LUND
UNIVERSITY

Responses from societal sectors in Sweden concerning the rise of cyberfraud

A quantitative document analysis using topic modelling

Viktor Henriksson

Lund University
Sociology of Law Department

Master Thesis (SOLM02)
Spring 2024

Supervisor: Matthias Baier

Examiner:

Abstract

Fraud is a crime that causes severe economic damage, and crime rates are rising because of technological advancements. There is a research gap regarding how fraud crime prevention is organised and which actors are important for it. This thesis aimed to explore how the rising occurrence of cyber fraud is addressed on a societal level, using the civil, banking, governmental, and business sectors. It also aimed to explore how this could be placed in the framework of risk society, governmentality, and nodal governance. It used a computational method called topic modelling to analyse a larger collection of documents. This method resulted in topics describing the most central themes of the documents, and it showed which topic was more important for each sector and how the sectors and topics related to each other. The key findings described that fraud prevention is largely seen as an issue for the individual, as they have to be informed and aware to prevent their own victimisation. The civil sector thought it was important to keep the banking sector responsible for fraud prevention, and the banking sector is the most central actor in the fraud prevention network, as it has demands and expectations on itself from the other actors. The findings imply that a new kind of fraud prevention is necessary, as technological advancements have made finding the perpetrator difficult, space irrelevant, and the larger scope of the crimes makes traditional law enforcement inefficient.

Keywords: fraud; fraud prevention; Sweden; topic modelling; risk society; governmentality; nodal governance.

Acknowledgements

I want to thank the University of Helsinki for giving me the chance to continue my exchange to write my thesis in Finland, as well as planting the inspiration to use computational methods already during the courses I took during the autumn. Especially I want to thank Karolina Stenlund at the University of Helsinki for her advice from a legal perspective. Further, I want to thank Alexander Stelzer, for his continuous support during the process from draft to final thesis. Finally, I want to thank my supervisor Matthias Baier, for his thoughtful feedback during the semester.

Table of Contents

1 Introduction

1.1 The issue of cyber fraud

Cybercrime is becoming a more persistent issue, as seen by cyber-attacks and rising crime rates. A part of cybercrime is fraud, which is on the rise. Reported fraud crimes have tripled in size since the 2000s in Sweden, and the economic damage of fraud worldwide was estimated to be around 1.026 trillion USD in 2022-2023, defrauding around a quarter of the world's population (GASA, 2023; Shannon, 2022). Cyberfraud is different from traditional forms of crime, as it does not take place physically in the sense that the victim and the perpetrator need to be in the same place simultaneously. In a sense, fraud is not as visible as other street crimes and is therefore often ignored in policy contexts, because it has not been seen as a concern (Digital Fraud Committee, 2022). Space and scope have gotten much more different meanings in fraud than with traditional crimes (Brenner, 2004). A scammer can be in another place, no matter if it is in the same country or not, and can easily automate the crime on a bigger scale (Brenner, 2004; Levi et al, 2017). When it comes to the transnational angle of cyber fraud, it heavily impedes the investigation of these crimes, as different legislations might not have the same law for these crimes, or law enforcement has other priorities (Grabosky, 2001). Further, cybercrime in general, but especially fraud, has not been seen as a social problem like traditional crimes do. The financial risks and the connection to terrorism were highlighted during the 2000s, alongside technological advancements and fraud has gotten more attention (Levi and Doig, 2020). Generally, crime has been declining globally, but when it comes to fraud, the crime rate is rising, which is seen in Sweden as well (Holst, et al, 2023; Prenzler, 2020). Fraud is one of the most common crime types in Sweden nowadays (Brottsofferjouren, n.d.).

