate whether a certain risk score shows connection between the nodes. By doing so, SNA can support tactical decisions on the deployment of enforcement assets in the most optimal locations worth inspecting. Yet, specifying risk factor boundaries still requires close cooperation with domain experts. Working fully data-driven is still an

intricate if not impossible endeavour in the food and consumer product domain. Secondly, using a network diagram and placing this data on a map gave a good picture of the geographical distances over which entities are connected. With the distance analysis capability based on radius-based selection one could quickly determine distances around a specified point on the map and begin the discovery and analysis that could lead to new insights. Again, this requires further input from domain experts. Finally, this research managed to track various network dynamics. However, taken the analysis of the network dynamics together suggests that there are two important considerations needed to draw further conclusions. First of all, one should think about what the time dimension means for the (suspicious) patterns are to be extracted from the network, which is strongly context dependent. Secondly, one should think about what (suspicious) patters can be extracted from a dynamic network.

### 8.2.2 Institutional Conclusions

The emergence of big data opens great opportunities in the public domain. Nonetheless, the analysis of big data is confronted with many challenges. From the two case studies executed in this research it is clear that data-driven fraud detection within the context of the food and consumer product domain is still a long way to go. Data-driven models that make up own rules are nothing if they are not data-intensive. In addition, the quality of the data that is used will inevitably have a huge effect on its chances for success. But, an argument that has not been cited previously in the context of fraud detection, is that even when a data-driven fraud detection system is capable of self-learning, it is still inevitable that people shape the process from data to value. And that is exactly where the decision-making process comes in. This research has thoroughly analysed the process from gathering data sources to building a model that can be used for Social Network Analysis to detect fraud. Based on two case studies, executed within the DSC of the NVWA, the following research can now be answered:

**How does the big data activity chain influence the potential value created by using Social Network Analysis for fraud detection?**

When going through the process from data understanding to modelling, it became apparent that various important assumptions and decisions have been made. Sometimes, these decisions appeared to be fundamental for the final outcome. Taking on a qualitative approach allowed an in-depth exploration of a topic on which little research has been performed. The analysis covered three steps of the CRISP-DM model: data understanding, data preparation, and modelling. It was shown that extracting value from big data is an intricate process, consisting of various operational complexities that should be sufficiently addressed. The complexities in each activity are the result of prerequisites, dilemmas, trade-offs, or path dependencies leading to institutional or technological lock-ins.

To begin, data understanding involves more than solely comprehending the meaning of the data. It requires an organisation to gather, select, acquire, interpret and verifying the data quality. Data understanding is characterised by an institutional lock-in. There is a tendency of organisations to become committed to develop in certain ways as a result of their organisational structure and practices or their beliefs and values. The act to think differently, unconventionally, or from a new perspective is likely to be constrained by ordinary cognitive and institutional processes.

Second, data preparation is a prerequisite for enriching the data and improving the accuracy of the final outcome. More specifically, it determines to what extent the model (correctly) reflects reality. In general, data cleaning takes on many forms and is considered to be time-consuming. This makes data preparation an expensive, yet inevitably important, phase. The data preparation is characterised with a lot of discretionary freedom in which various trade-offs are to be made. As a data scientist, there is much room to act and decide in selecting the data tables and ranges, given the absence of effective control and ambiguous rules.

Finally, in the modelling phase it is noticed that the available alternatives for Social Network Analysis are constrained by the capabilities of the software. The software selection is a key example of path dependency leading to a technological lock-in situation. Due to the selection of the software, the design and user practices have become fixated to such an extent that applying alternatives would be unfeasible (or requires substantive investment in resources and capabilities) – even if such alternatives were more desirable. Besides, when the analysis relies on the graphical representation of a network, there seems to be a trade-off between the accuracy of the network and its interpretability.

These complexities prevent the creation of options or lead to options sub-optimal. Neglecting them would be at the detriment of any future SNA-ambition an organisation may hold. The decisions made can be fundamental for knowledge-sharing, commitment, representativeness, accuracy, validity, operational efficiency and the effectiveness and efficiency of the deployment of enforcement assets.

## 8.3 Discussion

The final section of this chapter presented here will reflect back on the research and discuss its limitations and the recommendations for future research. This section is divided into two parts. The first part discusses the limitations and recommendations related to SNA in the context of fraud and consumer products and the second part discusses the limitations and recommendations related to the main research question.

### 8.3.1 Functional Limitations and Recommendations

SNA in the context of fraud detection, in particular in the food and consumer product domain, is still in its infancy. Given the absence of a comprehensive methodology to apply SNA in fraud detection, and the fact that there is no 'one-fits-all' solution, makes this research only exploratory in nature. Five important limitations and recommendations for further research are addressed: the first three issues are related to the execution of SNA, while the latter two have consider legal and ethical implications.

**Utilisation of multiple data sources**

One of the key characteristics of fraud is that it is perfectly concealed and that any evidence is often obfuscated. The value acquired from the utilisation of multiple datasets is often far higher than from individual data sets. This implies that other stakeholders and types of connections must be incorporated into the network to support the identification of suspicious patterns. For the dog trade network this means that at least the dog transfers within the Netherlands need to be integrated. In the current trade network, it is only possible to identify key importers and exporters. However, dogs are often translocated within the country as well. This suggest that the current network does not provide a solid foundation to assess the actual risk of rabies. Expanding the network of the manure case means adding other entities and stakeholders such as labs and agricultural advisors (see Table 3).

**Applying a variety of metrics**

A second limitation of this study is that the analysis was restricted to centrality metrics only. In the dog trade network it was difficult to identify the central actors given the fragmented network structure. This seems to imply that that SNA is not a suitable way to identify key players in the dog trade network. However, the software selection had a deterministic impact in exploring the opportunities of SNA for this case; only four metrics could be applied which all rely on shortest path. Therefore, it might be interesting to see if and what important entities could be identified when applying metrics that are not based on this assumption, such as Eigenvector Centrality or Degree Centrality. In addition, clustering algorithms could be used to identify (suspicious) behaviour in specific parts of the network. Also the option for collective interference metrics has not been

analysed. Since only a small part of the network entities has been inspected, collective interference could be used to infer a set of labels or probabilities for the uninspected nodes.

**Making results actionable: Collaboration with domain experts**

In general, the results of an SNA can be made actionable by doing one of the following; creating a holistic network and propagate behaviour on the network or disrupt the network. In this research, the network was located and some exploratory analysis was encouraged. Yet, neither active propagation of suspicious activities was done nor serious attempts were done to disrupt the network. Given both the network structures, the former could be an opportunity in the manure case and the latter could an opportunity in the dog trade case. However, both propagation and strategic disruption requires input from domain experts. Although domain experts were actively involved during the first phases of the process, the modelling and the analysis of the networks was executed fully independently. Given the operational complexities suggest the need to carefully consider the assumptions of the model with experts of the field. To illustrate, both the definition of 'important' and 'risky' are arbitrary yet fundamental for further actions. Finetuning these terms requires close cooperation with domain experts.

**Legal Considerations**

It is important that everyone involved in the analysis of social networks invests some time to reflect on the legal implications. Since social networks are in principle collections of natural individuals or organisations that are interrelated, a dominant legal consideration in SNA, which has not been addressed in this study, is the General Data Protection Regulation (GDPR). The GDPR is primarily intended to give control to individuals over their personal data and aims to simplify the regulatory environment for international organisations by unifying the regulation within the EU. This requires, among others, that personal data is relevant, not excessive, kept up to date and no longer than is necessary, and that the purposes of processing are specified, explicit, and legitimate (Smyth, 2018). Nonetheless, there is a lot of room and scope for law enforcement for the purpose of the prevention of threats and the safeguarding of public security. The Law Enforcement Directive (LED), a legislation parallel to the GDPR, deals with the processing and storage of personal data for 'law enforcement purposes' – which falls outside of the scope of the GDPR (European Commision, 2018). Although the regulation does not explicitly state inspectorates, it is very clear that such organisations should be given a lot of discretionary terms because of the need of a well-functioning society. Future work should concentrate on the exact implications of the GDPR and the LED on the use of SNA in the public domain.

**Ethical Considerations**

Next to the legal issues, SNA raises also ethical issues that often fall outside existing regulations and guidelines. A primary assumption behind SNA is that you are labelled by your friends, business partners, or close associations. That assumption does seem to be objectionable in terms of 'freedom of association' or 'freedom of commerce'. In addition, it is questionable whether a close contact can be interpret in the same way as an occasional connection. This is particular interesting in the context of inspection; the institution is usually separated into observation and assessment. One could state that analysts can use SNA for observing and devote assessing the network to a different department. However, SNA is often used to create appealing visualisations in which various assumptions have been made, and therefore possesses also an assessing quality; it is hard to draw the line here. To conclude, a key message from this study is that data does not speak for itself. The conditions of data production, processing, the methods of analysis, and the assessment of results should be questioned.

### 8.3.2 Institutional Limitations and Recommendations

There are several implications that put pressure on the quality of the results of this study. All steps taken in the research contain a certain level of subjectivity and influence from the researcher. Consequently, there is room for improvement on different aspects which are discussed below.

**CRISP-DM versus The Big Data Value Chain**

To begin, this research uses the CRISP-DM process to structure the activities of big data value creation. However, it could be argued that the CRISP-DM process and the big data value chain are in principle two separate concepts. In general, the CRISP-DM process describes the common approaches used by data analyst to translate business problems into data mining tasks, suggest appropriate data transformations, and provide means for evaluating the effectiveness of the results. The big data value chain, instead, is commonly used by a technology marketer to deliver value for the customer, often with a financial interest. Yet, in this research, the value chain is adapted with a focus on societal interest. In the context of inspection, this means that the data value chain can be described as the series of activities done to generate value in terms of safeguarding public well-being. This includes using the data as a raw material that needs to be processed, possibly based on the CRISP-DM model, to support decisions on the deployment of intelligence assets. However, as will be elaborated on in the next paragraph, this study did not consider any outcomes of the analysis put into practice – the act that provides the actual value for society.

**Extending the research scope**

Following from the previous section, a constraining factor pertains to the scope of this research. The scope is limited over two axes: the activity chain is confined to data understanding till modelling and the two cases analysed were taken from only one organisation. Future research can benefit from extending this scope along both axes. First of all, completing the activity chain allows to understand how not only data scientist shape value creation from big data, but also the users (i.e. inspectors) that are intended to act on the analysis. This research primarily challenged the assumption that the data revolution yields better information. Further research may be addressed to challenge the assumption that better information leads to better decision-making. Secondly, increasing the number of cases will increase the reliability of findings and allow to make more robust inferences concerning the generalisation of how various actors and their decisions shape the process from data to value. Given that this work was a qualitative study (although SNA itself is also quantitative in nature), it cannot be guaranteed that the identified organisational complexities are exhaustive. Moreover, the identified impact is not considered to be the be the sole answer. Different researchers may identify different decisions based on different projects. This was an R&D project. It could be questionable to what extent similar decisions are required for a more standardised process or for an implementation process. In the context of data driven fraud detection and inspection, other inspectorates can be included first, after which comparisons can be extended to include industries of different categories.

**Action research as intervention tool**

A second factor that may have influenced the outcomes of this research, is the close collaboration with the DSC. There is a potential lack of objectivity as a result of the researcher's stake in effecting a successful outcome for the organisation. Data collection tools in action research are themselves interventions that generate data. Interventions may evoke feelings of like anxiety, suspicion, and empathy making people to behave differently. This means that that the findings should be used, criticised, and complemented by future research. Additional empirical studies are required to determine the relative impacts on how decisions on operational complexities shape the process from data to value creation. More general, further research will be addressed to widen the empirical evidence on how big data affects public decision-making.

## References


Åkerman, M., Lundgren, C., Bärring, M., Folkesson, M., Berggren, V., Stahre, J., . . . Friis, M. (2018). Challenges Building a Data Value Chain to Enable Data-Driven Decisions: A Predictive Maintenance Case in 5G-Enabled Manufacturing. *Procedia Manufacturing, 17*, 411-418. doi:https://doi.org/10.1016/j.promfg.2018.10.064

Akoglu, L., Chandy, R., & Faloutsos, C. (2013). *Opinion fraud detection in online reviews by network effects*.

Algemene Rekenkamer. (2014). *Rijksbrede resultaten en thema's verantwoordingsonderzoek 2013*. Retrieved from https://www.rekenkamer.nl/onderwerpen/verantwoordingsonderzoek/documenten/rapporten/2014/05/21/staat-van-de-rijksverantwoording-2013

Almeida, M. P. S.-B. J. M. D. D., Engenharia Informática e de Computadores. (2009). Classification for fraud detection with social network analysis.

Anderson, C. (2015). *Creating a data-driven organization: Practical advice from the trenches*: " O'Reilly Media, Inc.".

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2014). Transformational issues of big data and analytics in networked business. *MIS quarterly : management information systems., 38*(2), 629-631.

Baesens, B., Vlasselaer, V. v., & Verbeke, W. (2015). *Fraud Analytics Using Descriptive Predictive and Social-Network Techniques*.

Barone, M., & Coscia, M. (2018). Birds of a feather scam together: Trustworthiness homophily in a business network. *Social Networks, 54*, 228-237. doi:https://doi.org/10.1016/j.socnet.2018.01.009

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly, 11*(3), 369-386.

Bhagat, S., Cormode, G., & Muthukrishnan, S. (2011). Node Classification in Social Networks. In C. C. Aggarwal (Ed.), *Social Network Data Analytics* (pp. 115-148). Boston, MA: Springer US.

Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: Four perspectives — Four challenges. *SIGMOD Record, 40*(4), 56-60.

Blatter, J. (2014). *Designing Case Studies: Explanatory Approaches in Small-N Research*.

Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. U. (2006). Complex networks: Structure and dynamics. *Physics Reports, 424*(4), 175-308. doi:https://doi.org/10.1016/j.physrep.2005.10.009

Bok, H. S., Kankanhalli, A., Raman, K. S., & Sambamurthy, V. J. I. J. o. E.-c. (2012). Revisiting media choice: a behavioral decision-making perspective. *8*(3), 19-35.

Bolton, G. (2010). *Reflective practice: writing and professional development*. Los Angeles: Sage.

Bolton, R. J., & Hand, D. J. (2002). Statistical Fraud Detection: a Review*, Statistic Science,* 235 - 249.

Borgatti, S. (2005). *Centrality and Network Flow* (Vol. 27).

Bradbury-Huang, H. (2010). What is good action research? Why the resurgent interest? *, 8*(1), 93-109.

Braganza, A. (2004). Rethinking the data–information–knowledge hierarchy: towards a case-based model. *International Journal of Information Management, 24*(4), 347-356 %@ 0268-4012.

Cachia, R. (2008). *Social Computing, Study on the Use and Impact of Online Social Networking*. Retrieved from Luxembourg:

Cane, P., Conaghan, J., & Walker, D. M. (2008). *The new Oxford companion to law*. Oxford: Oxford University Press.

Carrington, P. J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis* (Vol. 28): Cambridge university press.

Chakrabarti, S., Dom, B., & Indyk, P. (1998). Enhanced hypertext categorization using hyperlinks %J SIGMOD Rec. *27*(2), 307-318. doi:10.1145/276305.276332

Chang, Y.-C., Lai, K.-T., Chou, S.-C. T., & Chen, M.-S. (2017). *Mining the Networks of Telecommunication Fraud Groups using Social Network Analysis.* Paper presented at the Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017.

Chau, D. H., Pandit, S., & Faloutsos, C. (2006). *Detecting fraudulent personalities in networks of online auctioneers.* Paper presented at the European Conference on Principles of Data Mining and Knowledge Discovery.

Chauhan, S., & Panda, N. K. (2015). Chapter 12 - Basics of Social Networks Analysis. In S. Chauhan & N. K. Panda (Eds.), *Hacking Web Intelligence* (pp. 217-227). Boston: Syngress.

Chen, H., Chiang, R., & Storey, V. (2012). *Business Intelligence and Analytics: From Big Data to Big Impact* (Vol. 36).

Chen, H., Chung, W., Xu, J. J., Wang, G., Qin, Y., & Chau, M. (2004). Crime data mining: A general framework and some examples. *Computer, 37*(4), 50-56. doi: https://doi.org/10.1109/MC.2004.1297301

Chen, M., & Liu, S. M. Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171–209.

Chiu, C., Ku, Y., Lie, T., & Chen, Y. (2011). *Internet Auction Fraud Detection Using Social Network Analysis and Classification Tree Approaches* (Vol. 15).

Choi, H., & Varian, H. A. L. (2012). Predicting the Present with Google Trends. *Economic Record, 88*, 2-9. doi:10.1111/j.1475-4932.2012.00809.x

Consumentenbond. (2016). *Onderzoek voedselfraude*. Retrieved from Den Haag:

Cortes, C., Pregibon, D., & Volinsky, C. (2001). *Communities of Interest*. Paper presented at the Proceedings of the 4th International Conference on Advances in Intelligent Data Analysis.

Coughlan, P., & Coghlan, D. (2002). Action research for operations management. *International Journal of Operations Management, 22*(2), 220-240.

Cressey, D. R. (1953). *Other people's money; a study in the social psychology of embezzlement*. Glencoe, Ill.: Free Press.

Curry, E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. In J. M. Cavanillas, E. Curry, & W. Wahlster (Eds.), *New Horizons for a Data-Driven Economy: A Roadmap for Usage and Exploitation of Big Data in Europe* (pp. 29-37). Cham: Springer International Publishing.

Dazeley, R. P. (2006). *To the knowledge frontier and beyond: A hybrid system for incremental contextual-learning and prudence analysis.* University of Tasmania,

De Groot, F., Overgaauw, P., & Virginia, E. (n.d.). *De gezonde(re) en sociale hond in Nederland. Problematiek en plan van aanpak (internationale) hondenhandel.* Retrieved from Haarlem: http://dierenarts-haarlem.nl/Nieuwsbrieven/Problematiek(internationale)%20hondenhandel%20Def%20.pdf

Dreżewski, R., Sepielak, J., & Filipkowski, W. (2015). The application of social network analysis algorithms in a system supporting money laundering detection. *Information Sciences, 295*, 18-32. doi:10.1016/j.ins.2014.10.015

Easley, D., & Kleinberg, J. (2010). *Networks, Crowds, and Markets: Reasoning about a Highly Connected World* (Vol. 9): Cambridge University Press.

Elgendy, N., & Elragal, A. (2014). *Big Data Analytics: A Literature Review Paper*, Cham.

European Commision (2018). Data protection in the EU. Retrieved on 21 August 2019 from https://ec.europa.eu/info/law/law-topic/data-protection/data-protection-eu_en.

Fast, A., Friedland, L., Maier, M., Taylor, B., Jensen, D., Goldberg, H. G., & Komoroske, J. (2007). *Relational data pre-processing techniques for improved securities fraud detection.* Paper presented at the Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining.

Fronzetti Colladon, A., & Remondi, E. (2017). Using social network analysis to prevent money laundering. *Expert Systems with Applications, 67*, 49-58. doi:https://doi.org/10.1016/j.eswa.2016.09.029

Galloway, J., & Simoff, S. J. (2006). *Network Data Mining: Discovering Patterns of Interaction Between Attributes*, Berlin, Heidelberg.

Geman, S., & Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-6*(6), 721-741. doi:10.1109/TPAMI.1984.4767596

Ghoshal, A., Larson, E., Subramanyam, R., & Shaw, M. (2014). The impact of business analytics strategy on social, mobile, and cloud computing adoption.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., & Watts, D. J. (2010). Predicting consumer behavior with Web search. *Proc Natl Acad Sci U S A, 107*(41), 17486-17490. doi:10.1073/pnas.1005962107

Gray, R., Butler, S., Douglas, C., & Serpell, J. (2016). *Puppies from "puppy farms" show more temperament and behavioural problems than if acquired from other sources.* Paper presented at the UFAW Animal Welfare Conference York, UK, June 23.(Poster).

Grinsven, H. v., & Bleeker, A. (2017). *Evaluatie Meststoffenwet 2016: Syntheserapport*. Den Haag: Planbureau voor de Leefomgeving Retrieved from https://www.pbl.nl/publicaties/evaluatie-meststoffenwet-2016-syntheserapport

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems, 26*(3), 191-209. doi:10.1016/j.jsis.2017.07.003

Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods.

Henry, S., Lanier, M. M., Adler, M. J., Farr, K. A., Gertz, M., Gibbons, D. C., . . . Kleck, G. (2001). *What Is Crime?: Controversies over the Nature of Crime and What to Do about It*: Rowman & Littlefield Publishers.

Herriott, R. E., & Firestone, W. (1983). *Multisite Qualitative Policy Research: Optimizing Description and Generalizability* (Vol. 12).

Hogeschool HAS Den Bosch, & Universiteit Utrecht. (2015). *Feiten & Cijfers Gezelschapsdieren 2015.* Retrieved from Den Haag: Ministerie van Economische Zaken

Holzinger, A. (2018, 23-25 Aug. 2018). *From Machine Learning to Explainable AI.* Paper presented at the 2018 World Symposium on Digital Intelligence for Systems and Machines (DISA).

Janssen, M., Van der Voort, H., & Wahyudi, A. (2017). Factors influencing big data decision-making quality. *Journal of Business Research, 70*, 338-345. doi:10.1016/j.jbusres.2016.08.007

Jucevicius, G. (2014). *Strategic Tensions Of Smart Development*.

Jucevicius, G., & Juceviciene, R. (2015). Smart Development of Organizational Trust: Dilemmas and Paradoxes. *Procedia - Social and Behavioral Sciences, 213*, 860-866. doi:https://doi.org/10.1016/j.sbspro.2015.11.496

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications, 100*, 234-245. doi:https://doi.org/10.1016/j.eswa.2018.01.037

Kagan, R. A., & Scholz, J. T. (1980). The "Criminology of the Corporation" and Regulatory Enforcement Strategies. In E. Blankenburg & K. Lenk (Eds.), *Organisation und Recht: Organisatorische Bedingungen des Gesetzesvollzugs* (pp. 352-377). Wiesbaden: VS Verlag für Sozialwissenschaften.

Kamphuis, H. A. (2015). *Mestfraude*. Divisie Landbouw & Natuur NVWA.

Kayser, V., Nehrke, B., & Zubovic, D. (2018). Data Science as an Innovation Challenge: From Big Data to Value Proposition. *Technology Innovation Management Review, 8*(3), 16-25. doi:10.22215/timreview/1143

Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. In. Retrieved from https://methods.sagepub.com/book/the-data-revolution doi:10.4135/9781473909472

Kong, X., Shi, Y., Yu, S., Liu, J., & Xia, F. (2019). Academic social networks: Modeling, analysis, mining and applications. *Journal of Network and Computer Applications, 132*, 86-103. doi:https://doi.org/10.1016/j.jnca.2019.01.029

Leiden, I. v., Esseveldt, J. v., Wolsink, J., Wijk, A. v., & Endenburg, N. (2019). *Zo ziek als een hond? Gezondheids- en socialisatieproblemen bij puppy's in Nederland in relatie tot de herkomst*. Arnhem: Bekereeks.

Lettieri, N., Altamura, A., Malandrino, D., & Punzo, V. (2017). Agents Shaping Networks Shaping Agents Integrating Social Network Analysis and Agent-Based Modeling in Computational Crime Research. *Department of Computer Science - University of Salerno*, 15-27.

LICG. (2018). *Praktisch. De invloed van huisdieren op de ontwikkeling van kinderen*. Retrieved from Barneveld: https://www.licg.nl/media/3449/de-invloed-van-huisdieren-op-de-ontwikkeling-van-kinderen5_02.pdf

Lim, A., & Chai, D. S. (2015). Action Research Applied with Two Single Case Studies. In K. D. Strang (Ed.), *The Palgrave Handbook of Research Design in Business and Management* (pp. 375-392). New York: Palgrave Macmillan US.

Lim, C., Kim, K.-H., Kim, M.-J., Heo, J.-Y., Kim, K.-J., & Maglio, P. P. (2018). From data to value: A nine-factor framework for data-based value creation in information-intensive services. *International Journal of Information Management, 39*, 121-135. doi:10.1016/j.ijinfomgt.2017.12.007

Liu, J., Bier, E., Wilson, A., Honda, T., Sricharan Kumar, Gilpin, L., . . . Davies, D. (2015). Graph Analysis for Detecting Fraud, Waste, and Abuse in Healthcare Data. *Proceedings of the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence*.

Loughran, J. J. (2002). Effective Reflective Practice: In Search of Meaning in Learning about Teaching. *Journal of Teacher Education, 53*(1), 33-43. doi:10.1177/0022487102053001004

Lu, Q., & Getoor, L. (2003). *Link-based classification*. Paper presented at the Proceedings of the Twentieth International Conference on International Conference on Machine Learning, Washington, DC, USA.

Maciejewski, M. (2016). To do more, better, faster and more cheaply: using big data in public administration. *International Review of Administrative Sciences, 83*(1_suppl), 120-135. doi:10.1177/0020852316640058

Mathiassen, Chiasson, & Germonprez. (2012). Style Composition in Action Research Publication. *MIS Quarterly, 36*(2). doi:10.2307/41703459

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review, 90*.

McCue, C. (2015). Chapter 15 - Advanced Topics. In C. McCue (Ed.), *Data Mining and Predictive Analysis (Second Edition)* (pp. 349-365). Boston: Butterworth-Heinemann.

McGloho Bay, S., Anderle, M. G., Steier, D. M., & Faloutsos, C. (2009). SNARE: a link analytic system for graph labeling and risk detection. *KKD*.

McMillan, F. D., Duffy, D. L., & Serpell, J. A. J. A. A. B. S. (2011). Mental health of dogs formerly used as 'breeding stock'in commercial breeding establishments. *135*(1-2), 86-94.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *27*(1), 415-444. doi:10.1146/annurev.soc.27.1.415

McQuiston, J. H., Wilson, T., Harris, S., Bacon, R. M., Shapiro, S., Trevino, I., . . . Marano, N. (2008). Importation of dogs into the United States: risks from rabies and other zoonotic diseases. *Zoonoses and public health, 55*(8-10), 421-426.

Milovanović, S., Bogdanović, Z., Labus, A., Barać, D., & Despotović-Zrakić, M. (2019). An approach to identify user preferences based on social network analysis. *Future Generation Computer Systems, 93*, 121-129. doi:https://doi.org/10.1016/j.future.2018.10.028

Nash, R., Bouchard, M., & Malm, A. (2013). Investing in people: The role of social networks in the diffusion of a large-scale fraud. *Social Networks, 35*(4), 686-698. doi:https://doi.org/10.1016/j.socnet.2013.06.005

Neville, J., Jensen, D., Komoroske, J., Palmer, K., & Goldberg, H. (2005). *Using relational knowledge discovery to prevent securities fraud*. Paper presented at the Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, Chicago, Illinois, USA.

Nisbet, R., Miner, G., & Yale, K. (2018). Fraud Detection. In *Handbook of Statistical Analysis and Data Mining Applications* (Vol. 2, pp. 289-302): Academic Press.

Nooy, W. d. (2009). *Social network analysis* (Vol. 1). Oxford: Eolss Publishers.

Norvig, P. (1992). Chapter 16 - Expert Systems. In P. Norvig (Ed.), *Paradigms of Artificial Intelligence Programming* (pp. 530-563). San Francisco (CA): Morgan Kaufmann.

NVWA. (2016). *NVWA 2020 - Herijking van het Plan van Aanpak NVWA 2013*. Den Haag Retrieved from https://www.nvwa.nl/over-de-nvwa/organisatie/nvwa-2020

NVWA (2017a). Animatie 'Het miljoen van de NVWA' 2017. Retrieved on 20-05-2019 from https://nvwa.pleio.nl/videolist/watch/48381802/animatie-het-miljoen-van-de-nvwa-2017.

NVWA (2017b). Inspecties mestbeleid. Retrieved on 20-05-2019 from https://www.nvwa.nl/onderwerpen/mest/inspecties-mestbeleid.

NVWA. (2018). *Jaarverslag 2017 - feiten en cijfers*. Retrieved from https://magazines.nvwa.nl/jaarverslag/2017/01/feiten-en-cijfers

OM (2019, 5 April 2019). Gevangenisstraffen geëist tegen boeren in onderzoek naar stelselmatige fraude met mest in Friesland. Retrieved on May from https://www.om.nl/@105572/gevangenisstraffen/.

Oussous, A., Benjelloun, F.-Z., Ait Lahcen, A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University - Computer and Information Sciences, 30*(4), 431-448. doi:10.1016/j.jksuci.2017.06.001

Overton, S., & Zenick, B. (n.d.). *Visualizing Relationships and Connections in Complex Data Using Network Diagrams in SAS® Visual Analytics*. Retrieved from Cary: https://support.sas.com/resources/papers/proceedings15/3323-2015.pdf

Page, L., Brin, S., Motwani, R., & Winograd, T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web*.

Pandit, S., Chau, D. H., Wang, S., & Faloutsos, C. (2007). *Netprobe: a fast and scalable system for fraud detection in online auction networks.* Paper presented at the Proceedings of the 16th international conference on World Wide Web.

Park, J., & Barabási, A.-L. (2007). Distribution of node characteristics in complex networks. *104*(46), 17916-17920. doi:10.1073/pnas.0705081104 %J Proceedings of the National Academy of Sciences

Pearl, J. (1986). *Pearl, J.: Fusion, propagation, and structuring in belief networks. Artificial Intelligence 29, 241-288* (Vol. 29).
Peng, S., Yu, S., & Mueller, P. (2018). Social networking big data: Opportunities, solutions, and challenges. *Future Generation Computer Systems, 86*, 1456-1458. doi:10.1016/j.future.2018.05.040
Peng, Y., Kou, G., Sabatka, A., Chen, Z., Khazanchi, D., & Shi, Y. (2006, 25-27 Oct. 2006). *Application of Clustering Methods to Health Insurance Fraud Detection.* Paper presented at the 2006 International Conference on Service Systems and Service Management.
Pierre, R. (2018, 27 June 2018). Detecting Financial Fraud Using Machine Learning: Winning the War Against Imbalanced Data. Retrieved on 31 May from https://towardsdatascience.com/detecting-financial-fraud-using-machine-learning-three-ways-of-winning-the-war-against-imbalanced-a03f8815cce9.
Poulie, J., & Genugten, I. v. (2018). *Data analyse hondenhandel*. Retrieved from Utrecht: NVWA
Raad van Beheer. (2018). *Voortgangsrapportage Fairfok*. Den Haag Retrieved from https://www.rijksoverheid.nl/documenten/rapporten/2018/02/01/voortgangsrapportage-fairfok
Raad voor Dierenaangelegenheden. (2006). *Gedeelde zorg. Feiten & Cijfers.* Retrieved from Den Haag: https://www.rda.nl/binaries/raad-voor-dierenaangelegenheden/documenten/zienswijzen/2006/03/23/welzijn-gezelschapsdieren/RDA+2006_02.pdf
Ransbotham, S., Kiron, D., & Prentice, P. K. J. M. S. M. R. (2016). Beyond the hype: the hard work behind analytics success. *57*(3).
Rijksoverheid (2016). Mestproductie bij gebruiksnormen: bedrijven met overproductie, 2000-2015. Retrieved on 0606 from https://www.clo.nl/indicatoren/nl052811-mestproductie-bij-gebruiksnormen-bedrijven-met-overproductie.
Rijksoverheid (2019). Inspectieraad. Retrieved on 13 May from https://www.rijksinspecties.nl/over-de-inspectieraad/inspectieraad.
Rijskoverheid (2019). Minister Schouten stopt implementatie en ontwikkeling ICT systeem INSPECT bij NVWA. Retrieved on August 16 from https://www.rijksoverheid.nl/ministeries/ministerie-van-landbouw-natuur-en-voedselkwaliteit/nieuws/2019/04/15/minister-carola-schouten-stopt-implementatie-en-ontwikkeling-ict-systeem-inspect-bij-nvwa.
Rijt, W. v., Verhoeven, W., & Kok, R. (2016). *Beleid hondenfokkerij en -handel in Nederland. Beleidsdoorlichting en evaluatie I&R Hond*. Retrieved from Zoetermeer: https://ndg.nl/wp-content/uploads/Beleid-hondenfokkerij-en-handel-in-Nederland-beleidsdoorlichting-en-evaluatie-IR-1.pdf
Ross, J. W., Beath, C. M., & Quaadgras, A. J. H. B. R. (2013). You may not need big data after all. *91*(12), 90-+.
Ryman-Tubb, N. F., Krause, P., & Garn, W. (2018). How Artificial Intelligence and machine learning research impacts payment card fraud detection: A survey and industry benchmark. *Engineering Applications of Artificial Intelligence, 76*, 130-157. doi:https://doi.org/10.1016/j.engappai.2018.07.008
Sabo, A. S. (2018). *The illegal puppy trade from the Balkans. An explorative research on the nature, degree of organization and underlying reasons for the trade.* Retrieved from Utrecht: Utrecht University
Sapountzi, A., & Psannis, K. E. (2018). Social networking data analysis tools & challenges. *Future Generation Computer Systems, 86*, 893-913. doi:10.1016/j.future.2016.10.019
SAS (2019). SAS® Visual Investigator. Retrieved on 05/06 from https://www.sas.com/en_us/software/intelligence-analytics-visual-investigator.html.
Schaadt, G. (2013). Food Safety and the Challenge of Globalization. *Food Safety Magazine*.
Schilling, M. A. (2013). *Strategic management of Technological Innovation* (4 ed.). New York: McGraw-Hill.
Schmid, M. (2018). *'De stoere hond', wel of niet welkom? Onderzoek naar het aantal hoogrisicohonden in Nederland en de capaciteit en het aanbod vanuit kynologenclubs aan dit aantal honden.* . Retrieved from Amsterdam: Raad van Beheer
Schön, D. A. (1983). *The reflective practitioner : how professionals think in action*.
Scott, J. (2000). Social network analysis: A handbook. Sage London. (2nd Edition).
Scott, J. (2011). Social network analysis: developments, advances, and prospects. *Social Network Analysis and Mining, 1*(1), 21-26. doi:10.1007/s13278-010-0012-6
Scott, J., & Carrington, P. J. (2011). *The SAGE Handbook of Social Network Analysis*: SAGE Publications.
Scott, J., & Stokman, F. N. (2015). Social Networks. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 473-477).
Seawright, J., & Gerring, J. (2008). Case Selection Techniques in Case Study Research:A Menu of Qualitative and Quantitative Options. *61*(2), 294-308. doi:10.1177/1065912907313077
Shani, A., & Pasmore, W. (1982). *Towards a New Model of the Action Research Process* (Vol. 1982).
Sharma, R., Mithas, S., & Kankanhalli, A. J. E. J. o. I. S. (2014). Transforming decision-making processes: a research agenda for understanding the impact of business analytics on organisations. *23*(4), 433-441. doi:10.1057/ejis.2014.17
Shimshack, J. P., & Batten, F. (2014). The Economics of Environmental Monitoring and Enforcement: A Review. *Annual Review of Resource Economics, 6*, 339-360. doi:https://doi.org/10.1146/annurev-resource-091912-151821
Silva, J. S., & Saraiva, A. M. (2015, 25-28 Aug. 2015). *A methodology for applying social network analysis metrics to biological interaction networks.* Paper presented at the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
Smyth, E. (2018). Data Protection Act 2018 and law enforcement: an introduction. Retrieved on 21 August 2019 from https://www.kingsleynapley.co.uk/insights/blogs/data-protection-blog/data-protection-act-2018-and-law-enforcement-an-introduction.
Soulie Fogelman, F., Mekki, A., & Sean, S. (2011). *Using Social Networks for On-line Credit Card Fraud Analysis*.

Sparrow, M. K. (1991). The application of network analysis to criminal intelligence: An assessment of the prospects. *Social Networks, 13*(3), 251-274. doi:https://doi.org/10.1016/0378-8733(91)90008-H

Steketee, M., Miyaoka, A., & Spiegelman, M. (2015). Social Network Analysis. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences (Second Edition)* (pp. 461-467). Oxford: Elsevier.

Šubelj, L., Furlan, Š., & Bajec, M. (2011). An expert system for detecting automobile insurance fraud using social network analysis. *Expert Systems with Applications, 38*(1), 1039-1052. doi:10.1016/j.eswa.2010.07.143

Sutherland, E. H. (1939). *Principles of criminology*. Chicago, Philadelphia: J.B. Lippincott Company.

Suyker, P. (2018). *Overtredingen Hoofdspoor - NOTA*. Organisation document. NVWA. Utrecht.

Tayebi, M. A. (2015). *Predictive Models for Public Safety Using Social Network Analysis.* Applied Sciences: School of Computing Science,

Teyebi, & Glässer. (2016). *Social Network Analysis in predictive policing - concepts, models, and methods*. Calgary: Springer.

Tweede Kamer. (2015). *Beantwoording Kamervragen over legaal en gezond fokbeleid.* Den Haag Retrieved from https://zoek.officielebekendmakingen.nl/kst-28286-818.html

Uchida, C. D. (2010). *A National Discussion on Predictive Policing: Defining Our Terms and Mapping Successful Implementation Strategies*. Retrieved from Los Angeles: https://www.ncjrs.gov/pdffiles1/nij/grants/230404.pdf

Van der Hulst, R. C. (2008). *Introduction to Social Network Analysis (SNA) as an investigative tool* (Vol. 12).

Van der Voort, H. G., Klievink, A. J., Arnaboldi, M., & Meijer, A. J. (2019). Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? *Government Information Quarterly, 36*(1), 27-38. doi:10.1016/j.giq.2018.10.011

van Lint, R. (2018). Even wachten. Retrieved on 13 May from https://www.nvwa.nl/nieuws-en-media/weblogs/weblog/2018/even-wachten.

van Uhm, D. P. (2010). *Onderzoek illegale hondenhandel. De puppydossiers: een koppeling tussen theorie en praktijk.* Retrieved from Den Haag: International Fund for Animal Welfare.

Van Vlasselaer, V., Bravo, C., Caelen, O., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). APATE: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems, 75*, 38-48. doi:https://doi.org/10.1016/j.dss.2015.04.013

Van Vlasselaer, V., Eliassi-Rad, T., Akoglu, L., Snoeck, M., & Baesens, B. (2015). GOTCHA! Network-based Fraud Detection for Social Security Fraud. *Management Science*.

Vandrevala, T., Hampson, S. E., Daly, T., Arber, S., & Thomas, H. (2006). Dilemmas in decision-making about resuscitation—a focus group study of older people. *Social Science & Medicine, 62*(7), 1579-1593. doi:https://doi.org/10.1016/j.socscimed.2005.08.038

Velthof, G. L., T. de Koeijer, J.J. Schröder, M. T., A. Hooijboer, J. Rozemeijer, Bruggen, C. v., & Groenendijk, P. (2017). *Effecten van het mestbeleid op landbouw en milieu: beantwoording van de ex-postvragen in het kader van de evaluatie van deMeststoffenwet.* . Retrieved from https://www.wur.nl/upload_mm/a/f/4/fa1f62c7-4b34-4a59-8c42-731924fdcb90_WENR-rapport%202782_Totaal_LR.pdf

Wang, J.-C., & Chiu, C.-C. (2008). Recommending trusted online auction sellers using social network analysis. *Expert Systems with Applications, 34*(3), 1666-1679. doi:10.1016/j.eswa.2007.01.045

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*: Cambridge University Press.

Wauthier, L. M., & Williams, J. M. J. A. A. B. S. (2018). Using the mini C-BARQ to investigate the effects of puppy farming on dog behaviour. *206*, 75-86.

Webster, C. M., & Morrison, P. D. (2004). Network Analysis in Marketing. *Australasian Marketing Journal (AMJ), 12*(2), 8-18. doi:https://doi.org/10.1016/S1441-3582(04)70094-4

Wei, R., Liu, X., & Liu, X. (2019). Examining the Perceptual and Behavioral Effects of Mobile Internet Fraud: A Social Network Approach. *Telematics and Informatics*. doi:https://doi.org/10.1016/j.tele.2019.04.002

West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: A comprehensive review. *Computers & Security, 57*, 47-66. doi:https://doi.org/10.1016/j.cose.2015.09.005

Wirth, R., & Hipp, J. (2000). *CRISP-DM: Towards a standard process model for data mining.* Paper presented at the Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining.

Yanchun, Z., Wei, Z., & Changhai, Y. (2011, 25-27 May 2011). *Detection of Feedback Reputation Fraud in Taobao Using Social Network Theory.* Paper presented at the 2011 International Joint Conference on Service Sciences.

Yin, R. K. (2012). *Applications of Case Study Research* (3 ed. Vol. 34): SAGE Publications.

Yufeng, K., Chang-Tien, L., Sirwongwattana, S., & Yo-Ping, H. (2004, 21-23 March 2004). *Survey of fraud detection techniques.* Paper presented at the IEEE International Conference on Networking, Sensing and Control, 2004.