Fraud is an example of a new risk, as Beck (1992) explained when he coined the term "risk society". This means that with technological advancements, new types of threats appear, and law enforcement is not equipped for them (ibid.). Handling risks, especially of transnational nature, poses many challenges that policymakers deal with (Zedner, 2009). One of the challenges is finding proactive ways to combat fraud (Brenner, 2004). By implementing proactive approaches, risks can be mitigated, and the root causes can be addressed. Crime prevention is important, especially in fraud, as the victim and perpetrator may be in different legislations, making it more difficult for law enforcement to act (Button and Cross, 2017). However, there exists a critical knowledge gap in understanding how measures aimed at

combating cyber fraud are organised and implemented (Leppänen and Kankaanranta, 2017; Näsi et al., 2023). That is why this thesis aims to find out how the issue of cyber fraud is addressed in Sweden by analysing political documents relevant to fraud prevention. It will provide knowledge and an overview of how fraud prevention is organised on a national level, and it will give insight into how the principles of risk society are connected to fraud prevention. Utilising a Natural Language Processing (NLP) tool called topic modelling, which generates topics from a larger collection of documents, as a methodological approach, this thesis aims to analyse and interpret political documents to enhance the understanding of how policies adapt to emerging cyber threats, contributing to both theoretical advancements and practical implications for cyber fraud prevention.

1.2 Relevance to Sociology of Law

Sociology of Law examines the interplay between legal systems and societal structures. Cyber fraud is an issue that goes beyond conventional legal and social categories, involving technological, financial, and societal dimensions, representing a complex and dynamic risk in modern society. This can be connected to the concept of risk society, as coined by sociologist Ulrich Beck (1992), who suggests that contemporary societies are characterised by increased awareness of and responsiveness to various risks. These risks often transcend traditional boundaries, challenging established legal and social frameworks. This affects policymaking and the implementation of fraud prevention measures. Levi and Maguire (2004) claim that explicit policies, plans and strategies heavily influence which crimes Law Enforcement focuses on, explaining why a top-down perspective is important when examining crime prevention. Further, power and knowledge are inherent in Foucault's governmentality, which provides a framework to analyse how power and knowledge can legitimise policies, which can control the population, and normalise the practices of crime prevention (Madsen, 2014; Lawlor and Nale, 2014; Torres-Cuello and Pinzon-Salcedo, 2023). Examining the policy responses to cyber fraud from a top-down perspective aligns with the sociology of law's interest in understanding how legal systems, policies, and enforcement mechanisms are structured and implemented at the societal level. In addition, topic modelling can uncover hidden themes and explore the complexity of cyber fraud from a different angle than usual, which can navigate the many dimensions of cyber fraud and reveal the interplay between social structures and legal systems.

1.3 Aim and research questions

Fraud is a rising problem, made possible by technological advancements. This thesis aims to produce knowledge that can help to develop fraud prevention techniques and understand the actors important for fraud prevention, as well as what these actors deem important in prevention. In detail, this thesis aims to explore how the rising occurrence of cyber fraud is addressed on a societal level. This is done by using topic modelling to generate topics from a large collection of political documents, which are documents that outline past, present and/or future actions within or information about cyber fraud prevention. It will investigate how documents from four societal sectors frame cyber fraud as a societal risk, considering the concepts of risk society, governmentality, and nodal governance. The societal sectors in this thesis are the governmental, banking, civil, and business sectors.

- Which topics regarding cyber fraud in Sweden can be identified in political documents?
- Which similarities and/or differences can be found in the sectors in Sweden regarding topics about cyber fraud?
- How do the findings regarding cyber fraud reflect the concepts of risk society, governmentality, and nodal governance?

2 Fraud within the context of criminal justice

2.1 Fraud as a concept in scientific literature

Fraud is a broad category with many different types of crimes. Kemp et al (2020) define fraud as "an act of wilful deception that produces an economic benefit (or evasion of a loss) for the deceiver and a loss for the victim" (pg. 294). Burgard and Schlembach (2013) claim that fraud is a type of fabrication, where the deceiving party manipulates the victim into taking actions that hurt themselves. They compare cyber fraud with angling, where the deceiver throws out the hook and waits until someone gets caught. For this reason, fraud victims often feel shame and guilt over their part in the crime, and often hide their victimisation (ibid.). Fraud rarely gets reported to the police, and the reason for that can be that the victim does not want to appear as "foolish" (Smith, 2007). Fraud does however get more often reported to banks and other agencies (ibid.), but these numbers are not included in the official statistics. Individuals might not report because they do not understand that they have been victimised, or because they feel ashamed about it (ibid.). Companies might not report the crime because they do not want to taint their public image by revealing weaknesses in their security (ibid.). Most victims of cyber fraud are individuals, but companies are losing more monetary value (Levi et al., 2017). This shows how multifaceted the concept of fraud is.