# Appendices

# Appendix A: Summary of Literature Review

* *Collective Interference Algorithm*

** *Clustering method*

| Reference of the Paper and Country | Category | SNA Application | Algorithms / Metrics / Analysis | Dataset/Size of network |
|---|---|---|---|---|
| **(Cortes et al., 2001), UK** | Telecom fraud | A dynamic graph representation of nodes and edges in a telecom network which appear and disappear from the graph through time. | Edge weight<br>Neighbourhood metrics | Network IDs (nodes) and the communication between network IDs (edges). Dataset consisted of hundreds of millions of nodes and edges. To make network analytics feasible, the "top-k" (highest weight) edges and nodes have been selected. Dataset not further specified. |
| **(H. Chen et al., 2004), USA** | Insurance fraud | Identifying subgroups and key members in insurance networks and then studying interaction patterns to develop effective strategies for disrupting the networks. | Edge weight<br>Degree centrality<br>Betweenness centrality<br>Closeness centrality<br>Hierarchical clustering** | Incident summaries involving 164 crimes committed from 1985 through May 2002. The criminals (nodes) are connected (edges) through the crimes (nodes) they have been involved in. |
| **(Galloway & Simoff, 2006), Australia** | Insurance fraud | Twelve months of motor vehicle insurance claims from one company have been analysed by identifying irregularities (i.e., persons linked to multiple claims and/or addresses). Similar patterns were found which appeared to be one and the same person. | Eliminating the 'regular' small triangles of data comprised of a person, a claim number and an address<br>Visual analysis | Records of 12 months of insurance claims have been analysed. Not further specified. Edges between persons, addresses, claim numbers, telephone numbers, and bank accounts into which claims monies had been paid. |
| **(Neville et al., 2005), USA** | Security fraud | Focussing NASD's limited regulatory resources on the brokers who are most likely to engage in fraudulent behaviour | Neighbourhood metrics based on 55 attributes (not further specified) | A subset of the entire dataset was selected: 384,944 disclosures (node) filed on (edge) 1,245,919 brokers (node) worked for (edge) 16,047 firms (node), belonging to (edge) 295,455 branches (node) |

| | | | | |
|---|---|---|---|---|
| **(Chau et al., 2006), USA** | Online auction fraud | User level features (e.g., number of transactions, average price of goods exchanged, etc.) are mined to get an initial belief for spotting fraudsters, and network level features which capture the interactions between different users and use it to detect suspicious patterns. | 2LFS algorithm (based on The Markov Random Field Mode and Belief Propagation) * | Users (nodes) and edges represent a transaction between two users. Dataset not specified. |
| **(Fast et al., 2007), USA** | Security fraud | Developing statistical models that combine patterns of past behaviour, social structure among employees (reps) and firms, and the current risk environment to identify branches and reps that are at high-risk for future misconduct. | Clustering Collective interference algorithm | The dataset contained historical records on over 3.4 million reps (nodes), 360,000 branches (label) and over 25,000 firms (nodes). The nodes are connected through job changes. 70% of the dataset could be matched and used. |
| **(Pandit et al., 2007), USA** | Online auction fraud | The design and implementation of NetProbe, a system to systematically tackle the problem of fraud detection in large scale online auction network. | NetProbe algorithm (based on The Markov Random Field Mode and Belief Propagation) * | Synthetic data set: 7,000 nodes and 30,000 edges; Real dataset: a graph of eBay users with approximately 66,000 nodes and 800,000 edges. |
| **(Wang & Chiu, 2008), Taiwan** | Online auction fraud | Recommendation system that uses trading relationships to calculate level of recommendation for trusted online auction sellers. | Degree K-core** | 53,788 edges among all 20,528 accounts (only links that occurred more than once were selected) |
| **(Chiu et al., 2011), China** | Online auction fraud | Utilising network metrics and data-mining techniques to detect and cluster fraudsters based on Internet auction transaction records. | In- and out-degree K-core ** k-plex ** n-clique ** Betweenness | 3,886 accounts, transactions (edges) not specified. |
| **(McGloho Bay et al., 2009), USA** | Accounting Fraud | Using link analytics, risks of an account can be reranked not only based a single account, but also in other accounts with which it shares transactions. Also, a group of accounts that are closely related and have distributed risk may be identified while under individual flags they would fall below the threshold. | Propagation algorithm * | Two datasets (A+B): A: 1,380 accounts (nodes), 3,820 edges, and 11, 532 red flags B: 1, 678 accounts (nodes), 18, 720 edges |
| **(Yanchun et al., 2011), China** | Online auction fraud | Analysing online selling behaviour to detect the relationship of potential fraudsters and characteristic | Weighted in- and out-degree | Users (nodes) and edges represent a transaction between two users. Dataset not specified. |

| | | | | |
|---|---|---|---|---|
| | | features in fraudsters' behaviour. | | |
| **(Soulie Fogelman et al., 2011), France** | Credit card fraud | Using SNA for problems where one wants to detect and investigate risky events | Not specified | Millions of nodes. Edges are transactions. Not numerically specified. "All accepted transactions in one month". |
| **(Šubelj et al., 2011), Slovenia** | Insurance fraud | Identifying key participants using IAA based on individual and relational attributes of the nodes. A suspicion score is assigned to each participant, which corresponds to the likelihood of it being fraudulent. Instead of learning from initial labelled data, the system allows simple incorporation of the domain knowledge. Used to detect fraudulent groups. | Iterative Assessment Algorithm (IAA) *; a variant of Iterative Classification Algorithm (ICA) | 3,451 participants involved in 1,561 car collisions in Slovenia (1999-2008) |
| **(Nash et al., 2013), Canada** | Mortgage fraud | Using SNA to establish that an illicit innovation (mortgage fraud) can spread through a population of victims much like a legitimate innovation spreads through a population of consumers | Degree centrality Diffusion curves | 475 individuals (nodes) with 450 edges. |
| **(Akoglu et al., 2013), USA** | Opinion fraud | It exploits the network effect among reviewers and products to detect fraudulent users and fake reviews in online review networks and consists of two complementary steps; scoring users and reviews for fraud detection, and grouping for visualisation and sensemaking. It requires no labelled data and is scalable to large datasets. | Signed Inference Algorithm (SIA) * Cross-associations (CA) ** | 966,842 users (also called customers), 15,094 apps for products (e.g., hotels, restaurants, etc.) and 1,132,373 reviews. |
| **(Drężewski et al., 2015), Poland** | Money laundering | Using data from bank statements and the National Court Register to construct and analyse social networks regarding money laundering. The system is used to assign roles to persons from the network and allows for analysis of connections between them. | Authoritativeness Betweenness centrality Closeness centrality Degree centrality Hubness calculates Page Rank | Network consisted of 20 nodes and 20 edges. Dataset not specified |
| **(Liu et al., 2015), USA** | Health care fraud | Suspicious individuals, suspicious relationships, anomalous temporal changes and geospatial characteristics, structures in the graph of doctors, pharmacies, and patients | Degree Weight Entropy ratio Community size | 64 million claims from 5.2 million patients, more than 52,000 doctors, and nearly 9,000 pharmacies |

| | | | | |
|---|---|---|---|---|
| **(Van Vlasselaer, Eliassi-Rad, et al., 2015), Belgium** | Social security fraud | Used for considering which resources are often involved in fraud and impose a high risk to other companies to commit fraud as well. A score indicates which resources are coincidentally related (low risk) with fraudulent companies and which resources systematically pop up when fraud is detected (high-risk). | Extended PageRank * | Approximately 350,000 active and non-active companies and 5 million resources (nodes; bipartite network). Dataset not specified. |
| **(Van Vlasselaer, Eliassi-Rad, et al., 2015), Belgium** | Credit card fraud | Network-based APATE exploits the relationships between credit card holders and merchants by means of transactions. Starting from a limited set of labelled edges (i.e., fraudulent trans-actions) the set of network-related features measures the exposure of each network component (i.e., credit card holders, merchants and transactions) to fraud. | Collective interference algorithms | Company data set with more than 3 million transactions |
| **(Fronzetti Colladon & Remondi, 2017), Italy** | Money laundering | Analysis of the internal transaction database of an Italian factoring company by the mapping and the identification of risk profiles of the companies involved in the factoring business. | Weighted in- and out-degree<br>Closeness<br>Betweenness<br>Network constraint<br>Visual analysis | 559 nodes (sellers/debtors and nodes with a double role) and 33,670 links (money transfers in Italy and abroad) |
| **(Chang et al., 2017), Taiwan** | Telecom fraud | Unveil the underlying structure of fraud groups and identify the roles of the fraudsters in telecommunication fraud. | Degree<br>Betweenness<br>Closeness<br>Eigenvector<br>Structural holes | Two telecom fraud networks (A + B). Group members and their cohorts (A 211; B 504) are defined as nodes in the network. Both flight (total 699) and co-offending records (A 4,800; B 4,905) are used to link all members) |
| **(Wei et al., 2019), China** | Mobile internet fraud | Examining the perceived impact of mobile Internet fraud by the strength of the user's social relationships and network homophily based on survey results. | Network size<br>Network homogeneity | 816 survey respondents with each their network |

# Appendix B: Case Comparison

**MA = Manure Case DT = Dog Trade Case**

| | |
|---|---|
| | **1. Political attention** |
| **MA** | The Netherlands has one of the most productive and efficient agricultural sectors in the world, which is also ecologically efficient (Grinsven & Bleeker, 2017). Nevertheless, the pressure on the surrounding living environment remains high, both because of the intensity of the use of space in the Dutch Delta and because of the scale of agriculture. Around 65% of the Dutch territory is used by any form of agriculture (Grinsven & Bleeker, 2017). Dutch livestock farming is one of them and plays a major role in the European meat market. The consequence of this is that the country is left with residues from the production chain, in particular manure-related nutrients (phosphate, nitrogen and ammonia). Manure is a societal problem with a high position on the political agenda. Due to over-fertilisation the soil, drinking water, and air become polluted. This results in a decline in biodiversity with far-reaching consequences for both humans and animals. In addition, the Dutch livestock farming is dependent on European derogation (partial suppression of the law). When the country loses its derogation, the economic consequences for the farmers will be disastrous. |
| **DT** | Dogs can have a positive influence on people's health; they can contribute to the socio-emotional and cognitive development of a child and reduce absenteeism among adults. Moreover, dogs are of added value for (lonely) elderly people; to keep company, relieve stress or give a purpose in life (LICG, 2018; Raad voor Dierenaangelegenheden, 2006). It is therefore not surprising that the demand for puppies is very high. Due to the limited number of bona fide puppies in The Netherlands, many puppies come from the illegal market (van Uhm, 2010). One could speak of illegal import if European regulations are not met. This includes the registration, chipping and vaccinating, as well as having a valid health certificate and passport.<br>Dog trade is a lucrative criminal business that generates millions of euros (Poulie & Genugten, 2018). Dogs enter the Netherlands via "unknown flows" and are usually imported from Eastern European and non-EU countries, which imposes risks to public health and animal welfare. Rabies still occurs in these countries (high risk countries), and if treatment is not initiated very quickly after infection, it may cause death. The spread of rabies is facilitated by the long incubation period of four to eight weeks after infection. As a result, puppies enter the country in good health, but after a certain number of weeks they may spread the rabies virus (McQuiston et al., 2008). To prevent this, dogs are only allowed to enter the Netherlands from the age of fifteen weeks (Tweede Kamer, 2015)<br>Even though the number of annual European victims is fairly limited, (1 to 10 cases per year) (Poulie & Genugten, 2018), the consequences can be vital. In addition to the risk of public health, also animal welfare is usually at risk. The conditions during breeding, distribution, and trade ensure that the physical and mental health of the puppies are at stake. Research by McMillan, Duffy, and Serpell (2011) shows that more than a quarter of the dogs from the so-called “puppy factories” have health problems. Besides, the circumstances in a puppy factory can lead to behavioural problems. McMillan et al. (2011) show that eight out of ten dogs suffer from behavioural problems, varying from minor problems (five out of eight) to moderate and to serious behavioural problems. The studies by both Wauthier and Williams (2018) and Gray, Butler, Douglas, and Serpell (2016) show that dogs raised in a puppy factory more often exhibit undesirable behaviour compared to dogs raised elsewhere.<br>In general, it can be stated that it is a relatively small topic within the organisation but with considerable public and political attention. Although illegal dog trade may not be as high on the political agenda as manure fraud, significant attention is given to this problem due to public interest (bottom-up). This is caused by the human empathy related to animal suffering. |
| | **2. Size of Fraud & The Fraud Triangle** |
| **MA** | An estimated 25-40% of the manure is traded or dumped illegally (OM, 2019). In circa 25% of the cases one could speak of "serious violation of the law" (Kamphuis, 2015). Due to the concealment of illegitimate activities (i.e., the numbers on paper are “correct”), precise numbers on fraud are lacking (Grinsven & Bleeker, 2017). In addition, due to the set-up of selected inspection within the NVWA, the percentage of non-compliance says little about the actual scope of fraud (Velthof et al., 2017). However, these numbers still suggest that farmers commit fraud on large scale and that it has become the standard in the industry (i.e., rationalism). Livestock farming is a competitive industry with small margins and farmers are operating in difficult times (i.e., financial pressure). Besides, there is much complexity involved in the regulation and lawsuits often remain open-ended for a long time, giving a strong incentive for fraud (i.e., opportunity). In 2016, the NVWA executed 4.064 inspections in which 527 administrative fines were imposed and 124 charge sheets were documented (NVWA, 2017b). More details about the size of fraud can be found in Appendix C. |

| | |
|---|---|
| **DT** | The main driver for illegal dog trade is assumed to be money (i.e., financial pressure). A lot of profit could be made with trading puppy's illegally. For example, the profit on a small dog can be as high as € 760[5] (De Groot, Overgaauw, & Virginia, n.d.). Internet has gained an increasingly important role in facilitating illegal dog trade. To illustrate, a daily count of the range of dogs offered on Marktplaats in 2015 (both puppies and second-hand dogs) revealed a total of 13,206 dogs (Leiden, Esseveldt, Wolsink, Wijk, & Endenburg, 2019).<br>Several estimations have been made obtain the number of puppies in the Netherlands. For example, the Raad voor Dierenaangelegenheden (2006) estimated 180,000 puppies on an annual basis. In 2015, an annual increase of 150,000 puppies was estimated (De Groot et al., n.d.; Hogeschool HAS Den Bosch & Universiteit Utrecht, 2015). Van Rijt, Verhoeven and Kok (2016) estimated the new growth at 140,000 to 160,000 puppies and Schmid (2018) comes with an estimation of 158,000 puppies. These estimates are based on the number of puppies required to maintain the existing population. It is then assumed that dogs have an average lifespan of ten years.<br>An estimated 25% of the dogs in the Netherlands are traded via "unknown flows" (Poulie & Genugten, 2018). Illegally is then defined as a blind spot consisting of dogs which neither have a registered birth date nor a registered trade number (import) (Poulie & Genugten, 2018). All in all, the exact size of the international puppy trade is difficult to establish, but the aforementioned estimates show that it is highly likely that tens of thousands of puppies per year are traded internationally (and partly illegally) (Leiden et al., 2019). In doing so, dog traders capitalise on the buyer's emotion. Trading dogs is "easy", as puppy's sell themselves through their affection (Poulie & Genugten, 2018). Many people buy a dog impulsively and are not critical on the health and origin of the dog (Leiden et al., 2019). This keeps the market attractive while the chance of getting caught is perceived as relatively low (i.e., opportunity) (Raad van Beheer, 2018; Rijt, Verhoeven, & Kok, 2016). In addition, illegal dog trade largely takes place outside the control of the authorities due to the complex organisation enforcement activities (De Groot et al., n.d.). Dog trade is often combined with other criminal activities, such as drugs dealing or illegal trade in veterinary medicines (Poulie & Genugten, 2018). Illegal dog traders hardly show empathy for the animals (i.e., rationalism).<br>In total, from 2015 to the first half of 2018, 375 inspections were carried out by the NVWA with regard to dogs[6] (Leiden et al., 2019). The majority of the inspections (52.8%) were carried out at companies, more than a third part (38.5%) among private individuals and a small part is unknown (8.5%) (Leiden et al., 2019). Violation of one of the laws or rules is registered as "non-agreement". For the total of 375 inspections, 75% was classified as non-agreement (Leiden et al., 2019). |
| | **3. Types and Variations of Fraud** |
| **MA** | On the one hand, there is an overproduction of manure. On the other hand, there are areas with a high need for manure. Transportation of manure costs money (e.g., a pig farm with 10,000 pigs without own land pays about €200,000 for manure disposal on yearly basis) (Kamphuis, 2015). These cost impose a major financial burden on the farmers who already have to deal with low margins. For that reason, farmers try to find many different ways to circumvent the costs of manure. Examples include cheating with the weights and volumes of transported and stored manure, the nitrate and phosphate content, the VDMs (transport licenses) whether or not combined with the AGR/GPS systems, or the declaration of "non-existing" pieces of land. To illustrate, the NVWA has identified over 40 different forms of fraud related to manure transportation (Suyker, 2018). |
| **DT** | Illegal dog trade occurs in variety of forms, like the age of the dog (the shorter a breeder has them, the cheaper) or lacking vaccinations through cheating with chips and passports. Whether it is a true fraud case is usually difficult to prove. Nonetheless, violations can be divided into three categories. More than a third of the violations (37.1%) relates to an article from the "Regulation on Trade of Living Animals and Living products" (Leiden et al., 2019). It appears that in those cases there is often something wrong with the trade recognition, the health certificate, or the vaccination against rabies. In addition, violations have been found regarding the (incorrectly registered) age of the puppies, the identification document, the trade register or the chip. The second category is concerned with violation of "The Decree on Animal Holders" (35.3%) (Leiden et al., 2019). This includes violations with regard to the registration and the professional skills of the trader or breeder. Articles regarding animal behaviour also fall under this legislation. The supervision of this aspect is still under development, so violations of these articles are not yet very common in the data. The last part of the violations found (27.6%) relates to "The Identification and Registration of Animals Decree" (Leiden et al., 2019). Within this legislation, the violation usually involves the failure of registering and report dogs. The table in Appendix D.1 contains an overview of the distribution of the violations found among the individual articles within the legislation and regulations. |
| | **4. (Un)common** |

[5] The cost price of a dog is €190 and the transport cost is €50, while the puppy is sold for €1000.
[6] Of which 144 (38%) in 2015, 65 (17%) in 2016, 107 (29%) in 2017 and 59 (16%) in the first half of 2018.

| | |
|---|---|
| **MA** | Although fraud is generally considered to be uncommon (i.e., the majority of transactions is legitimate), fraud in the livestock industry occurs on large scale. An estimated 25-40% of the manure is traded or dumped illegally (OM, 2019), which is substantially higher than other fraud contexts that have applied SNA. |
| **DT** | An estimated 25% of unknown dog flows (neither birth registered nor imported) make up a blind spot in the dog trade market (Poulie & Genugten, 2018). Similarly, this makes illegal dog trade not uncommon. |
| | **5. (Un)intentional** |
| **MA** | Within the manure industry, inventiveness is paramount and firms do not shy away from offending the laws. Farmers are likely to make a cost-benefit analysis and as long as the benefits of illegitimate behaviour exceed the costs, fraud is likely to occur. The 'amoral calculator' seems to be present within this industry. Since only a small portion (0,13% of the total number of transportations) is physically inspected there is plenty of room to appear normal, yet to operate "below the radar".<br>Next to the amoral calculator, also the 'political citizen' seems to be applicable; the margins in the livestock industry are rather low and the regulations that impose manure transportation result huge financial burdens on the farmers. Organisational incompetence, instead, is rather used as an argument for violating, due to the complex manure regulations[7]. In practice, framers are usually advised by intermediaries on how to hide their fraudulent activities. |
| **DT** | Since the aim of illegal dog trade is mostly financial gain, it can considered to be purposive. Unawareness does (i.e. organisational incompetence) not seem to play a major role within this context. According to Van Uhm (2010), a small number of vets (estimated to be 5% to 10% in the Netherlands) consciously act as facilitators. These veterinarians provide vaccination documentation (against extra payment) to traders, and fraud with, among other things, the identification chips, age and vaccinations in the passport. This makes it easier for puppies to be sold (Van Uhm, 2010). However, it must be taken into account that these findings date from 2010 and that the current situation is unknown. |
| | **6. (Un)concealed** |
| **MA** | Fraudsters actively conceal their activities. Accountants and intermediaries are hired to make sure that the financial statements and administrations are "correct". Besides, farmers adapt their transportation to strategical times based on the inspection time and frequency (i.e., they know when at what time inspectors usually work), and try to reduce the risk of being inspected for instance by transporting their manure at other times or via different routes. Fraudulent activities are camouflaged in such a way that "selected inspection (non-randomised) achieve almost the same results as randomised (a-selected) inspections" [8]. This makes the fraud inherently difficult to track. |
| **DT** | In a similar vein, concealment of illegitimate activities occurs when trading dogs. On paper, data is usually correct, but in fact these are false representatives of reality. Violations can only be found after thorough inspections or when inspectors act based on their tacit expertise. |
| | **7. Time-evolving** |
| **MA** | The way in which fraud occurs is time-evolving in the manure industry. The different actors in the fraud network are specialised in finding novel ways to circumvent the law. For that reason, the NVWA wants to implement more real-time inspection. Again, as long as the benefits of fraud exceed the costs (which is usually the case), fraud is likely to occur in one way or another. Due to limited inspection capacity and the low chances of getting caught, farmers are even willing to commit the same fraud even if they have already been charged for being fraudulent. |
| **DT** | Violations of the law can sometimes be recovered during the inspection. In other cases, the NVWA may decide to carry out a re-inspection later. In the period from 2015 up to mid-2018, the NVWA carried out 34 re-inspections[9]. In half of the cases (50.0%) violations were (again) detected, the majority (88.2%) with was related to "The Decree on Animal Holders"(Leiden et al., 2019). Nonetheless, it remains hard to establish to what extent this data represents the real fraud. Similarly, hard data on the way in which violations and/or fraud change over time are lacking. |
| | **8. Individual vs. Organised** |
| **MA** | Methods to reduce costs lead, in most cases, also to evasion of the law and ideas are usually exchanged quickly among the farmers[10]. Also when a particular farm is inspected this is communicated through the network. Besides, many farmers are dependent on intermediaries to dispose their manure. Not only registered intermediaries are used to coordinate fraud, but manure is also attributed to private bodies that are not officially registered at RVO. This indicates that farmers need other parties to hide their illegitimate activities. |

[7] Statement of a Domain Expert
[8] Statement of a Domain Expert
[9] Of which 13 (38%) in 2015, 12 (35%) in 2016, 8 (24%) in 2017 and 1 (3%) in the first half of 2018.
[10] Statement of a Domain Expert

| | |
|---|---|
| **DT** | Networks are of great importance in dog trade (Leiden et al., 2019). The network is very flexible, since it can consist of both registered companies and individuals who arrange everything themselves (Sabo, 2018; van Uhm, 2010)There is also some interdependence between illegal dog trade and other forms of crime. For example, there traders who are guilty of theft, property crimes, violent crimes and / or drug trafficking (van Uhm, 2010). |
| | **9. Stage of data-driven inspection** |
| **MA** | Within the manure chain, inspections are carried out based on two inputs: a notification system (external; i.e., notifications received from citizens) and an data driven outlier system (internal). The NVWA spends ±85 FTE on manure inspection (Kamphuis, 2015). Both the NVWA and the RVO are responsible for manure-related inspections, focusing on physical and administrative inspections respectively. |
| **DT** | The NVWA, the LID and the National Police are involved in the enforcement related to animal welfare. Together they make efforts to tackle, among other things, the illegal dog trade. The NVWA concluded that inspections are carried out almost fully notification-driven (~90%) (Poulie & Genugten, 2018), so in a reactive rather than proactive manner. A maximum of 70% of the notifications could be tackled, and the notifications with the highest risk profile are being followed up first (Poulie & Genugten, 2018). Once a notification enters the system it is evaluated through a number of decision criteria based on risk indicators, e.g., country of origin and risk of rabies (for the full list see appendix D.2). A lower priority is usually given to an organisation which has been inspected recently, or when the notification is repeatedly obtained through the same reporter. The NVWA spends ±5 FTE on dog trade inspection. |
| | **10. Goals of SNA** |
| **MA** | Within the manure chain, SNA has the primary goal extract patterns and identify different roles within the network. |
| **DT** | The primary goal within the dog trade case is to identify the actors that have a high possess a high risk. |
| | **11. Geographical concentration** |
| **MA** | The network analysis is based on the region Oost-Brabant/Noord-Limburg. This region is selected is based on its concentration of livestock farms and the number of manure transportations, imposing a higher risk for pollution. |
| **DT** | The project aims to get insight into the illegal import of dogs in the Netherlands. This means that the network is concentrated on the entire country. |
| | **12. Stakeholders (facilitators)** |
| **MA** | In 2017, the manure sector consisted of 25.700 manure producers (livestock farmers), 25.000 users (agrarians), and 11.00 intermediaries (e.g., transporters) (NVWA, 2018). In the barrier model (Appendix E) the facilitators of fraud within the manure chain have been identified. For each step in the chain the facilitators are as following:<br>*Source (Storage/Processing):*<br>- Farmers<br>- Licensing authority<br>- Agricultural advisor<br>- Lawyers<br>- Accountants<br>- Equipment supplier<br>- Financial institutions<br>- Government<br>- Wageningen University and Research (WUR)<br>- Social connections (friends and families)<br>- Veterinarians<br>- Landlords<br>*Transport:*<br>- Garages-Independent samplers<br>- Labs<br>- Intermediaries (transport companies)<br>- Truck manufacturers<br>- Contractors<br>*Processing:*<br>- Advisors<br>- Inspection agencies<br>- Labs<br>*Customer:*<br>- Grant providers<br>- Accountants<br>- Financial institutions<br>- Equipment supplier"<br>- Farmers<br>- Landlords<br>*Overarching:*<br>- Government |
| **DT** | Fraud facilitators related to dog trade are not identified based on a particular step in the chain. The following actors have been identified:<br>- Breeder<br>- Intermediary ("tussenhandelaar")<br>- Transporter<br>- Passport Producer<br>- Veterinarians<br>- Chipper<br>- Chip producer<br>- Buyer<br>- Dog trader |

# Appendix C: Fraud Details Manure

## C.1 – Details on supervision in 2014 and January-August 2015

| **NVWA (fysiek toezicht)** | 2014 | Stand Augustus 2015 |
|---|---|---|
| FTE | 84 | 86 |
| Kosten toezicht (mln) | €10,4 | €11,3 |
| Aantal controles | 4537 | 2608 |
| Aantal boetes (Bestuursrecht) | 527 | 340 |
| Opgelegde boetes | € 5.061.644 | € 7.065.645 |
| Proces verbalen (Strafrecht) | 124 | 143 |
| **RVO.nl (administratieve toezicht)** | | |
| Kosten toezicht (mln) | 16,4 | € 17,3 |
| Aantal bestuurlijke boetes | 513 | 370 |
| Opgelegde boetes | € 1.967.214 | € 2.188.995 |
| Totaal | | |
| Totaal opgelegde boetes | € 7.028.858 | € 9.254.640 |
| Waarvan geïnd (mln) | € 2,1 | In procedure |
| Milieubelasting (kg N) | 330.000 | 300.000 |
| Milieubelasting (kg P) | 390.000 | 600.000 |

## C2 – Type of offense in the period 2014-2016

| **Soort overtreding opgelegd in de periode 2014-2016** | **Bedrag in €** | **Aantal beschikkingen** | **Aantal boetes** |
|---|---|---|---|
| **Overtreding administratieve verplichtingen** | 173.440 | 599 | 601 |
| **Overtreding administratieve verplichtingen intermediair** | 17.400 | 49 | 56 |
| **Administratieve verplichtingen overige leveranciers en afnemers bedrijven** | 900 | 3 | 3 |
| **Vervoer van dierlijke meststoffen** | 118.800 | 106 | 396 |
| **Vervoersbewijs dierlijke meststoffen** | 774.670 | 604 | 2.725 |
| **Vervoersbewijs zuiveringsslib en compost** | 6.000 | 6 | 26 |
| **Grensoverschrijdende overbrenging** | 94.192 | 148 | 345 |
| **Hoeveelheidsbepaling** | 472.800 | 142 | 1.492 |
| **Overige bepalingen** | 1.500 | 5 | 5 |
| **Mestverwerkingsplicht** | 85.835 | 13 | 13 |
| **Gebruiksnormen dierlijke mest** | 10.675.801 | 1.025 | 1.025 |
| **Verantwoordingsplicht** | 10.215.645 | 71 | 71 |
| **Totaal** | 22.636.983 | 2.771 | 6.758 |

# Appendix D: Fraud Details Illegal Dog Trade

## D.1 – Registered violations after inspections related to dogs executed by the NVWA between 2015 and medio (n=859)*

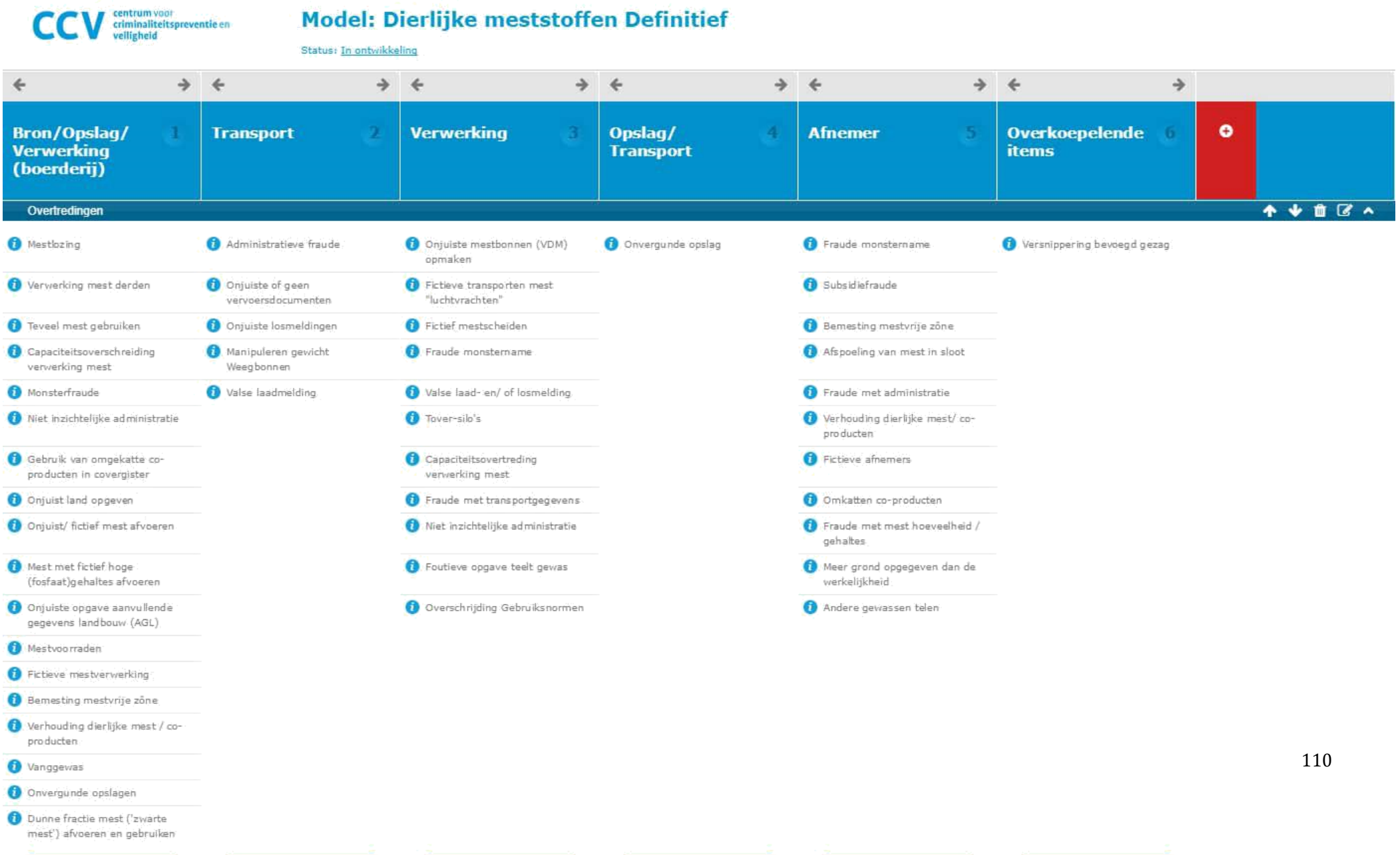
| **Besluit houders van dieren n %** | n | % |
|---|---|---|
| Aanmelden | 98 | 11,4% |
| Vakbekwaamheid | 95 | 11,1% |
| Huisvesting en verzorging | 19 | 2,2% |
| Inenting | 16 | 1,9% |
| Administratie | 13 | 1,5% |
| Verzorgen van dieren | 11 | 1,3% |
| Behuizing | 10 | 1,2% |
| Fokken 10 | 10 | 1,2% |
| Houden van dieren | 7 | 0,8% |
| Gezondheid | 6 | 0,7% |
| Diergeneeskundige ingrepen | 4 | 0,5% |
| Informatieverstrekking bij verkoop of aflevering | 4 | 0,5% |
| Socialisatie | 4 | 0,5% |
| Huisvesting honden en katten buiten de inrichting | 4 | 0,5% |
| Vastleggen of in ren houden | 2 | 0,2% |
| **Regeling handel levende dieren en levende producten** | | |
| Handelserkenning | 77 | 9,0% |
| Gezondheidscertificaat | 67 | 7,8% |
| Inenting rabiës | 66 | 7,7% |
| Leeftijd pups | 38 | 4,4% |
| Identificatiedocument | 24 | 2,8% |
| Handelsregister | 23 | 2,7% |
| Identificatie met chip | 20 | 2,3% |
| Paspoort | 2 | 0,2% |
| Overige | 2 | 0,2% |
| **Besluit identificatie en registratie van dieren** | | |
| Registreren en melden | 179 | 20,8% |
| Identificatie | 30 | 3,5% |
| Administratieplicht chipper | 28 | 3,3% |
| **Totaal** | 859 | 100% |

* In some cases multiple violations have been established, making the number of violations greater than the number of inspections executed.

## D.2 – Risk indicators for evaluating MOS notifications

| **Risk-indicators** | **Explanation** |
|---|---|
| **Country of origin** | Dogs from abroad possess a higher risk |
| **Risk of rabies** | High risk countries include Romania, Bulgaria, Turkey, Russia, Spain |
| **Non-registered traders** | Traders should be either registered in the trade register, I&R CDD and / or UBN |
| **Completeness of notification** | Only complete notifications can be followed up. Anonymous notifications usually contain too little information |
| **Reporter** | DA or sector, call from an enforcement partner (e.g. DIPO or LID), or politically sensitive notifications have higher priority |
| **Frequency of reporting** | Some people report very frequent (sometimes even under a different name), giving them a too big stake in the inspections executed |
| **Size** | Breeding and trading of more than twenty dogs on an annual basis is considered to be commercial trading. These businesses are given higher priority |
| **Location** | Consideration whether the business active just one or more locations |
| **Registration of family** | Only puppies and no mother registered animal |
| **Passport information** | Passport information is incorrect |

# Appendix E: Barrier Model Manure



- Agrarisch adviseur
- Voerleverancier
- Advocaten
- Accountant
- Apparatuur leveranciers
- Financiële instellingen
- Overheid
- Wageningen University & Research (WUR)
- Familie/ vrienden
- Dierenarts
- Verhuurder
- Boer

- Toezichthouders
- Garagebedrijven
- Onafhankelijke monsternemers
- Laboratoria
- Transportbedrijven
- Vrachtwagenbouwers
- Loonwerkers

- Toezichthouders
- Laboratoria

- Subsidieverstrekkers
- Accountant
- Financiële instellingen
- Leverancier apparatuur

- Overheid

## Gelegenheid

- Versnippering
- Onduidelijke wetgeving
- Onoverzichtelijke wetgevingscombinaties
- Beperkte zichtbaarheid handhaving
- Vaktechnische kennis ontbreekt
- Monstername door chauffeur
- Omgeving
- Gebrek samenwerking in de keten door handhavers
- Lage pakkans
- Samenspanning (collusie)met verklikkers

- Onduidelijke en complexe wetgeving
- Positie monsternemers
- Fraudegevoelige apparatuur
- Certificering

- Toelatingseisen apparatuur
- Lastig vaststellen dat er met monsters wordt gefraudeerd
- Rol intermediair
- Grote invloed balangenbehartiger (sector) richting wetgever

- Onduidelijke wetgeving
- Onduidelijk proces vergisters
- Mobiele scheider

- Economisch belang
- Cultuur in de sector
- Samenspanning (collusie) met verklikkers
- Gebrek samenwerking in de keten door handhavers
- Vaktechnische kennis ontbreekt
- Handhavingstekort in de keten
- Vrije bewijsleer
- Registratie biedt gelegenheid om te frauderen
- Lobby sector richting poltiek/ beleid

## Ketenpartner

- Politie
- NVWA
- ILT
- Waterschap

- Politie
- NVWA
- ILT
- Douane

- NVWA
- Politie
- Omgevingsdiensten (provincie, gemeente)
- Waterschap

- Politie
- NVWA
- ILT
- Douane

- Waterschap
- NVWA
- ILT
- Belastingdienst

- Bestuurstafel
- Link en overlap ketenpartners niet altijd duidelijk

- Omgevingsdiensten (provincie, gemeente)
- Belastingdienst
- RVO
- RIEC
- LTO
- Certificeringsburo's
- CUMULA

- RVO
- RIEC

- Gemeente/ Provincie
- RIEC

- RVO
- RIEC

- RVO
- RIEC
- Politie
- Omgevingsdiensten (provincie, gemeente)
- Kamer van Koophandel

## Indicatoren

- Buitenlandse invoer diervoer en kunstmest
- Data overloop opslagcapaciteit
- Mest scheiders klachten geur- en geluidsoverlast
- Fraude administratie
- Bedrijven in financiële problemen
- Co-producten
- Verhouding inrichting versus grond
- Hoge gehalten stikstof en fosfaat in afvoer mest
- Eerdere overtredingen
- Bedrijf in grensgebied
- Grondwater vervuiling per watergang/gebied

- Opvallende data en tijdstippen
- Opvallende locaties
- Veel storingen bemonsteringsapparatuur
- Te lage gewichtslading
- Attributen om nitraat toe te dienen aan mest
- Te lage of te hoge gehaltes nitraat mest
- Onjuiste voorraad administratie mest
- Vervoersbewijs Dierlijke Meststoffen (VDM)
- Bandenspanning
- Aantal ritten korte periode zelfde boer
- Laad- en losregistraties koppelen
- Veel storingen AGR/GPS
- Fictieve laad- en losmeldingen

- Verkeerde/sjoemelen met mestcode
- Netwerkanalyse covergisters
- Profilering adviesbureau

- LMA analyse. Met name bij grote vergisters
- Gehaltes nitraat in oppervlaktewater
- Het uiterlijk van de mais
- Ongewone voorvallen
- Opvallende data KvK
- Fosfaattoestand van de bodem

- Indicatoren op basis van analyse bestanden diverse instanties
- Eerdere overtredingen
- Ketsplaat niet emissiearm uitrijden

## Barrieres

- Melding buitenlandse invoer voer
- Fysieke controles NVWA/OD bij data overloop mestopslag
- Registratieplicht mestscheider
- Data verschillende bronnen

- Controleren op onverwachte data en tijdstippen
- Real time electronische administratie
- VDM querys achteraf info obv codes
- Erkenning lab intrekken

- Onopvallender controleren en zichtbaar zijn
- Milieuvluchten
- GDI mailadres RVO, wie is de adviseur

- Versterken wetgeving. Minder complex
- Uitwisseling informatie makkelijker maken
- Oog- en oorfunctie
- Steekproeven

- Coördinator bijzonderheden gebiedsgericht
- Openbaarmaking
- Stelselwijziging

samenvoegen en analyseren

- Banken minder makkelijk financieren
- Bemonsteren vracht, omgekeerde bewijslast
- Afspraken over percelen vastleggen
- Forfaitair stelsel
- GGH
- Gedragscode sector
- Perceptie pakkans vergroten
- Boeren royeren door afnemers als Campina, AH, Vion, etc
- Snellere analysemethoden toepassen
- Expertiseteam mestfraude
- Verscherpt toezicht export mest
- APK apparatuur mestauto's
- Bibob aanvraag
- Openbaarmaking
- Goed gedrag beïnvloeden of belonen
- Bij grote subsidies intensieve, materiele controle
- Nachtelijke/weekendcontroles en op regionale en B- wegen
- Cross compliance invoeren

# Appendix F: Inspection Data 2016-2018

| | Cumulative | Percentage |
|---|---|---|
| **N.o. inspections** | 3937 | 100% |
| **N.o. approved inspections** | 3357 | 85,27% |
| **N.o. disapproved inspections** | 580 | 14,73% |
| **Total n.o. companies (unique BRS numbers)** | 6017 | 100% |
| **N.o. inspected companies** | 305 | 5,07% |
| **Average compliance** | - | 77,81% |
| *Standard deviation* | - | 37% |
| **Weighted average compliance** | - | 85,27% |

# Appendix G: Data Description Illegal Dog Trade

**DM_CCD_MELDING**: I&R data (Basis voor netwerkanalyse)

Deze data bevat alle geregistreerde verplaatsing meldingen over honden. Elke regel bevat een eigen melding ten aanzien van aanvoer, vermissing, afvoer, adreswijziging, vervanging, import, dood, intrekking, gevonden, contact, geboorte en export. Voor het hondenhandel project zijn de import meldingen van belang. Chipnummer is de identificatie van een individuele hond. De combinatie van postcode en huisnummer identificeert de persoon/klant/bedrijf/stichting die de melding doet.

**DM_TCS_CERTIFIATEN**: TRACES data (Basis voor netwerkanalyse)

Deze data bevat de certificaat informatie van aangevoerde honden van het buitenland naar Nederland. Er zijn twee typen certificaten:

- Intra-Trade certificaten: dit is een certificaat per hond en kan met behulp van de tabel **DM_TCS_DIER_IDENTITEIT** gekoppeld worden aan het chipnummer
- GDB voor dieren: bij import van levende dieren of dierlijke producten in de Europese Unie moet de lading voor binnenkomst worden aangemeld bij een erkende Buitengrens Inspectie Post (BIP). Daarvoor is een Gemeenschappelijk Veterinair Document van Binnenkomst (GDB) nodig. Dit is een certificaat per partij honden die wordt ingevoerd en kan niet gekoppeld worden aan een unieke hond/chipnummer. De herkomst en bestemming van de invoer (certificaat) kunnen bepaald worden aan de hand van postcode en huisnummer.

**DM_VTE_ACTIVITEITEN**: SPIN data

Deze data bevat alle inspecties gedaan binnen de NVWA op een locatie (postcode, huisnummer). Door de filteren op het verificatie programma kunnen de inspecties gerelateerd aan honden worden geselecteerd. Er wordt aangegeven of de inspectie akkoord is of niet en of er verder acties worden ondernomen.

**DM_NHR_ONDERNEMING**: KVK data

Deze data bevat gegevens van alle bedrijven die bij KVK zijn ingeschreven met hun gegevens dus ook postcode en huisnummer

**'Overzicht meldingen gezelschapsdieren tm 20 nov 2018 incl risicokolom.xlsx'**: MOS meldingen

Deze data bevat alle MOS meldingen die gedaan zijn over gezelschapsdieren bij de NVWA. Dit is een Excel bestand, in DWH Oracle Algemeen is ook een tabel DM_MOS_MELDINGEN beschikbaar. Er moet gekeken worden of deze twee hetzelfde zijn en of het Excel bestand kan worden vervangen door de datamart.

**'Lenie_pdl_huisdieren_22-11-2018.xlsx'**: UBN data

Deze data bevat de UBN en BRS gegevens (ook postcode en huisnummer) van huisdieren opgevraagd bij RVO. Bepalen of deze dat vervangen kan worden door een datamart die al aanwezig is binnen de NVWA.

**'Handelserkenningen overige diersoorten 20 november 2018.xlsx'**: Handelserkennngen data

Deze data bevat handels erkenningen met informatie zoals soms KvK nummer en postcode, huisnummer. Dit is een Excel bestand, in DWH Oracle Algemeen is ook een tabel DM_MOS_ERKENNINGEN beschikbaar. Er moet gekeken worden of deze twee hetzelfde zijn en of het Excel bestand kan worden vervangen door de datamart.