2.2 Legal framework regarding fraud

The legal framework describes, first and foremost, what fraud is in the eyes of the Swedish law. This gives another perspective to the concept of fraud, as outlined in the previous section, and it gives an overview of which rules and regulations are foundational for fraud prevention. Fraud can be a form of volume crime, meaning that a high volume of crime gets committed and often has low clearance rates (Brown and Smith, 2018). In a way, criminal law only describes the outliers of the cases that do get reported and settled, but the legal framework is nonetheless important as it does not only describe the additional laws that describe how and who should handle fraud.

The criminal law about fraud is described in chapter 9 of Brottsbalken (1962:700). General scams are described in the first paragraph, where it is defined as a person who is deceiving another person and causes them to act or neglect to act in a way which brings profit for the perpetrator and damage for the victim or someone there is disposing of (Brottsbalken ch. 9 §

1). The paragraph does not describe the way fraud happens, and is therefore neutral in the technology used, meaning that they can refer to online as well as offline fraud. Incomplete or aborted attempts of fraud are seen as crimes as well (Shannon et al, 2016).

The second part of the first paragraph of chapter 9 in Brottsbalken describes fraud in the sense that someone is leaving false or incomplete information in a program or recording to affect the result of automatic information processing, which causes profit for the perpetrator and damage for someone else. This paragraph is connected to cybercrime, as it concerns computer-related crime. It can concern everything from computers and phones, as well as ATMs and vending machines (See Friberg, Brottsbalk (1962:700), 9 ch. § 1, Karnov (JUNO) (Visited 2024-02-19). The Swedish law only states damage and does not connect it to damage in an economic sense.

Fraud is connected to money laundering, as the money gained from it is not legitimate. In Sweden, the newest law about measures against money laundering was implemented in 2017 (SFS 2017:630), which will henceforth be referred to as the "anti-money laundering law". The purpose of the law is to reduce money laundering by putting responsibility on financial institutes to regulate their operations and monitor their customers (Prop. 2016/17:173). The definition of money laundering is stated in Ch. 1 § 6, where it is about measures regarding money that derives from criminal activity that has the purpose of hiding the crime, making the money available for use for someone else, or facilitating someone to escape prosecution for the crime. In chapter 3, the law states that financial institutes need to have knowledge about the customer and what their intention with starting a business relationship with them is, and they shall continuously revise it. This is called Know Your Customer (KYC), which is a part of fighting money laundering and financial crime (Arner et al, 2018). The banks must know who the customer is and what the purpose of them having an account at the bank is, which is different from the anonymity the internet provides (ibid.). Chapter 4 of the anti-money laundering law states that financial institutes must report suspicious and deviant transactions. These can be gains from fraudulent activities that still have not been acted on in a money laundering sense (Prop. 2016/17:173). Ch. 4 §3 states that if money can be supposed to be connected to money laundering, financing of terrorism, or it can be assumed to come from criminal activities, the financial institute shall act on the customer and report the event to the Swedish police authority. Even though fraud is not strictly money laundering, it is very much connected to that field and the law can be applied to fraud as well.

An EU Directive, the Second Payment Services Directive, or PSD2 as it is commonly known, was issued by the EU in 2015 (Steennot, 2018). PSD2 calls for strong customer authentication (SCA) as a way to combat unauthorised payments, which are payments that the payer did not agree to (ibid.). The issue might arise in what can be seen as an unauthorised payment and where the banks' responsibilities lay. PSD2 provides consumer protection in the way that the payment service cannot hold the payer liable for their own damages if no SCA was used, or the payer's SCA and payment instruments were not in their possession (ibid.). PSD2 is executed in Swedish law by changes in the law (2010:751) about payment services. Ch. 5 a §1 follows from PSD2, as it states that it is the financial institutes' responsibility to restore an account where there have been unauthorized transactions. Ch 5 a § 5 describes SCA, where the party that did not require SCA is the one that is responsible for restoring the account. However, even if SCA was used but the account owner acted with negligence according to Ch. 5a § 3, the bank's responsibility to restore the account is no longer valid. Further, if there have been unauthorised transactions and the account holder does not contact the bank about them as soon as they got to know about them, the responsibility falls on to the account owner according to Ch. 5a §6. However, this becomes a difficult procedure and difficult to decide where the responsibility lies. A court ruling in the Swedish Supreme Court changed the praxis, as many fraud victims did not get their money back as they were deemed to have acted with negligence (SVT, 2023). In Högsta Domstolen verdict 2022-06-21 in case nr. T 4623-21, the court ruled that the victim could not be seen as having acted with negligence within the context: he was called multiple times by someone who said they were from the bank, and the second time he was asked for bank identification codes. The scammers used the new laws regarding GDPR and said that the victim had not answered a letter about it. The court ruled that the victim did not purposefully give out codes to an unauthorised person as he thought it was the bank, especially when it happened over multiple phone calls. They built a relationship and manipulated the victim, and therefore the victim could not be blamed, according to the verdict. Since that court ruling, the banks' responsibility in fraud cases has been widely debated, as to which circumstances the account owner should be responsible for the damage, and when the banks should compensate them (SVT, 2023). This ruling is, however, not applicable to companies (Krantz, 2024).