# Appendix H: Results SNA Manure

## H.1 Results Reach Centrality

* Due to privacy considerations the ID code presented here is not the original BRS number.

| ID | Reach Centrality | Ranking Reach | No. Approved Inspections | No. Disapproved Inspections | No. Inspections | Percentage of Compliance | RF National Holidays | RF Time of Transp. | RF VDM | Firm Type |
|---|---|---|---|---|---|---|---|---|---|---|
| **453732578** | 5,0000 | 1 | 40 | 5 | 45 | 88,89% | 0,47% | 1,63% | 5,43% | Supplier/Transporter/Customer |
| **453727705** | 6,0000 | 2 | 52 | 3 | 55 | 94,55% | 0,54% | 1,49% | 2,48% | Supplier/Transporter/Customer |
| **453657449** | 6,0000 | 3 | 89 | 3 | 92 | 96,74% | 0,99% | 0,91% | 8,67% | Transporter/Customer |
| **453705707** | 6,0000 | 4 | 7 | 5 | 12 | 58,33% | 0,29% | 0,80% | 2,55% | Supplier/Customer |
| **453664354** | 6,0000 | 5 | 35 | 1 | 36 | 97,22% | 0,61% | 0,27% | 2,22% | Supplier/Transporter/Customer |
| **455072273** | 6,0000 | 6 | 64 | 5 | 69 | 92,75% | 1,01% | 0,25% | 3,21% | Supplier/Transporter/Customer |
| **475372381** | 6,0000 | 7 | 1 | 0 | 1 | 100,00% | 0,90% | 3,23% | 5,15% | Transporter |
| **455329858** | 6,0000 | 8 | 35 | 7 | 42 | 83,33% | 1,01% | 0,32% | 1,66% | Supplier/Transporter/Customer |
| **455073047** | 6,0000 | 9 | 23 | 5 | 28 | 82,14% | 0,88% | 0,62% | 10,02% | Supplier/Transporter/Customer |
| **466151559** | 6,0000 | 10 | 0 | 0 | 0 | No inspection | 0,00% | 2,04% | 11,36% | Supplier/Customer |
| **453740591** | 6,0000 | 11 | 14 | 0 | 14 | 100,00% | 1,06% | 0,00% | 3,29% | Supplier/Transporter/Customer |
| **453730396** | 6,0000 | 12 | 3 | 2 | 5 | 60,00% | 2,34% | 3,96% | 7,19% | Transporter |
| **453659656** | 6,0000 | 13 | 3 | 0 | 3 | 100,00% | 2,40% | 2,81% | 3,01% | Transporter |
| **457965193** | 6,0000 | 14 | 39 | 9 | 48 | 81,25% | 0,98% | 2,56% | 2,64% | Supplier/Transporter/Customer |
| **453763772** | 6,0000 | 15 | 104 | 11 | 115 | 90,43% | 0,39% | 18,82% | 5,92% | Supplier/Transporter/Customer |
| **475784315** | 6,0000 | 16 | 8 | 6 | 14 | 57,14% | 3,95% | 0,67% | 20,58% | Supplier/Transporter/Customer |
| **453733323** | 6,0000 | 17 | 0 | 1 | 1 | 0,00% | 0,00% | 1,59% | 2,92% | Supplier/Customer |
| **458157077** | 6,0000 | 18 | 11 | 1 | 12 | 91,67% | 1,03% | 7,38% | 4,29% | Supplier/Transporter/Customer |
| **458076829** | 6,0000 | 19 | 38 | 1 | 39 | 97,44% | 0,00% | 0,17% | 1,69% | Supplier/Transporter/Customer |
| **453666102** | 6,0000 | 20 | 70 | 3 | 73 | 95,89% | 0,61% | 0,54% | 3,81% | Transporter/Customer |
| **460681901** | 6,0000 | 21 | 8 | 0 | 8 | 100,00% | 0,49% | 0,38% | 2,81% | Supplier/Customer |
| **460953609** | 6,0000 | 22 | 10 | 1 | 11 | 90,91% | 4,66% | 7,07% | 3,26% | Transporter/Customer |
| **455072249** | 6,0000 | 23 | 1 | 0 | 1 | 100,00% | 0,28% | 0,19% | 5,50% | Transporter |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **454075775** | 6,0000 | 24 | 0 | 0 | 0 | No inspection | 0,00% | 5,31% | 11,84% | Supplier/Customer |
| **458737629** | 6,0000 | 25 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 4,26% | Supplier/Customer |
| **458557766** | 6,0000 | 26 | 1 | 2 | 3 | 33,33% | 0,84% | 1,18% | 2,93% | Supplier/Transporter/Customer |
| **460187954** | 6,0000 | 27 | 50 | 15 | 65 | 76,92% | 0,00% | 7,98% | 2,60% | Supplier/Transporter/Customer |
| **460211053** | 6,0000 | 28 | 84 | 3 | 87 | 96,55% | 0,25% | 5,74% | 6,65% | Transporter/Customer |
| **453714126** | 6,0000 | 29 | 11 | 6 | 17 | 64,71% | 0,20% | 0,71% | 3,76% | Transporter/Customer |
| **453687736** | 6,0000 | 30 | 15 | 3 | 18 | 83,33% | 2,99% | 7,83% | 4,39% | Supplier/Transporter/Customer |
| **454425412** | 6,0000 | 31 | 30 | 8 | 38 | 78,95% | 0,00% | 7,31% | 11,42% | Supplier/Transporter/Customer |
| **454043792** | 6,0000 | 32 | 1 | 0 | 1 | 100,00% | 0,00% | 0,49% | 4,16% | Supplier |
| **453508287** | 6,0000 | 33 | 14 | 0 | 14 | 100,00% | 0,00% | 0,34% | 3,38% | Transporter/Customer |
| **458028422** | 6,0000 | 34 | 0 | 0 | 0 | No inspection | 0,81% | 2,42% | 2,42% | Supplier |
| **457698856** | 6,0000 | 35 | 9 | 0 | 9 | 100,00% | 0,00% | 0,06% | 0,89% | Supplier/Transporter/Customer |
| **453707783** | 6,0000 | 36 | 50 | 14 | 64 | 78,13% | 0,87% | 0,79% | 7,80% | Supplier/Transporter/Customer |
| **455414006** | 6,0000 | 37 | 1 | 0 | 1 | 100,00% | 0,00% | 0,67% | 6,44% | Supplier/Customer |
| **455003738** | 6,0000 | 38 | 3 | 1 | 4 | 75,00% | 0,00% | 1,39% | 0,00% | Supplier |
| **453706451** | 6,0000 | 39 | 11 | 1 | 12 | 91,67% | 1,09% | 1,09% | 6,20% | Transporter/Customer |
| **457881106** | 6,0000 | 40 | 5 | 1 | 6 | 83,33% | 1,78% | 1,63% | 14,30% | Supplier/Transporter |
| **455077561** | 6,0000 | 41 | 2 | 0 | 2 | 100,00% | 0,53% | 0,00% | 5,06% | Supplier/Transporter/Customer |
| **453706296** | 6,0000 | 42 | 16 | 2 | 18 | 88,89% | 0,14% | 0,68% | 5,77% | Transporter/Customer |
| **459820374** | 6,0000 | 43 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,83% | Customer |
| **450884222** | 6,0000 | 44 | 1 | 0 | 1 | 100,00% | 4,59% | 0,00% | 0,00% | Supplier |
| **453279446** | 6,0000 | 45 | 1 | 0 | 1 | 100,00% | 2,43% | 1,62% | 0,00% | Supplier/Customer |
| **459892734** | 6,0000 | 46 | 2 | 0 | 2 | 100,00% | 1,76% | 0,65% | 4,57% | Supplier/Customer |
| **450874955** | 6,0000 | 47 | 0 | 0 | 0 | No inspection | 1,63% | 0,30% | 1,78% | Supplier |
| **453732734** | 6,0000 | 48 | 15 | 2 | 17 | 88,24% | 0,00% | 1,80% | 11,42% | Transporter |
| **474386564** | 6,0000 | 49 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,95% | Supplier |
| **248409917** | 6,0000 | 50 | 4 | 0 | 4 | 100,00% | 2,96% | 4,03% | 15,56% | Supplier |
| **479020900** | 6,0000 | 51 | 2 | 1 | 3 | 66,67% | 0,00% | 1,80% | 18,92% | Transporter |
| **247522533** | 6,0000 | 52 | 0 | 0 | 0 | No inspection | 3,19% | 3,19% | 0,00% | Supplier |
| **270352196** | 6,0000 | 53 | 0 | 0 | 0 | No inspection | 2,95% | 0,00% | 1,31% | Supplier |
| **248307905** | 6,0000 | 54 | 1 | 0 | 1 | 100,00% | 0,00% | 0,00% | 29,51% | Supplier |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **458366762** | 6,0000 | 55 | 4 | 0 | 4 | 100,00% | 0,11% | 5,04% | 0,22% | Supplier/Customer |
| **453689124** | 6,0000 | 56 | 2 | 0 | 2 | 100,00% | 3,64% | 4,47% | 14,78% | Supplier/Transporter/Customer |
| **450557122** | 6,0000 | 57 | 0 | 0 | 0 | No inspection | 0,74% | 2,59% | 1,85% | Supplier |
| **479052679** | 6,0000 | 58 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 100,00% | Supplier |
| **458814478** | 6,0000 | 59 | 1 | 1 | 2 | 50,00% | 1,46% | 0,73% | 3,16% | Supplier |
| **248074652** | 6,0000 | 60 | 0 | 0 | 0 | No inspection | 0,00% | 2,76% | 0,00% | Supplier |
| **459785051** | 6,0000 | 61 | 1 | 0 | 1 | 100,00% | 0,54% | 0,42% | 1,45% | Supplier/Customer |
| **453983651** | 6,0000 | 62 | 0 | 0 | 0 | No inspection | 0,00% | 1,16% | 8,09% | Supplier |
| **450775334** | 6,0000 | 63 | 0 | 0 | 0 | No inspection | 5,10% | 0,00% | 27,55% | Supplier |
| **248209076** | 6,0000 | 64 | 1 | 0 | 1 | 100,00% | 0,00% | 1,89% | 0,00% | Supplier |
| **248210491** | 6,0000 | 65 | 0 | 0 | 0 | No inspection | 1,36% | 1,81% | 0,45% | Supplier |
| **453126995** | 6,0000 | 66 | 0 | 0 | 0 | No inspection | 0,82% | 0,00% | 0,00% | Supplier |
| **455277397** | 6,0000 | 67 | 0 | 0 | 0 | No inspection | 0,00% | 1,45% | 0,73% | Supplier |
| **458631074** | 6,0000 | 68 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,00% | Customer |
| **455078173** | 6,0000 | 69 | 0 | 0 | 0 | No inspection | 2,74% | 0,27% | 11,23% | Customer |
| **453181776** | 6,0000 | 70 | 0 | 0 | 0 | No inspection | 11,18% | 0,00% | 0,00% | Supplier |
| **451641068** | 6,0000 | 71 | 0 | 0 | 0 | No inspection | 2,35% | 15,29% | 4,71% | Supplier |
| **453658963** | 6,0000 | 72 | 0 | 0 | 0 | No inspection | 0,00% | 1,42% | 0,00% | Customer |
| **450648886** | 6,0000 | 73 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,00% | Customer |
| **458138515** | 6,0000 | 74 | 2 | 0 | 2 | 100,00% | 0,00% | 1,00% | 0,00% | Supplier |
| **476051174** | 6,0000 | 75 | 0 | 0 | 0 | No inspection | 0,00% | 2,27% | 0,76% | Supplier |
| **476258451** | 6,0000 | 76 | 0 | 0 | 0 | No inspection | 0,00% | 4,20% | 47,90% | Supplier |
| **474380786** | 6,0000 | 77 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 5,26% | Supplier |
| **460472210** | 6,0000 | 78 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,00% | Supplier |
| **460253690** | 6,0000 | 79 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 3,60% | Supplier |
| **270318599** | 6,0000 | 80 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 3,33% | Customer |
| **248025984** | 6,0000 | 81 | 1 | 0 | 1 | 100,00% | 11,67% | 0,00% | 6,67% | Supplier |
| **248208412** | 6,0000 | 82 | 0 | 0 | 0 | No inspection | 0,00% | 0,33% | 0,00% | Supplier |
| **451469090** | 6,0000 | 83 | 1 | 0 | 1 | 100,00% | 1,25% | 0,50% | 7,77% | Supplier/Customer |
| **450748844** | 6,0000 | 84 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 9,21% | Supplier |
| **476998996** | 6,0000 | 85 | 0 | 0 | 0 | No inspection | 0,00% | 1,50% | 89,00% | Supplier |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **455329604** | 6,0000 | 86 | 1 | 0 | 1 | 100,00% | 0,43% | 1,18% | 2,45% | Supplier/Customer |
| **452793739** | 6,0000 | 87 | 0 | 1 | 1 | 0,00% | 0,00% | 0,00% | 12,00% | Supplier |
| **458356187** | 6,0000 | 88 | 0 | 0 | 0 | No inspection | 8,43% | 1,92% | 1,53% | Supplier |
| **460710778** | 6,0000 | 89 | 0 | 0 | 0 | No inspection | 0,00% | 4,21% | 23,16% | Supplier |
| **458142236** | 6,0000 | 90 | 0 | 0 | 0 | No inspection | 1,68% | 1,68% | 8,40% | Supplier |
| **450048417** | 6,0000 | 91 | 3 | 0 | 3 | 100,00% | 0,00% | 0,00% | 0,00% | Supplier/Customer |
| **452917412** | 6,0000 | 92 | 0 | 0 | 0 | No inspection | 0,00% | 1,96% | 0,00% | Supplier |
| **458387324** | 6,0000 | 93 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,00% | Supplier |
| **474385949** | 6,0000 | 94 | 0 | 0 | 0 | No inspection | 1,11% | 0,74% | 0,74% | Supplier |
| **247901307** | 6,0000 | 95 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,94% | Supplier |
| **460198302** | 6,0000 | 96 | 0 | 0 | 0 | No inspection | 1,36% | 0,68% | 1,36% | Supplier |
| **455315206** | 6,0000 | 97 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,00% | Customer |
| **465512703** | 6,0000 | 98 | 0 | 1 | 1 | 0,00% | 0,39% | 10,94% | 3,52% | Supplier/Customer |
| **455212271** | 6,0000 | 99 | 0 | 0 | 0 | No inspection | 0,92% | 1,45% | 0,00% | Supplier/Transporter/Customer |
| **474337561** | 6,0000 | 100 | 0 | 0 | 0 | No inspection | 2,04% | 0,00% | 8,16% | Supplier |

### H.2 Results Closeness Centrality

* Due to privacy considerations the ID code presented here is not the original BRS number.

| ID | Closeness Centrality | Ranking Closeness | No. Approved Inspections | No. Disapproved Inspections | No. Inspections | Percentage of Compliance | RF National Holidays | RF Time of Transp. | RF VDM | Firm type |
|---|---|---|---|---|---|---|---|---|---|---|
| **453727705** | 1,0000 | 1 | 52 | 3 | 55 | 94,55% | 0,54% | 1,49% | 2,48% | Supplier/Transporter/Customer |
| **453732578** | 0,9964 | 2 | 40 | 5 | 45 | 88,89% | 0,47% | 1,63% | 5,43% | Supplier/Transporter/Customer |
| **453705707** | 0,9673 | 3 | 7 | 5 | 12 | 58,33% | 0,29% | 0,80% | 2,55% | Supplier/Customer |
| **453730396** | 0,9653 | 4 | 3 | 2 | 5 | 60,00% | 2,34% | 3,96% | 7,19% | Transporter |
| **455084180** | 0,9608 | 5 | 59 | 5 | 64 | 92,19% | 0,76% | 0,50% | 2,08% | Transporter |
| **460681901** | 0,9504 | 6 | 8 | 0 | 8 | 100,00% | 0,49% | 0,38% | 2,81% | Supplier/Customer |
| **453657449** | 0,9486 | 7 | 89 | 3 | 92 | 96,74% | 0,99% | 0,91% | 8,67% | Transporter/Customer |
| **452772998** | 0,9468 | 8 | 31 | 5 | 36 | 86,11% | 0,33% | 5,11% | 6,27% | Transporter/Customer |
| **453659656** | 0,9452 | 9 | 3 | 0 | 3 | 100,00% | 2,40% | 2,81% | 3,01% | Transporter |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **457698856** | 0,9407 | 10 | 9 | 0 | 9 | 100,00% | 0,00% | 0,06% | 0,89% | Supplier/Transporter/Customer |
| **453664354** | 0,9401 | 11 | 35 | 1 | 36 | 97,22% | 0,61% | 0,27% | 2,22% | Supplier/Transporter/Customer |
| **458737629** | 0,9393 | 12 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 4,26% | Supplier/Customer |
| **455073047** | 0,9386 | 13 | 23 | 5 | 28 | 82,14% | 0,88% | 0,62% | 10,02% | Supplier/Transporter/Customer |
| **460953609** | 0,9376 | 14 | 10 | 1 | 11 | 90,91% | 4,66% | 7,07% | 3,26% | Transporter/Customer |
| **453508287** | 0,9358 | 15 | 14 | 0 | 14 | 100,00% | 0,00% | 0,34% | 3,38% | Transporter/Customer |
| **454342781** | 0,9339 | 16 | 17 | 1 | 18 | 94,44% | 0,00% | 0,00% | 4,03% | Transporter/Customer |
| **457965193** | 0,9318 | 17 | 39 | 9 | 48 | 81,25% | 0,98% | 2,56% | 2,64% | Supplier/Transporter/Customer |
| **458076829** | 0,9315 | 18 | 38 | 1 | 39 | 97,44% | 0,00% | 0,17% | 1,69% | Supplier/Transporter/Customer |
| **453707783** | 0,9309 | 19 | 50 | 14 | 64 | 78,13% | 0,87% | 0,79% | 7,80% | Supplier/Transporter/Customer |
| **458724077** | 0,9278 | 20 | 28 | 1 | 29 | 96,55% | 1,08% | 0,54% | 4,74% | Transporter/Customer |
| **453763772** | 0,9272 | 21 | 104 | 11 | 115 | 90,43% | 0,39% | 18,82% | 5,92% | Supplier/Transporter/Customer |
| **466151559** | 0,9268 | 22 | 0 | 0 | 0 | No inspection | 0,00% | 2,04% | 11,36% | Supplier/Customer |
| **475784315** | 0,9264 | 23 | 8 | 6 | 14 | 57,14% | 3,95% | 0,67% | 20,58% | Supplier/Transporter/Customer |
| **453714126** | 0,9247 | 24 | 11 | 6 | 17 | 64,71% | 0,20% | 0,71% | 3,76% | Transporter/Customer |
| **455329858** | 0,9222 | 25 | 35 | 7 | 42 | 83,33% | 1,01% | 0,32% | 1,66% | Supplier/Transporter/Customer |
| **453740591** | 0,9169 | 26 | 14 | 0 | 14 | 100,00% | 1,06% | 0,00% | 3,29% | Supplier/Transporter/Customer |
| **460211053** | 0,9161 | 27 | 84 | 3 | 87 | 96,55% | 0,25% | 5,74% | 6,65% | Transporter/Customer |
| **453666102** | 0,9153 | 28 | 70 | 3 | 73 | 95,89% | 0,61% | 0,54% | 3,81% | Transporter/Customer |
| **458557766** | 0,9148 | 29 | 1 | 2 | 3 | 33,33% | 0,84% | 1,18% | 2,93% | Supplier/Transporter/Customer |
| **453727676** | 0,9107 | 30 | 2 | 1 | 3 | 66,67% | 0,00% | 1,59% | 8,76% | Supplier/Transporter/Customer |
| **453687736** | 0,9106 | 31 | 15 | 3 | 18 | 83,33% | 2,99% | 7,83% | 4,39% | Supplier/Transporter/Customer |
| **454075775** | 0,9097 | 32 | 0 | 0 | 0 | No inspection | 0,00% | 5,31% | 11,84% | Supplier/Customer |
| **454990495** | 0,9088 | 33 | 6 | 0 | 6 | 100,00% | 1,34% | 2,16% | 8,32% | Transporter |
| **455072273** | 0,9052 | 34 | 64 | 5 | 69 | 92,75% | 1,01% | 0,25% | 3,21% | Supplier/Transporter/Customer |
| **458028422** | 0,9052 | 35 | 0 | 0 | 0 | No inspection | 0,81% | 2,42% | 2,42% | Supplier |
| **453706451** | 0,9051 | 36 | 11 | 1 | 12 | 91,67% | 1,09% | 1,09% | 6,20% | Transporter/Customer |
| **455414006** | 0,9040 | 37 | 1 | 0 | 1 | 100,00% | 0,00% | 0,67% | 6,44% | Supplier/Customer |
| **454425412** | 0,9039 | 38 | 30 | 8 | 38 | 78,95% | 0,00% | 7,31% | 11,42% | Supplier/Transporter/Customer |
| **453733323** | 0,8981 | 39 | 0 | 1 | 1 | 0,00% | 0,00% | 1,59% | 2,92% | Supplier/Customer |
| **474180774** | 0,8975 | 40 | 8 | 1 | 9 | 88,89% | 0,78% | 1,55% | 2,91% | Supplier/Transporter/Customer |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **460492159** | 0,8941 | 41 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 3,37% | Transporter |
| **460238188** | 0,8932 | 42 | 24 | 0 | 24 | 100,00% | 1,67% | 0,00% | 21,04% | Transporter |
| **455367028** | 0,8908 | 43 | 4 | 1 | 5 | 80,00% | 0,00% | 0,00% | 8,16% | Transporter |
| **453710045** | 0,8903 | 44 | 12 | 0 | 12 | 100,00% | 0,12% | 0,00% | 3,41% | Supplier/Transporter |
| **475605458** | 0,8875 | 45 | 4 | 1 | 5 | 80,00% | 1,56% | 4,29% | 7,01% | Supplier/Transporter/Customer |
| **454839050** | 0,8842 | 46 | 9 | 1 | 10 | 90,00% | 0,07% | 0,63% | 7,99% | Supplier/Transporter/Customer |
| **458366762** | 0,8842 | 47 | 4 | 0 | 4 | 100,00% | 0,11% | 5,04% | 0,22% | Supplier/Customer |
| **455458337** | 0,8834 | 48 | 7 | 4 | 11 | 63,64% | 0,86% | 2,15% | 2,58% | Supplier/Transporter/Customer |
| **453706296** | 0,8792 | 49 | 16 | 2 | 18 | 88,89% | 0,14% | 0,68% | 5,77% | Transporter/Customer |
| **460187954** | 0,8787 | 50 | 50 | 15 | 65 | 76,92% | 0,00% | 7,98% | 2,60% | Supplier/Transporter/Customer |
| **454566057** | 0,8778 | 51 | 64 | 1 | 65 | 98,46% | 0,05% | 0,19% | 6,09% | Transporter/Customer |
| **460501757** | 0,8743 | 52 | 2 | 0 | 2 | 100,00% | 0,00% | 31,58% | 0,00% | Transporter |
| **455113496** | 0,8742 | 53 | 5 | 0 | 5 | 100,00% | 0,00% | 0,00% | 0,00% | Transporter |
| **455258864** | 0,8733 | 54 | 8 | 1 | 9 | 88,89% | 0,91% | 0,00% | 4,55% | Transporter/Customer |
| **459826865** | 0,8729 | 55 | 23 | 4 | 27 | 85,19% | 0,54% | 0,00% | 0,00% | Transporter/Customer |
| **459988503** | 0,8723 | 56 | 8 | 2 | 10 | 80,00% | 1,26% | 0,32% | 3,03% | Supplier/Transporter/Customer |
| **458063714** | 0,8712 | 57 | 2 | 1 | 3 | 66,67% | 0,00% | 0,00% | 7,23% | Transporter/Customer |
| **455029562** | 0,8706 | 58 | 1 | 0 | 1 | 100,00% | 0,53% | 1,24% | 2,08% | Supplier/Customer |
| **453665897** | 0,8705 | 59 | 0 | 0 | 0 | No inspection | 0,00% | 0,77% | 10,00% | Transporter |
| **455003738** | 0,8699 | 60 | 3 | 1 | 4 | 75,00% | 0,00% | 1,39% | 0,00% | Supplier |
| **474316280** | 0,8699 | 61 | 4 | 1 | 5 | 80,00% | 0,00% | 0,00% | 7,28% | Transporter |
| **475372381** | 0,8698 | 62 | 1 | 0 | 1 | 100,00% | 0,90% | 3,23% | 5,15% | Transporter |
| **459765751** | 0,8688 | 63 | 6 | 1 | 7 | 85,71% | 0,00% | 0,00% | 0,00% | Transporter |
| **473882885** | 0,8687 | 64 | 1 | 0 | 1 | 100,00% | 0,00% | 0,00% | 2,33% | Transporter |
| **454634592** | 0,8675 | 65 | 9 | 0 | 9 | 100,00% | 0,00% | 0,00% | 3,54% | Transporter/Customer |
| **457878669** | 0,8674 | 66 | 2 | 0 | 2 | 100,00% | 0,22% | 0,11% | 0,86% | Supplier/Transporter/Customer |
| **457881106** | 0,8674 | 67 | 5 | 1 | 6 | 83,33% | 1,78% | 1,63% | 14,30% | Supplier/Transporter |
| **455386306** | 0,8672 | 68 | 29 | 1 | 30 | 96,67% | 0,61% | 6,14% | 2,72% | Supplier/Transporter/Customer |
| **475435041** | 0,8666 | 69 | 4 | 1 | 5 | 80,00% | 0,50% | 0,17% | 2,91% | Transporter |
| **458157077** | 0,8665 | 70 | 11 | 1 | 12 | 91,67% | 1,03% | 7,38% | 4,29% | Supplier/Transporter/Customer |
| **453703295** | 0,8662 | 71 | 12 | 0 | 12 | 100,00% | 0,00% | 1,43% | 7,14% | Transporter/Customer |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **454581512** | 0,8650 | 72 | 0 | 0 | 0 | No inspection | 1,80% | 0,13% | 2,57% | Transporter |
| **451334994** | 0,8648 | 73 | 0 | 1 | 1 | 0,00% | 0,00% | 3,21% | 0,17% | Supplier |
| **453707709** | 0,8643 | 74 | 8 | 3 | 11 | 72,73% | 0,00% | 11,11% | 0,00% | Supplier/Transporter/Customer |
| **453470501** | 0,8626 | 75 | 11 | 1 | 12 | 91,67% | 0,99% | 0,47% | 2,37% | Supplier/Transporter/Customer |
| **461653523** | 0,8618 | 76 | 2 | 3 | 5 | 40,00% | 2,22% | 2,42% | 4,75% | Supplier/Transporter/Customer |
| **458237186** | 0,8613 | 77 | 1 | 0 | 1 | 100,00% | 8,40% | 0,00% | 0,00% | Transporter |
| **460201947** | 0,8610 | 78 | 2 | 1 | 3 | 66,67% | 0,05% | 0,10% | 5,15% | Supplier/Transporter/Customer |
| **450557122** | 0,8600 | 79 | 0 | 0 | 0 | No inspection | 0,74% | 2,59% | 1,85% | Supplier |
| **453663069** | 0,8599 | 80 | 13 | 0 | 13 | 100,00% | 0,00% | 0,14% | 4,47% | Transporter/Customer |
| **248236674** | 0,8592 | 81 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,00% | Supplier |
| **458074363** | 0,8572 | 82 | 1 | 3 | 4 | 25,00% | 2,06% | 3,09% | 5,15% | Transporter |
| **450683439** | 0,8565 | 83 | 0 | 0 | 0 | No inspection | 0,00% | 2,02% | 0,00% | Supplier |
| **474386564** | 0,8560 | 84 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 0,95% | Supplier |
| **454603301** | 0,8551 | 85 | 4 | 1 | 5 | 80,00% | 0,44% | 0,44% | 11,95% | Transporter/Customer |
| **458505245** | 0,8551 | 86 | 15 | 6 | 21 | 71,43% | 0,00% | 0,13% | 37,12% | Supplier/Transporter/Customer |
| **461087907** | 0,8545 | 87 | 4 | 0 | 4 | 100,00% | 0,00% | 1,30% | 0,43% | Supplier/Transporter/Customer |
| **461431570** | 0,8541 | 88 | 1 | 1 | 2 | 50,00% | 0,25% | 1,51% | 7,54% | Transporter/Customer |
| **460253690** | 0,8534 | 89 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 3,60% | Supplier |
| **450083396** | 0,8519 | 90 | 1 | 0 | 1 | 100,00% | 1,08% | 16,13% | 2,15% | Supplier |
| **451225361** | 0,8519 | 91 | 0 | 0 | 0 | No inspection | 5,41% | 0,00% | 4,32% | Supplier |
| **458138515** | 0,8505 | 92 | 2 | 0 | 2 | 100,00% | 0,00% | 1,00% | 0,00% | Supplier |
| **461115528** | 0,8503 | 93 | 0 | 0 | 0 | No inspection | 0,00% | 0,75% | 0,75% | Transporter |
| **453726728** | 0,8492 | 94 | 10 | 0 | 10 | 100,00% | 0,00% | 0,00% | 0,39% | Transporter/Customer |
| **454079785** | 0,8491 | 95 | 26 | 10 | 36 | 72,22% | 0,00% | 6,01% | 4,37% | Transporter/Customer |
| **452949835** | 0,8463 | 96 | 1 | 0 | 1 | 100,00% | 11,11% | 5,56% | 2,78% | Supplier |
| **455072249** | 0,8461 | 97 | 1 | 0 | 1 | 100,00% | 0,28% | 0,19% | 5,50% | Transporter |
| **453732734** | 0,8461 | 98 | 15 | 2 | 17 | 88,24% | 0,00% | 1,80% | 11,42% | Transporter |
| **460301741** | 0,8456 | 99 | 0 | 0 | 0 | No inspection | 0,00% | 12,54% | 4,18% | Supplier |
| **453775574** | 0,8454 | 100 | 7 | 2 | 9 | 77,78% | 6,12% | 0,00% | 4,08% | Transporter |

### H.3 Results Stress Centrality

* Due to privacy considerations the ID code presented here is not the original BRS number.

| ID | Stress Centrality | Ranking Stress | No. Approved Inspections | No. Disapproved Inspections | No. Inspections | Percentage of Compliance | RF National Holidays | RF Time of Transp. | RF VDM | Firm type |
|---|---|---|---|---|---|---|---|---|---|---|
| **455084180** | 1,0000 | 1 | 59 | 5 | 64 | 92,19% | 0,76% | 0,50% | 2,08% | Transporter |
| **453727705** | 0,9683 | 2 | 52 | 3 | 55 | 94,55% | 0,54% | 1,49% | 2,48% | Supplier/Transporter/Customer |
| **453732578** | 0,5367 | 3 | 40 | 5 | 45 | 88,89% | 0,47% | 1,63% | 5,43% | Supplier/Transporter/Customer |
| **453657449** | 0,4459 | 4 | 89 | 3 | 92 | 96,74% | 0,99% | 0,91% | 8,67% | Transporter/Customer |
| **454566057** | 0,4249 | 5 | 64 | 1 | 65 | 98,46% | 0,05% | 0,19% | 6,09% | Transporter/Customer |
| **453705707** | 0,3765 | 6 | 7 | 5 | 12 | 58,33% | 0,29% | 0,80% | 2,55% | Supplier/Customer |
| **453664354** | 0,3230 | 7 | 35 | 1 | 36 | 97,22% | 0,61% | 0,27% | 2,22% | Supplier/Transporter/Customer |
| **458792050** | 0,3196 | 8 | 12 | 2 | 14 | 85,71% | 2,28% | 0,27% | 4,56% | Transporter |
| **455072273** | 0,3164 | 9 | 64 | 5 | 69 | 92,75% | 1,01% | 0,25% | 3,21% | Supplier/Transporter/Customer |
| **475372381** | 0,3158 | 10 | 1 | 0 | 1 | 100,00% | 0,90% | 3,23% | 5,15% | Transporter |
| **455329858** | 0,3043 | 11 | 35 | 7 | 42 | 83,33% | 1,01% | 0,32% | 1,66% | Supplier/Transporter/Customer |
| **455073047** | 0,2973 | 12 | 23 | 5 | 28 | 82,14% | 0,88% | 0,62% | 10,02% | Supplier/Transporter/Customer |
| **466151559** | 0,2537 | 13 | 0 | 0 | 0 | No inspection | 0,00% | 2,04% | 11,36% | Supplier/Customer |
| **453740591** | 0,2338 | 14 | 14 | 0 | 14 | 100,00% | 1,06% | 0,00% | 3,29% | Supplier/Transporter/Customer |
| **453730396** | 0,2318 | 15 | 3 | 2 | 5 | 60,00% | 2,34% | 3,96% | 7,19% | Transporter |
| **454839050** | 0,2295 | 16 | 9 | 1 | 10 | 90,00% | 0,07% | 0,63% | 7,99% | Supplier/Transporter/Customer |
| **460492159** | 0,2286 | 17 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 3,37% | Transporter |
| **453659656** | 0,2258 | 18 | 3 | 0 | 3 | 100,00% | 2,40% | 2,81% | 3,01% | Transporter |
| **454990495** | 0,2190 | 19 | 6 | 0 | 6 | 100,00% | 1,34% | 2,16% | 8,32% | Transporter |
| **457965193** | 0,2168 | 20 | 39 | 9 | 48 | 81,25% | 0,98% | 2,56% | 2,64% | Supplier/Transporter/Customer |
| **474180774** | 0,2114 | 21 | 8 | 1 | 9 | 88,89% | 0,78% | 1,55% | 2,91% | Supplier/Transporter/Customer |
| **270126771** | 0,2035 | 22 | 12 | 1 | 13 | 92,31% | 0,95% | 0,09% | 0,38% | Transporter |
| **453763772** | 0,2029 | 23 | 104 | 11 | 115 | 90,43% | 0,39% | 18,82% | 5,92% | Supplier/Transporter/Customer |
| **475784315** | 0,2027 | 24 | 8 | 6 | 14 | 57,14% | 3,95% | 0,67% | 20,58% | Supplier/Transporter/Customer |
| **459988503** | 0,1931 | 25 | 8 | 2 | 10 | 80,00% | 1,26% | 0,32% | 3,03% | Supplier/Transporter/Customer |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **453733323** | 0,1910 | 26 | 0 | 1 | 1 | 0,00% | 0,00% | 1,59% | 2,92% | Supplier/Customer |
| **453764927** | 0,1890 | 27 | 4 | 1 | 5 | 80,00% | 1,06% | 0,56% | 4,61% | Transporter |
| **458157077** | 0,1855 | 28 | 11 | 1 | 12 | 91,67% | 1,03% | 7,38% | 4,29% | Supplier/Transporter/Customer |
| **458076829** | 0,1848 | 29 | 38 | 1 | 39 | 97,44% | 0,00% | 0,17% | 1,69% | Supplier/Transporter/Customer |
| **461582936** | 0,1792 | 30 | 15 | 0 | 15 | 100,00% | 3,27% | 0,00% | 9,07% | Transporter |
| **453666102** | 0,1774 | 31 | 70 | 3 | 73 | 95,89% | 0,61% | 0,54% | 3,81% | Transporter/Customer |
| **460681901** | 0,1735 | 32 | 8 | 0 | 8 | 100,00% | 0,49% | 0,38% | 2,81% | Supplier/Customer |
| **453688944** | 0,1701 | 33 | 15 | 1 | 16 | 93,75% | 2,40% | 0,68% | 5,39% | Transporter |
| **460953609** | 0,1660 | 34 | 10 | 1 | 11 | 90,91% | 4,66% | 7,07% | 3,26% | Transporter/Customer |
| **455072249** | 0,1660 | 35 | 1 | 0 | 1 | 100,00% | 0,28% | 0,19% | 5,50% | Transporter |
| **454342781** | 0,1607 | 36 | 17 | 1 | 18 | 94,44% | 0,00% | 0,00% | 4,03% | Transporter/Customer |
| **454075775** | 0,1606 | 37 | 0 | 0 | 0 | No inspection | 0,00% | 5,31% | 11,84% | Supplier/Customer |
| **458383524** | 0,1590 | 38 | 11 | 2 | 13 | 84,62% | 0,49% | 0,12% | 3,68% | Transporter/Customer |
| **453710045** | 0,1588 | 39 | 12 | 0 | 12 | 100,00% | 0,12% | 0,00% | 3,41% | Supplier/Transporter |
| **453713359** | 0,1569 | 40 | 3 | 0 | 3 | 100,00% | 3,96% | 7,23% | 13,43% | Supplier/Transporter |
| **453726728** | 0,1533 | 41 | 10 | 0 | 10 | 100,00% | 0,00% | 0,00% | 0,39% | Transporter/Customer |
| **460942854** | 0,1515 | 42 | 7 | 1 | 8 | 87,50% | 7,05% | 1,24% | 8,30% | Transporter |
| **458737629** | 0,1460 | 43 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 4,26% | Supplier/Customer |
| **458557766** | 0,1439 | 44 | 1 | 2 | 3 | 33,33% | 0,84% | 1,18% | 2,93% | Supplier/Transporter/Customer |
| **460187954** | 0,1414 | 45 | 50 | 15 | 65 | 76,92% | 0,00% | 7,98% | 2,60% | Supplier/Transporter/Customer |
| **455386306** | 0,1411 | 46 | 29 | 1 | 30 | 96,67% | 0,61% | 6,14% | 2,72% | Supplier/Transporter/Customer |
| **460238188** | 0,1411 | 47 | 24 | 0 | 24 | 100,00% | 1,67% | 0,00% | 21,04% | Transporter |
| **460211053** | 0,1359 | 48 | 84 | 3 | 87 | 96,55% | 0,25% | 5,74% | 6,65% | Transporter/Customer |
| **453732808** | 0,1314 | 49 | 32 | 2 | 34 | 94,12% | 0,12% | 0,00% | 0,48% | Transporter/Customer |
| **454869722** | 0,1312 | 50 | 4 | 0 | 4 | 100,00% | 0,31% | 5,50% | 3,36% | Transporter/Customer |
| **461008977** | 0,1292 | 51 | 6 | 1 | 7 | 85,71% | 3,10% | 0,00% | 22,77% | Transporter/Customer |
| **453667205** | 0,1259 | 52 | 13 | 1 | 14 | 92,86% | 4,94% | 0,00% | 5,35% | Transporter |
| **453755452** | 0,1234 | 53 | 3 | 0 | 3 | 100,00% | 1,52% | 0,00% | 31,56% | Transporter |
| **453706402** | 0,1226 | 54 | 4 | 0 | 4 | 100,00% | 2,77% | 0,49% | 9,30% | Transporter/Customer |
| **453714126** | 0,1219 | 55 | 11 | 6 | 17 | 64,71% | 0,20% | 0,71% | 3,76% | Transporter/Customer |
| **455407564** | 0,1217 | 56 | 1 | 0 | 1 | 100,00% | 4,11% | 1,93% | 1,93% | Transporter |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **455690049** | 0,1216 | 57 | 0 | 1 | 1 | 0,00% | 0,41% | 0,00% | 9,45% | Transporter |
| **453687736** | 0,1183 | 58 | 15 | 3 | 18 | 83,33% | 2,99% | 7,83% | 4,39% | Supplier/Transporter/Customer |
| **458294222** | 0,1165 | 59 | 2 | 0 | 2 | 100,00% | 9,01% | 0,00% | 29,01% | Transporter |
| **454581512** | 0,1149 | 60 | 0 | 0 | 0 | No inspection | 1,80% | 0,13% | 2,57% | Transporter |
| **452772998** | 0,1084 | 61 | 31 | 5 | 36 | 86,11% | 0,33% | 5,11% | 6,27% | Transporter/Customer |
| **454425412** | 0,1063 | 62 | 30 | 8 | 38 | 78,95% | 0,00% | 7,31% | 11,42% | Supplier/Transporter/Customer |
| **460235184** | 0,1018 | 63 | 2 | 0 | 2 | 100,00% | 0,58% | 0,00% | 0,58% | Transporter |
| **454043792** | 0,1018 | 64 | 1 | 0 | 1 | 100,00% | 0,00% | 0,49% | 4,16% | Supplier |
| **453508287** | 0,1016 | 65 | 14 | 0 | 14 | 100,00% | 0,00% | 0,34% | 3,38% | Transporter/Customer |
| **458028422** | 0,1016 | 66 | 0 | 0 | 0 | No inspection | 0,81% | 2,42% | 2,42% | Supplier |
| **475605458** | 0,1005 | 67 | 4 | 1 | 5 | 80,00% | 1,56% | 4,29% | 7,01% | Supplier/Transporter/Customer |
| **454616907** | 0,0969 | 68 | 0 | 0 | 0 | No inspection | 6,44% | 3,35% | 0,26% | Supplier |
| **475435041** | 0,0955 | 69 | 4 | 1 | 5 | 80,00% | 0,50% | 0,17% | 2,91% | Transporter |
| **457698856** | 0,0931 | 70 | 9 | 0 | 9 | 100,00% | 0,00% | 0,06% | 0,89% | Supplier/Transporter/Customer |
| **461653523** | 0,0896 | 71 | 2 | 3 | 5 | 40,00% | 2,22% | 2,42% | 4,75% | Supplier/Transporter/Customer |
| **453707783** | 0,0880 | 72 | 50 | 14 | 64 | 78,13% | 0,87% | 0,79% | 7,80% | Supplier/Transporter/Customer |
| **458724077** | 0,0878 | 73 | 28 | 1 | 29 | 96,55% | 1,08% | 0,54% | 4,74% | Transporter/Customer |
| **455414006** | 0,0876 | 74 | 1 | 0 | 1 | 100,00% | 0,00% | 0,67% | 6,44% | Supplier/Customer |
| **475092926** | 0,0858 | 75 | 3 | 0 | 3 | 100,00% | 1,91% | 2,70% | 3,34% | Transporter |
| **454869591** | 0,0847 | 76 | 1 | 3 | 4 | 25,00% | 0,00% | 0,00% | 8,24% | Supplier/Transporter/Customer |
| **455003738** | 0,0843 | 77 | 3 | 1 | 4 | 75,00% | 0,00% | 1,39% | 0,00% | Supplier |
| **457878669** | 0,0840 | 78 | 2 | 0 | 2 | 100,00% | 0,22% | 0,11% | 0,86% | Supplier/Transporter/Customer |
| **453741333** | 0,0807 | 79 | 2 | 0 | 2 | 100,00% | 4,09% | 0,14% | 1,63% | Transporter |
| **453546456** | 0,0792 | 80 | 2 | 0 | 2 | 100,00% | 0,00% | 0,00% | 0,46% | Transporter |
| **453663069** | 0,0784 | 81 | 13 | 0 | 13 | 100,00% | 0,00% | 0,14% | 4,47% | Transporter/Customer |
| **458671940** | 0,0762 | 82 | 5 | 0 | 5 | 100,00% | 0,00% | 0,43% | 7,83% | Transporter/Customer |
| **460840279** | 0,0736 | 83 | 6 | 4 | 10 | 60,00% | 0,79% | 0,99% | 0,79% | Supplier/Transporter/Customer |
| **453659786** | 0,0735 | 84 | 1 | 0 | 1 | 100,00% | 4,29% | 0,00% | 2,86% | Transporter/Customer |
| **455325290** | 0,0735 | 85 | 5 | 0 | 5 | 100,00% | 1,30% | 0,00% | 2,61% | Transporter |
| **455258864** | 0,0729 | 86 | 8 | 1 | 9 | 88,89% | 0,91% | 0,00% | 4,55% | Transporter/Customer |
| **455458337** | 0,0706 | 87 | 7 | 4 | 11 | 63,64% | 0,86% | 2,15% | 2,58% | Supplier/Transporter/Customer |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **452841509** | 0,0700 | 88 | 2 | 0 | 2 | 100,00% | 0,00% | 0,31% | 1,89% | Transporter |
| **474029482** | 0,0698 | 89 | 1 | 0 | 1 | 100,00% | 0,00% | 0,00% | 7,58% | Transporter |
| **453706451** | 0,0697 | 90 | 11 | 1 | 12 | 91,67% | 1,09% | 1,09% | 6,20% | Transporter/Customer |
| **454603301** | 0,0695 | 91 | 4 | 1 | 5 | 80,00% | 0,44% | 0,44% | 11,95% | Transporter/Customer |
| **454909997** | 0,0687 | 92 | 2 | 0 | 2 | 100,00% | 0,37% | 0,00% | 30,63% | Transporter/Customer |
| **453741412** | 0,0676 | 93 | 4 | 1 | 5 | 80,00% | 0,94% | 0,00% | 4,47% | Transporter |
| **453705371** | 0,0675 | 94 | 4 | 0 | 4 | 100,00% | 5,56% | 0,00% | 6,94% | Transporter/Customer |
| **454634592** | 0,0674 | 95 | 9 | 0 | 9 | 100,00% | 0,00% | 0,00% | 3,54% | Transporter/Customer |
| **458063714** | 0,0673 | 96 | 2 | 1 | 3 | 66,67% | 0,00% | 0,00% | 7,23% | Transporter/Customer |
| **453727676** | 0,0647 | 97 | 2 | 1 | 3 | 66,67% | 0,00% | 1,59% | 8,76% | Supplier/Transporter/Customer |
| **455367028** | 0,0645 | 98 | 4 | 1 | 5 | 80,00% | 0,00% | 0,00% | 8,16% | Transporter |
| **458610284** | 0,0632 | 99 | 12 | 1 | 13 | 92,31% | 1,00% | 1,33% | 0,00% | Transporter/Customer |
| **453708321** | 0,0611 | 100 | 5z | 3 | 8 | 62,50% | 2,33% | 0,33% | 6,00% | Transporter |

### H.4 Results Betweenness

* Due to privacy considerations the ID code presented here is not the original BRS number.