The legal framework is relevant as the law steers what is deemed as criminal. However, the issue with cases that never get reported does make the reported cases unique in a sense. The

legal system does not have the resources to take care of all the fraud cases, leading to a lot of them never being solved.

2.3 Types of frauds

2.3.1 General types of fraud

Levi et al (2017) differentiate between multiple types of cyber fraud. Cyber-dependent crimes would not exist without the Internet (ibid.). Cyber-enabled and cyber-assisted crimes are quite similar as they describe crimes that could take place without the internet, but the difference is that cyber-enabled crime uses the internet as the main platform and can scale up the scope of the crime, while cyber-assisted crimes integrate technology in the criminal act via using encrypted communication, but the internet is not the main platform (ibid.). Technical tools reduced the cost and effort for scammers to reach out to potential victims (ibid.). 80 % of reported fraud crimes in the UK are cyber-enabled ones (Digital Fraud Committee, 2022).

Shannon et al (2016) identified five main types of scams: card fraud, credit fraud, advertisement fraud, invoice fraud, and miscellaneous telephone and internet fraud. Card fraud, also called Card-Not-Present (CNP) fraud, is one of the most common cyber frauds and has risen dramatically because of technological advancements (Bodker et al., 2023). It means that a fraudster gets access to someone's credit or debit card details and uses them to make purchases (ibid.). Credit fraud is when someone uses someone else's identity to buy something or take loans (Fjelkegård and Horgby, 2023). Advertisement fraud entails a scammer putting up a fake advertisement, selling something, but then never delivering it (Shannon et al, 2016). Invoice fraud is when an individual or business gets an invoice for something they did not order (Fjelkegård and Horgby, 2023). The miscellaneous telephone and internet frauds will be described below.

Most of these scams are so-called Authorised Push Payment (APP) fraud. It is when an individual or a company authorises a payment to someone who is posing as a legitimate receiver (Taylor and Galica, 2020). These transactions are difficult for the bank to act on as they have been signed and confirmed by the victim and therefore seen as legitimate (Maher, 2021). This is used in a variety of ways, from advertisement frauds to social engineering scams.

2.3.2 Social engineering

Social engineering is a manipulation tactic that deceives people into disclosing confidential information (Atkins and Huang, 2013). The tactic exploits the weaknesses and emotions of a person to make them more vulnerable (ibid.). Quite often this is in the form of stressing and scaring the victim by telling them that something serious has happened to their bank account, but the scammer can also focus on building a relationship with the victim (Digital Fraud Committee, 2022; Fjelkegård and Horgby, 2023). Current events like the COVID-19 pandemic and the Russo-Ukrainian war get exploited, as fraudsters use them to add credibility to their claims (Digital Fraud Committee, 2022). Atkins and Huang (2013) argue that social engineering scams are difficult to prevent, as they exploit human factors and vulnerabilities. This, coupled with the fact that most people do not believe this would happen to them, makes it difficult to defend against (ibid.).

Phishing is a type of fraud using social engineering, where the scammer pretends to be a trustworthy source, such as a bank or an authority, and tricks the victim into either downloading malware or giving out information (Atkins and Huang, 2013). Phishing usually refers to emails, but it can come in the form of text messages and phone calls, called smishing and vishing respectively (Fjelkegård and Horgby, 2023). Phishing and smishing usually contain links that the user can click, but with vishing, the scammer usually deceives the victim by pretending to be from the bank or some authority to report that something serious has happened and that urgent actions are needed (Fjelkegård and Horgby, 2023). This, coupled with the scammer often manipulating the phone number to make it look legitimate, scares the victim and leads them to follow the scammer's instruction in hope that they will help (ibid.).