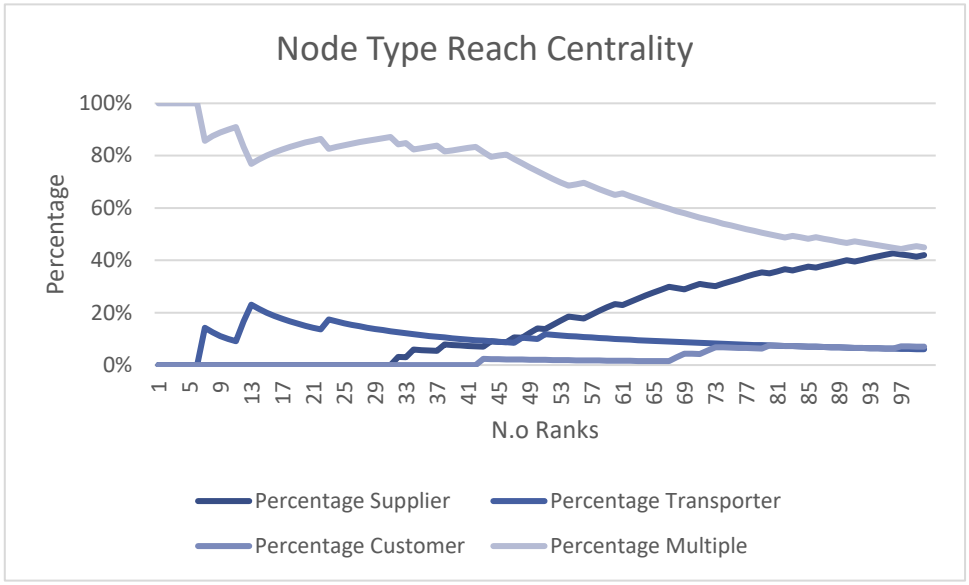
| ID | Between-ness Centrality | Ranking Between-ness | No. Approved Inspections | No. Disapproved Inspections | No. Inspections | Percentage of Compliance | RF National Holidays | RF Time of Transp. | RF VDM | Firm type |
|---|---|---|---|---|---|---|---|---|---|---|
| **453727705** | 1,0000 | 1 | 52 | 3 | 55 | 94,55% | 0,54% | 1,49% | 2,48% | Supplier/Transporter/Customer |
| **453732578** | 0,5825 | 2 | 40 | 5 | 45 | 88,89% | 0,47% | 1,63% | 5,43% | Supplier/Transporter/Customer |
| **455084180** | 0,5186 | 3 | 59 | 5 | 64 | 92,19% | 0,76% | 0,50% | 2,08% | Transporter |
| **453657449** | 0,5130 | 4 | 89 | 3 | 92 | 96,74% | 0,99% | 0,91% | 8,67% | Transporter/Customer |
| **455073047** | 0,3400 | 5 | 23 | 5 | 28 | 82,14% | 0,88% | 0,62% | 10,02% | Supplier/Transporter/Customer |
| **453664354** | 0,3380 | 6 | 35 | 1 | 36 | 97,22% | 0,61% | 0,27% | 2,22% | Supplier/Transporter/Customer |
| **455072273** | 0,3201 | 7 | 64 | 5 | 69 | 92,75% | 1,01% | 0,25% | 3,21% | Supplier/Transporter/Customer |
| **455329858** | 0,3168 | 8 | 35 | 7 | 42 | 83,33% | 1,01% | 0,32% | 1,66% | Supplier/Transporter/Customer |
| **453705707** | 0,3121 | 9 | 7 | 5 | 12 | 58,33% | 0,29% | 0,80% | 2,55% | Supplier/Customer |
| **453666102** | 0,2991 | 10 | 70 | 3 | 73 | 95,89% | 0,61% | 0,54% | 3,81% | Transporter/Customer |
| **453763772** | 0,2845 | 11 | 104 | 11 | 115 | 90,43% | 0,39% | 18,82% | 5,92% | Supplier/Transporter/Customer |

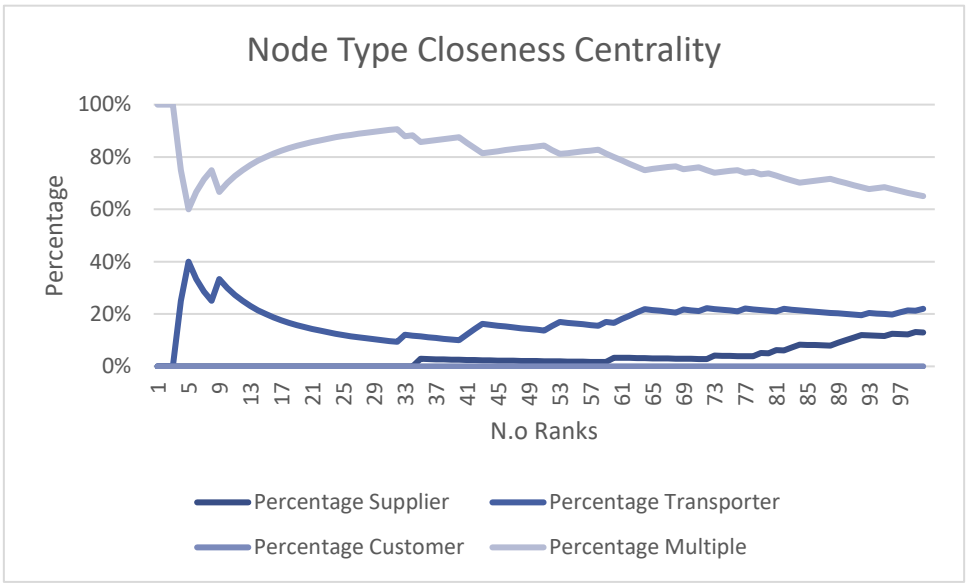
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **454566057** | 0,2727 | 12 | 64 | 1 | 65 | 98,46% | 0,05% | 0,19% | 6,09% | Transporter/Customer |
| **457965193** | 0,2664 | 13 | 39 | 9 | 48 | 81,25% | 0,98% | 2,56% | 2,64% | Supplier/Transporter/Customer |
| **453740591** | 0,2403 | 14 | 14 | 0 | 14 | 100,00% | 1,06% | 0,00% | 3,29% | Supplier/Transporter/Customer |
| **460953609** | 0,2356 | 15 | 10 | 1 | 11 | 90,91% | 4,66% | 7,07% | 3,26% | Transporter/Customer |
| **474180774** | 0,2068 | 16 | 8 | 1 | 9 | 88,89% | 0,78% | 1,55% | 2,91% | Supplier/Transporter/Customer |
| **458157077** | 0,1985 | 17 | 11 | 1 | 12 | 91,67% | 1,03% | 7,38% | 4,29% | Supplier/Transporter/Customer |
| **460681901** | 0,1910 | 18 | 8 | 0 | 8 | 100,00% | 0,49% | 0,38% | 2,81% | Supplier/Customer |
| **460211053** | 0,1850 | 19 | 84 | 3 | 87 | 96,55% | 0,25% | 5,74% | 6,65% | Transporter/Customer |
| **475784315** | 0,1838 | 20 | 8 | 6 | 14 | 57,14% | 3,95% | 0,67% | 20,58% | Supplier/Transporter/Customer |
| **453730396** | 0,1826 | 21 | 3 | 2 | 5 | 60,00% | 2,34% | 3,96% | 7,19% | Transporter |
| **453726728** | 0,1818 | 22 | 10 | 0 | 10 | 100,00% | 0,00% | 0,00% | 0,39% | Transporter/Customer |
| **458076829** | 0,1814 | 23 | 38 | 1 | 39 | 97,44% | 0,00% | 0,17% | 1,69% | Supplier/Transporter/Customer |
| **454342781** | 0,1807 | 24 | 17 | 1 | 18 | 94,44% | 0,00% | 0,00% | 4,03% | Transporter/Customer |
| **475372381** | 0,1794 | 25 | 1 | 0 | 1 | 100,00% | 0,90% | 3,23% | 5,15% | Transporter |
| **454869722** | 0,1729 | 26 | 4 | 0 | 4 | 100,00% | 0,31% | 5,50% | 3,36% | Transporter/Customer |
| **453659656** | 0,1728 | 27 | 3 | 0 | 3 | 100,00% | 2,40% | 2,81% | 3,01% | Transporter |
| **453687736** | 0,1678 | 28 | 15 | 3 | 18 | 83,33% | 2,99% | 7,83% | 4,39% | Supplier/Transporter/Customer |
| **270126771** | 0,1634 | 29 | 12 | 1 | 13 | 92,31% | 0,95% | 0,09% | 0,38% | Transporter |
| **466151559** | 0,1598 | 30 | 0 | 0 | 0 | No inspection | 0,00% | 2,04% | 11,36% | Supplier/Customer |
| **453710045** | 0,1586 | 31 | 12 | 0 | 12 | 100,00% | 0,12% | 0,00% | 3,41% | Supplier/Transporter |
| **454990495** | 0,1491 | 32 | 6 | 0 | 6 | 100,00% | 1,34% | 2,16% | 8,32% | Transporter |
| **453714126** | 0,1371 | 33 | 11 | 6 | 17 | 64,71% | 0,20% | 0,71% | 3,76% | Transporter/Customer |
| **460238188** | 0,1368 | 34 | 24 | 0 | 24 | 100,00% | 1,67% | 0,00% | 21,04% | Transporter |
| **453733323** | 0,1340 | 35 | 0 | 1 | 1 | 0,00% | 0,00% | 1,59% | 2,92% | Supplier/Customer |
| **454839050** | 0,1330 | 36 | 9 | 1 | 10 | 90,00% | 0,07% | 0,63% | 7,99% | Supplier/Transporter/Customer |
| **459988503** | 0,1267 | 37 | 8 | 2 | 10 | 80,00% | 1,26% | 0,32% | 3,03% | Supplier/Transporter/Customer |
| **453688944** | 0,1246 | 38 | 15 | 1 | 16 | 93,75% | 2,40% | 0,68% | 5,39% | Transporter |
| **452772998** | 0,1225 | 39 | 31 | 5 | 36 | 86,11% | 0,33% | 5,11% | 6,27% | Transporter/Customer |
| **453508287** | 0,1209 | 40 | 14 | 0 | 14 | 100,00% | 0,00% | 0,34% | 3,38% | Transporter/Customer |
| **458792050** | 0,1191 | 41 | 12 | 2 | 14 | 85,71% | 2,28% | 0,27% | 4,56% | Transporter |

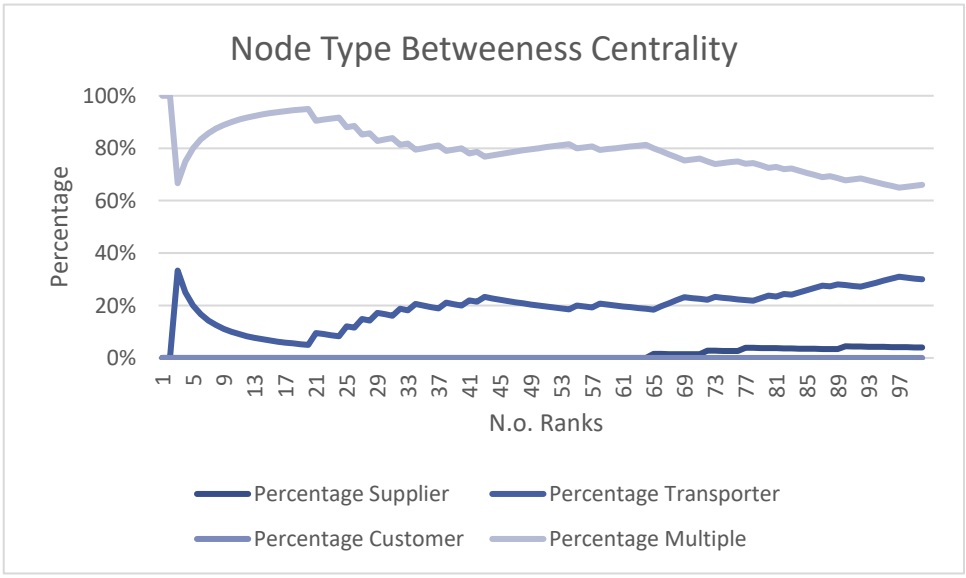
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **458737629** | 0,1178 | 42 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 4,26% | Supplier/Customer |
| **454581512** | 0,1175 | 43 | 0 | 0 | 0 | No inspection | 1,80% | 0,13% | 2,57% | Transporter |
| **453707783** | 0,1170 | 44 | 50 | 14 | 64 | 78,13% | 0,87% | 0,79% | 7,80% | Supplier/Transporter/Customer |
| **454634592** | 0,1165 | 45 | 9 | 0 | 9 | 100,00% | 0,00% | 0,00% | 3,54% | Transporter/Customer |
| **455386306** | 0,1155 | 46 | 29 | 1 | 30 | 96,67% | 0,61% | 6,14% | 2,72% | Supplier/Transporter/Customer |
| **453713359** | 0,1136 | 47 | 3 | 0 | 3 | 100,00% | 3,96% | 7,23% | 13,43% | Supplier/Transporter |
| **475605458** | 0,1102 | 48 | 4 | 1 | 5 | 80,00% | 1,56% | 4,29% | 7,01% | Supplier/Transporter/Customer |
| **454075775** | 0,1094 | 49 | 0 | 0 | 0 | No inspection | 0,00% | 5,31% | 11,84% | Supplier/Customer |
| **457878669** | 0,1071 | 50 | 2 | 0 | 2 | 100,00% | 0,22% | 0,11% | 0,86% | Supplier/Transporter/Customer |
| **458724077** | 0,1060 | 51 | 28 | 1 | 29 | 96,55% | 1,08% | 0,54% | 4,74% | Transporter/Customer |
| **454425412** | 0,1056 | 52 | 30 | 8 | 38 | 78,95% | 0,00% | 7,31% | 11,42% | Supplier/Transporter/Customer |
| **461653523** | 0,1055 | 53 | 2 | 3 | 5 | 40,00% | 2,22% | 2,42% | 4,75% | Supplier/Transporter/Customer |
| **455077561** | 0,1041 | 54 | 2 | 0 | 2 | 100,00% | 0,53% | 0,00% | 5,06% | Supplier/Transporter/Customer |
| **460235184** | 0,1023 | 55 | 2 | 0 | 2 | 100,00% | 0,58% | 0,00% | 0,58% | Transporter |
| **458557766** | 0,0980 | 56 | 1 | 2 | 3 | 33,33% | 0,84% | 1,18% | 2,93% | Supplier/Transporter/Customer |
| **458383524** | 0,0963 | 57 | 11 | 2 | 13 | 84,62% | 0,49% | 0,12% | 3,68% | Transporter/Customer |
| **460492159** | 0,0941 | 58 | 0 | 0 | 0 | No inspection | 0,00% | 0,00% | 3,37% | Transporter |
| **461008977** | 0,0904 | 59 | 6 | 1 | 7 | 85,71% | 3,10% | 0,00% | 22,77% | Transporter/Customer |
| **460187954** | 0,0883 | 60 | 50 | 15 | 65 | 76,92% | 0,00% | 7,98% | 2,60% | Supplier/Transporter/Customer |
| **453663069** | 0,0864 | 61 | 13 | 0 | 13 | 100,00% | 0,00% | 0,14% | 4,47% | Transporter/Customer |
| **457881106** | 0,0858 | 62 | 5 | 1 | 6 | 83,33% | 1,78% | 1,63% | 14,30% | Supplier/Transporter |
| **460840279** | 0,0852 | 63 | 6 | 4 | 10 | 60,00% | 0,79% | 0,99% | 0,79% | Supplier/Transporter/Customer |
| **453732808** | 0,0837 | 64 | 32 | 2 | 34 | 94,12% | 0,12% | 0,00% | 0,48% | Transporter/Customer |
| **454616907** | 0,0832 | 65 | 0 | 0 | 0 | No inspection | 6,44% | 3,35% | 0,26% | Supplier |
| **453667205** | 0,0825 | 66 | 13 | 1 | 14 | 92,86% | 4,94% | 0,00% | 5,35% | Transporter |
| **461582936** | 0,0824 | 67 | 15 | 0 | 15 | 100,00% | 3,27% | 0,00% | 9,07% | Transporter |
| **475092926** | 0,0813 | 68 | 3 | 0 | 3 | 100,00% | 1,91% | 2,70% | 3,34% | Transporter |
| **453764927** | 0,0799 | 69 | 4 | 1 | 5 | 80,00% | 1,06% | 0,56% | 4,61% | Transporter |
| **458671940** | 0,0773 | 70 | 5 | 0 | 5 | 100,00% | 0,00% | 0,43% | 7,83% | Transporter/Customer |
| **453706451** | 0,0762 | 71 | 11 | 1 | 12 | 91,67% | 1,09% | 1,09% | 6,20% | Transporter/Customer |

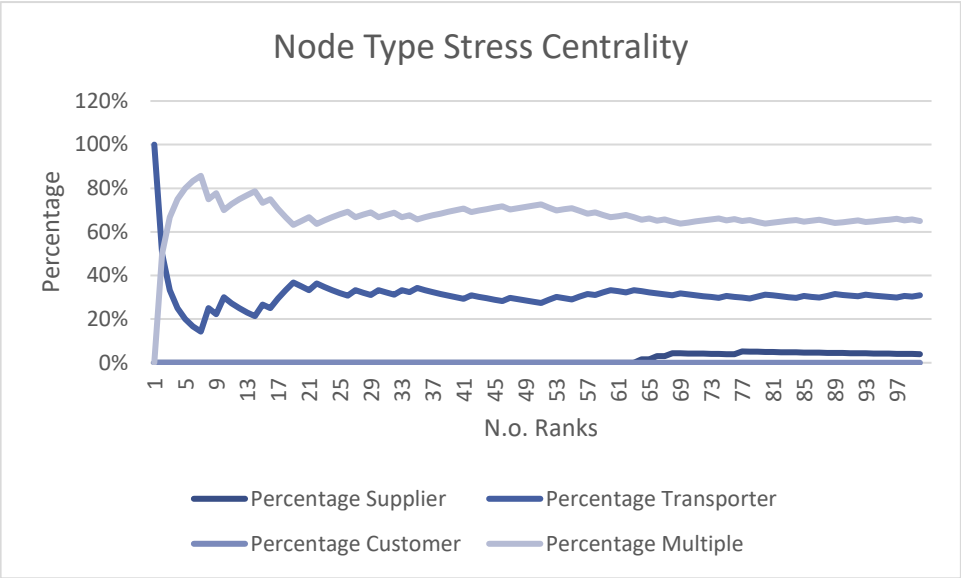
| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **458028422** | 0,0754 | 72 | 0 | 0 | 0 | No inspection | 0,81% | 2,42% | 2,42% | Supplier |
| **455072249** | 0,0735 | 73 | 1 | 0 | 1 | 100,00% | 0,28% | 0,19% | 5,50% | Transporter |
| **453706402** | 0,0723 | 74 | 4 | 0 | 4 | 100,00% | 2,77% | 0,49% | 9,30% | Transporter/Customer |
| **457698856** | 0,0720 | 75 | 9 | 0 | 9 | 100,00% | 0,00% | 0,06% | 0,89% | Supplier/Transporter/Customer |
| **460201947** | 0,0712 | 76 | 2 | 1 | 3 | 66,67% | 0,05% | 0,10% | 5,15% | Supplier/Transporter/Customer |
| **454043792** | 0,0696 | 77 | 1 | 0 | 1 | 100,00% | 0,00% | 0,49% | 4,16% | Supplier |
| **461431570** | 0,0696 | 78 | 1 | 1 | 2 | 50,00% | 0,25% | 1,51% | 7,54% | Transporter/Customer |
| **453732734** | 0,0696 | 79 | 15 | 2 | 17 | 88,24% | 0,00% | 1,80% | 11,42% | Transporter |
| **460942854** | 0,0685 | 80 | 7 | 1 | 8 | 87,50% | 7,05% | 1,24% | 8,30% | Transporter |
| **455414006** | 0,0683 | 81 | 1 | 0 | 1 | 100,00% | 0,00% | 0,67% | 6,44% | Supplier/Customer |
| **453741333** | 0,0672 | 82 | 2 | 0 | 2 | 100,00% | 4,09% | 0,14% | 1,63% | Transporter |
| **454603301** | 0,0666 | 83 | 4 | 1 | 5 | 80,00% | 0,44% | 0,44% | 11,95% | Transporter/Customer |
| **453755452** | 0,0642 | 84 | 3 | 0 | 3 | 100,00% | 1,52% | 0,00% | 31,56% | Transporter |
| **458294222** | 0,0618 | 85 | 2 | 0 | 2 | 100,00% | 9,01% | 0,00% | 29,01% | Transporter |
| **453665481** | 0,0608 | 86 | 26 | 1 | 27 | 96,30% | 0,00% | 0,28% | 0,00% | Transporter |
| **475435041** | 0,0593 | 87 | 4 | 1 | 5 | 80,00% | 0,50% | 0,17% | 2,91% | Transporter |
| **455258864** | 0,0584 | 88 | 8 | 1 | 9 | 88,89% | 0,91% | 0,00% | 4,55% | Transporter/Customer |
| **452841509** | 0,0583 | 89 | 2 | 0 | 2 | 100,00% | 0,00% | 0,31% | 1,89% | Transporter |
| **455003738** | 0,0567 | 90 | 3 | 1 | 4 | 75,00% | 0,00% | 1,39% | 0,00% | Supplier |
| **454869591** | 0,0533 | 91 | 1 | 3 | 4 | 25,00% | 0,00% | 0,00% | 8,24% | Supplier/Transporter/Customer |
| **453689124** | 0,0523 | 92 | 2 | 0 | 2 | 100,00% | 3,64% | 4,47% | 14,78% | Supplier/Transporter/Customer |
| **455690049** | 0,0517 | 93 | 0 | 1 | 1 | 0,00% | 0,41% | 0,00% | 9,45% | Transporter |
| **459765751** | 0,0507 | 94 | 6 | 1 | 7 | 85,71% | 0,00% | 0,00% | 0,00% | Transporter |
| **455325290** | 0,0506 | 95 | 5 | 0 | 5 | 100,00% | 1,30% | 0,00% | 2,61% | Transporter |
| **455407564** | 0,0497 | 96 | 1 | 0 | 1 | 100,00% | 4,11% | 1,93% | 1,93% | Transporter |
| **453708321** | 0,0471 | 97 | 5 | 3 | 8 | 62,50% | 2,33% | 0,33% | 6,00% | Transporter |
| **453660479** | 0,0470 | 98 | 13 | 1 | 14 | 92,86% | 1,02% | 1,22% | 0,25% | Supplier/Transporter/Customer |
| **453749960** | 0,0466 | 99 | 0 | 1 | 1 | 0,00% | 0,00% | 0,00% | 25,41% | Transporter/Customer |
| **248364479** | 0,0462 | 100 | 0 | 0 | 0 | No inspection | 1,87% | 0,39% | 0,69% | Transporter/Customer |

### H.5 Node Type Distribution Centrality Metrics



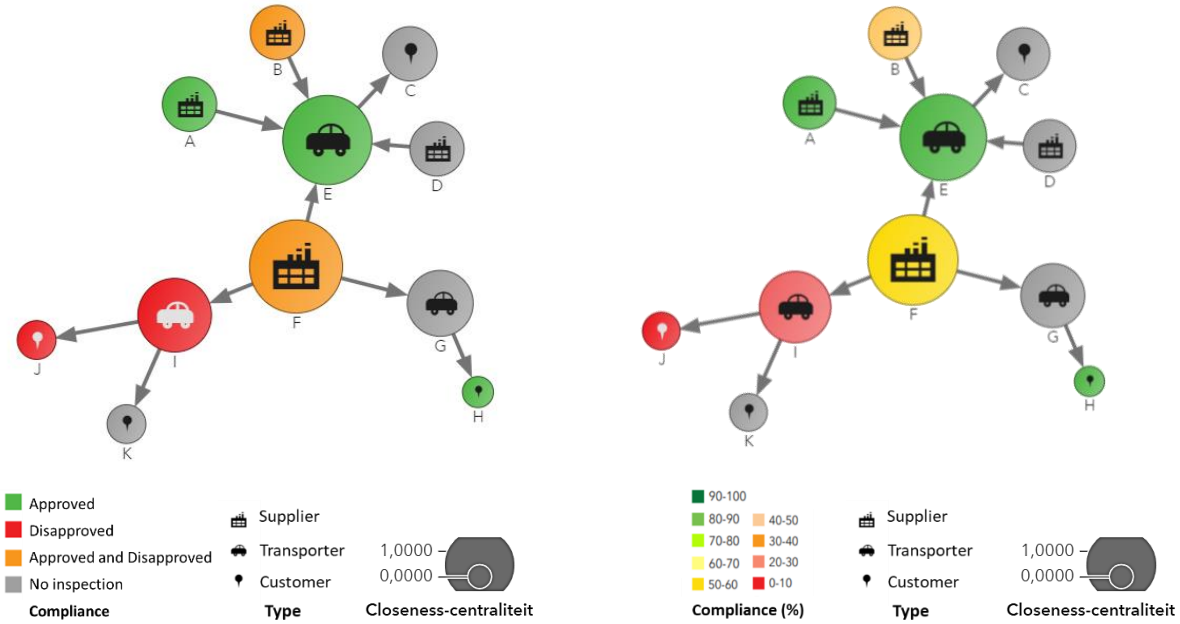






### H.6 Dutch National Holidays 2018

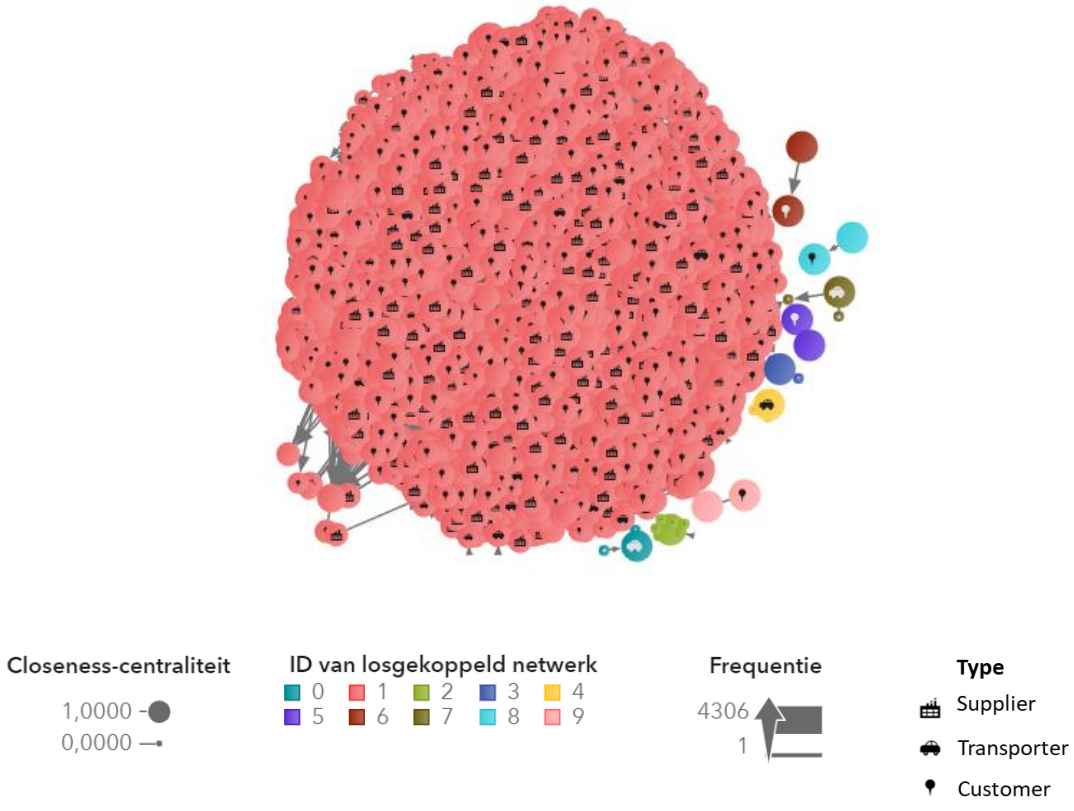
| Date | Day | Name | Type |
|---|---|---|---|
| **January 1** | Tuesday | New Year's Day | National holiday |
| **March 3** | Friday | Good Friday | National holiday |
| **April 4** | Sunday | Easter Sunday | National holiday |
| **April 4** | Monday | Easter Monday | National holiday |
| **April 4** | Friday | King's Birthday | National holiday |
| **May 5** | Saturday | Liberation Day | National holiday |
| **May 6** | Thursday | Ascension Day | National holiday |
| **May 20** | Sunday | Whit Sunday | National holiday |
| **May 21** | Monday | Whit Monday | National holiday |
| **December 25** | Tuesday | Christmas Day | National holiday |
| **December 26** | Wednesday | Second Day of Christmas | National holiday |

## H.7 Sample Network Diagram Manure (Traffic light + Compliance Intervals)

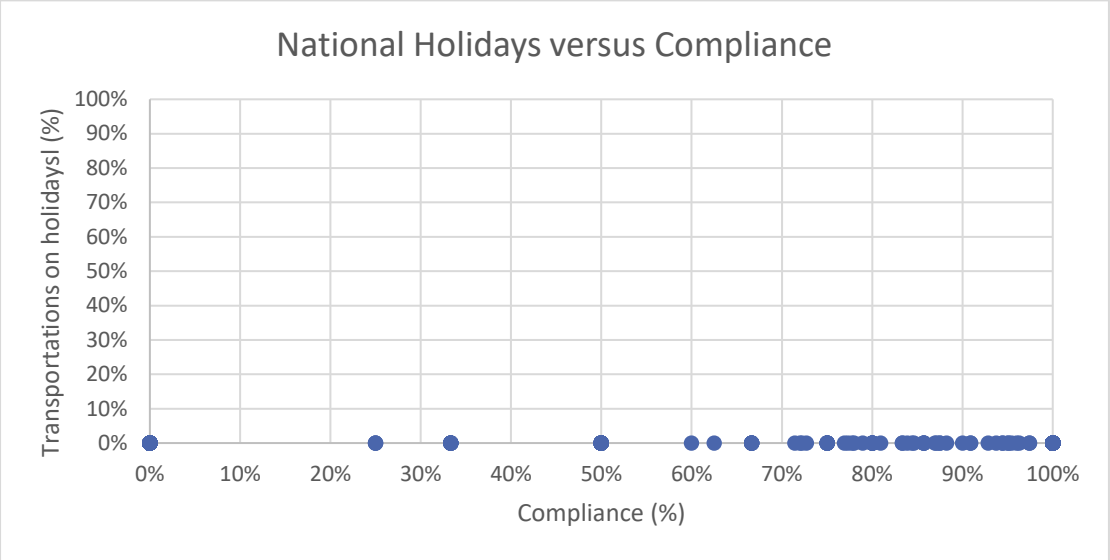


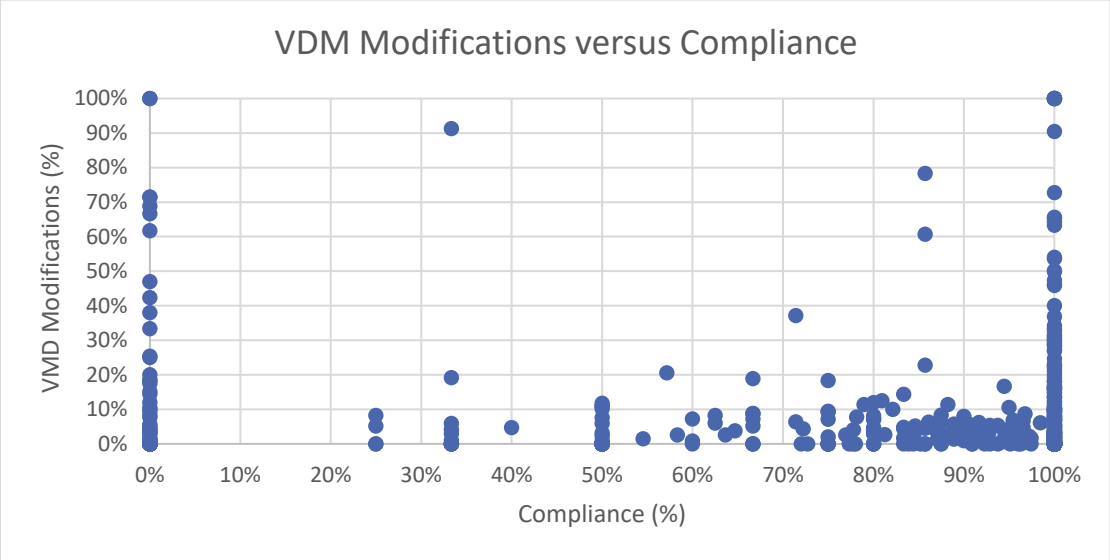

**Traffic Light Scenario** **Compliance Intervals**

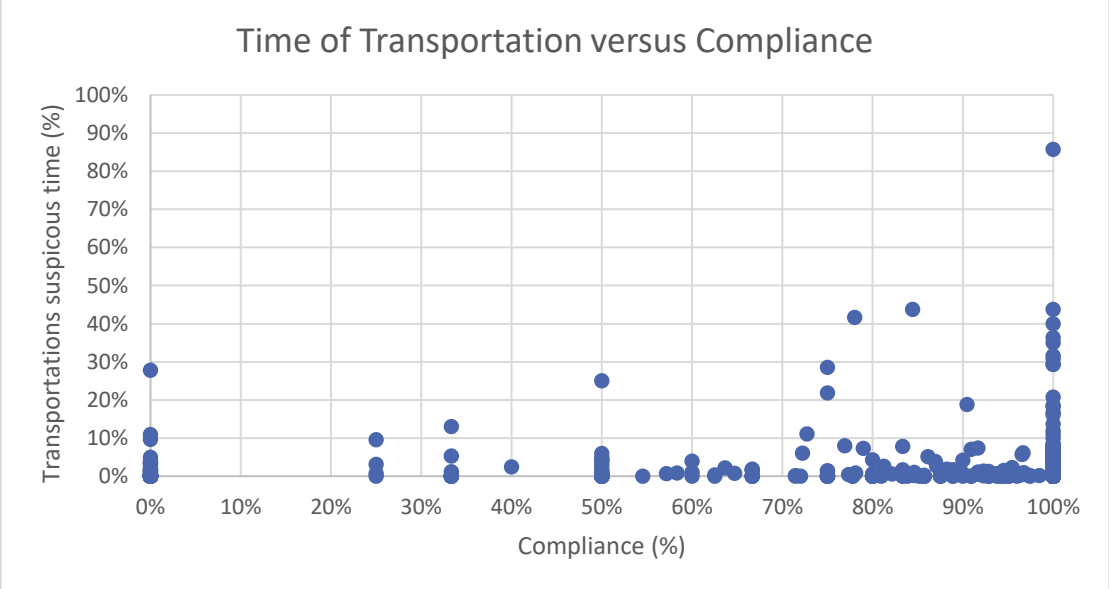
## H.8 Disconnected Network Manure



**H.9 Risk Factor versus Compliance**



## H.10 Frequency Table Risk Factor National Holidays

| % Of transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent | % f transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent | % Of transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 5340 | 90,1 | 5340 | 90,1 | 0,606 | 1 | 0,02 | 5397 | 91,06 | 1,075 | 2 | 0,03 | 5461 | 92,14 |
| 0,048 | 1 | 0,02 | 5341 | 90,11 | 0,609 | 1 | 0,02 | 5398 | 91,07 | 1,078 | 1 | 0,02 | 5462 | 92,15 |
| 0,055 | 1 | 0,02 | 5342 | 90,13 | 0,611 | 1 | 0,02 | 5399 | 91,09 | 1,091 | 2 | 0,03 | 5464 | 92,19 |
| 0,07 | 1 | 0,02 | 5343 | 90,15 | 0,614 | 1 | 0,02 | 5400 | 91,11 | 1,095 | 1 | 0,02 | 5465 | 92,21 |
| 0,11 | 1 | 0,02 | 5344 | 90,16 | 0,654 | 1 | 0,02 | 5401 | 91,13 | 1,111 | 3 | 0,05 | 5468 | 92,26 |
| 0,118 | 1 | 0,02 | 5345 | 90,18 | 0,662 | 1 | 0,02 | 5402 | 91,14 | 1,117 | 1 | 0,02 | 5469 | 92,27 |
| 0,119 | 1 | 0,02 | 5346 | 90,2 | 0,685 | 1 | 0,02 | 5403 | 91,16 | 1,156 | 1 | 0,02 | 5470 | 92,29 |
| 0,144 | 1 | 0,02 | 5347 | 90,21 | 0,704 | 1 | 0,02 | 5404 | 91,18 | 1,163 | 1 | 0,02 | 5471 | 92,31 |
| 0,203 | 1 | 0,02 | 5348 | 90,23 | 0,735 | 1 | 0,02 | 5405 | 91,19 | 1,19 | 2 | 0,03 | 5473 | 92,34 |
| 0,221 | 1 | 0,02 | 5349 | 90,25 | 0,741 | 2 | 0,03 | 5407 | 91,23 | 1,196 | 1 | 0,02 | 5474 | 92,36 |
| 0,224 | 1 | 0,02 | 5350 | 90,26 | 0,746 | 1 | 0,02 | 5408 | 91,24 | 1,198 | 1 | 0,02 | 5475 | 92,37 |
| 0,251 | 1 | 0,02 | 5351 | 90,28 | 0,752 | 1 | 0,02 | 5409 | 91,26 | 1,205 | 1 | 0,02 | 5476 | 92,39 |
| 0,254 | 1 | 0,02 | 5352 | 90,3 | 0,76 | 1 | 0,02 | 5410 | 91,28 | 1,24 | 1 | 0,02 | 5477 | 92,41 |
| 0,28 | 1 | 0,02 | 5353 | 90,32 | 0,763 | 1 | 0,02 | 5411 | 91,29 | 1,242 | 1 | 0,02 | 5478 | 92,42 |
| 0,28 | 1 | 0,02 | 5354 | 90,33 | 0,773 | 1 | 0,02 | 5412 | 91,31 | 1,25 | 4 | 0,07 | 5482 | 92,49 |
| 0,285 | 1 | 0,02 | 5355 | 90,35 | 0,775 | 1 | 0,02 | 5413 | 91,33 | 1,253 | 1 | 0,02 | 5483 | 92,51 |
| 0,294 | 1 | 0,02 | 5356 | 90,37 | 0,781 | 1 | 0,02 | 5414 | 91,34 | 1,261 | 1 | 0,02 | 5484 | 92,53 |
| 0,299 | 1 | 0,02 | 5357 | 90,38 | 0,791 | 1 | 0,02 | 5415 | 91,36 | 1,266 | 1 | 0,02 | 5485 | 92,54 |
| 0,302 | 1 | 0,02 | 5358 | 90,4 | 0,806 | 1 | 0,02 | 5416 | 91,38 | 1,282 | 1 | 0,02 | 5486 | 92,56 |
| 0,304 | 1 | 0,02 | 5359 | 90,42 | 0,813 | 3 | 0,05 | 5419 | 91,43 | 1,299 | 2 | 0,03 | 5488 | 92,59 |
| 0,306 | 1 | 0,02 | 5360 | 90,43 | 0,82 | 2 | 0,03 | 5421 | 91,46 | 1,304 | 1 | 0,02 | 5489 | 92,61 |
| 0,312 | 1 | 0,02 | 5361 | 90,45 | 0,826 | 1 | 0,02 | 5422 | 91,48 | 1,316 | 1 | 0,02 | 5490 | 92,63 |
| 0,313 | 1 | 0,02 | 5362 | 90,47 | 0,83 | 1 | 0,02 | 5423 | 91,5 | 1,327 | 1 | 0,02 | 5491 | 92,64 |
| 0,314 | 1 | 0,02 | 5363 | 90,48 | 0,84 | 2 | 0,03 | 5425 | 91,53 | 1,336 | 1 | 0,02 | 5492 | 92,66 |
| 0,33 | 1 | 0,02 | 5364 | 90,5 | 0,841 | 1 | 0,02 | 5426 | 91,55 | 1,337 | 1 | 0,02 | 5493 | 92,68 |
| 0,369 | 2 | 0,03 | 5366 | 90,53 | 0,855 | 1 | 0,02 | 5427 | 91,56 | 1,357 | 1 | 0,02 | 5494 | 92,69 |
| 0,373 | 1 | 0,02 | 5367 | 90,55 | 0,858 | 1 | 0,02 | 5428 | 91,58 | 1,361 | 1 | 0,02 | 5495 | 92,71 |
| 0,386 | 1 | 0,02 | 5368 | 90,57 | 0,866 | 1 | 0,02 | 5429 | 91,6 | 1,37 | 2 | 0,03 | 5497 | 92,75 |
| 0,391 | 1 | 0,02 | 5369 | 90,59 | 0,867 | 1 | 0,02 | 5430 | 91,61 | 1,379 | 1 | 0,02 | 5498 | 92,76 |
| 0,408 | 1 | 0,02 | 5370 | 90,6 | 0,881 | 1 | 0,02 | 5431 | 91,63 | 1,418 | 1 | 0,02 | 5499 | 92,78 |
| 0,411 | 1 | 0,02 | 5371 | 90,62 | 0,882 | 1 | 0,02 | 5432 | 91,65 | 1,423 | 1 | 0,02 | 5500 | 92,8 |
| 0,412 | 1 | 0,02 | 5372 | 90,64 | 0,902 | 1 | 0,02 | 5433 | 91,67 | 1,429 | 1 | 0,02 | 5501 | 92,81 |
| 0,428 | 1 | 0,02 | 5373 | 90,65 | 0,91 | 1 | 0,02 | 5434 | 91,68 | 1,435 | 1 | 0,02 | 5502 | 92,83 |
| 0,442 | 1 | 0,02 | 5374 | 90,67 | 0,921 | 1 | 0,02 | 5435 | 91,7 | 1,449 | 3 | 0,05 | 5505 | 92,88 |
| 0,462 | 1 | 0,02 | 5375 | 90,69 | 0,941 | 1 | 0,02 | 5436 | 91,72 | 1,456 | 1 | 0,02 | 5506 | 92,9 |
| 0,465 | 1 | 0,02 | 5376 | 90,7 | 0,943 | 1 | 0,02 | 5437 | 91,73 | 1,46 | 1 | 0,02 | 5507 | 92,91 |
| 0,474 | 1 | 0,02 | 5377 | 90,72 | 0,947 | 1 | 0,02 | 5438 | 91,75 | 1,515 | 1 | 0,02 | 5508 | 92,93 |
| 0,488 | 1 | 0,02 | 5378 | 90,74 | 0,948 | 1 | 0,02 | 5439 | 91,77 | 1,521 | 1 | 0,02 | 5509 | 92,95 |
| 0,491 | 1 | 0,02 | 5379 | 90,75 | 0,957 | 1 | 0,02 | 5440 | 91,78 | 1,531 | 1 | 0,02 | 5510 | 92,96 |
| 0,503 | 1 | 0,02 | 5380 | 90,77 | 0,971 | 1 | 0,02 | 5441 | 91,8 | 1,558 | 1 | 0,02 | 5511 | 92,98 |
| 0,53 | 1 | 0,02 | 5381 | 90,79 | 0,976 | 1 | 0,02 | 5442 | 91,82 | 1,6 | 1 | 0,02 | 5512 | 93 |
| 0,532 | 1 | 0,02 | 5382 | 90,8 | 0,99 | 1 | 0,02 | 5443 | 91,83 | 1,63 | 1 | 0,02 | 5513 | 93,02 |
| 0,537 | 1 | 0,02 | 5383 | 90,82 | 0,992 | 1 | 0,02 | 5444 | 91,85 | 1,63 | 1 | 0,02 | 5514 | 93,03 |
| 0,538 | 2 | 0,03 | 5385 | 90,86 | 0,994 | 1 | 0,02 | 5445 | 91,87 | 1,638 | 1 | 0,02 | 5515 | 93,05 |
| 0,543 | 1 | 0,02 | 5386 | 90,87 | 0,997 | 1 | 0,02 | 5446 | 91,88 | 1,639 | 3 | 0,05 | 5518 | 93,1 |
| 0,543 | 1 | 0,02 | 5387 | 90,89 | 1 | 1 | 0,02 | 5447 | 91,9 | 1,653 | 1 | 0,02 | 5519 | 93,12 |
| 0,556 | 1 | 0,02 | 5388 | 90,91 | 1,003 | 2 | 0,03 | 5449 | 91,94 | 1,667 | 3 | 0,05 | 5522 | 93,17 |
| 0,559 | 1 | 0,02 | 5389 | 90,92 | 1,011 | 1 | 0,02 | 5450 | 91,95 | 1,681 | 1 | 0,02 | 5523 | 93,18 |
| 0,573 | 1 | 0,02 | 5390 | 90,94 | 1,013 | 1 | 0,02 | 5451 | 91,97 | 1,695 | 3 | 0,05 | 5526 | 93,23 |
| 0,575 | 1 | 0,02 | 5391 | 90,96 | 1,015 | 3 | 0,05 | 5454 | 92,02 | 1,717 | 1 | 0,02 | 5527 | 93,25 |
| 0,577 | 1 | 0,02 | 5392 | 90,97 | 1,026 | 1 | 0,02 | 5455 | 92,04 | 1,724 | 1 | 0,02 | 5528 | 93,27 |
| 0,577 | 1 | 0,02 | 5393 | 90,99 | 1,03 | 1 | 0,02 | 5456 | 92,05 | 1,754 | 2 | 0,03 | 5530 | 93,3 |
| 0,578 | 1 | 0,02 | 5394 | 91,01 | 1,031 | 1 | 0,02 | 5457 | 92,07 | 1,762 | 1 | 0,02 | 5531 | 93,32 |
| 0,581 | 1 | 0,02 | 5395 | 91,02 | 1,056 | 1 | 0,02 | 5458 | 92,09 | 1,77 | 2 | 0,03 | 5533 | 93,35 |
| 0,592 | 1 | 0,02 | 5396 | 91,04 | 1,056 | 1 | 0,02 | 5459 | 92,1 | 1,779 | 1 | 0,02 | 5534 | 93,37 |

| 1,786 | 1 | 0,02 | 5535 | 93,39 | 3,175 | 1 | 0,02 | 5614 | 94,72 | 5,785 | 1 | 0,02 | 5701 | 96,19 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,799 | 1 | 0,02 | 5536 | 93,4 | 3,191 | 1 | 0,02 | 5615 | 94,74 | 5,882 | 3 | 0,05 | 5704 | 96,24 |
| 1,802 | 1 | 0,02 | 5537 | 93,42 | 3,205 | 1 | 0,02 | 5616 | 94,75 | 6,04 | 1 | 0,02 | 5705 | 96,25 |
| 1,807 | 1 | 0,02 | 5538 | 93,44 | 3,226 | 5 | 0,08 | 5621 | 94,84 | 6,061 | 4 | 0,07 | 5709 | 96,32 |
| 1,835 | 1 | 0,02 | 5539 | 93,45 | 3,247 | 1 | 0,02 | 5622 | 94,85 | 6,122 | 2 | 0,03 | 5711 | 96,36 |
| 1,874 | 1 | 0,02 | 5540 | 93,47 | 3,252 | 1 | 0,02 | 5623 | 94,87 | 6,13 | 1 | 0,02 | 5712 | 96,37 |
| 1,887 | 2 | 0,03 | 5542 | 93,5 | 3,267 | 1 | 0,02 | 5624 | 94,89 | 6,25 | 4 | 0,07 | 5716 | 96,44 |
| 1,908 | 1 | 0,02 | 5543 | 93,52 | 3,279 | 1 | 0,02 | 5625 | 94,9 | 6,349 | 1 | 0,02 | 5717 | 96,46 |
| 1,914 | 1 | 0,02 | 5544 | 93,54 | 3,297 | 1 | 0,02 | 5626 | 94,92 | 6,41 | 1 | 0,02 | 5718 | 96,47 |
| 1,923 | 1 | 0,02 | 5545 | 93,55 | 3,306 | 1 | 0,02 | 5627 | 94,94 | 6,443 | 1 | 0,02 | 5719 | 96,49 |
| 2,041 | 2 | 0,03 | 5548 | 93,61 | 3,55 | 1 | 0,02 | 5630 | 94,99 | 6,452 | 2 | 0,03 | 5721 | 96,52 |
| 2,062 | 1 | 0,02 | 5549 | 93,62 | 3,571 | 1 | 0,02 | 5631 | 95,01 | 6,522 | 1 | 0,02 | 5722 | 96,54 |
| 2,069 | 1 | 0,02 | 5550 | 93,64 | 3,614 | 1 | 0,02 | 5632 | 95,02 | 6,589 | 1 | 0,02 | 5723 | 96,56 |
| 2,083 | 2 | 0,03 | 5552 | 93,67 | 3,636 | 1 | 0,02 | 5633 | 95,04 | 6,757 | 1 | 0,02 | 5724 | 96,57 |
| 2,155 | 1 | 0,02 | 5553 | 93,69 | 3,642 | 1 | 0,02 | 5634 | 95,06 | 6,765 | 1 | 0,02 | 5725 | 96,59 |
| 2,174 | 3 | 0,05 | 5556 | 93,74 | 3,704 | 3 | 0,05 | 5637 | 95,11 | 6,78 | 1 | 0,02 | 5726 | 96,61 |
| 2,183 | 1 | 0,02 | 5557 | 93,76 | 3,727 | 1 | 0,02 | 5638 | 95,12 | 6,838 | 1 | 0,02 | 5727 | 96,63 |
| 2,222 | 2 | 0,03 | 5559 | 93,79 | 3,745 | 1 | 0,02 | 5639 | 95,14 | 6,897 | 2 | 0,03 | 5729 | 96,66 |
| 2,273 | 1 | 0,02 | 5560 | 93,81 | 3,774 | 1 | 0,02 | 5640 | 95,16 | 6,944 | 1 | 0,02 | 5730 | 96,68 |
| 2,28 | 1 | 0,02 | 5561 | 93,82 | 3,782 | 1 | 0,02 | 5641 | 95,17 | 6,957 | 1 | 0,02 | 5731 | 96,69 |
| 2,304 | 1 | 0,02 | 5562 | 93,84 | 3,79 | 1 | 0,02 | 5642 | 95,19 | 7,054 | 1 | 0,02 | 5732 | 96,71 |
| 2,306 | 1 | 0,02 | 5563 | 93,86 | 3,825 | 1 | 0,02 | 5643 | 95,21 | 7,059 | 1 | 0,02 | 5733 | 96,73 |
| 2,326 | 1 | 0,02 | 5564 | 93,88 | 3,846 | 1 | 0,02 | 5644 | 95,23 | 7,087 | 1 | 0,02 | 5734 | 96,74 |
| 2,333 | 1 | 0,02 | 5565 | 93,89 | 3,896 | 1 | 0,02 | 5645 | 95,24 | 7,143 | 2 | 0,03 | 5736 | 96,78 |
| 2,338 | 1 | 0,02 | 5566 | 93,91 | 3,906 | 1 | 0,02 | 5646 | 95,26 | 7,317 | 1 | 0,02 | 5737 | 96,79 |
| 2,344 | 1 | 0,02 | 5567 | 93,93 | 3,937 | 1 | 0,02 | 5647 | 95,28 | 7,407 | 1 | 0,02 | 5738 | 96,81 |
| 2,353 | 1 | 0,02 | 5568 | 93,94 | 3,952 | 1 | 0,02 | 5648 | 95,29 | 7,5 | 1 | 0,02 | 5739 | 96,83 |
| 2,362 | 2 | 0,03 | 5570 | 93,98 | 3,959 | 1 | 0,02 | 5649 | 95,31 | 7,692 | 3 | 0,05 | 5742 | 96,88 |
| 2,381 | 4 | 0,07 | 5574 | 94,04 | 4 | 4 | 0,07 | 5653 | 95,38 | 7,813 | 1 | 0,02 | 5743 | 96,9 |
| 2,397 | 1 | 0,02 | 5575 | 94,06 | 4,082 | 1 | 0,02 | 5654 | 95,39 | 7,965 | 1 | 0,02 | 5744 | 96,91 |
| 2,405 | 1 | 0,02 | 5576 | 94,08 | 4,087 | 1 | 0,02 | 5655 | 95,41 | 8 | 1 | 0,02 | 5745 | 96,93 |
| 2,424 | 1 | 0,02 | 5577 | 94,09 | 4,108 | 1 | 0,02 | 5656 | 95,43 | 8,108 | 1 | 0,02 | 5746 | 96,95 |
| 2,429 | 1 | 0,02 | 5578 | 94,11 | 4,167 | 3 | 0,05 | 5659 | 95,48 | 8,197 | 1 | 0,02 | 5747 | 96,96 |
| 2,439 | 2 | 0,03 | 5580 | 94,15 | 4,286 | 1 | 0,02 | 5660 | 95,5 | 8,235 | 1 | 0,02 | 5748 | 96,98 |
| 2,469 | 1 | 0,02 | 5581 | 94,16 | 4,348 | 1 | 0,02 | 5661 | 95,51 | 8,333 | 2 | 0,03 | 5750 | 97,01 |
| 2,489 | 1 | 0,02 | 5582 | 94,18 | 4,386 | 1 | 0,02 | 5662 | 95,53 | 8,397 | 1 | 0,02 | 5751 | 97,03 |
| 2,5 | 1 | 0,02 | 5583 | 94,2 | 4,39 | 1 | 0,02 | 5663 | 95,55 | 8,429 | 1 | 0,02 | 5752 | 97,05 |
| 2,511 | 1 | 0,02 | 5584 | 94,21 | 4,437 | 1 | 0,02 | 5664 | 95,56 | 8,547 | 1 | 0,02 | 5753 | 97,06 |
| 2,553 | 1 | 0,02 | 5585 | 94,23 | 4,534 | 1 | 0,02 | 5665 | 95,58 | 8,889 | 1 | 0,02 | 5754 | 97,08 |
| 2,597 | 2 | 0,03 | 5587 | 94,26 | 4,545 | 4 | 0,07 | 5669 | 95,65 | 9,014 | 1 | 0,02 | 5755 | 97,1 |
| 2,632 | 1 | 0,02 | 5588 | 94,28 | 4,586 | 1 | 0,02 | 5670 | 95,66 | 9,091 | 1 | 0,02 | 5756 | 97,11 |
| 2,667 | 1 | 0,02 | 5589 | 94,3 | 4,615 | 1 | 0,02 | 5671 | 95,68 | 9,302 | 1 | 0,02 | 5757 | 97,13 |
| 2,732 | 1 | 0,02 | 5590 | 94,31 | 4,639 | 1 | 0,02 | 5672 | 95,7 | 9,524 | 2 | 0,03 | 5759 | 97,17 |
| 2,74 | 1 | 0,02 | 5591 | 94,33 | 4,658 | 1 | 0,02 | 5673 | 95,71 | 9,756 | 2 | 0,03 | 5761 | 97,2 |
| 2,752 | 1 | 0,02 | 5592 | 94,35 | 4,717 | 1 | 0,02 | 5674 | 95,73 | 9,901 | 1 | 0,02 | 5762 | 97,22 |
| 2,773 | 1 | 0,02 | 5593 | 94,36 | 4,762 | 3 | 0,05 | 5677 | 95,78 | 9,924 | 1 | 0,02 | 5763 | 97,23 |
| 2,778 | 3 | 0,05 | 5596 | 94,42 | 4,785 | 1 | 0,02 | 5678 | 95,8 | 10 | 2 | 0,03 | 5765 | 97,27 |
| 2,809 | 1 | 0,02 | 5597 | 94,43 | 4,893 | 1 | 0,02 | 5679 | 95,82 | 10,11 | 1 | 0,02 | 5766 | 97,28 |
| 2,817 | 1 | 0,02 | 5598 | 94,45 | 4,918 | 1 | 0,02 | 5680 | 95,83 | 10,17 | 1 | 0,02 | 5767 | 97,3 |
| 2,837 | 1 | 0,02 | 5599 | 94,47 | 4,928 | 1 | 0,02 | 5681 | 95,85 | 10,18 | 1 | 0,02 | 5768 | 97,32 |
| 2,89 | 1 | 0,02 | 5600 | 94,48 | 4,933 | 1 | 0,02 | 5682 | 95,87 | 10,34 | 1 | 0,02 | 5769 | 97,33 |
| 2,941 | 1 | 0,02 | 5601 | 94,5 | 4,938 | 1 | 0,02 | 5683 | 95,88 | 10,59 | 1 | 0,02 | 5770 | 97,35 |
| 2,951 | 1 | 0,02 | 5602 | 94,52 | 5 | 2 | 0,03 | 5685 | 95,92 | 10,98 | 1 | 0,02 | 5771 | 97,37 |
| 2,957 | 1 | 0,02 | 5603 | 94,53 | 5,102 | 1 | 0,02 | 5686 | 95,93 | 11,11 | 4 | 0,07 | 5775 | 97,44 |
| 2,963 | 1 | 0,02 | 5604 | 94,55 | 5,128 | 1 | 0,02 | 5687 | 95,95 | 11,18 | 1 | 0,02 | 5776 | 97,45 |
| 2,97 | 1 | 0,02 | 5605 | 94,57 | 5,172 | 1 | 0,02 | 5688 | 95,97 | 11,25 | 1 | 0,02 | 5777 | 97,47 |
| 2,985 | 1 | 0,02 | 5606 | 94,58 | 5,263 | 2 | 0,03 | 5690 | 96 | 11,57 | 1 | 0,02 | 5778 | 97,49 |
| 2,986 | 1 | 0,02 | 5607 | 94,6 | 5,376 | 1 | 0,02 | 5691 | 96,02 | 11,67 | 1 | 0,02 | 5779 | 97,5 |
| 3,03 | 1 | 0,02 | 5608 | 94,62 | 5,405 | 3 | 0,05 | 5694 | 96,07 | 11,76 | 2 | 0,03 | 5781 | 97,54 |
| 3,086 | 1 | 0,02 | 5609 | 94,63 | 5,556 | 4 | 0,07 | 5698 | 96,14 | 12 | 2 | 0,03 | 5783 | 97,57 |
| 3,097 | 1 | 0,02 | 5610 | 94,65 | 5,607 | 1 | 0,02 | 5699 | 96,15 | 12,28 | 2 | 0,03 | 5785 | 97,6 |
| 3,125 | 3 | 0,05 | 5613 | 94,7 | 5,753 | 1 | 0,02 | 5700 | 96,17 | 12,5 | 4 | 0,07 | 5789 | 97,67 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **12,6** | **1** | **0,02** | **579** | **97,6** | **21,0** | **1** | **0,02** | **583** | **98,4** | **36,3** | **1** | **0,02** | **586** | **98,9** |
| **12,7** | 1 | 0,02 | 579 | 97,7 | **21,2** | 1 | 0,02 | 583 | 98,4 | **37,5** | 3 | 0,05 | 586 | 99,0 |
| **12,9** | 1 | 0,02 | 579 | 97,7 | **21,4** | 1 | 0,02 | 583 | 98,4 | **39,3** | 1 | 0,02 | 587 | 99,0 |
| **13,0** | 2 | 0,03 | 579 | 97,7 | **21,5** | 1 | 0,02 | 583 | 98,4 | **40** | 5 | 0,08 | 587 | 99,1 |
| **13,2** | 1 | 0,02 | 579 | 97,7 | **22,2** | 2 | 0,03 | 583 | 98,5 | **41,1** | 1 | 0,02 | 587 | 99,1 |
| **13,3** | 3 | 0,05 | 579 | 97,8 | **22,7** | 1 | 0,02 | 583 | 98,5 | **42,8** | 2 | 0,03 | 587 | 99,1 |
| **13,6** | 2 | 0,03 | 580 | 97,8 | **23,5** | 1 | 0,02 | 584 | 98,5 | **44** | 1 | 0,02 | 587 | 99,1 |
| **13,7** | 1 | 0,02 | 580 | 97,8 | **23,6** | 1 | 0,02 | 584 | 98,5 | **47,3** | 1 | 0,02 | 588 | 99,2 |
| **14,1** | 1 | 0,02 | 580 | 97,8 | **23,8** | 1 | 0,02 | 584 | 98,5 | **47,6** | 1 | 0,02 | 588 | 99,2 |
| **14,2** | 6 | 0,1 | 580 | 97,9 | **24,1** | 1 | 0,02 | 584 | 98,5 | **50** | 9 | 0,15 | 589 | 99,3 |
| **14,6** | 1 | 0,02 | 580 | 98,0 | **25** | 4 | 0,07 | 584 | 98,6 | **54,7** | 1 | 0,02 | 589 | 99,3 |
| **14,8** | 1 | 0,02 | 581 | 98,0 | **25,5** | 1 | 0,02 | 584 | 98,6 | **58,8** | 1 | 0,02 | 589 | 99,4 |
| **15,3** | 1 | 0,02 | 581 | 98,0 | **26,4** | 1 | 0,02 | 584 | 98,6 | **60** | 1 | 0,02 | 589 | 99,4 |
| **15,4** | 1 | 0,02 | 581 | 98,0 | **28,5** | 2 | 0,03 | 585 | 98,7 | **64,2** | 1 | 0,02 | 589 | 99,4 |
| **16,6** | 3 | 0,05 | 581 | 98,1 | **29,1** | 1 | 0,02 | 585 | 98,7 | **66,6** | 2 | 0,03 | 589 | 99,4 |
| **17,3** | 1 | 0,02 | 581 | 98,1 | **30** | 1 | 0,02 | 585 | 98,7 | **71,4** | 1 | 0,02 | 589 | 99,4 |
| **18,1** | 4 | 0,07 | 582 | 98,1 | **30,1** | 1 | 0,02 | 585 | 98,7 | **75** | 1 | 0,02 | 589 | 99,5 |
| **18,6** | 1 | 0,02 | 582 | 98,2 | **30,7** | 2 | 0,03 | 585 | 98,8 | **80** | 3 | 0,05 | 590 | 99,5 |
| **18,7** | 1 | 0,02 | 582 | 98,2 | **32,1** | 1 | 0,02 | 585 | 98,8 | **92,8** | 1 | 0,02 | 590 | 99,5 |
| **19,2** | 1 | 0,02 | 582 | 98,2 | **32,4** | 1 | 0,02 | 585 | 98,8 | **100** | 25 | 0,42 | 592 | 100 |
| **19,7** | 1 | 0,02 | 582 | 98,2 | **33,3** | 6 | 0,1 | 586 | 98,9 | | | | | |
| **20** | 8 | 0,13 | 583 | 98,4 | **36,1** | 1 | 0,02 | 586 | 98,9 | | | | | |

## H.11 Frequency Table Risk Factor Time of Transportation

| Of % transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent | Of % transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent | Of % transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 5085 | 85,79 | 5085 | 85,79 | **0,26** | 1 | 0,02 | 5109 | 86,2 | **0,41** | 1 | 0,02 | 5135 | 86,64 |
| **0,06** | 1 | 0,02 | 5086 | 85,81 | **0,27** | 1 | 0,02 | 5110 | 86,22 | **0,41** | 1 | 0,02 | 5136 | 86,65 |
| **0,09** | 1 | 0,02 | 5087 | 85,83 | **0,27** | 1 | 0,02 | 5111 | 86,23 | **0,42** | 1 | 0,02 | 5137 | 86,67 |
| **0,09** | 1 | 0,02 | 5088 | 85,84 | **0,27** | 2 | 0,03 | 5113 | 86,27 | **0,43** | 1 | 0,02 | 5138 | 86,69 |
| **0,11** | 1 | 0,02 | 5089 | 85,86 | **0,28** | 1 | 0,02 | 5114 | 86,28 | **0,44** | 1 | 0,02 | 5139 | 86,7 |
| **0,12** | 1 | 0,02 | 5090 | 85,88 | **0,28** | 1 | 0,02 | 5115 | 86,3 | **0,45** | 1 | 0,02 | 5140 | 86,72 |
| **0,12** | 1 | 0,02 | 5091 | 85,9 | **0,28** | 1 | 0,02 | 5116 | 86,32 | **0,47** | 1 | 0,02 | 5141 | 86,74 |
| **0,13** | 1 | 0,02 | 5092 | 85,91 | **0,29** | 1 | 0,02 | 5117 | 86,33 | **0,47** | 1 | 0,02 | 5142 | 86,76 |
| **0,13** | 1 | 0,02 | 5093 | 85,93 | **0,29** | 1 | 0,02 | 5118 | 86,35 | **0,47** | 1 | 0,02 | 5143 | 86,77 |
| **0,13** | 1 | 0,02 | 5094 | 85,95 | **0,31** | 1 | 0,02 | 5119 | 86,37 | **0,48** | 1 | 0,02 | 5144 | 86,79 |
| **0,14** | 1 | 0,02 | 5095 | 85,96 | **0,31** | 1 | 0,02 | 5120 | 86,38 | **0,48** | 1 | 0,02 | 5145 | 86,81 |
| **0,16** | 1 | 0,02 | 5096 | 85,98 | **0,32** | 1 | 0,02 | 5121 | 86,4 | **0,48** | 1 | 0,02 | 5146 | 86,82 |
| **0,16** | 1 | 0,02 | 5097 | 86 | **0,32** | 2 | 0,03 | 5123 | 86,43 | **0,48** | 1 | 0,02 | 5147 | 86,84 |
| **0,16** | 1 | 0,02 | 5098 | 86,01 | **0,33** | 1 | 0,02 | 5124 | 86,45 | **0,48** | 1 | 0,02 | 5148 | 86,86 |
| **0,16** | 1 | 0,02 | 5099 | 86,03 | **0,33** | 1 | 0,02 | 5125 | 86,47 | **0,50** | 1 | 0,02 | 5149 | 86,87 |
| **0,18** | 1 | 0,02 | 5100 | 86,05 | **0,33** | 1 | 0,02 | 5126 | 86,49 | **0,50** | 1 | 0,02 | 5150 | 86,89 |
| **0,19** | 1 | 0,02 | 5101 | 86,06 | **0,35** | 1 | 0,02 | 5127 | 86,5 | **0,50** | 1 | 0,02 | 5151 | 86,91 |
| **0,22** | 1 | 0,02 | 5102 | 86,08 | **0,36** | 1 | 0,02 | 5128 | 86,52 | **0,51** | 1 | 0,02 | 5152 | 86,92 |
| **0,24** | 1 | 0,02 | 5103 | 86,1 | **0,37** | 1 | 0,02 | 5129 | 86,54 | **0,52** | 1 | 0,02 | 5153 | 86,94 |
| **0,25** | 1 | 0,02 | 5104 | 86,11 | **0,37** | 1 | 0,02 | 5130 | 86,55 | **0,52** | 1 | 0,02 | 5154 | 86,96 |
| **0,25** | 1 | 0,02 | 5105 | 86,13 | **0,38** | 1 | 0,02 | 5131 | 86,57 | **0,53** | 1 | 0,02 | 5155 | 86,97 |
| **0,25** | 1 | 0,02 | 5106 | 86,15 | **0,38** | 1 | 0,02 | 5132 | 86,59 | **0,53** | 1 | 0,02 | 5156 | 86,99 |
| **0,26** | 1 | 0,02 | 5107 | 86,17 | **0,39** | 1 | 0,02 | 5133 | 86,6 | **0,54** | 2 | 0,03 | 5158 | 87,03 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0,26** | **1** | **0,02** | **510** | **86,1** | 0,395 | **1** | **0,02** | **513** | **86,6** | 0,542 | **1** | **0,02** | **515** | **87,0** |
| **0,55** | 1 | 0,02 | 5160 | 87,06 | 0,957 | 2 | 0,03 | 5235 | 88,32 | 1,546 | 1 | 0,02 | 5320 | 89,76 |
| **0,55** | 1 | 0,02 | 5161 | 87,08 | **0,96** | 1 | 0,02 | 5236 | 88,34 | **1,55** | 1 | 0,02 | 5321 | 89,78 |
| **0,55** | 1 | 0,02 | 5162 | 87,09 | **0,97** | 1 | 0,02 | 5237 | 88,36 | **1,56** | 4 | 0,07 | 5325 | 89,84 |
| **0,56** | 1 | 0,02 | 5163 | 87,11 | **0,98** | 1 | 0,02 | 5238 | 88,38 | **1,58** | 1 | 0,02 | 5326 | 89,86 |
| **0,58** | 1 | 0,02 | 5164 | 87,13 | **0,98** | 1 | 0,02 | 5239 | 88,39 | **1,59** | 1 | 0,02 | 5327 | 89,88 |
| **0,59** | 1 | 0,02 | 5165 | 87,14 | **0,98** | 1 | 0,02 | 5240 | 88,41 | **1,59** | 1 | 0,02 | 5328 | 89,89 |
| **0,59** | 1 | 0,02 | 5166 | 87,16 | **0,98** | 1 | 0,02 | 5241 | 88,43 | **1,6** | 1 | 0,02 | 5329 | 89,91 |
| **0,59** | 1 | 0,02 | 5167 | 87,18 | **1** | 6 | 0,1 | 5247 | 88,53 | **1,61** | 5 | 0,08 | 5334 | 89,99 |
| **0,60** | 1 | 0,02 | 5168 | 87,19 | **1,00** | 1 | 0,02 | 5248 | 88,54 | **1,61** | 1 | 0,02 | 5335 | 90,01 |
| **0,61** | 1 | 0,02 | 5169 | 87,21 | **1,01** | 1 | 0,02 | 5249 | 88,56 | **1,62** | 2 | 0,03 | 5337 | 90,05 |
| **0,61** | 1 | 0,02 | 5170 | 87,23 | **1,02** | 1 | 0,02 | 5250 | 88,58 | **1,63** | 1 | 0,02 | 5338 | 90,06 |
| **0,61** | 2 | 0,03 | 5172 | 87,26 | **1,02** | 1 | 0,02 | 5251 | 88,59 | **1,63** | 1 | 0,02 | 5339 | 90,08 |
| **0,62** | 1 | 0,02 | 5173 | 87,28 | **1,03** | 1 | 0,02 | 5252 | 88,61 | **1,65** | 1 | 0,02 | 5340 | 90,1 |
| **0,62** | 1 | 0,02 | 5174 | 87,3 | **1,09** | 1 | 0,02 | 5253 | 88,63 | **1,66** | 1 | 0,02 | 5341 | 90,11 |
| **0,62** | 1 | 0,02 | 5175 | 87,31 | **1,09** | 1 | 0,02 | 5254 | 88,65 | **1,68** | 1 | 0,02 | 5342 | 90,13 |
| **0,62** | 1 | 0,02 | 5176 | 87,33 | **1,12** | 1 | 0,02 | 5255 | 88,66 | **1,69** | 1 | 0,02 | 5343 | 90,15 |
| **0,63** | 1 | 0,02 | 5177 | 87,35 | **1,13** | 1 | 0,02 | 5256 | 88,68 | **1,72** | 1 | 0,02 | 5344 | 90,16 |
| **0,63** | 1 | 0,02 | 5178 | 87,36 | **1,13** | 1 | 0,02 | 5257 | 88,7 | **1,73** | 1 | 0,02 | 5345 | 90,18 |
| **0,64** | 1 | 0,02 | 5179 | 87,38 | **1,14** | 1 | 0,02 | 5258 | 88,71 | **1,73** | 1 | 0,02 | 5346 | 90,2 |
| **0,65** | 1 | 0,02 | 5180 | 87,4 | **1,15** | 1 | 0,02 | 5259 | 88,73 | **1,75** | 3 | 0,05 | 5349 | 90,25 |
| **0,66** | 1 | 0,02 | 5181 | 87,41 | **1,16** | 3 | 0,05 | 5262 | 88,78 | **1,80** | 1 | 0,02 | 5350 | 90,26 |
| **0,66** | 1 | 0,02 | 5182 | 87,43 | **1,17** | 1 | 0,02 | 5263 | 88,8 | **1,80** | 1 | 0,02 | 5351 | 90,28 |
| **0,67** | 2 | 0,03 | 5184 | 87,46 | **1,17** | 1 | 0,02 | 5264 | 88,81 | **1,80** | 1 | 0,02 | 5352 | 90,3 |
| **0,67** | 1 | 0,02 | 5185 | 87,48 | **1,18** | 1 | 0,02 | 5265 | 88,83 | **1,81** | 1 | 0,02 | 5353 | 90,32 |
| **0,68** | 1 | 0,02 | 5186 | 87,5 | **1,19** | 2 | 0,03 | 5267 | 88,86 | **1,81** | 3 | 0,05 | 5356 | 90,37 |
| **0,68** | 2 | 0,03 | 5188 | 87,53 | **1,20** | 1 | 0,02 | 5268 | 88,88 | **1,83** | 1 | 0,02 | 5357 | 90,38 |
| **0,69** | 1 | 0,02 | 5189 | 87,55 | **1,21** | 1 | 0,02 | 5269 | 88,9 | **1,85** | 2 | 0,03 | 5359 | 90,42 |
| **0,69** | 1 | 0,02 | 5190 | 87,57 | **1,21** | 1 | 0,02 | 5270 | 88,92 | **1,88** | 3 | 0,05 | 5362 | 90,47 |
| **0,70** | 1 | 0,02 | 5191 | 87,58 | **1,22** | 2 | 0,03 | 5272 | 88,95 | **1,91** | 1 | 0,02 | 5363 | 90,48 |
| **0,70** | 1 | 0,02 | 5192 | 87,6 | **1,23** | 1 | 0,02 | 5273 | 88,97 | **1,92** | 1 | 0,02 | 5364 | 90,5 |
| **0,71** | 1 | 0,02 | 5193 | 87,62 | **1,23** | 1 | 0,02 | 5274 | 88,98 | **1,92** | 1 | 0,02 | 5365 | 90,52 |
| **0,71** | 1 | 0,02 | 5194 | 87,63 | **1,24** | 1 | 0,02 | 5275 | 89 | **1,93** | 1 | 0,02 | 5366 | 90,53 |
| **0,72** | 2 | 0,03 | 5196 | 87,67 | **1,25** | 2 | 0,03 | 5277 | 89,03 | **1,93** | 1 | 0,02 | 5367 | 90,55 |
| **0,73** | 1 | 0,02 | 5197 | 87,68 | **1,26** | 1 | 0,02 | 5278 | 89,05 | **1,94** | 1 | 0,02 | 5368 | 90,57 |
| **0,74** | 3 | 0,05 | 5200 | 87,73 | **1,28** | 2 | 0,03 | 5280 | 89,08 | **1,96** | 3 | 0,05 | 5371 | 90,62 |
| **0,74** | 1 | 0,02 | 5201 | 87,75 | **1,29** | 1 | 0,02 | 5281 | 89,1 | **1,98** | 2 | 0,03 | 5373 | 90,65 |
| **0,74** | 1 | 0,02 | 5202 | 87,77 | **1,29** | 3 | 0,05 | 5284 | 89,15 | **2** | 1 | 0,02 | 5374 | 90,67 |
| **0,75** | 1 | 0,02 | 5203 | 87,78 | **1,30** | 1 | 0,02 | 5285 | 89,17 | **2,02** | 3 | 0,05 | 5377 | 90,72 |
| **0,76** | 1 | 0,02 | 5204 | 87,8 | **1,31** | 1 | 0,02 | 5286 | 89,19 | **2,03** | 1 | 0,02 | 5378 | 90,74 |
| **0,76** | 1 | 0,02 | 5205 | 87,82 | **1,32** | 1 | 0,02 | 5287 | 89,2 | **2,04** | 3 | 0,05 | 5381 | 90,79 |
| **0,77** | 2 | 0,03 | 5207 | 87,85 | **1,32** | 1 | 0,02 | 5288 | 89,22 | **2,04** | 1 | 0,02 | 5382 | 90,8 |
| **0,78** | 1 | 0,02 | 5208 | 87,87 | **1,33** | 3 | 0,05 | 5291 | 89,27 | **2,05** | 1 | 0,02 | 5383 | 90,82 |
| **0,78** | 2 | 0,03 | 5210 | 87,9 | **1,35** | 2 | 0,03 | 5293 | 89,3 | **2,08** | 2 | 0,03 | 5385 | 90,86 |
| **0,8** | 1 | 0,02 | 5211 | 87,92 | **1,37** | 1 | 0,02 | 5294 | 89,32 | **2,12** | 1 | 0,02 | 5386 | 90,87 |
| **0,80** | 1 | 0,02 | 5212 | 87,94 | **1,38** | 3 | 0,05 | 5297 | 89,37 | **2,14** | 1 | 0,02 | 5387 | 90,89 |
| **0,80** | 1 | 0,02 | 5213 | 87,95 | **1,39** | 1 | 0,02 | 5298 | 89,39 | **2,14** | 1 | 0,02 | 5388 | 90,91 |
| **0,80** | 1 | 0,02 | 5214 | 87,97 | **1,40** | 2 | 0,03 | 5300 | 89,42 | **2,15** | 1 | 0,02 | 5389 | 90,92 |
| **0,82** | 1 | 0,02 | 5215 | 87,99 | **1,41** | 1 | 0,02 | 5301 | 89,44 | **2,16** | 1 | 0,02 | 5390 | 90,94 |
| **0,82** | 1 | 0,02 | 5216 | 88 | **1,42** | 1 | 0,02 | 5302 | 89,46 | **2,19** | 2 | 0,03 | 5392 | 90,97 |
| **0,83** | 1 | 0,02 | 5217 | 88,02 | **1,44** | 1 | 0,02 | 5303 | 89,47 | **2,22** | 4 | 0,07 | 5396 | 91,04 |
| **0,84** | 2 | 0,03 | 5219 | 88,05 | **1,44** | 2 | 0,03 | 5305 | 89,51 | **2,27** | 3 | 0,05 | 5399 | 91,09 |
| **0,84** | 3 | 0,05 | 5222 | 88,11 | **1,45** | 1 | 0,02 | 5306 | 89,52 | **2,29** | 1 | 0,02 | 5400 | 91,11 |
| **0,85** | 1 | 0,02 | 5223 | 88,12 | **1,46** | 1 | 0,02 | 5307 | 89,54 | **2,31** | 1 | 0,02 | 5401 | 91,13 |
| **0,86** | 2 | 0,03 | 5225 | 88,16 | **1,47** | 4 | 0,07 | 5311 | 89,61 | **2,32** | 2 | 0,03 | 5403 | 91,16 |
| **0,87** | 2 | 0,03 | 5227 | 88,19 | **1,48** | 1 | 0,02 | 5312 | 89,62 | **2,35** | 2 | 0,03 | 5405 | 91,19 |
| **0,90** | 1 | 0,02 | 5228 | 88,21 | **1,49** | 3 | 0,05 | 5315 | 89,67 | **2,38** | 4 | 0,07 | 5409 | 91,26 |
| **0,90** | 1 | 0,02 | 5229 | 88,22 | **1,5** | 1 | 0,02 | 5316 | 89,69 | **2,41** | 1 | 0,02 | 5410 | 91,28 |
| **0,91** | 1 | 0,02 | 5230 | 88,24 | **1,50** | 1 | 0,02 | 5317 | 89,71 | **2,41** | 1 | 0,02 | 5411 | 91,29 |
| **0,93** | 2 | 0,03 | 5232 | 88,27 | **1,52** | 1 | 0,02 | 5318 | 89,72 | **2,42** | 1 | 0,02 | 5412 | 91,31 |
| **0,95** | 1 | 0,02 | 5233 | 88,29 | **1,53** | 1 | 0,02 | 5319 | 89,74 | **2,43** | 3 | 0,05 | 5415 | 91,36 |
| **2,45** | 1 | 0,02 | 5416 | 91,38 | 4,237 | 1 | 0,02 | 5534 | 93,37 | 7,229 | 1 | 0,02 | 5648 | 95,29 |
| **2,45** | 1 | 0,02 | 5417 | 91,4 | **4,25** | 1 | 0,02 | 5535 | 93,39 | **7,26** | 1 | 0,02 | 5649 | 95,31 |
| **2,46** | 1 | 0,02 | 5418 | 91,41 | **4,26** | 1 | 0,02 | 5536 | 93,4 | **7,29** | 1 | 0,02 | 5650 | 95,33 |
| **2,48** | 1 | 0,02 | 5419 | 91,43 | **4,28** | 1 | 0,02 | 5537 | 93,42 | **7,30** | 1 | 0,02 | 5651 | 95,34 |
| **2,5** | 4 | 0,07 | 5423 | 91,5 | **4,30** | 1 | 0,02 | 5538 | 93,44 | **7,38** | 1 | 0,02 | 5652 | 95,36 |
| **2,55** | 1 | 0,02 | 5424 | 91,51 | **4,34** | 2 | 0,03 | 5540 | 93,47 | **7,40** | 2 | 0,03 | 5654 | 95,39 |
| **2,56** | 3 | 0,05 | 5427 | 91,56 | **4,46** | 2 | 0,03 | 5542 | 93,5 | **7,65** | 1 | 0,02 | 5655 | 95,41 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **2,59** | 1 | 0,02 | 5428 | 91,58 | **4,47** | 1 | 0,02 | 5543 | 93,52 | **7,69** | 9 | 0,15 | 5664 | 95,56 |
| **2,63** | 1 | 0,02 | 5429 | 91,6 | **4,54** | 1 | 0,02 | 5544 | 93,54 | **7,75** | 1 | 0,02 | 5665 | 95,58 |
| **2,64** | 1 | 0,02 | 5430 | 91,61 | **4,61** | 1 | 0,02 | 5545 | 93,55 | **7,79** | 1 | 0,02 | 5666 | 95,6 |
| **2,66** | 1 | 0,02 | 5431 | 91,63 | **4,67** | 1 | 0,02 | 5546 | 93,57 | **7,83** | 1 | 0,02 | 5667 | 95,61 |
| **2,70** | 1 | 0,02 | 5432 | 91,65 | **4,76** | 5 | 0,08 | 5551 | 93,66 | **7,89** | 1 | 0,02 | 5668 | 95,63 |
| **2,75** | 1 | 0,02 | 5433 | 91,67 | **4,83** | 1 | 0,02 | 5552 | 93,67 | **7,93** | 1 | 0,02 | 5669 | 95,65 |
| **2,77** | 6 | 0,1 | 5439 | 91,77 | **4,91** | 1 | 0,02 | 5553 | 93,69 | **7,97** | 1 | 0,02 | 5670 | 95,66 |
| **2,79** | 1 | 0,02 | 5440 | 91,78 | **5** | 2 | 0,03 | 5555 | 93,72 | **8** | 2 | 0,03 | 5672 | 95,7 |
| **2,80** | 1 | 0,02 | 5441 | 91,8 | **5,04** | 1 | 0,02 | 5556 | 93,74 | **8,06** | 1 | 0,02 | 5673 | 95,71 |
| **2,81** | 1 | 0,02 | 5442 | 91,82 | **5,05** | 1 | 0,02 | 5557 | 93,76 | **8,21** | 1 | 0,02 | 5674 | 95,73 |
| **2,85** | 1 | 0,02 | 5443 | 91,83 | **5,08** | 2 | 0,03 | 5559 | 93,79 | **8,25** | 1 | 0,02 | 5675 | 95,75 |
| **2,89** | 2 | 0,03 | 5445 | 91,87 | **5,10** | 1 | 0,02 | 5560 | 93,81 | **8,33** | 10 | 0,17 | 5685 | 95,92 |
| **2,91** | 1 | 0,02 | 5446 | 91,88 | **5,10** | 1 | 0,02 | 5561 | 93,82 | **8,47** | 1 | 0,02 | 5686 | 95,93 |
| **2,94** | 1 | 0,02 | 5447 | 91,9 | **5,11** | 1 | 0,02 | 5562 | 93,84 | **8,51** | 1 | 0,02 | 5687 | 95,95 |
| **3** | 1 | 0,02 | 5448 | 91,92 | **5,17** | 1 | 0,02 | 5563 | 93,86 | **8,69** | 2 | 0,03 | 5689 | 95,98 |
| **3,03** | 6 | 0,1 | 5454 | 92,02 | **5,19** | 1 | 0,02 | 5564 | 93,88 | **8,72** | 1 | 0,02 | 5690 | 96 |
| **3,09** | 1 | 0,02 | 5455 | 92,04 | **5,20** | 1 | 0,02 | 5565 | 93,89 | **8,8** | 1 | 0,02 | 5691 | 96,02 |
| **3,12** | 4 | 0,07 | 5459 | 92,1 | **5,26** | 5 | 0,08 | 5570 | 93,98 | **8,82** | 1 | 0,02 | 5692 | 96,04 |
| **3,13** | 1 | 0,02 | 5460 | 92,12 | **5,30** | 1 | 0,02 | 5571 | 93,99 | **9,09** | 10 | 0,17 | 5702 | 96,2 |
| **3,19** | 1 | 0,02 | 5461 | 92,14 | **5,31** | 1 | 0,02 | 5572 | 94,01 | **9,30** | 2 | 0,03 | 5704 | 96,24 |
| **3,21** | 1 | 0,02 | 5462 | 92,15 | **5,33** | 1 | 0,02 | 5573 | 94,03 | **9,37** | 1 | 0,02 | 5705 | 96,25 |
| **3,22** | 3 | 0,05 | 5465 | 92,21 | **5,46** | 2 | 0,03 | 5575 | 94,06 | **9,43** | 1 | 0,02 | 5706 | 96,27 |
| **3,23** | 1 | 0,02 | 5466 | 92,22 | **5,47** | 1 | 0,02 | 5576 | 94,08 | **9,52** | 1 | 0,02 | 5707 | 96,29 |
| **3,25** | 1 | 0,02 | 5467 | 92,24 | **5,49** | 2 | 0,03 | 5578 | 94,11 | **9,63** | 1 | 0,02 | 5708 | 96,31 |
| **3,27** | 1 | 0,02 | 5468 | 92,26 | **5,50** | 1 | 0,02 | 5579 | 94,13 | **9,67** | 1 | 0,02 | 5709 | 96,32 |
| **3,33** | 9 | 0,15 | 5477 | 92,41 | **5,55** | 4 | 0,07 | 5583 | 94,2 | **9,95** | 1 | 0,02 | 5710 | 96,34 |
| **3,35** | 1 | 0,02 | 5478 | 92,42 | **5,66** | 2 | 0,03 | 5585 | 94,23 | **10** | 8 | 0,13 | 5718 | 96,47 |
| **3,35** | 1 | 0,02 | 5479 | 92,44 | **5,67** | 1 | 0,02 | 5586 | 94,25 | **10,3** | 1 | 0,02 | 5719 | 96,49 |
| **3,37** | 1 | 0,02 | 5480 | 92,46 | **5,69** | 1 | 0,02 | 5587 | 94,26 | **10,4** | 1 | 0,02 | 5720 | 96,51 |
| **3,39** | 1 | 0,02 | 5481 | 92,48 | **5,71** | 1 | 0,02 | 5588 | 94,28 | **10,5** | 1 | 0,02 | 5721 | 96,52 |
| **3,40** | 2 | 0,03 | 5483 | 92,51 | **5,73** | 1 | 0,02 | 5589 | 94,3 | **10,6** | 2 | 0,03 | 5723 | 96,56 |
| **3,44** | 3 | 0,05 | 5486 | 92,56 | **5,76** | 1 | 0,02 | 5590 | 94,31 | **10,8** | 1 | 0,02 | 5724 | 96,57 |
| **3,50** | 1 | 0,02 | 5487 | 92,58 | **5,79** | 1 | 0,02 | 5591 | 94,33 | **10,9** | 1 | 0,02 | 5725 | 96,59 |
| **3,53** | 1 | 0,02 | 5488 | 92,59 | **5,88** | 12 | 0,2 | 5603 | 94,53 | **10,9** | 1 | 0,02 | 5726 | 96,61 |
| **3,57** | 4 | 0,07 | 5492 | 92,66 | **5,92** | 1 | 0,02 | 5604 | 94,55 | **11,1** | 9 | 0,15 | 5735 | 96,76 |
| **3,63** | 2 | 0,03 | 5494 | 92,69 | **5,95** | 1 | 0,02 | 5605 | 94,57 | **11,7** | 2 | 0,03 | 5737 | 96,79 |
| **3,70** | 4 | 0,07 | 5498 | 92,76 | **6,01** | 1 | 0,02 | 5606 | 94,58 | **11,8** | 1 | 0,02 | 5738 | 96,81 |
| **3,75** | 3 | 0,05 | 5501 | 92,81 | **6,14** | 1 | 0,02 | 5607 | 94,6 | **11,9** | 1 | 0,02 | 5739 | 96,83 |
| **3,77** | 1 | 0,02 | 5502 | 92,83 | **6,17** | 1 | 0,02 | 5608 | 94,62 | **11,9** | 1 | 0,02 | 5740 | 96,84 |
| **3,82** | 1 | 0,02 | 5503 | 92,85 | **6,25** | 9 | 0,15 | 5617 | 94,77 | **12,2** | 2 | 0,03 | 5742 | 96,88 |
| **3,84** | 4 | 0,07 | 5507 | 92,91 | **6,38** | 1 | 0,02 | 5618 | 94,79 | **12,5** | 11 | 0,19 | 5753 | 97,06 |
| **3,92** | 4 | 0,07 | 5511 | 92,98 | **6,45** | 1 | 0,02 | 5619 | 94,8 | **12,5** | 1 | 0,02 | 5754 | 97,08 |
| **3,93** | 1 | 0,02 | 5512 | 93 | **6,52** | 1 | 0,02 | 5620 | 94,82 | **12,5** | 1 | 0,02 | 5755 | 97,1 |
| **3,95** | 1 | 0,02 | 5513 | 93,02 | **6,66** | 6 | 0,1 | 5626 | 94,92 | **13,0** | 2 | 0,03 | 5757 | 97,13 |
| **4** | 4 | 0,07 | 5517 | 93,08 | **6,68** | 1 | 0,02 | 5627 | 94,94 | **13,3** | 5 | 0,08 | 5762 | 97,22 |
| **4,03** | 1 | 0,02 | 5518 | 93,1 | **6,74** | 1 | 0,02 | 5628 | 94,96 | **13,4** | 1 | 0,02 | 5763 | 97,23 |
| **4,04** | 1 | 0,02 | 5519 | 93,12 | **6,76** | 1 | 0,02 | 5629 | 94,97 | **13,5** | 1 | 0,02 | 5764 | 97,25 |
| **4,05** | 1 | 0,02 | 5520 | 93,13 | **6,81** | 1 | 0,02 | 5630 | 94,99 | **13,6** | 1 | 0,02 | 5765 | 97,27 |
| **4,08** | 2 | 0,03 | 5522 | 93,17 | **6,84** | 1 | 0,02 | 5631 | 95,01 | **13,8** | 1 | 0,02 | 5766 | 97,28 |
| **4,11** | 1 | 0,02 | 5523 | 93,18 | **6,89** | 3 | 0,05 | 5634 | 95,06 | **14,2** | 8 | 0,13 | 5774 | 97,42 |
| **4,16** | 5 | 0,08 | 5528 | 93,27 | **6,94** | 1 | 0,02 | 5635 | 95,07 | **14,8** | 1 | 0,02 | 5775 | 97,44 |
| **4,19** | 1 | 0,02 | 5529 | 93,28 | **7,01** | 1 | 0,02 | 5636 | 95,09 | **14,9** | 1 | 0,02 | 5776 | 97,45 |
| **4,20** | 1 | 0,02 | 5530 | 93,3 | **7,06** | 1 | 0,02 | 5637 | 95,11 | **14,9** | 1 | 0,02 | 5777 | 97,47 |
| **4,21** | 2 | 0,03 | 5532 | 93,34 | **7,09** | 1 | 0,02 | 5638 | 95,12 | **15,2** | 1 | 0,02 | 5778 | 97,49 |
| **4,21** | 1 | 0,02 | 5533 | 93,35 | **7,14** | 9 | 0,15 | 5647 | 95,28 | **15,2** | 1 | 0,02 | 5779 | 97,5 |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **15,38** | **2** | **0,03** | **5781** | **97,54** | **20,83** | **1** | **0,02** | **5824** | **98,26** | **36,36** | **2** | **0,03** | **5883** | **99,26** |
| **15,79** | 1 | 0,02 | 5782 | 97,55 | **21,43** | 1 | 0,02 | 5825 | 98,28 | **37,25** | 1 | 0,02 | 5884 | 99,27 |
| **16** | 2 | 0,03 | 5784 | 97,59 | **21,49** | 1 | 0,02 | 5826 | 98,3 | **40** | 5 | 0,08 | 5889 | 99,36 |
| **16,05** | 1 | 0,02 | 5785 | 97,6 | **21,62** | 1 | 0,02 | 5827 | 98,31 | **41,03** | 1 | 0,02 | 5890 | 99,38 |
| **16,07** | 1 | 0,02 | 5786 | 97,62 | **21,79** | 2 | 0,03 | 5829 | 98,35 | **41,18** | 1 | 0,02 | 5891 | 99,39 |
| **16,13** | 1 | 0,02 | 5787 | 97,64 | **22,22** | 3 | 0,05 | 5832 | 98,4 | **41,6** | 1 | 0,02 | 5892 | 99,41 |
| **16,67** | 10 | 0,17 | 5797 | 97,81 | **23,08** | 1 | 0,02 | 5833 | 98,41 | **41,67** | 1 | 0,02 | 5893 | 99,43 |
| **17,07** | 1 | 0,02 | 5798 | 97,82 | **25** | 16 | 0,27 | 5849 | 98,68 | **42,86** | 1 | 0,02 | 5894 | 99,44 |
| **17,14** | 2 | 0,03 | 5800 | 97,86 | **26,92** | 1 | 0,02 | 5850 | 98,7 | **43,75** | 2 | 0,03 | 5896 | 99,48 |
| **17,65** | 1 | 0,02 | 5801 | 97,87 | **27,78** | 1 | 0,02 | 5851 | 98,72 | **45,21** | 1 | 0,02 | 5897 | 99,49 |
| **17,78** | 1 | 0,02 | 5802 | 97,89 | **27,91** | 1 | 0,02 | 5852 | 98,73 | **46,15** | 1 | 0,02 | 5898 | 99,51 |
| **17,86** | 1 | 0,02 | 5803 | 97,91 | **28,57** | 3 | 0,05 | 5855 | 98,79 | **50** | 13 | 0,22 | 5911 | 99,73 |
| **18,02** | 1 | 0,02 | 5804 | 97,92 | **28,95** | 1 | 0,02 | 5856 | 98,8 | **52,17** | 1 | 0,02 | 5912 | 99,75 |
| **18,18** | 2 | 0,03 | 5806 | 97,96 | **29,33** | 1 | 0,02 | 5857 | 98,82 | **52,38** | 1 | 0,02 | 5913 | 99,76 |
| **18,52** | 2 | 0,03 | 5808 | 97,99 | **29,41** | 1 | 0,02 | 5858 | 98,84 | **56,25** | 1 | 0,02 | 5914 | 99,78 |
| **18,56** | 1 | 0,02 | 5809 | 98,01 | **30** | 1 | 0,02 | 5859 | 98,85 | **60** | 1 | 0,02 | 5915 | 99,8 |
| **18,75** | 1 | 0,02 | 5810 | 98,03 | **30,77** | 2 | 0,03 | 5861 | 98,89 | **62,5** | 2 | 0,03 | 5917 | 99,83 |
| **18,82** | 1 | 0,02 | 5811 | 98,04 | **31,58** | 1 | 0,02 | 5862 | 98,9 | **66,67** | 1 | 0,02 | 5918 | 99,85 |
| **19,44** | 1 | 0,02 | 5812 | 98,06 | **33,33** | 16 | 0,27 | 5878 | 99,17 | **75** | 1 | 0,02 | 5919 | 99,87 |
| **20** | 9 | 0,15 | 5821 | 98,21 | **34,62** | 1 | 0,02 | 5879 | 99,19 | **85,71** | 1 | 0,02 | 5920 | 99,88 |
| **20,34** | 1 | 0,02 | 5822 | 98,23 | **35** | 1 | 0,02 | 5880 | 99,21 | **100** | 7 | 0,12 | 5927 | 100 |
| **20,69** | 1 | 0,02 | 5823 | 98,25 | **35,06** | 1 | 0,02 | 5881 | 99,22 | | | | | |