Atkins and Huang (2013) mention advance fee scams, which are when a scammer promises the victim a reward if they pay a fee. This can be in the form of an inheritance, where the scammer claims that money is available for the victim and that they need to act fast if they want to claim it (ibid.). Quite often, the scammer says that they will need to act with secrecy, or the authorities and banks will not allow the transaction, instilling doubt in the authorities for the victim (ibid.). This is one way scammers use people's hope and give them false pretences. This is common in investment fraud. Investment fraud entails an investment opportunity that turns out to be fake (Digital Fraud Committee, 2022). Quite often, cryptocurrency is involved in these scams, as they are seen as valuable new technology, but the real reason is often the encrypted nature of cryptocurrency making it almost impossible to track (ibid.). Romance fraud is different from the aforementioned scams, as it focuses on the

relationship between the victim and a persona the scammer has created (Fjelkegård and Horgby, 2023). The usual modus is the victim encountering a scammer on a dating website, and they start a relationship (ibid.). Soon, the "lover" encounters problems and needs the victim to send them money, and this behaviour escalates (ibid.).

Social manipulation scams yield large damage to their victims, monetary as well as psychological (Fjelkegård and Horgby, 2023). In contrast, victims of card and advertisement scams do usually not lose as much money (ibid.). Generally, in all fraud types except for card frauds, the money goes through multiple bank accounts to make the money more difficult to track by the banks and authorities (Fjelkegård and Horgby, 2023). The individuals who receive money in a scam in that manner are called money mules (ibid.). Quite often, money mules are younger people who have been recruited via social media in exchange for a cut of the profit (Digital Fraud Committee, 2022).

2.4 Fraud crime rates

This section explains the fraud crime rates, to give context to the rising occurrence of fraud. Fraud crime rates have risen dramatically in the last decade in the whole world (Junger, Wang and Schlömer, 2020). The same picture can be seen in Sweden, where fraud crime rates have risen during the 2000s and 2010s, and now telephone scams are described as the most serious problem (Fjelkegård and Horgby, 2023). Reported crimes and profits of crime give different pictures, as card frauds are the most reported ones, while fraud via social manipulation had the highest amount of money lost (ibid.). The rise in fraud rates could also be attributed to digitalisation, which makes it easier and faster for scammers to reach their victims and make off with their money (Digital Fraud Committee, 2022; Junger, Wang and Schlömer, 2020). As of 2022, card scams were 40 % of all reported crimes, crimes through social manipulation 20 %, and advertisement frauds 14 %, with other types of frauds having lower percentages (Fjelkegård and Horgby, 2023).

2.5 Cybercrime and Law Enforcement

Technology has been attributed as a factor for rising fraud crime rates (Junger, Wang and Schlömer, 2020), which makes it a relevant part to examine to understand how fraud is addressed. Technology advances fast and the methods for crime develop faster than the

traditional detection methods can keep up, quite often because traditional staff lacks knowledge about technology (McCord, et. al., 2022). Fraudsters often take on fake or stolen identities, making policing online more difficult as it is not known who is supposed to be prosecuted (Cross, 2016). Many cybercrimes are small-scale crimes, spread out in different jurisdictions, with many victims, which is a reason they are down-prioritized when it comes to Police investigation (Wall, 2007). Also, under-reporting makes it difficult to investigate cybercrimes (ibid.). Space has also become more irrelevant, as the scammer can commit the crime from anywhere in the world (Van Nguyen, 2022). When it comes to different nation-states, the legal angle can become difficult as an action can be under civil law in one country, criminal law in another, or not illegal in a third country (ibid.). For these reasons, the clearance rates of fraud crimes are quite low (Brown and Smith, 2018).