## H.12 Frequency Table Risk Factor VDM Modifications

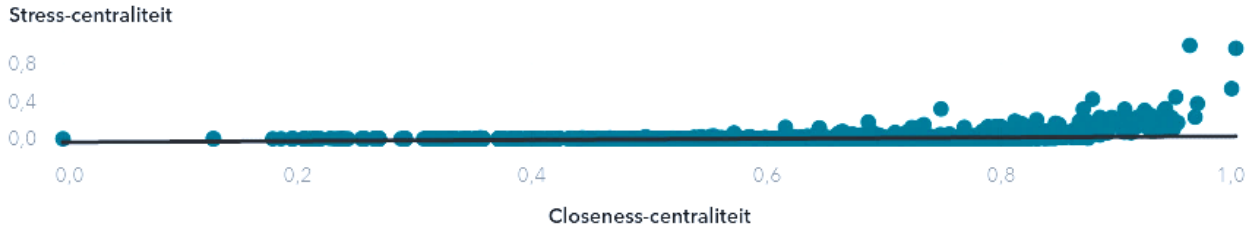
| % Of transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent | % Of transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent | % Of transportations | Frequency | Percent | Cumulative Frequency | Cumulative Percent |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 4757 | 80,26 | 4757 | 80,26 | **0,625** | 1 | 0,02 | 4787 | 80,77 | **0,935** | 1 | 0,02 | 4820 | 81,32 |
| 0,169 | 1 | 0,02 | 4758 | 80,28 | **0,631** | 1 | 0,02 | 4788 | 80,78 | **0,94** | 1 | 0,02 | 4821 | 81,34 |
| 0,21 | 1 | 0,02 | 4759 | 80,29 | **0,633** | 1 | 0,02 | 4789 | 80,8 | **0,943** | 3 | 0,05 | 4824 | 81,39 |
| 0,219 | 1 | 0,02 | 4760 | 80,31 | **0,667** | 1 | 0,02 | 4790 | 80,82 | **0,952** | 3 | 0,05 | 4827 | 81,44 |
| 0,254 | 1 | 0,02 | 4761 | 80,33 | **0,676** | 1 | 0,02 | 4791 | 80,83 | **1,02** | 1 | 0,02 | 4828 | 81,46 |
| 0,258 | 1 | 0,02 | 4762 | 80,34 | **0,678** | 1 | 0,02 | 4792 | 80,85 | **1,031** | 2 | 0,03 | 4830 | 81,49 |
| 0,278 | 1 | 0,02 | 4763 | 80,36 | **0,69** | 1 | 0,02 | 4793 | 80,87 | **1,066** | 1 | 0,02 | 4831 | 81,51 |
| 0,289 | 1 | 0,02 | 4764 | 80,38 | **0,69** | 1 | 0,02 | 4794 | 80,88 | **1,099** | 1 | 0,02 | 4832 | 81,53 |
| 0,292 | 1 | 0,02 | 4765 | 80,39 | **0,725** | 1 | 0,02 | 4795 | 80,9 | **1,111** | 1 | 0,02 | 4833 | 81,54 |
| 0,327 | 1 | 0,02 | 4766 | 80,41 | **0,727** | 1 | 0,02 | 4796 | 80,92 | **1,124** | 1 | 0,02 | 4834 | 81,56 |
| 0,333 | 1 | 0,02 | 4767 | 80,43 | **0,741** | 1 | 0,02 | 4797 | 80,93 | **1,139** | 1 | 0,02 | 4835 | 81,58 |
| 0,368 | 1 | 0,02 | 4768 | 80,45 | **0,746** | 1 | 0,02 | 4798 | 80,95 | **1,163** | 1 | 0,02 | 4836 | 81,59 |
| 0,379 | 1 | 0,02 | 4769 | 80,46 | **0,758** | 1 | 0,02 | 4799 | 80,97 | **1,176** | 1 | 0,02 | 4837 | 81,61 |
| 0,389 | 1 | 0,02 | 4770 | 80,48 | **0,763** | 1 | 0,02 | 4800 | 80,99 | **1,19** | 3 | 0,05 | 4840 | 81,66 |
| 0,395 | 1 | 0,02 | 4771 | 80,5 | **0,764** | 1 | 0,02 | 4801 | 81 | **1,205** | 1 | 0,02 | 4841 | 81,68 |
| 0,408 | 1 | 0,02 | 4772 | 80,51 | **0,778** | 1 | 0,02 | 4802 | 81,02 | **1,22** | 1 | 0,02 | 4842 | 81,69 |
| 0,435 | 1 | 0,02 | 4773 | 80,53 | **0,781** | 1 | 0,02 | 4803 | 81,04 | **1,235** | 1 | 0,02 | 4843 | 81,71 |
| 0,452 | 1 | 0,02 | 4774 | 80,55 | **0,791** | 1 | 0,02 | 4804 | 81,05 | **1,282** | 1 | 0,02 | 4844 | 81,73 |
| 0,463 | 1 | 0,02 | 4775 | 80,56 | **0,813** | 2 | 0,03 | 4806 | 81,09 | **1,29** | 1 | 0,02 | 4845 | 81,74 |
| 0,476 | 1 | 0,02 | 4776 | 80,58 | **0,826** | 2 | 0,03 | 4808 | 81,12 | **1,299** | 2 | 0,03 | 4847 | 81,78 |
| 0,501 | 1 | 0,02 | 4777 | 80,6 | **0,83** | 1 | 0,02 | 4809 | 81,14 | **1,311** | 1 | 0,02 | 4848 | 81,8 |
| 0,538 | 1 | 0,02 | 4778 | 80,61 | **0,847** | 1 | 0,02 | 4810 | 81,15 | **1,327** | 1 | 0,02 | 4849 | 81,81 |
| 0,541 | 1 | 0,02 | 4779 | 80,63 | **0,857** | 1 | 0,02 | 4811 | 81,17 | **1,333** | 2 | 0,03 | 4851 | 81,85 |
| 0,546 | 1 | 0,02 | 4780 | 80,65 | **0,862** | 1 | 0,02 | 4812 | 81,19 | **1,361** | 2 | 0,03 | 4853 | 81,88 |
| 0,552 | 1 | 0,02 | 4781 | 80,66 | **0,87** | 1 | 0,02 | 4813 | 81,2 | **1,387** | 1 | 0,02 | 4854 | 81,9 |
| 0,559 | 2 | 0,03 | 4783 | 80,7 | **0,893** | 2 | 0,03 | 4815 | 81,24 | **1,395** | 1 | 0,02 | 4855 | 81,91 |
| 0,578 | 1 | 0,02 | 4784 | 80,72 | **0,894** | 1 | 0,02 | 4816 | 81,26 | **1,408** | 1 | 0,02 | 4856 | 81,93 |
| 0,585 | 1 | 0,02 | 4785 | 80,73 | **0,898** | 1 | 0,02 | 4817 | 81,27 | **1,429** | 1 | 0,02 | 4857 | 81,95 |
| 0,606 | 1 | 0,02 | 4786 | 80,75 | **0,901** | 2 | 0,03 | 4819 | 81,31 | **1,435** | 1 | 0,02 | 4858 | 81,96 |

| 1,448 | 1 | 0,02 | 4859 | 81,98 | 2,609 | 1 | 0,02 | 4934 | 83,25 | 3,846 | 3 | 0,05 | 5024 | 84,76 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1,449 | 1 | 0,02 | 4860 | 82 | 2,622 | 1 | 0,02 | 4935 | 83,26 | 3,922 | 1 | 0,02 | 5025 | 84,78 |
| 1,452 | 1 | 0,02 | 4861 | 82,01 | 2,625 | 1 | 0,02 | 4936 | 83,28 | 3,926 | 1 | 0,02 | 5026 | 84,8 |
| 1,471 | 1 | 0,02 | 4862 | 82,03 | 2,639 | 1 | 0,02 | 4937 | 83,3 | 4 | 1 | 0,02 | 5027 | 84,82 |
| 1,485 | 1 | 0,02 | 4863 | 82,05 | 2,646 | 1 | 0,02 | 4938 | 83,31 | 4,031 | 1 | 0,02 | 5028 | 84,83 |
| 1,493 | 1 | 0,02 | 4864 | 82,07 | 2,655 | 1 | 0,02 | 4939 | 83,33 | 4,082 | 2 | 0,03 | 5030 | 84,87 |
| 1,504 | 1 | 0,02 | 4865 | 82,08 | 2,703 | 2 | 0,03 | 4941 | 83,36 | 4,094 | 1 | 0,02 | 5031 | 84,88 |
| 1,523 | 1 | 0,02 | 4866 | 82,1 | 2,712 | 1 | 0,02 | 4942 | 83,38 | 4,13 | 1 | 0,02 | 5032 | 84,9 |
| 1,533 | 1 | 0,02 | 4867 | 82,12 | 2,719 | 1 | 0,02 | 4943 | 83,4 | 4,156 | 1 | 0,02 | 5033 | 84,92 |
| 1,571 | 1 | 0,02 | 4868 | 82,13 | 2,74 | 3 | 0,05 | 4946 | 83,45 | 4,167 | 4 | 0,07 | 5037 | 84,98 |
| 1,575 | 1 | 0,02 | 4869 | 82,15 | 2,747 | 1 | 0,02 | 4947 | 83,47 | 4,181 | 1 | 0,02 | 5038 | 85 |
| 1,587 | 2 | 0,03 | 4871 | 82,18 | 2,778 | 4 | 0,07 | 4951 | 83,53 | 4,19 | 1 | 0,02 | 5039 | 85,02 |
| 1,6 | 1 | 0,02 | 4872 | 82,2 | 2,806 | 1 | 0,02 | 4952 | 83,55 | 4,225 | 1 | 0,02 | 5040 | 85,03 |
| 1,604 | 1 | 0,02 | 4873 | 82,22 | 2,813 | 1 | 0,02 | 4953 | 83,57 | 4,233 | 1 | 0,02 | 5041 | 85,05 |
| 1,613 | 1 | 0,02 | 4874 | 82,23 | 2,817 | 1 | 0,02 | 4954 | 83,58 | 4,263 | 1 | 0,02 | 5042 | 85,07 |
| 1,635 | 1 | 0,02 | 4875 | 82,25 | 2,837 | 1 | 0,02 | 4955 | 83,6 | 4,274 | 1 | 0,02 | 5043 | 85,09 |
| 1,657 | 1 | 0,02 | 4876 | 82,27 | 2,857 | 3 | 0,05 | 4958 | 83,65 | 4,292 | 1 | 0,02 | 5044 | 85,1 |
| 1,667 | 1 | 0,02 | 4877 | 82,28 | 2,871 | 1 | 0,02 | 4959 | 83,67 | 4,301 | 1 | 0,02 | 5045 | 85,12 |
| 1,689 | 1 | 0,02 | 4878 | 82,3 | 2,905 | 1 | 0,02 | 4960 | 83,68 | 4,324 | 1 | 0,02 | 5046 | 85,14 |
| 1,695 | 1 | 0,02 | 4879 | 82,32 | 2,907 | 2 | 0,03 | 4962 | 83,72 | 4,348 | 1 | 0,02 | 5047 | 85,15 |
| 1,724 | 3 | 0,05 | 4882 | 82,37 | 2,918 | 1 | 0,02 | 4963 | 83,74 | 4,372 | 1 | 0,02 | 5048 | 85,17 |
| 1,739 | 1 | 0,02 | 4883 | 82,39 | 2,934 | 1 | 0,02 | 4964 | 83,75 | 4,393 | 1 | 0,02 | 5049 | 85,19 |
| 1,775 | 1 | 0,02 | 4884 | 82,4 | 2,941 | 2 | 0,03 | 4966 | 83,79 | 4,444 | 2 | 0,03 | 5051 | 85,22 |
| 1,778 | 1 | 0,02 | 4885 | 82,42 | 2,965 | 1 | 0,02 | 4967 | 83,8 | 4,464 | 1 | 0,02 | 5052 | 85,24 |
| 1,786 | 1 | 0,02 | 4886 | 82,44 | 2,976 | 1 | 0,02 | 4968 | 83,82 | 4,471 | 1 | 0,02 | 5053 | 85,25 |
| 1,802 | 1 | 0,02 | 4887 | 82,45 | 3,006 | 1 | 0,02 | 4969 | 83,84 | 4,472 | 2 | 0,03 | 5055 | 85,29 |
| 1,81 | 1 | 0,02 | 4888 | 82,47 | 3,026 | 1 | 0,02 | 4970 | 83,85 | 4,545 | 2 | 0,03 | 5057 | 85,32 |
| 1,835 | 1 | 0,02 | 4889 | 82,49 | 3,03 | 3 | 0,05 | 4973 | 83,9 | 4,552 | 1 | 0,02 | 5058 | 85,34 |
| 1,852 | 1 | 0,02 | 4890 | 82,5 | 3,077 | 1 | 0,02 | 4974 | 83,92 | 4,561 | 1 | 0,02 | 5059 | 85,36 |
| 1,887 | 2 | 0,03 | 4892 | 82,54 | 3,093 | 1 | 0,02 | 4975 | 83,94 | 4,569 | 1 | 0,02 | 5060 | 85,37 |
| 1,923 | 1 | 0,02 | 4893 | 82,55 | 3,125 | 3 | 0,05 | 4978 | 83,99 | 4,614 | 1 | 0,02 | 5061 | 85,39 |
| 1,926 | 1 | 0,02 | 4894 | 82,57 | 3,134 | 1 | 0,02 | 4979 | 84,01 | 4,706 | 1 | 0,02 | 5062 | 85,41 |
| 1,961 | 1 | 0,02 | 4895 | 82,59 | 3,163 | 1 | 0,02 | 4980 | 84,02 | 4,741 | 1 | 0,02 | 5063 | 85,42 |
| 2,02 | 1 | 0,02 | 4896 | 82,61 | 3,209 | 1 | 0,02 | 4981 | 84,04 | 4,747 | 1 | 0,02 | 5064 | 85,44 |
| 2,041 | 2 | 0,03 | 4898 | 82,64 | 3,226 | 2 | 0,03 | 4983 | 84,07 | 4,762 | 5 | 0,08 | 5069 | 85,52 |
| 2,055 | 1 | 0,02 | 4899 | 82,66 | 3,252 | 1 | 0,02 | 4984 | 84,09 | 4,765 | 1 | 0,02 | 5070 | 85,54 |
| 2,079 | 1 | 0,02 | 4900 | 82,67 | 3,261 | 1 | 0,02 | 4985 | 84,11 | 4,878 | 1 | 0,02 | 5071 | 85,56 |
| 2,083 | 2 | 0,03 | 4902 | 82,71 | 3,286 | 1 | 0,02 | 4986 | 84,12 | 5 | 6 | 0,1 | 5077 | 85,66 |
| 2,085 | 1 | 0,02 | 4903 | 82,72 | 3,333 | 4 | 0,07 | 4990 | 84,19 | 5,031 | 1 | 0,02 | 5078 | 85,68 |
| 2,128 | 1 | 0,02 | 4904 | 82,74 | 3,339 | 1 | 0,02 | 4991 | 84,21 | 5,056 | 1 | 0,02 | 5079 | 85,69 |
| 2,151 | 1 | 0,02 | 4905 | 82,76 | 3,364 | 1 | 0,02 | 4992 | 84,22 | 5,128 | 3 | 0,05 | 5082 | 85,74 |
| 2,174 | 1 | 0,02 | 4906 | 82,77 | 3,368 | 1 | 0,02 | 4993 | 84,24 | 5,147 | 1 | 0,02 | 5083 | 85,76 |
| 2,192 | 1 | 0,02 | 4907 | 82,79 | 3,378 | 1 | 0,02 | 4994 | 84,26 | 5,152 | 1 | 0,02 | 5084 | 85,78 |
| 2,218 | 1 | 0,02 | 4908 | 82,81 | 3,382 | 1 | 0,02 | 4995 | 84,28 | 5,155 | 1 | 0,02 | 5085 | 85,79 |
| 2,222 | 1 | 0,02 | 4909 | 82,82 | 3,39 | 3 | 0,05 | 4998 | 84,33 | 5,175 | 1 | 0,02 | 5086 | 85,81 |
| 2,256 | 1 | 0,02 | 4910 | 82,84 | 3,412 | 1 | 0,02 | 4999 | 84,34 | 5,199 | 1 | 0,02 | 5087 | 85,83 |
| 2,292 | 1 | 0,02 | 4911 | 82,86 | 3,448 | 2 | 0,03 | 5001 | 84,38 | 5,263 | 3 | 0,05 | 5090 | 85,88 |
| 2,326 | 2 | 0,03 | 4913 | 82,89 | 3,462 | 1 | 0,02 | 5002 | 84,39 | 5,298 | 1 | 0,02 | 5091 | 85,9 |
| 2,366 | 1 | 0,02 | 4914 | 82,91 | 3,516 | 1 | 0,02 | 5003 | 84,41 | 5,35 | 1 | 0,02 | 5092 | 85,91 |
| 2,381 | 2 | 0,03 | 4916 | 82,94 | 3,532 | 1 | 0,02 | 5004 | 84,43 | 5,394 | 1 | 0,02 | 5093 | 85,93 |
| 2,419 | 2 | 0,03 | 4918 | 82,98 | 3,544 | 1 | 0,02 | 5005 | 84,44 | 5,405 | 1 | 0,02 | 5094 | 85,95 |
| 2,439 | 4 | 0,07 | 4922 | 83,04 | 3,559 | 1 | 0,02 | 5006 | 84,46 | 5,425 | 1 | 0,02 | 5095 | 85,96 |
| 2,448 | 1 | 0,02 | 4923 | 83,06 | 3,571 | 4 | 0,07 | 5010 | 84,53 | 5,464 | 1 | 0,02 | 5096 | 85,98 |
| 2,482 | 1 | 0,02 | 4924 | 83,08 | 3,604 | 1 | 0,02 | 5011 | 84,55 | 5,495 | 1 | 0,02 | 5097 | 86 |
| 2,5 | 2 | 0,03 | 4926 | 83,11 | 3,67 | 1 | 0,02 | 5012 | 84,56 | 5,499 | 1 | 0,02 | 5098 | 86,01 |
| 2,55 | 1 | 0,02 | 4927 | 83,13 | 3,683 | 1 | 0,02 | 5013 | 84,58 | 5,556 | 1 | 0,02 | 5099 | 86,03 |
| 2,571 | 1 | 0,02 | 4928 | 83,14 | 3,704 | 5 | 0,08 | 5018 | 84,66 | 5,607 | 2 | 0,03 | 5101 | 86,06 |
| 2,575 | 2 | 0,03 | 4930 | 83,18 | 3,745 | 1 | 0,02 | 5019 | 84,68 | 5,634 | 1 | 0,02 | 5102 | 86,08 |
| 2,586 | 1 | 0,02 | 4931 | 83,2 | 3,76 | 1 | 0,02 | 5020 | 84,7 | 5,687 | 1 | 0,02 | 5103 | 86,1 |
| 2,597 | 2 | 0,03 | 4933 | 83,23 | 3,815 | 1 | 0,02 | 5021 | 84,71 | 5,714 | 2 | 0,03 | 5105 | 86,13 |

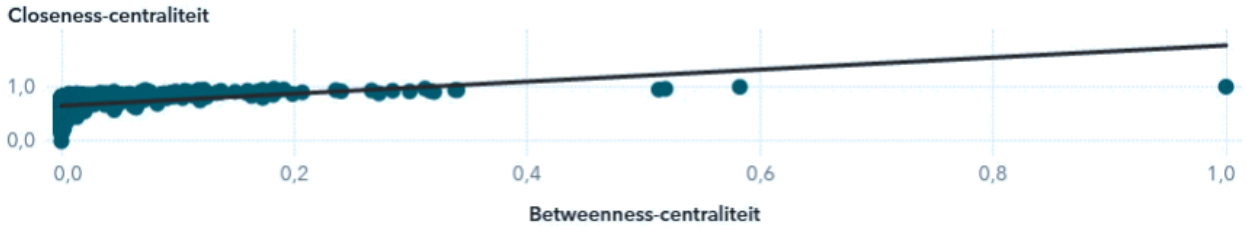
| 5,768 | 1 | 0,02 | 5106 | 86,15 | 8,092 | 1 | 0,02 | 5195 | 87,65 | 11,59 | 2 | 0,03 | 5289 | 89,24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 5,882 | 6 | 0,1 | 5112 | 86,25 | 8,156 | 1 | 0,02 | 5196 | 87,67 | 11,76 | 3 | 0,05 | 5292 | 89,29 |
| 5,924 | 1 | 0,02 | 5113 | 86,27 | 8,163 | 1 | 0,02 | 5197 | 87,68 | 11,84 | 1 | 0,02 | 5293 | 89,3 |
| 5,941 | 1 | 0,02 | 5114 | 86,28 | 8,197 | 3 | 0,05 | 5200 | 87,73 | 11,88 | 1 | 0,02 | 5294 | 89,32 |
| 5,952 | 1 | 0,02 | 5115 | 86,3 | 8,209 | 1 | 0,02 | 5201 | 87,75 | 11,95 | 1 | 0,02 | 5295 | 89,34 |
| 5,97 | 1 | 0,02 | 5116 | 86,32 | 8,239 | 1 | 0,02 | 5202 | 87,77 | 11,97 | 1 | 0,02 | 5296 | 89,35 |
| 5,983 | 1 | 0,02 | 5117 | 86,33 | 8,299 | 1 | 0,02 | 5203 | 87,78 | 12 | 4 | 0,07 | 5300 | 89,42 |
| 6 | 1 | 0,02 | 5118 | 86,35 | 8,325 | 1 | 0,02 | 5204 | 87,8 | 12,12 | 2 | 0,03 | 5302 | 89,46 |
| 6,061 | 1 | 0,02 | 5119 | 86,37 | 8,333 | 6 | 0,1 | 5210 | 87,9 | 12,14 | 1 | 0,02 | 5303 | 89,47 |
| 6,088 | 1 | 0,02 | 5120 | 86,38 | 8,403 | 2 | 0,03 | 5212 | 87,94 | 12,2 | 1 | 0,02 | 5304 | 89,49 |
| 6,198 | 1 | 0,02 | 5121 | 86,4 | 8,571 | 3 | 0,05 | 5215 | 87,99 | 12,42 | 1 | 0,02 | 5305 | 89,51 |
| 6,204 | 1 | 0,02 | 5122 | 86,42 | 8,673 | 1 | 0,02 | 5216 | 88 | 12,5 | 7 | 0,12 | 5312 | 89,62 |
| 6,25 | 4 | 0,07 | 5126 | 86,49 | 8,765 | 1 | 0,02 | 5217 | 88,02 | 12,69 | 1 | 0,02 | 5313 | 89,64 |
| 6,265 | 1 | 0,02 | 5127 | 86,5 | 8,8 | 1 | 0,02 | 5218 | 88,04 | 12,9 | 1 | 0,02 | 5314 | 89,66 |
| 6,289 | 1 | 0,02 | 5128 | 86,52 | 8,995 | 1 | 0,02 | 5219 | 88,05 | 12,95 | 1 | 0,02 | 5315 | 89,67 |
| 6,358 | 1 | 0,02 | 5129 | 86,54 | 9,023 | 1 | 0,02 | 5220 | 88,07 | 13 | 1 | 0,02 | 5316 | 89,69 |
| 6,383 | 1 | 0,02 | 5130 | 86,55 | 9,074 | 1 | 0,02 | 5221 | 88,09 | 13,04 | 2 | 0,03 | 5318 | 89,72 |
| 6,4 | 1 | 0,02 | 5131 | 86,57 | 9,091 | 3 | 0,05 | 5224 | 88,14 | 13,13 | 1 | 0,02 | 5319 | 89,74 |
| 6,444 | 1 | 0,02 | 5132 | 86,59 | 9,211 | 1 | 0,02 | 5225 | 88,16 | 13,33 | 3 | 0,05 | 5322 | 89,79 |
| 6,452 | 1 | 0,02 | 5133 | 86,6 | 9,231 | 2 | 0,03 | 5227 | 88,19 | 13,43 | 1 | 0,02 | 5323 | 89,81 |
| 6,486 | 1 | 0,02 | 5134 | 86,62 | 9,259 | 1 | 0,02 | 5228 | 88,21 | 13,56 | 1 | 0,02 | 5324 | 89,83 |
| 6,522 | 1 | 0,02 | 5135 | 86,64 | 9,299 | 1 | 0,02 | 5229 | 88,22 | 13,64 | 1 | 0,02 | 5325 | 89,84 |
| 6,557 | 1 | 0,02 | 5136 | 86,65 | 9,302 | 3 | 0,05 | 5232 | 88,27 | 13,7 | 1 | 0,02 | 5326 | 89,86 |
| 6,653 | 1 | 0,02 | 5137 | 86,67 | 9,375 | 2 | 0,03 | 5234 | 88,31 | 13,79 | 1 | 0,02 | 5327 | 89,88 |
| 6,667 | 5 | 0,08 | 5142 | 86,76 | 9,446 | 1 | 0,02 | 5235 | 88,32 | 13,81 | 1 | 0,02 | 5328 | 89,89 |
| 6,757 | 1 | 0,02 | 5143 | 86,77 | 9,449 | 1 | 0,02 | 5236 | 88,34 | 13,95 | 1 | 0,02 | 5329 | 89,91 |
| 6,78 | 1 | 0,02 | 5144 | 86,79 | 9,483 | 1 | 0,02 | 5237 | 88,36 | 14,07 | 1 | 0,02 | 5330 | 89,93 |
| 6,818 | 4 | 0,07 | 5148 | 86,86 | 9,524 | 2 | 0,03 | 5239 | 88,39 | 14,15 | 1 | 0,02 | 5331 | 89,94 |
| 6,87 | 1 | 0,02 | 5149 | 86,87 | 9,536 | 1 | 0,02 | 5240 | 88,41 | 14,29 | 4 | 0,07 | 5335 | 90,01 |
| 6,897 | 2 | 0,03 | 5151 | 86,91 | 9,574 | 1 | 0,02 | 5241 | 88,43 | 14,3 | 1 | 0,02 | 5336 | 90,03 |
| 6,944 | 1 | 0,02 | 5152 | 86,92 | 9,586 | 1 | 0,02 | 5242 | 88,44 | 14,45 | 1 | 0,02 | 5337 | 90,05 |
| 6,977 | 1 | 0,02 | 5153 | 86,94 | 9,756 | 1 | 0,02 | 5243 | 88,46 | 14,52 | 1 | 0,02 | 5338 | 90,06 |
| 7,013 | 1 | 0,02 | 5154 | 86,96 | 9,898 | 1 | 0,02 | 5244 | 88,48 | 14,78 | 1 | 0,02 | 5339 | 90,08 |
| 7,031 | 1 | 0,02 | 5155 | 86,97 | 10 | 12 | 0,2 | 5256 | 88,68 | 14,81 | 2 | 0,03 | 5341 | 90,11 |
| 7,053 | 1 | 0,02 | 5156 | 86,99 | 10,02 | 1 | 0,02 | 5257 | 88,7 | 14,94 | 1 | 0,02 | 5342 | 90,13 |
| 7,143 | 5 | 0,08 | 5161 | 87,08 | 10,06 | 1 | 0,02 | 5258 | 88,71 | 15 | 2 | 0,03 | 5344 | 90,16 |
| 7,194 | 1 | 0,02 | 5162 | 87,09 | 10,14 | 1 | 0,02 | 5259 | 88,73 | 15,04 | 1 | 0,02 | 5345 | 90,18 |
| 7,234 | 1 | 0,02 | 5163 | 87,11 | 10,14 | 1 | 0,02 | 5260 | 88,75 | 15,22 | 1 | 0,02 | 5346 | 90,2 |
| 7,285 | 1 | 0,02 | 5164 | 87,13 | 10,17 | 1 | 0,02 | 5261 | 88,76 | 15,28 | 1 | 0,02 | 5347 | 90,21 |
| 7,33 | 1 | 0,02 | 5165 | 87,14 | 10,2 | 1 | 0,02 | 5262 | 88,78 | 15,29 | 1 | 0,02 | 5348 | 90,23 |
| 7,393 | 1 | 0,02 | 5166 | 87,16 | 10,24 | 1 | 0,02 | 5263 | 88,8 | 15,38 | 1 | 0,02 | 5349 | 90,25 |
| 7,407 | 3 | 0,05 | 5169 | 87,21 | 10,26 | 1 | 0,02 | 5264 | 88,81 | 15,56 | 2 | 0,03 | 5351 | 90,28 |
| 7,416 | 1 | 0,02 | 5170 | 87,23 | 10,3 | 1 | 0,02 | 5265 | 88,83 | 15,79 | 2 | 0,03 | 5353 | 90,32 |
| 7,5 | 1 | 0,02 | 5171 | 87,24 | 10,32 | 1 | 0,02 | 5266 | 88,85 | 15,83 | 2 | 0,03 | 5355 | 90,35 |
| 7,534 | 1 | 0,02 | 5172 | 87,26 | 10,34 | 2 | 0,03 | 5268 | 88,88 | 15,91 | 2 | 0,03 | 5357 | 90,38 |
| 7,538 | 1 | 0,02 | 5173 | 87,28 | 10,42 | 1 | 0,02 | 5269 | 88,9 | 16 | 1 | 0,02 | 5358 | 90,4 |
| 7,547 | 2 | 0,03 | 5175 | 87,31 | 10,49 | 1 | 0,02 | 5270 | 88,92 | 16,07 | 1 | 0,02 | 5359 | 90,42 |
| 7,576 | 1 | 0,02 | 5176 | 87,33 | 10,53 | 3 | 0,05 | 5273 | 88,97 | 16,3 | 1 | 0,02 | 5360 | 90,43 |
| 7,599 | 1 | 0,02 | 5177 | 87,35 | 10,76 | 1 | 0,02 | 5274 | 88,98 | 16,34 | 1 | 0,02 | 5361 | 90,45 |
| 7,634 | 1 | 0,02 | 5178 | 87,36 | 10,9 | 1 | 0,02 | 5275 | 89 | 16,42 | 1 | 0,02 | 5362 | 90,47 |
| 7,692 | 5 | 0,08 | 5183 | 87,45 | 10,94 | 1 | 0,02 | 5276 | 89,02 | 16,67 | 11 | 0,19 | 5373 | 90,65 |
| 7,759 | 1 | 0,02 | 5184 | 87,46 | 11,11 | 3 | 0,05 | 5279 | 89,07 | 16,84 | 1 | 0,02 | 5374 | 90,67 |
| 7,769 | 1 | 0,02 | 5185 | 87,48 | 11,23 | 1 | 0,02 | 5280 | 89,08 | 17,48 | 1 | 0,02 | 5375 | 90,69 |
| 7,795 | 1 | 0,02 | 5186 | 87,5 | 11,27 | 1 | 0,02 | 5281 | 89,1 | 17,65 | 1 | 0,02 | 5376 | 90,7 |
| 7,826 | 1 | 0,02 | 5187 | 87,51 | 11,36 | 1 | 0,02 | 5282 | 89,12 | 18,03 | 1 | 0,02 | 5377 | 90,72 |
| 7,843 | 1 | 0,02 | 5188 | 87,53 | 11,42 | 1 | 0,02 | 5283 | 89,13 | 18,07 | 1 | 0,02 | 5378 | 90,74 |
| 7,895 | 3 | 0,05 | 5191 | 87,58 | 11,42 | 1 | 0,02 | 5284 | 89,15 | 18,18 | 3 | 0,05 | 5381 | 90,79 |
| 7,994 | 1 | 0,02 | 5192 | 87,6 | 11,42 | 1 | 0,02 | 5285 | 89,17 | 18,28 | 1 | 0,02 | 5382 | 90,8 |
| 8 | 1 | 0,02 | 5193 | 87,62 | 11,48 | 1 | 0,02 | 5286 | 89,19 | 18,34 | 1 | 0,02 | 5383 | 90,82 |
| 8,054 | 1 | 0,02 | 5194 | 87,63 | 11,54 | 1 | 0,02 | 5287 | 89,2 | 18,42 | 1 | 0,02 | 5384 | 90,84 |

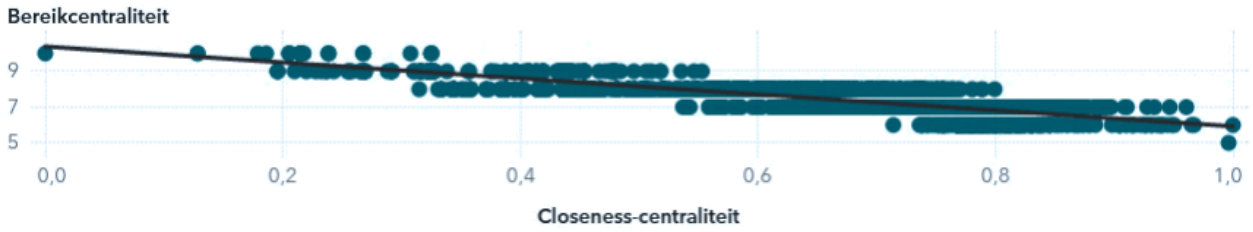
| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **18,52** | **1** | **0,02** | **5385** | **90,86** | **30,77** | **1** | **0,02** | **5480** | **92,46** | **56,1** | **1** | **0,02** | **5606** | **94,58** |
| **18,6** | 1 | 0,02 | 5386 | 90,87 | **31,25** | 1 | 0,02 | 5481 | 92,48 | **57,14** | 3 | 0,05 | 5609 | 94,63 |
| **18,63** | 1 | 0,02 | 5387 | 90,89 | **31,56** | 1 | 0,02 | 5482 | 92,49 | **58,06** | 1 | 0,02 | 5610 | 94,65 |
| **18,64** | 1 | 0,02 | 5388 | 90,91 | **31,71** | 1 | 0,02 | 5483 | 92,51 | **58,33** | 2 | 0,03 | 5612 | 94,69 |
| **18,75** | 1 | 0,02 | 5389 | 90,92 | **32,22** | 1 | 0,02 | 5484 | 92,53 | **59,26** | 1 | 0,02 | 5613 | 94,7 |
| **18,87** | 1 | 0,02 | 5390 | 90,94 | **32,26** | 1 | 0,02 | 5485 | 92,54 | **60** | 5 | 0,08 | 5618 | 94,79 |
| **18,92** | 2 | 0,03 | 5392 | 90,97 | **32,31** | 1 | 0,02 | 5486 | 92,56 | **60,33** | 1 | 0,02 | 5619 | 94,8 |
| **19,15** | 3 | 0,05 | 5395 | 91,02 | **32,35** | 1 | 0,02 | 5487 | 92,58 | **60,61** | 1 | 0,02 | 5620 | 94,82 |
| **19,48** | 1 | 0,02 | 5396 | 91,04 | **32,61** | 1 | 0,02 | 5488 | 92,59 | **60,66** | 1 | 0,02 | 5621 | 94,84 |
| **19,77** | 1 | 0,02 | 5397 | 91,06 | **32,89** | 1 | 0,02 | 5489 | 92,61 | **60,67** | 1 | 0,02 | 5622 | 94,85 |
| **20** | 7 | 0,12 | 5404 | 91,18 | **33,33** | 19 | 0,32 | 5508 | 92,93 | **61,11** | 1 | 0,02 | 5623 | 94,87 |
| **20,41** | 1 | 0,02 | 5405 | 91,19 | **34,17** | 1 | 0,02 | 5509 | 92,95 | **61,54** | 1 | 0,02 | 5624 | 94,89 |
| **20,51** | 1 | 0,02 | 5406 | 91,21 | **34,21** | 1 | 0,02 | 5510 | 92,96 | **61,67** | 1 | 0,02 | 5625 | 94,9 |
| **20,58** | 1 | 0,02 | 5407 | 91,23 | **34,38** | 1 | 0,02 | 5511 | 92,98 | **61,7** | 1 | 0,02 | 5626 | 94,92 |
| **20,61** | 1 | 0,02 | 5408 | 91,24 | **34,78** | 1 | 0,02 | 5512 | 93 | **62,07** | 1 | 0,02 | 5627 | 94,94 |
| **20,93** | 1 | 0,02 | 5409 | 91,26 | **35,29** | 1 | 0,02 | 5513 | 93,02 | **62,5** | 4 | 0,07 | 5631 | 95,01 |
| **21,04** | 1 | 0,02 | 5410 | 91,28 | **35,48** | 1 | 0,02 | 5514 | 93,03 | **63,33** | 1 | 0,02 | 5632 | 95,02 |
| **21,15** | 1 | 0,02 | 5411 | 91,29 | **35,56** | 1 | 0,02 | 5515 | 93,05 | **64,29** | 1 | 0,02 | 5633 | 95,04 |
| **21,43** | 2 | 0,03 | 5413 | 91,33 | **35,66** | 1 | 0,02 | 5516 | 93,07 | **64,41** | 1 | 0,02 | 5634 | 95,06 |
| **21,54** | 1 | 0,02 | 5414 | 91,34 | **35,71** | 3 | 0,05 | 5519 | 93,12 | **64,71** | 1 | 0,02 | 5635 | 95,07 |
| **21,74** | 1 | 0,02 | 5415 | 91,36 | **36,36** | 2 | 0,03 | 5521 | 93,15 | **64,81** | 1 | 0,02 | 5636 | 95,09 |
| **21,79** | 1 | 0,02 | 5416 | 91,38 | **36,84** | 1 | 0,02 | 5522 | 93,17 | **65,38** | 1 | 0,02 | 5637 | 95,11 |
| **21,88** | 2 | 0,03 | 5418 | 91,41 | **37,04** | 1 | 0,02 | 5523 | 93,18 | **65,63** | 1 | 0,02 | 5638 | 95,12 |
| **22,22** | 4 | 0,07 | 5422 | 91,48 | **37,12** | 1 | 0,02 | 5524 | 93,2 | **66,67** | 7 | 0,12 | 5645 | 95,24 |
| **22,41** | 1 | 0,02 | 5423 | 91,5 | **37,5** | 2 | 0,03 | 5526 | 93,23 | **67,86** | 1 | 0,02 | 5646 | 95,26 |
| **22,64** | 2 | 0,03 | 5425 | 91,53 | **37,96** | 1 | 0,02 | 5527 | 93,25 | **68,38** | 1 | 0,02 | 5647 | 95,28 |
| **22,77** | 1 | 0,02 | 5426 | 91,55 | **38,3** | 1 | 0,02 | 5528 | 93,27 | **68,63** | 1 | 0,02 | 5648 | 95,29 |
| **23,08** | 2 | 0,03 | 5428 | 91,58 | **38,46** | 1 | 0,02 | 5529 | 93,28 | **68,85** | 1 | 0,02 | 5649 | 95,31 |
| **23,16** | 1 | 0,02 | 5429 | 91,6 | **38,71** | 1 | 0,02 | 5530 | 93,3 | **70** | 2 | 0,03 | 5651 | 95,34 |
| **23,53** | 1 | 0,02 | 5430 | 91,61 | **38,78** | 1 | 0,02 | 5531 | 93,32 | **70,45** | 1 | 0,02 | 5652 | 95,36 |
| **23,79** | 1 | 0,02 | 5431 | 91,63 | **40** | 10 | 0,17 | 5541 | 93,49 | **71,43** | 4 | 0,07 | 5656 | 95,43 |
| **24,07** | 1 | 0,02 | 5432 | 91,65 | **40,35** | 1 | 0,02 | 5542 | 93,5 | **72,73** | 1 | 0,02 | 5657 | 95,44 |
| **24,11** | 1 | 0,02 | 5433 | 91,67 | **40,63** | 1 | 0,02 | 5543 | 93,52 | **73,08** | 1 | 0,02 | 5658 | 95,46 |
| **24,14** | 2 | 0,03 | 5435 | 91,7 | **41,32** | 1 | 0,02 | 5544 | 93,54 | **75** | 2 | 0,03 | 5660 | 95,5 |
| **24,44** | 1 | 0,02 | 5436 | 91,72 | **41,67** | 1 | 0,02 | 5545 | 93,55 | **76,67** | 1 | 0,02 | 5661 | 95,51 |
| **24,64** | 2 | 0,03 | 5438 | 91,75 | **42,31** | 2 | 0,03 | 5547 | 93,59 | **77,27** | 1 | 0,02 | 5662 | 95,53 |
| **25** | 6 | 0,1 | 5444 | 91,85 | **42,86** | 3 | 0,05 | 5550 | 93,64 | **78,26** | 1 | 0,02 | 5663 | 95,55 |
| **25,41** | 1 | 0,02 | 5445 | 91,87 | **43,75** | 5 | 0,08 | 5555 | 93,72 | **78,57** | 3 | 0,05 | 5666 | 95,6 |
| **25,74** | 1 | 0,02 | 5446 | 91,88 | **44,44** | 1 | 0,02 | 5556 | 93,74 | **79,17** | 2 | 0,03 | 5668 | 95,63 |
| **25,93** | 2 | 0,03 | 5448 | 91,92 | **45,1** | 1 | 0,02 | 5557 | 93,76 | **80** | 3 | 0,05 | 5671 | 95,68 |
| **26,09** | 1 | 0,02 | 5449 | 91,94 | **45,45** | 1 | 0,02 | 5558 | 93,77 | **81,25** | 1 | 0,02 | 5672 | 95,7 |
| **26,14** | 1 | 0,02 | 5450 | 91,95 | **45,9** | 1 | 0,02 | 5559 | 93,79 | **81,52** | 1 | 0,02 | 5673 | 95,71 |
| **26,38** | 1 | 0,02 | 5451 | 91,97 | **46,15** | 2 | 0,03 | 5561 | 93,82 | **81,82** | 1 | 0,02 | 5674 | 95,73 |
| **26,67** | 2 | 0,03 | 5453 | 92 | **46,51** | 1 | 0,02 | 5562 | 93,84 | **82,83** | 1 | 0,02 | 5675 | 95,75 |
| **26,92** | 1 | 0,02 | 5454 | 92,02 | **46,99** | 1 | 0,02 | 5563 | 93,86 | **83,33** | 3 | 0,05 | 5678 | 95,8 |
| **27,03** | 1 | 0,02 | 5455 | 92,04 | **47,06** | 2 | 0,03 | 5565 | 93,89 | **85** | 1 | 0,02 | 5679 | 95,82 |
| **27,16** | 1 | 0,02 | 5456 | 92,05 | **47,37** | 1 | 0,02 | 5566 | 93,91 | **85,42** | 1 | 0,02 | 5680 | 95,83 |
| **27,27** | 2 | 0,03 | 5458 | 92,09 | **47,62** | 1 | 0,02 | 5567 | 93,93 | **85,71** | 1 | 0,02 | 5681 | 95,85 |
| **27,55** | 1 | 0,02 | 5459 | 92,1 | **47,9** | 1 | 0,02 | 5568 | 93,94 | **87,04** | 1 | 0,02 | 5682 | 95,87 |
| **27,78** | 1 | 0,02 | 5460 | 92,12 | **50** | 23 | 0,39 | 5591 | 94,33 | **87,5** | 1 | 0,02 | 5683 | 95,88 |
| **27,92** | 1 | 0,02 | 5461 | 92,14 | **51,61** | 1 | 0,02 | 5592 | 94,35 | **88,24** | 1 | 0,02 | 5684 | 95,9 |
| **28,13** | 1 | 0,02 | 5462 | 92,15 | **52** | 2 | 0,03 | 5594 | 94,38 | **88,89** | 1 | 0,02 | 5685 | 95,92 |
| **28,57** | 8 | 0,13 | 5470 | 92,29 | **52,94** | 1 | 0,02 | 5595 | 94,4 | **89** | 1 | 0,02 | 5686 | 95,93 |
| **29,01** | 1 | 0,02 | 5471 | 92,31 | **53,66** | 1 | 0,02 | 5596 | 94,42 | **90,48** | 1 | 0,02 | 5687 | 95,95 |
| **29,21** | 1 | 0,02 | 5472 | 92,32 | **53,85** | 1 | 0,02 | 5597 | 94,43 | **90,91** | 2 | 0,03 | 5689 | 95,98 |
| **29,41** | 1 | 0,02 | 5473 | 92,34 | **54,05** | 1 | 0,02 | 5598 | 94,45 | **91,3** | 1 | 0,02 | 5690 | 96 |
| **29,51** | 1 | 0,02 | 5474 | 92,36 | **54,35** | 2 | 0,03 | 5600 | 94,48 | **93,33** | 1 | 0,02 | 5691 | 96,02 |
| **29,63** | 1 | 0,02 | 5475 | 92,37 | **54,55** | 1 | 0,02 | 5601 | 94,5 | **95,24** | 1 | 0,02 | 5692 | 96,04 |
| **30** | 2 | 0,03 | 5477 | 92,41 | **55,05** | 1 | 0,02 | 5602 | 94,52 | **100** | 235 | 3,96 | 5927 | 100 |
| **30,34** | 1 | 0,02 | 5478 | 92,42 | **55,56** | 2 | 0,03 | 5604 | 94,55 | | | | | |
| **30,63** | 1 | 0,02 | 5479 | 92,44 | **56** | 1 | 0,02 | 5605 | 94,57 | | | | | |

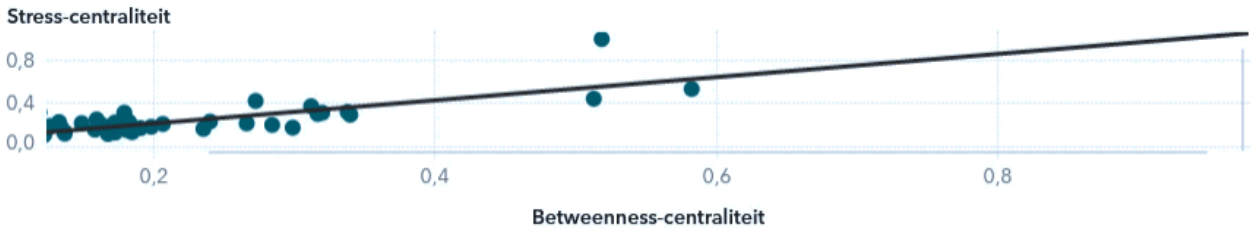
## H.13 Correlations between Centralities

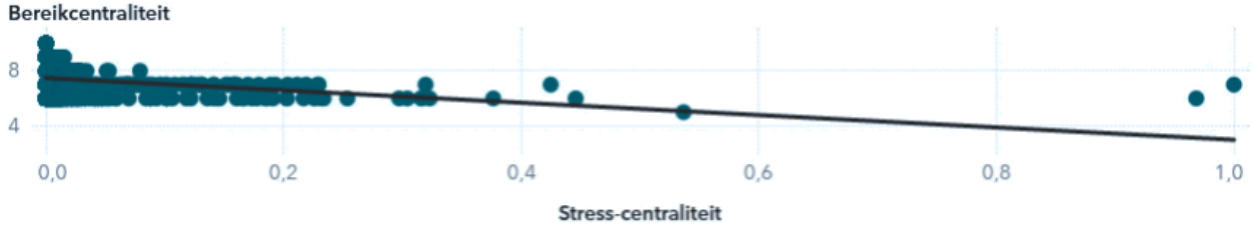




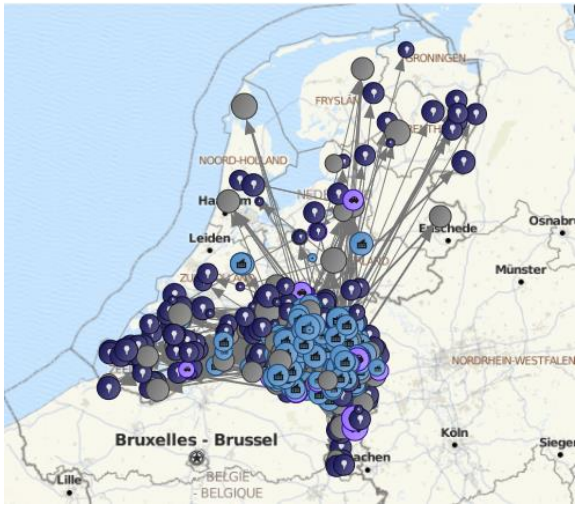


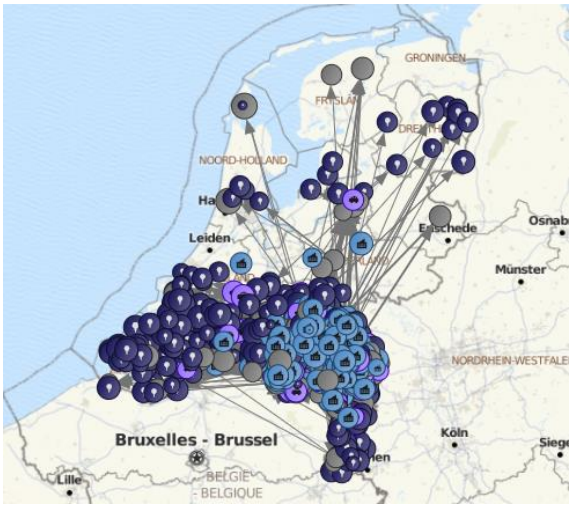



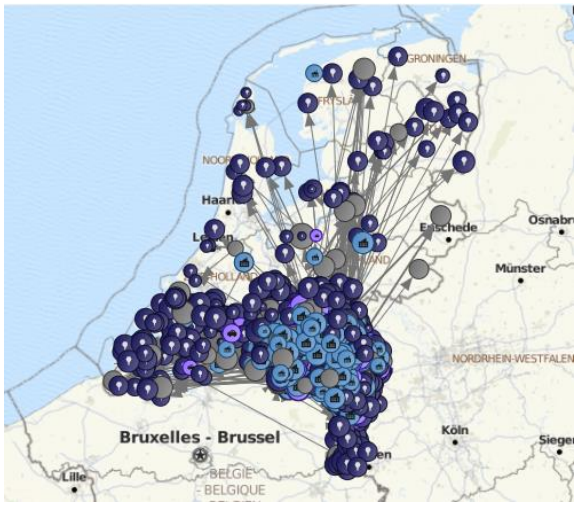
## H.14 Tracking Changes Over Time



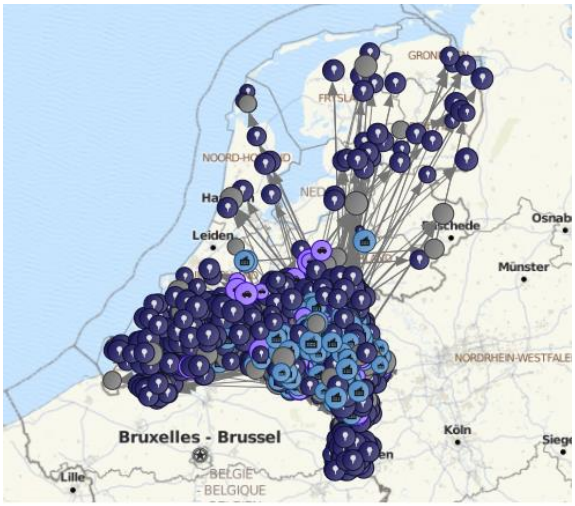

**January 2018**



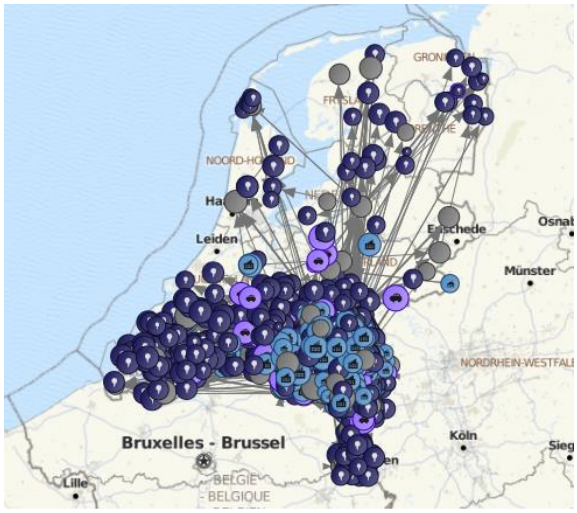

**February 2018**



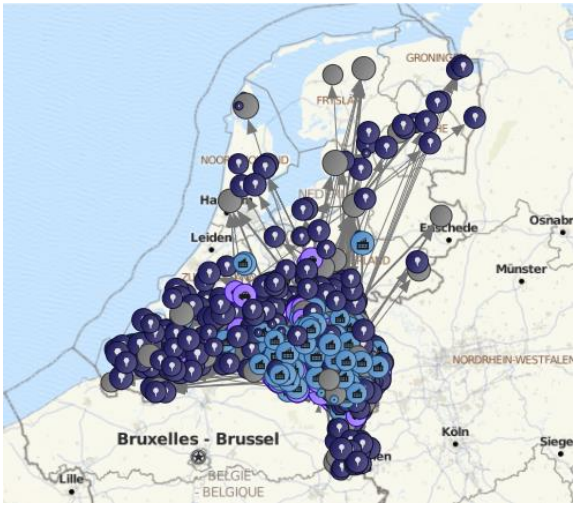

**March 2018**



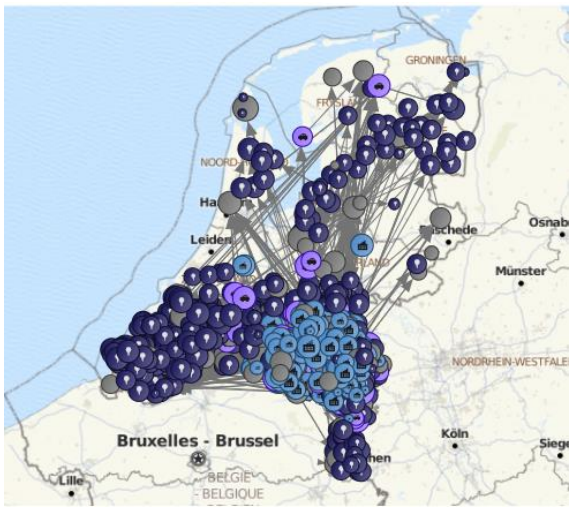

**April 2018**



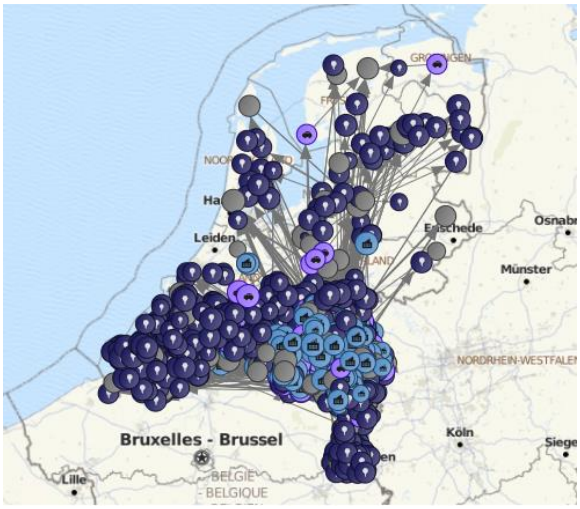

**May 2018**



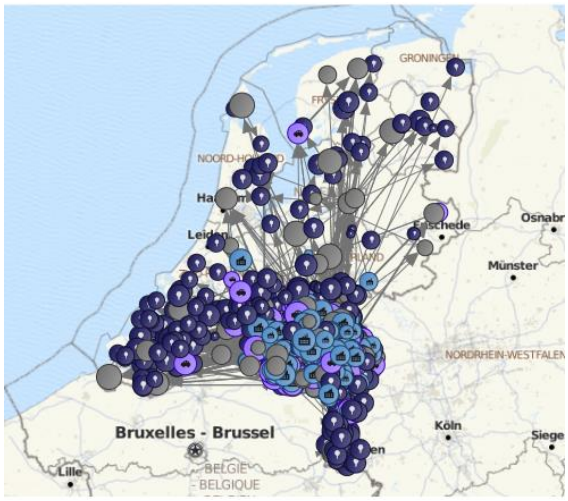

**June 2018**



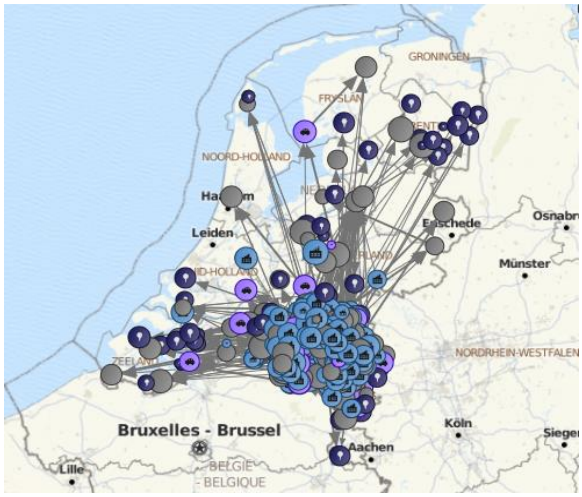

**July 2018**



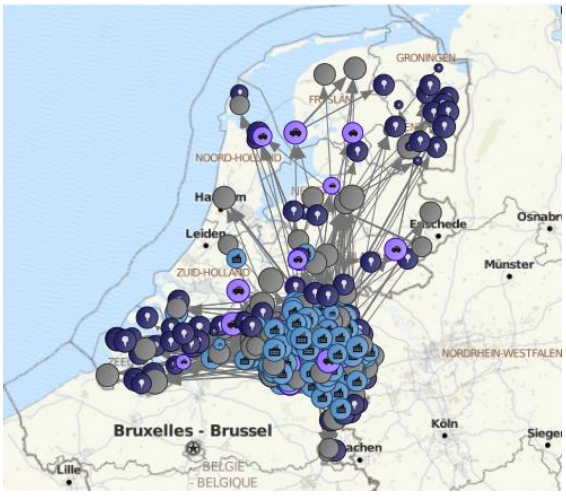

**August 2018**



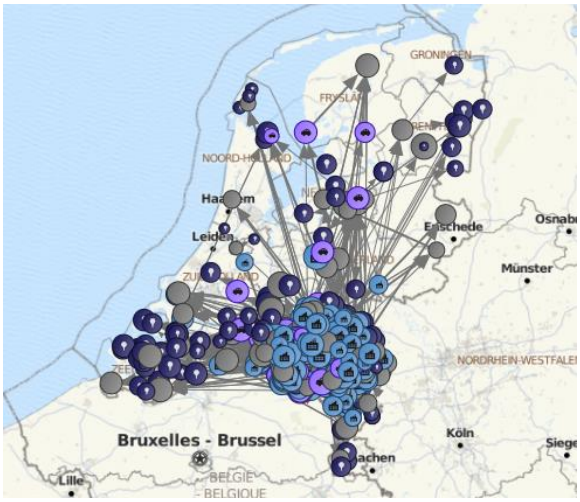

**September 2018**




**October 2018**




**November 2018**



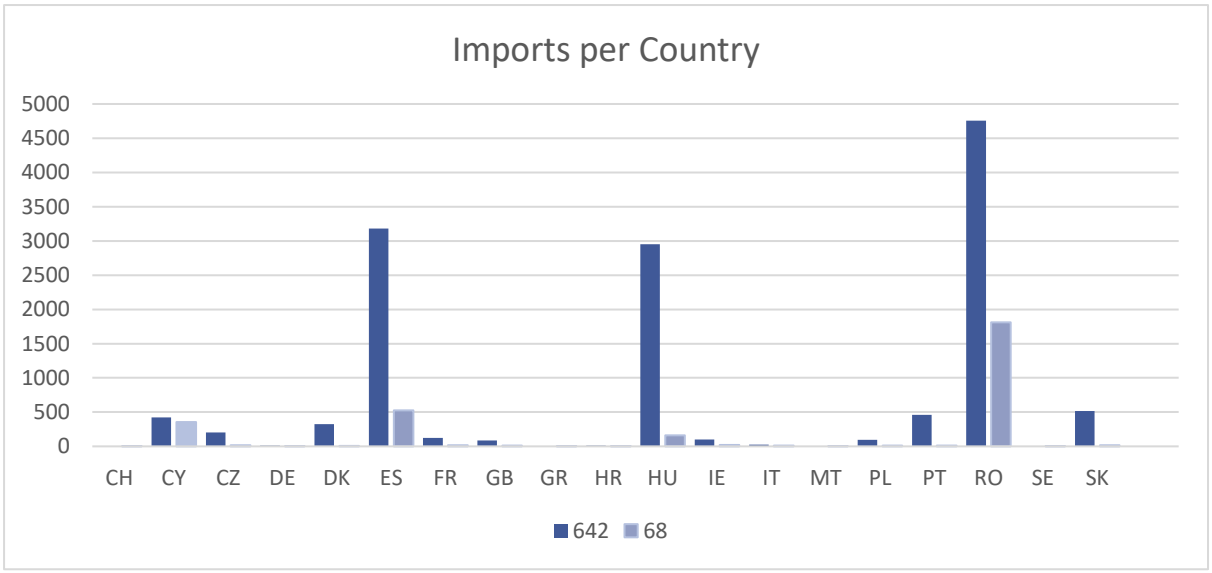

**December 2018**

# Appendix I: Results SNA Illegal Dog Trade

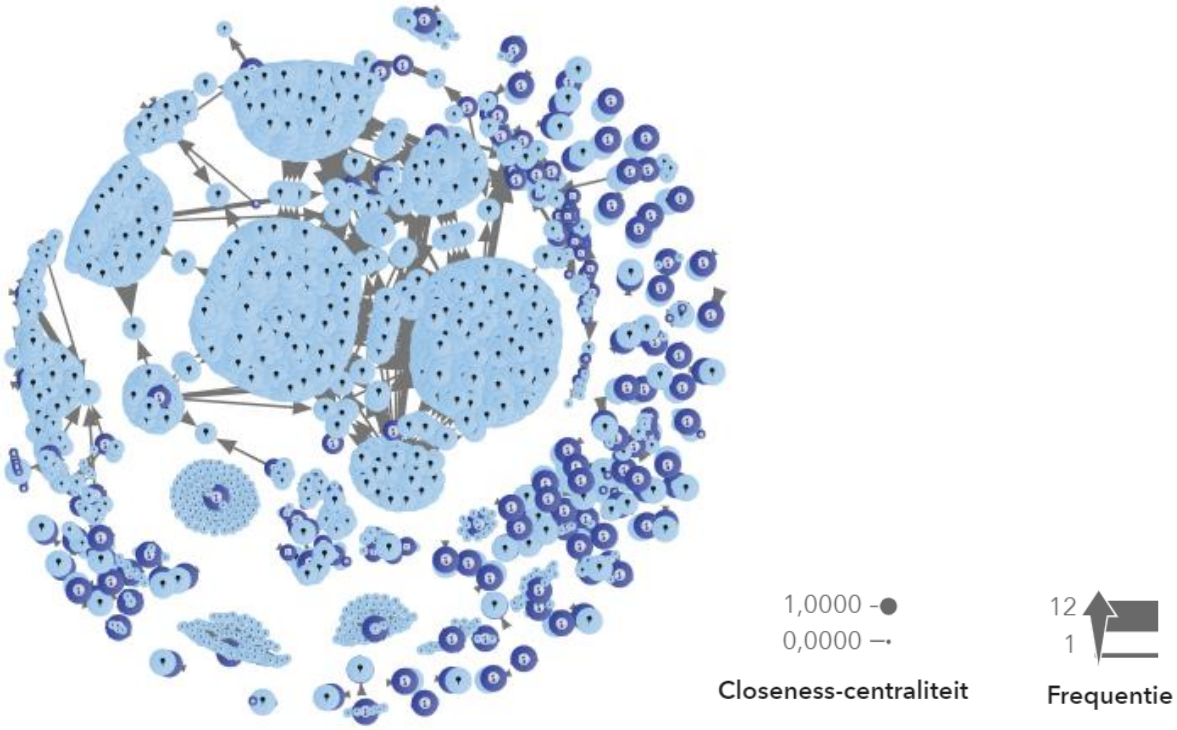
## I.1 Import per Country 2018

| Country | Total number of imported dogs | Number of imports |
|---|---|---|
| **BG** | 642 | 68 |
| **CH** | 4 | 2 |
| **CY** | 424 | 355 |
| **CZ** | 202 | 15 |
| **DE** | 13 | 2 |
| **DK** | 325 | 5 |
| **ES** | 3181 | 525 |
| **FR** | 125 | 14 |
| **GB** | 87 | 11 |
| **GR** | 1 | 1 |
| **HR** | 9 | 2 |
| **HU** | 2951 | 159 |
| **IE** | 102 | 20 |
| **IT** | 25 | 13 |
| **MT** | 1 | 1 |
| **PL*** | 96 | 12 |
| **PT** | 460 | 9 |
| **RO*** | 4757 | 1812 |
| **SE** | 2 | 1 |
| **SK** | 518 | 17 |
| Total | **13.925** | **3044** |

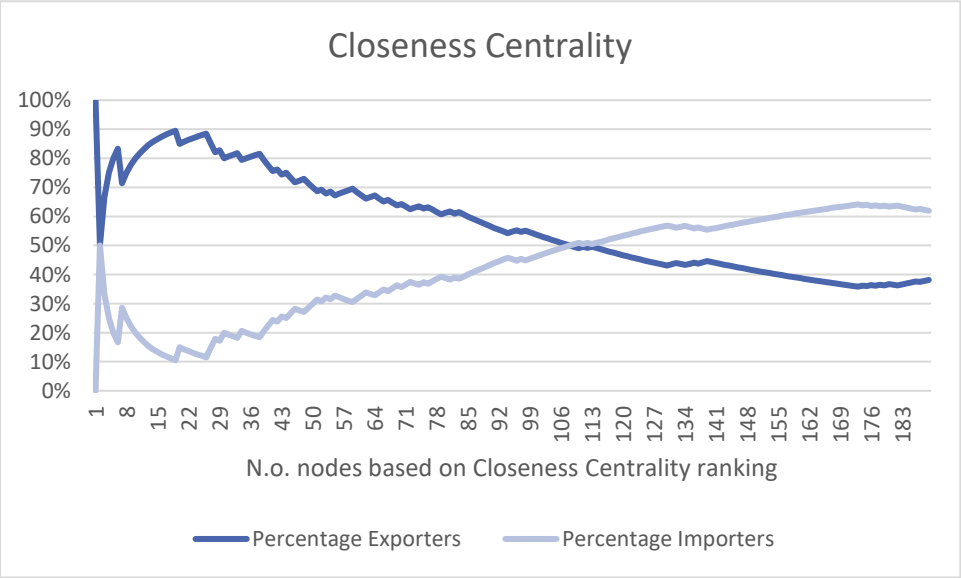
** Rabies Risk Country*

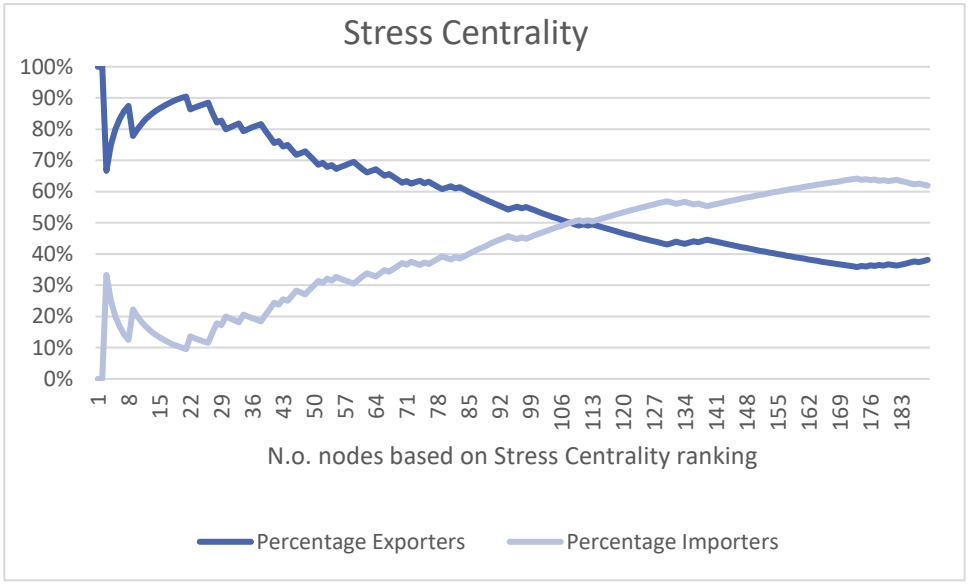


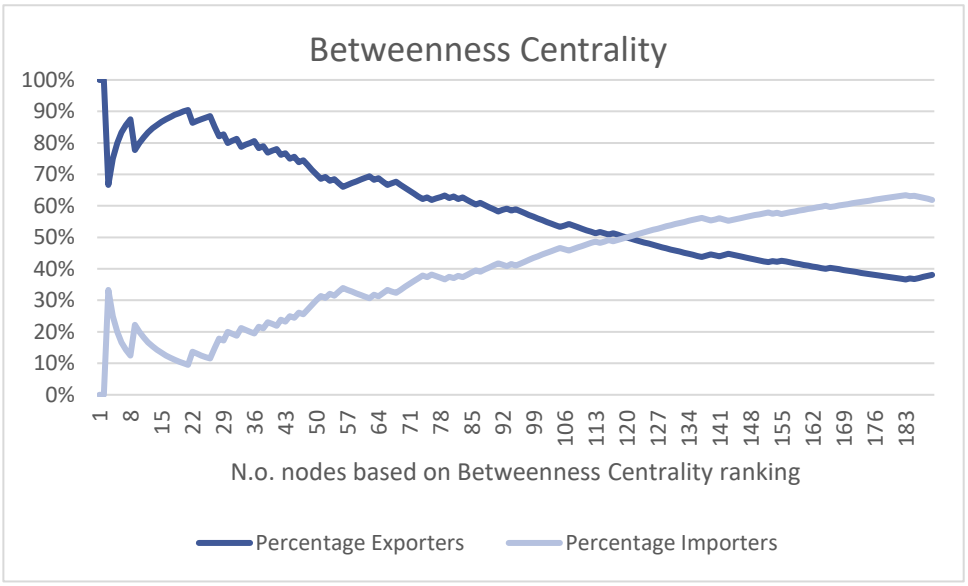
## I.2 Full Network Visualisation Based on Type



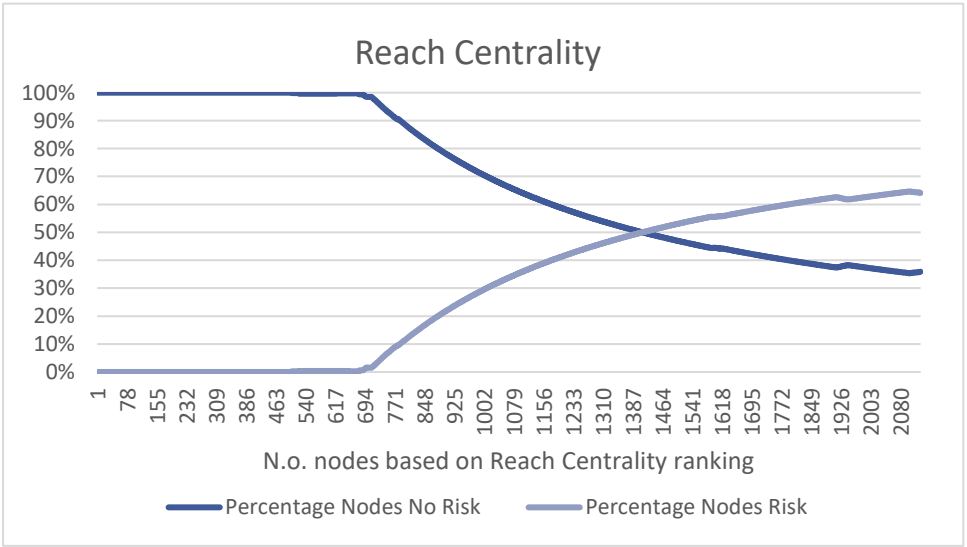

## I.3 Composition of Exporters and Importers Related to Centrality

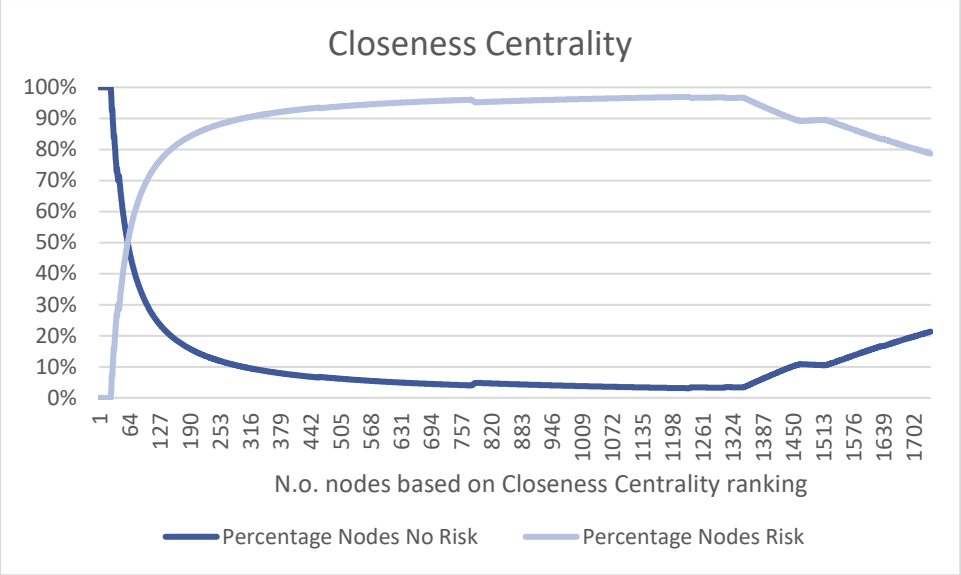


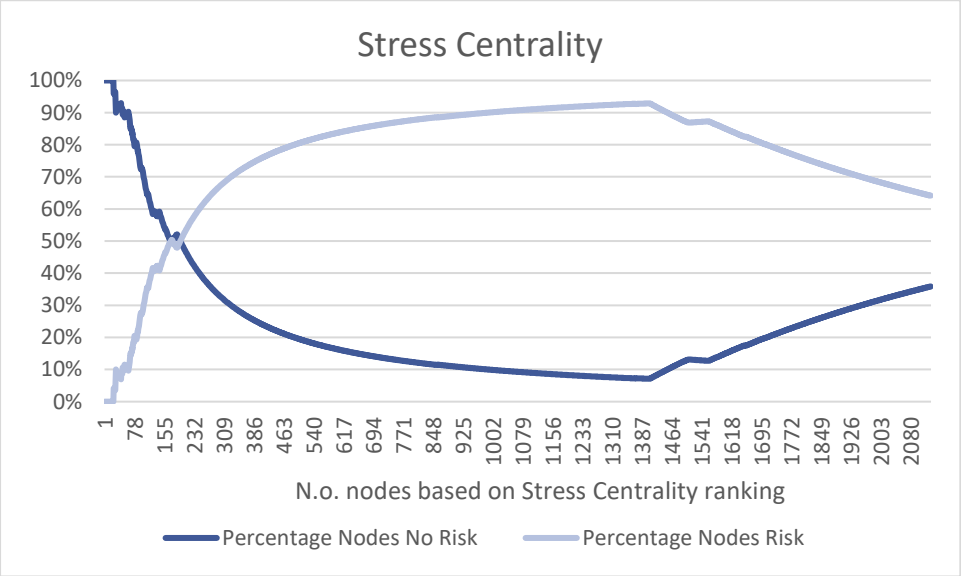


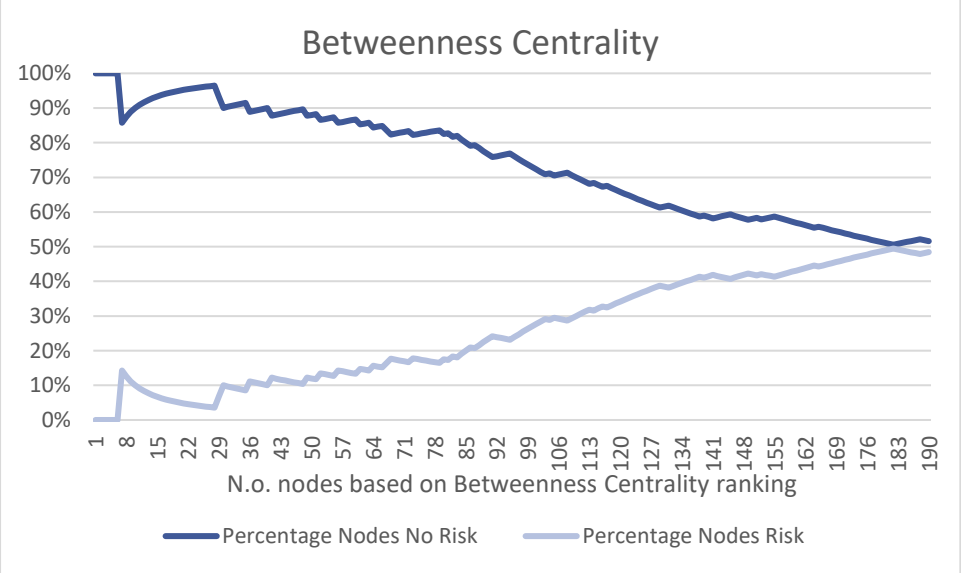




## I.4 Composition of Risk and No Risk Related to Centrality




Closeness Centrality
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
64
127
190
253
316
379
442
505
568
631
694
757
820
883
946
1009
1072
1135
1198
1261
1324
1387
1450
1513
1576
1639
1702
N.o. nodes based on Closeness Centrality ranking
Percentage Nodes No Risk
Percentage Nodes Risk


Stress Centrality
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
78
155
232
309
386
463
540
617
694
771
848
925
1002
1079
1156
1233
1310
1387
1464
1541
1618
1695
1772
1849
1926
2003
2080
N.o. nodes based on Stress Centrality ranking
Percentage Nodes No Risk
Percentage Nodes Risk


Betweenness Centrality
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
1
8
15
22
29
36
43
50
57
64
71
78
85
92
99
106
113
120
127
134
141
148
155
162
169
176
183
190
N.o. nodes based on Betweenness Centrality ranking
Percentage Nodes No Risk
Percentage Nodes Risk

## I.5 Results Reach Centrality

* Due to privacy considerations the ID code presented here is not the original business number.

| ID Code | Reach Centrality | Ranking Reach | Rabies Risk | Type |
|---|---|---|---|---|
| **487979** | 1,000 | 1 | 0,00 | Importer |
| **805048** | 1,000 | 2 | 0,00 | Exporter |
| **805048** | 1,000 | 3 | 0,00 | Exporter |
| **455480** | 1,000 | 4 | 0,00 | Exporter |
| **934846** | 1,000 | 5 | 0,00 | Exporter |
| **370471** | 1,000 | 6 | 0,00 | Exporter |
| **155395** | 1,000 | 7 | 0,00 | Exporter |
| **566229** | 1,000 | 8 | 0,00 | Exporter |
| **433256** | 1,000 | 9 | 0,00 | Exporter |
| **742927** | 1,000 | 10 | 0,00 | Exporter |
| **681887** | 1,000 | 11 | 0,00 | Exporter |
| **724201** | 1,000 | 12 | 0,00 | Exporter |
| **260892** | 2,000 | 13 | 0,00 | Exporter |
| **757725** | 2,000 | 14 | 0,00 | Exporter |
| **428333** | 2,000 | 15 | 0,00 | Exporter |
| **297625** | 2,000 | 16 | 0,00 | Importer |
| **268011** | 2,000 | 17 | 0,00 | Exporter |
| **677154** | 2,000 | 18 | 0,00 | Importer |
| **875061** | 2,000 | 19 | 0,00 | Importer |
| **389641** | 2,000 | 20 | 0,00 | Importer |
| **677154** | 2,000 | 21 | 0,00 | Importer |
| **389641** | 2,000 | 22 | 0,00 | Importer |
| **723466** | 2,000 | 23 | 0,00 | Importer |
| **592049** | 2,000 | 24 | 0,00 | Importer |
| **566229** | 2,000 | 25 | 0,00 | Importer |
| **677154** | 2,000 | 26 | 0,00 | Importer |
| **319036** | 2,000 | 27 | 0,00 | Importer |
| **560492** | 2,000 | 28 | 0,00 | Exporter |
| **389641** | 2,000 | 29 | 0,00 | Importer |
| **389641** | 2,000 | 30 | 0,00 | Importer |
| **389641** | 2,000 | 31 | 0,00 | Importer |
| **389641** | 2,000 | 32 | 0,00 | Importer |
| **389641** | 2,000 | 33 | 0,00 | Importer |
| **389641** | 2,000 | 34 | 0,00 | Importer |
| **389641** | 2,000 | 35 | 0,00 | Importer |
| **389641** | 2,000 | 36 | 0,00 | Exporter |
| **389641** | 2,000 | 37 | 0,00 | Importer |
| **389641** | 2,000 | 38 | 0,00 | Importer |
| **389641** | 2,000 | 39 | 0,00 | Importer |
| **389641** | 2,000 | 40 | 0,00 | Importer |
| **389641** | 2,000 | 41 | 0,00 | Importer |
| **389641** | 2,000 | 42 | 0,00 | Importer |
| **389641** | 2,000 | 43 | 0,00 | Importer |
| **389641** | 2,000 | 44 | 0,00 | Importer |
| **389641** | 2,000 | 45 | 0,00 | Importer |
| **389641** | 2,000 | 46 | 0,00 | Importer |
| **487979** | 2,000 | 47 | 0,00 | Importer |
| **487979** | 2,000 | 48 | 0,00 | Importer |
| **487979** | 2,000 | 49 | 0,00 | Importer |

| | | | | |
|---|---|---|---|---|
| **201431** | 2,000 | 50 | 0,00 | Importer |
| **201431** | 2,000 | 51 | 0,00 | Importer |
| **404344** | 2,000 | 52 | 0,00 | Importer |
| **334289** | 2,000 | 53 | 0,00 | Importer |
| **989631** | 2,000 | 54 | 0,00 | Importer |
| **503468** | 2,000 | 55 | 0,00 | Importer |
| **260892** | 2,000 | 56 | 0,00 | Importer |
| **260892** | 2,000 | 57 | 0,00 | Importer |
| **805048** | 2,000 | 58 | 0,00 | Importer |
| **805048** | 2,000 | 59 | 0,00 | Exporter |
| **501895** | 2,000 | 60 | 0,00 | Importer |
| **723466** | 2,000 | 61 | 0,00 | Importer |
| **723466** | 2,000 | 62 | 0,00 | Importer |
| **723466** | 2,000 | 63 | 0,00 | Importer |
| **696355** | 2,000 | 64 | 0,00 | Importer |
| **635153** | 2,000 | 65 | 0,00 | Importer |
| **583586** | 2,000 | 66 | 0,00 | Importer |
| **583586** | 2,000 | 67 | 0,00 | Importer |
| **583586** | 2,000 | 68 | 0,00 | Importer |
| **851912** | 2,000 | 69 | 0,00 | Importer |
| **851912** | 2,000 | 70 | 0,00 | Importer |
| **592049** | 2,000 | 71 | 0,00 | Importer |
| **490004** | 2,000 | 72 | 0,00 | Importer |
| **490004** | 2,000 | 73 | 0,00 | Importer |
| **490004** | 2,000 | 74 | 0,00 | Importer |
| **490004** | 2,000 | 75 | 0,00 | Importer |
| **490004** | 2,000 | 76 | 0,00 | Importer |
| **490004** | 2,000 | 77 | 0,00 | Importer |
| **490004** | 2,000 | 78 | 0,00 | Importer |
| **853401** | 2,000 | 79 | 0,00 | Importer |
| **853401** | 2,000 | 80 | 0,00 | Importer |
| **853401** | 2,000 | 81 | 0,00 | Importer |
| **853401** | 2,000 | 82 | 0,00 | Importer |
| **166732** | 2,000 | 83 | 0,00 | Importer |
| **166732** | 2,000 | 84 | 0,00 | Importer |
| **148460** | 2,000 | 85 | 0,00 | Importer |
| **940960** | 2,000 | 86 | 0,00 | Importer |
| **844941** | 2,000 | 87 | 0,00 | Exporter |
| **604267** | 2,000 | 88 | 0,00 | Importer |
| **924153** | 2,000 | 89 | 0,00 | Importer |
| **300690** | 2,000 | 90 | 0,00 | Importer |
| **300690** | 2,000 | 91 | 0,00 | Importer |
| **300690** | 2,000 | 92 | 0,00 | Importer |
| **744826** | 2,000 | 93 | 0,00 | Importer |
| **291410** | 2,000 | 94 | 0,00 | Importer |
| **488424** | 2,000 | 95 | 0,00 | Importer |
| **488424** | 2,000 | 96 | 0,00 | Importer |
| **987414** | 2,000 | 97 | 0,00 | Exporter |
| **126286** | 2,000 | 98 | 0,00 | Importer |
| **126286** | 2,000 | 99 | 0,00 | Importer |
| **126286** | 2,000 | 100 | 0,00 | Importer |

## I.6 Results Closeness Centrality

* Due to privacy considerations the ID code presented here is not the original business number.

| ID Code | Closeness Centrality | Ranking Closeness | Rabies Risk | Type |
|---|---|---|---|---|
| **389641** | 1,000 | 1 | 0,00 | Exporter |
| **487979** | 1,000 | 2 | 0,00 | Importer |
| **260892** | 1,000 | 3 | 0,00 | Exporter |
| **805048** | 1,000 | 4 | 0,00 | Exporter |
| **805048** | 1,000 | 5 | 0,00 | Exporter |
| **853401** | 1,000 | 6 | 0,00 | Exporter |
| **291410** | 1,000 | 7 | 0,00 | Importer |
| **455480** | 1,000 | 8 | 0,00 | Exporter |
| **934846** | 1,000 | 9 | 0,00 | Exporter |
| **166366** | 1,000 | 10 | 0,00 | Exporter |
| **757725** | 1,000 | 11 | 0,00 | Exporter |
| **271747** | 1,000 | 12 | 0,00 | Exporter |
| **370471** | 1,000 | 13 | 0,00 | Exporter |
| **155395** | 1,000 | 14 | 0,00 | Exporter |
| **566229** | 1,000 | 15 | 0,00 | Exporter |
| **433256** | 1,000 | 16 | 0,00 | Exporter |
| **742927** | 1,000 | 17 | 0,00 | Exporter |
| **681887** | 1,000 | 18 | 0,00 | Exporter |
| **428333** | 1,000 | 19 | 0,00 | Exporter |
| **297625** | 1,000 | 20 | 0,00 | Importer |
| **724201** | 1,000 | 21 | 0,00 | Exporter |
| **268011** | 1,000 | 22 | 0,00 | Exporter |
| **838266** | 1,000 | 23 | 0,00 | Exporter |
| **455480** | 1,000 | 24 | 0,00 | Exporter |
| **677154** | 1,000 | 25 | 0,00 | Exporter |
| **677154** | 1,000 | 26 | 1,00 | Exporter |
| **805048** | 0,984 | 27 | 1,00 | Importer |
| **861540** | 0,981 | 28 | 0,00 | Importer |
| **805048** | 0,962 | 29 | 1,00 | Exporter |
| **677154** | 0,952 | 30 | 1,00 | Importer |
| **677154** | 0,950 | 31 | 0,90 | Exporter |
| **201431** | 0,945 | 32 | 0,00 | Exporter |
| **270300** | 0,944 | 33 | 1,00 | Exporter |
| **805048** | 0,942 | 34 | 1,00 | Importer |
| **727314** | 0,931 | 35 | 1,00 | Exporter |
| **143194** | 0,931 | 36 | 1,00 | Exporter |
| **433256** | 0,930 | 37 | 0,27 | Exporter |
| **596831** | 0,929 | 38 | 0,00 | Exporter |
| **389641** | 0,928 | 39 | 1,00 | Importer |
| **383661** | 0,928 | 40 | 1,00 | Importer |
| **258566** | 0,923 | 41 | 0,00 | Importer |
| **155395** | 0,923 | 42 | 0,00 | Exporter |
| **805048** | 0,920 | 43 | 1,00 | Importer |
| **805048** | 0,916 | 44 | 1,00 | Exporter |
| **924153** | 0,915 | 45 | 1,00 | Importer |
| **455480** | 0,915 | 46 | 1,00 | Importer |
| **875964** | 0,915 | 47 | 1,00 | Exporter |
| **599142** | 0,915 | 48 | 1,00 | Exporter |
| **428595** | 0,915 | 49 | 1,00 | Importer |

| | | | | |
|---|---|---|---|---|
| **594584** | 0,915 | 50 | 1,00 | Importer |
| **628229** | 0,915 | 51 | 1,00 | Importer |
| **721965** | 0,915 | 52 | 1,00 | Exporter |
| **675075** | 0,915 | 53 | 1,00 | Importer |
| **433256** | 0,915 | 54 | 1,00 | Exporter |
| **637868** | 0,915 | 55 | 1,00 | Importer |
| **268011** | 0,915 | 56 | 1,00 | Exporter |
| **389641** | 0,910 | 57 | 1,00 | Exporter |
| **503468** | 0,910 | 58 | 1,00 | Exporter |
| **677154** | 0,910 | 59 | 1,00 | Exporter |
| **677154** | 0,908 | 60 | 1,00 | Importer |
| **663136** | 0,893 | 61 | 1,00 | Importer |
| **221697** | 0,892 | 62 | 1,00 | Importer |
| **246387** | 0,892 | 63 | 1,00 | Exporter |
| **698441** | 0,892 | 64 | 1,00 | Exporter |
| **691015** | 0,892 | 65 | 1,00 | Importer |
| **744826** | 0,884 | 66 | 1,00 | Importer |
| **698138** | 0,884 | 67 | 1,00 | Exporter |
| **841908** | 0,882 | 68 | 1,00 | Importer |
| **805048** | 0,876 | 69 | 1,00 | Importer |
| **260892** | 0,870 | 70 | 1,00 | Exporter |
| **744826** | 0,870 | 71 | 1,00 | Importer |
| **526105** | 0,870 | 72 | 1,00 | Importer |
| **161976** | 0,870 | 73 | 1,00 | Exporter |
| **662914** | 0,870 | 74 | 1,00 | Exporter |
| **468923** | 0,870 | 75 | 1,00 | Importer |
| **270300** | 0,870 | 76 | 1,00 | Exporter |
| **677154** | 0,867 | 77 | 1,00 | Importer |
| **487979** | 0,867 | 78 | 1,00 | Importer |
| **743983** | 0,867 | 79 | 1,00 | Importer |
| **389641** | 0,865 | 80 | 1,00 | Exporter |
| **175031** | 0,865 | 81 | 1,00 | Exporter |
| **677154** | 0,865 | 82 | 1,00 | Importer |
| **471681** | 0,862 | 83 | 1,00 | Exporter |
| **677154** | 0,860 | 84 | 1,00 | Importer |
| **677154** | 0,860 | 85 | 1,00 | Importer |
| **441390** | 0,859 | 86 | 1,00 | Importer |
| **712897** | 0,852 | 87 | 1,00 | Importer |
| **389641** | 0,852 | 88 | 1,00 | Importer |
| **389641** | 0,852 | 89 | 1,00 | Importer |
| **389641** | 0,852 | 90 | 1,00 | Importer |
| **389641** | 0,852 | 91 | 1,00 | Importer |
| **389641** | 0,852 | 92 | 1,00 | Importer |
| **389641** | 0,852 | 93 | 1,00 | Importer |
| **389641** | 0,852 | 94 | 1,00 | Importer |
| **389641** | 0,852 | 95 | 1,00 | Exporter |
| **389641** | 0,852 | 96 | 1,00 | Exporter |
| **389641** | 0,852 | 97 | 1,00 | Importer |
| **389641** | 0,852 | 98 | 1,00 | Exporter |
| **389641** | 0,852 | 99 | 1,00 | Importer |
| **389641** | 0,852 | 100 | 1,00 | Importer |

## I.7 Results Stress Centrality

* Due to privacy considerations the ID code presented here is not the original business number.

| ID Code | Stress Centrality | Ranking Stress | Rabies Risk | Type |
|---|---|---|---|---|
| **389641** | 1,000 | 1 | 0,00 | Exporter |
| **487979** | 1,000 | 2 | 0,00 | Exporter |
| **260892** | 1,000 | 3 | 0,00 | Importer |
| **805048** | 1,000 | 4 | 0,00 | Exporter |
| **805048** | 1,000 | 5 | 0,00 | Exporter |
| **853401** | 1,000 | 6 | 0,00 | Exporter |
| **291410** | 1,000 | 7 | 0,00 | Exporter |
| **455480** | 1,000 | 8 | 0,00 | Exporter |
| **934846** | 1,000 | 9 | 0,00 | Importer |
| **166366** | 1,000 | 10 | 0,00 | Exporter |
| **757725** | 1,000 | 11 | 0,00 | Exporter |
| **271747** | 1,000 | 12 | 0,00 | Exporter |
| **370471** | 1,000 | 13 | 0,00 | Exporter |
| **155395** | 1,000 | 14 | 0,00 | Exporter |
| **566229** | 1,000 | 15 | 0,00 | Exporter |
| **433256** | 1,000 | 16 | 0,00 | Exporter |
| **742927** | 1,000 | 17 | 0,00 | Exporter |
| **681887** | 1,000 | 18 | 0,00 | Exporter |
| **428333** | 1,000 | 19 | 0,00 | Exporter |
| **297625** | 1,000 | 20 | 0,00 | Exporter |
| **724201** | 1,000 | 21 | 0,00 | Exporter |
| **268011** | 1,000 | 22 | 0,00 | Importer |
| **838266** | 1,000 | 23 | 0,00 | Exporter |
| **805048** | 1,000 | 24 | 1,00 | Exporter |
| **258566** | 1,000 | 25 | 0,00 | Exporter |
| **155395** | 0,944 | 26 | 0,00 | Exporter |
| **455480** | 0,889 | 27 | 0,00 | Importer |
| **677154** | 0,857 | 28 | 0,00 | Importer |
| **805048** | 0,763 | 29 | 1,00 | Exporter |
| **433256** | 0,721 | 30 | 0,27 | Importer |
| **678473** | 0,643 | 31 | 0,00 | Exporter |
| **201431** | 0,629 | 32 | 0,00 | Exporter |
| **830853** | 0,619 | 33 | 0,00 | Exporter |
| **875061** | 0,600 | 34 | 0,00 | Importer |
| **938007** | 0,595 | 35 | 0,00 | Exporter |
| **987414** | 0,587 | 36 | 0,00 | Exporter |
| **853401** | 0,571 | 37 | 0,00 | Exporter |
| **455480** | 0,530 | 38 | 0,00 | Exporter |
| **433256** | 0,497 | 39 | 0,00 | Importer |
| **389641** | 0,444 | 40 | 0,00 | Importer |
| **861540** | 0,433 | 41 | 0,00 | Importer |
| **375249** | 0,402 | 42 | 0,00 | Exporter |
| **801487** | 0,389 | 43 | 0,00 | Importer |
| **805048** | 0,388 | 44 | 1,00 | Exporter |
| **677154** | 0,346 | 45 | 0,00 | Importer |
| **201431** | 0,342 | 46 | 0,00 | Importer |
| **805048** | 0,334 | 47 | 1,00 | Exporter |
| **566229** | 0,330 | 48 | 0,00 | Exporter |
| **678473** | 0,318 | 49 | 0,00 | Importer |

| | | | | |
|---|---|---|---|---|
| **389641** | 0,300 | 50 | 0,00 | Importer |
| **723466** | 0,300 | 51 | 0,00 | Importer |
| **805048** | 0,300 | 52 | 1,00 | Exporter |
| **596831** | 0,293 | 53 | 0,00 | Importer |
| **742927** | 0,238 | 54 | 0,00 | Exporter |
| **677154** | 0,238 | 55 | 0,00 | Importer |
| **456696** | 0,229 | 56 | 0,00 | Exporter |
| **456407** | 0,217 | 57 | 0,00 | Exporter |
| **455480** | 0,217 | 58 | 0,00 | Exporter |
| **742927** | 0,217 | 59 | 0,00 | Exporter |
| **592049** | 0,209 | 60 | 0,00 | Importer |
| **566229** | 0,209 | 61 | 0,00 | Importer |
| **677154** | 0,209 | 62 | 0,00 | Importer |
| **805048** | 0,193 | 63 | 1,00 | Exporter |
| **389641** | 0,192 | 64 | 0,00 | Exporter |
| **677154** | 0,184 | 65 | 1,00 | Importer |
| **677154** | 0,157 | 66 | 0,90 | Importer |
| **441390** | 0,138 | 67 | 1,00 | Exporter |
| **675075** | 0,133 | 68 | 0,00 | Importer |
| **487979** | 0,117 | 69 | 0,00 | Importer |
| **127801** | 0,117 | 70 | 1,00 | Importer |
| **389641** | 0,117 | 71 | 0,00 | Exporter |
| **677154** | 0,115 | 72 | 1,00 | Importer |
| **488424** | 0,115 | 73 | 0,00 | Exporter |
| **805048** | 0,107 | 74 | 1,00 | Exporter |
| **471681** | 0,085 | 75 | 1,00 | Importer |
| **924153** | 0,069 | 76 | 0,00 | Exporter |
| **663136** | 0,068 | 77 | 1,00 | Importer |
| **270300** | 0,066 | 78 | 1,00 | Importer |
| **224838** | 0,056 | 79 | 0,00 | Importer |
| **391377** | 0,053 | 80 | 0,00 | Exporter |
| **389641** | 0,051 | 81 | 0,00 | Exporter |
| **677154** | 0,051 | 82 | 0,00 | Importer |
| **852847** | 0,050 | 83 | 0,00 | Exporter |
| **389641** | 0,045 | 84 | 1,00 | Importer |
| **383661** | 0,045 | 85 | 1,00 | Importer |
| **677154** | 0,045 | 86 | 1,00 | Importer |
| **342448** | 0,045 | 87 | 0,00 | Importer |
| **677154** | 0,044 | 88 | 0,46 | Importer |
| **727314** | 0,039 | 89 | 1,00 | Importer |
| **143194** | 0,039 | 90 | 1,00 | Importer |
| **389641** | 0,035 | 91 | 1,00 | Importer |
| **503468** | 0,035 | 92 | 1,00 | Importer |
| **677154** | 0,035 | 93 | 1,00 | Importer |
| **663136** | 0,034 | 94 | 0,08 | Importer |
| **844941** | 0,032 | 95 | 0,00 | Exporter |
| **848052** | 0,031 | 96 | 0,00 | Exporter |
| **841908** | 0,029 | 97 | 1,00 | Importer |
| **678473** | 0,027 | 98 | 0,00 | Exporter |
| **924153** | 0,026 | 99 | 1,00 | Importer |
| **455480** | 0,026 | 100 | 1,00 | Importer |

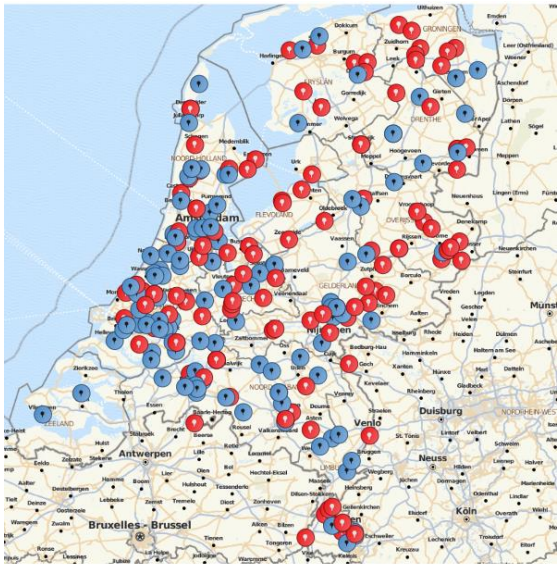
## I.8 Results Betweenness Centrality

* Due to privacy considerations the ID code presented here is not the original business number.

| ID Code | Betweenness Centrality | Ranking Betweenness | Rabies Risk | Type |
|---|---|---|---|---|
| **258566** | 1,000 | 1 | 0,00 | Exporter |
| **389641** | 1,000 | 2 | 0,00 | Exporter |
| **487979** | 1,000 | 3 | 0,00 | Importer |
| **260892** | 1,000 | 4 | 0,00 | Exporter |
| **805048** | 1,000 | 5 | 0,00 | Exporter |
| **805048** | 1,000 | 6 | 0,00 | Exporter |
| **805048** | 1,000 | 7 | 1,00 | Exporter |
| **853401** | 1,000 | 8 | 0,00 | Exporter |
| **291410** | 1,000 | 9 | 0,00 | Importer |
| **455480** | 1,000 | 10 | 0,00 | Exporter |
| **934846** | 1,000 | 11 | 0,00 | Exporter |
| **166366** | 1,000 | 12 | 0,00 | Exporter |
| **757725** | 1,000 | 13 | 0,00 | Exporter |
| **271747** | 1,000 | 14 | 0,00 | Exporter |
| **370471** | 1,000 | 15 | 0,00 | Exporter |
| **155395** | 1,000 | 16 | 0,00 | Exporter |
| **566229** | 1,000 | 17 | 0,00 | Exporter |
| **433256** | 1,000 | 18 | 0,00 | Exporter |
| **742927** | 1,000 | 19 | 0,00 | Exporter |
| **681887** | 1,000 | 20 | 0,00 | Exporter |
| **428333** | 1,000 | 21 | 0,00 | Exporter |
| **297625** | 1,000 | 22 | 0,00 | Importer |
| **724201** | 1,000 | 23 | 0,00 | Exporter |
| **268011** | 1,000 | 24 | 0,00 | Exporter |
| **838266** | 1,000 | 25 | 0,00 | Exporter |
| **155395** | 0,944 | 26 | 0,00 | Exporter |
| **455480** | 0,889 | 27 | 0,00 | Importer |
| **677154** | 0,857 | 28 | 0,00 | Importer |
| **805048** | 0,803 | 29 | 1,00 | Exporter |
| **433256** | 0,721 | 30 | 0,27 | Importer |
| **678473** | 0,643 | 31 | 0,00 | Exporter |
| **830853** | 0,619 | 32 | 0,00 | Exporter |
| **875061** | 0,600 | 33 | 0,00 | Importer |
| **987414** | 0,587 | 34 | 0,00 | Exporter |
| **853401** | 0,571 | 35 | 0,00 | Exporter |
| **805048** | 0,560 | 36 | 1,00 | Exporter |
| **433256** | 0,497 | 37 | 0,00 | Importer |
| **201431** | 0,452 | 38 | 0,00 | Exporter |
| **389641** | 0,444 | 39 | 0,00 | Importer |
| **938007** | 0,426 | 40 | 0,00 | Exporter |
| **805048** | 0,420 | 41 | 1,00 | Exporter |
| **861540** | 0,402 | 42 | 0,00 | Importer |
| **375249** | 0,402 | 43 | 0,00 | Exporter |
| **801487** | 0,389 | 44 | 0,00 | Importer |
| **455480** | 0,347 | 45 | 0,00 | Exporter |
| **677154** | 0,346 | 46 | 0,00 | Importer |
| **566229** | 0,330 | 47 | 0,00 | Exporter |
| **678473** | 0,318 | 48 | 0,00 | Importer |

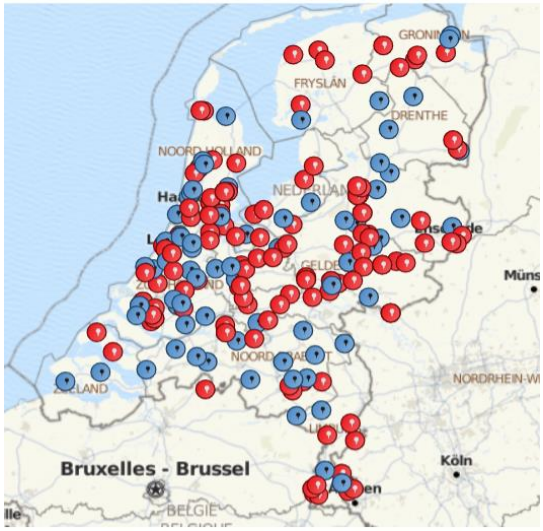
| | | | | |
|---|---|---|---|---|
| **677154** | 0,305 | 49 | 1,00 | Importer |
| **389641** | 0,300 | 50 | 0,00 | Importer |
| **723466** | 0,300 | 51 | 0,00 | Importer |
| **805048** | 0,258 | 52 | 1,00 | Exporter |
| **201431** | 0,253 | 53 | 0,00 | Importer |
| **742927** | 0,238 | 54 | 0,00 | Exporter |
| **677154** | 0,238 | 55 | 0,00 | Importer |
| **677154** | 0,234 | 56 | 0,90 | Importer |
| **456696** | 0,229 | 57 | 0,00 | Exporter |
| **456407** | 0,217 | 58 | 0,00 | Exporter |
| **455480** | 0,217 | 59 | 0,00 | Exporter |
| **742927** | 0,217 | 60 | 0,00 | Exporter |
| **805048** | 0,193 | 61 | 1,00 | Exporter |
| **389641** | 0,192 | 62 | 0,00 | Exporter |
| **596831** | 0,188 | 63 | 0,00 | Importer |
| **805048** | 0,138 | 64 | 1,00 | Exporter |
| **389641** | 0,117 | 65 | 0,00 | Importer |
| **487979** | 0,117 | 66 | 0,00 | Importer |
| **127801** | 0,117 | 67 | 1,00 | Exporter |
| **441390** | 0,117 | 68 | 1,00 | Exporter |
| **592049** | 0,114 | 69 | 0,00 | Importer |
| **566229** | 0,114 | 70 | 0,00 | Importer |
| **677154** | 0,114 | 71 | 0,00 | Importer |
| **675075** | 0,102 | 72 | 0,00 | Importer |
| **471681** | 0,097 | 73 | 1,00 | Importer |
| **224838** | 0,094 | 74 | 0,00 | Importer |
| **391377** | 0,091 | 75 | 0,00 | Exporter |
| **677154** | 0,089 | 76 | 0,00 | Importer |
| **488424** | 0,089 | 77 | 0,00 | Exporter |
| **852847** | 0,089 | 78 | 0,00 | Exporter |
| **924153** | 0,080 | 79 | 0,00 | Exporter |
| **677154** | 0,077 | 80 | 1,00 | Importer |
| **389641** | 0,062 | 81 | 0,00 | Exporter |
| **677154** | 0,053 | 82 | 0,46 | Importer |
| **844941** | 0,051 | 83 | 0,00 | Exporter |
| **841908** | 0,047 | 84 | 1,00 | Importer |
| **744826** | 0,046 | 85 | 1,00 | Importer |
| **698138** | 0,046 | 86 | 1,00 | Importer |
| **512467** | 0,046 | 87 | 0,00 | Exporter |
| **677154** | 0,042 | 88 | 1,00 | Importer |
| **541703** | 0,041 | 89 | 1,00 | Importer |
| **663136** | 0,039 | 90 | 1,00 | Importer |
| **270300** | 0,038 | 91 | 1,00 | Importer |
| **588966** | 0,037 | 92 | 0,00 | Exporter |
| **678473** | 0,034 | 93 | 0,00 | Exporter |
| **191127** | 0,030 | 94 | 0,00 | Importer |
| **848052** | 0,028 | 95 | 0,00 | Exporter |
| **663136** | 0,027 | 96 | 0,08 | Importer |
| **389641** | 0,026 | 97 | 1,00 | Importer |
| **503468** | 0,026 | 98 | 1,00 | Importer |
| **677154** | 0,026 | 99 | 1,00 | Importer |
| **389641** | 0,026 | 100 | 1,00 | Importer |

## I.9 Tracking Changes Over Time

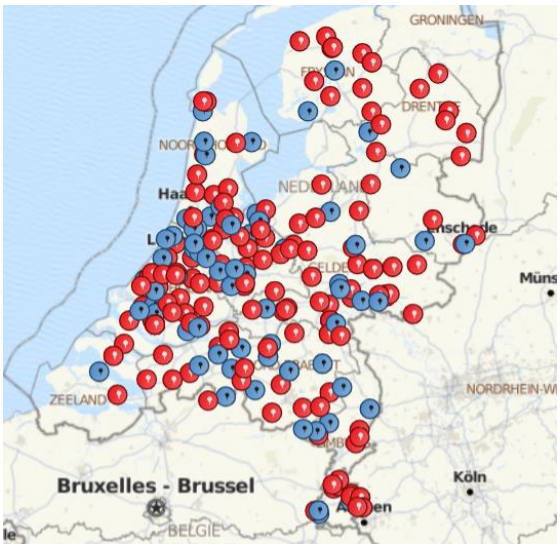



**January 2018**


**February 2018**

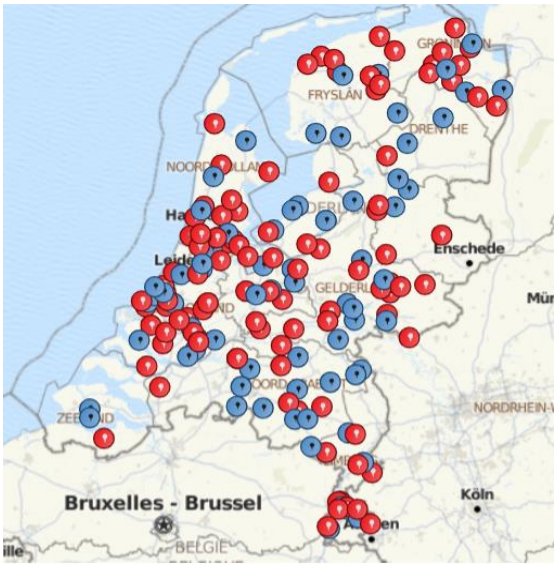



**March 2018**

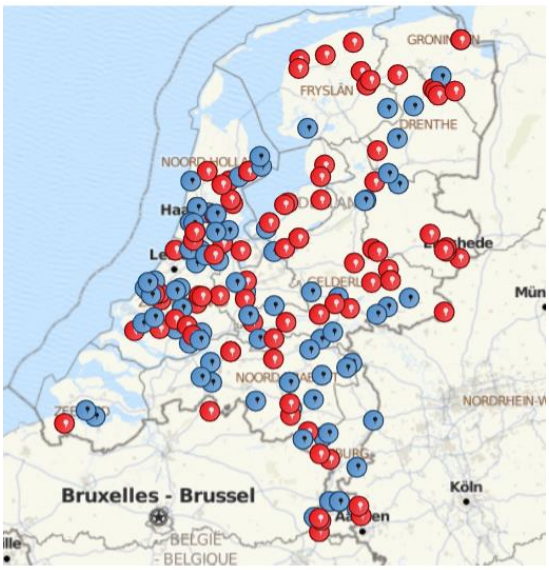



**April 2018**

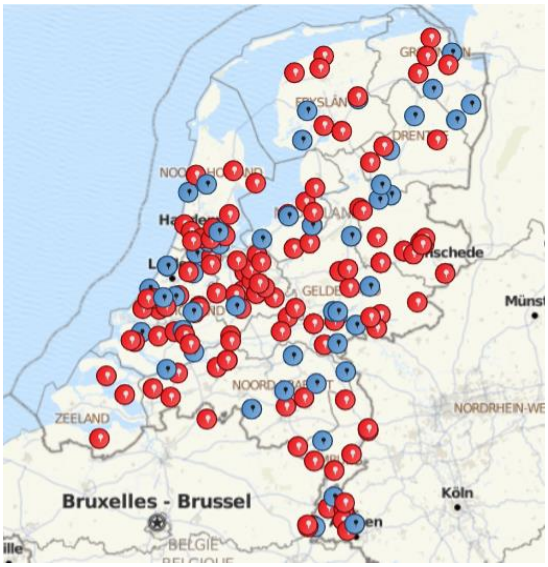

**May 2018**




**June 2018**

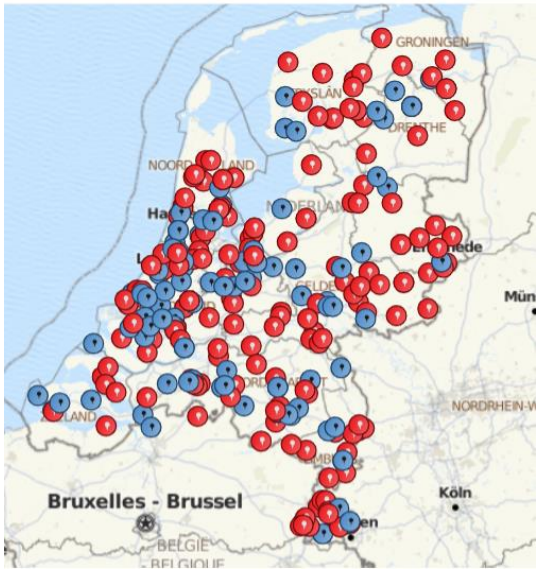



**July 2018**



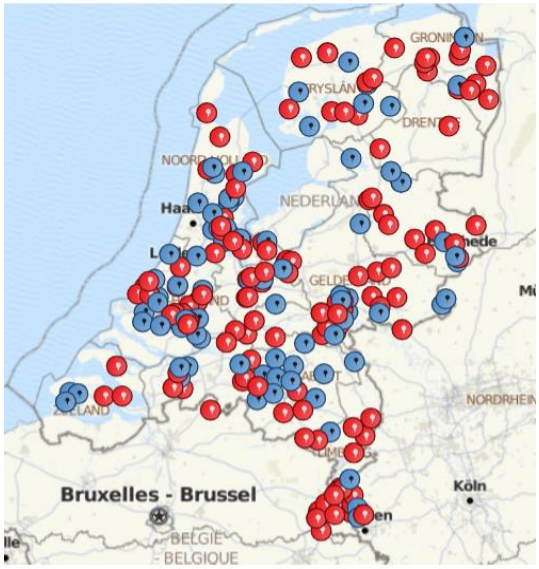

**August 2018**


**September 2018**




**October 2018**


**November 2018**




**December 2018**