Cybercrimes are usually seen as unique and not as part of the police's work tasks (Wall, 2007). Law Enforcement is based on traditional crimes, with a focus on the physical space and scale of the crime, as a single perpetrator only has a limited number of resources and possibilities to commit crimes (Brenner, 2004). The traditional law enforcement has a reactive angle and is more focused on applying knowledge from the past (ibid.). Cybercrime differs from traditional crime in that it does not have the same physical limitations, and the scope can become much bigger thanks to the opportunity to use automated tools (ibid.). The bigger scope can be another reason that law enforcement has issues investigating cybercrimes, and the perpetrator can gain perfect anonymity, meaning that they cannot be traced (ibid.). Gathering evidence is also an issue in cybercrime, as the traces of the crime can disappear depending on the nature of the tools being used, and it can also be connected to the legislation, if the perpetrator and the investigators are in different countries (ibid.).

Wall (2007) claims that the police is a conservative organisation that values traditional methods, which makes it more difficult for them to keep up with technological advancements. Brenner (2004) argues that the traditional model of law enforcement is not suitable for cybercriminality, and that a new model should focus on collaboration between commercial entities and law enforcement. Dupont (2004) and Wall (2002) discuss the importance of networking in a risk society, as the resources of the police are not enough to effectively combat cybercrime. Therefore, other stakeholders such as banks and voluntary groups are needed to surveil, handle, and prevent cybercrimes (ibid.).

3 Fraud within the context of crime prevention and risk

3.1 Crime prevention

3.1.1 What is crime prevention?

Prevention is a big part of combatting crime, where a proactive approach is emphasised in the risk society (Brenner, 2004). Crime prevention research can sometimes clash with the expectations of politicians, as they want to implement measures quickly and effectively, while this seldom is the case with science (Wikström and Torstensson, 1999). This can make the proactive angle more difficult when measures that have no basis or proven effectiveness are used. This section will outline the overall approaches in crime prevention and describe how crime prevention is handled in Sweden generally and with the prevention of fraud and cybercrime. The prevention angle is important to understand to examine how cyber fraud rates are being dealt with and topic modelling can be used to uncover this from a large collection of documents.

Different approaches to crime prevention exist, where the overall strategies are situational and social crime prevention. Situational crime prevention focuses on opportunity and has its basis in rational choice and routine activity theories. and in a way ignores societal structures and sees crime as one-off events (Alvesalo et al., 2006; Button and Cross, 2017). Situational crime prevention is connected to the risk society, as the risks are seen as immediate causes rather than underlying factors (Zedner, 2009). This also has the side-effect of putting the responsibility for security on victims, as they in this framework should take more care to manage their risks (ibid.). The criticism towards situational crime prevention is that it does not take the offender's behaviour into account, and therefore ignores other factors that cause crime (Button and Cross, 2017). Social crime prevention addresses those issues and focuses more on the perpetrator and the risk factors surrounding them (Grant, 2015). It is a long-term strategy and requires more resources, but research has shown that it is necessary to prevent individuals from getting into long-term criminal careers (Savolainen, 2005). However, fraud prevention is more connected to situational crime prevention, as it focuses on the situation rather than the perpetrator (Prenzler, 2016).

3.1.2 Crime prevention policies in Europe and Sweden

Policies regarding fraud prevention are important to understand as they regulate the work on a macro level, and they are important to understand how the issue of fraud is addressed. Crime

prevention policies in Europe vary, where the Nordics usually have separate legislation and policies for crime prevention while Eastern Europe does not distinguish between criminal policy and crime prevention policy (Graham, 1993). Crime prevention in Europe tends to not just target the action itself, but rather wider social problems such as fear of crime, quality of life, and anti-social behaviour (Di Ronco, 2016). Generally, the Nordics are similar in crime prevention as they focus more on collaboration between the citizens and welfare policies, having the focus on welfare rather than law enforcement (Lappi-Seppälä and Tonry, 2011; Snortum, 1983). The Nordic crime prevention model collaborates with actors outside of the justice system, as the system has the purpose of creating norms and morally educating, rather than to deter (Aromaa and Takala, 2005). Welfare plays a big part in crime prevention, but unlike other countries, welfare is seen as a basic right rather than having the strict purpose of preventing crime, such as equal opportunities for education (ibid.). Also, increasing the citizens' perceived safety is part of crime prevention strategies (ibid.).

Sweden has national crime prevention programmes, where an interdisciplinary approach with both situational and social crime prevention methods is needed (Takala, 2005). Further, measures on a local level and cooperation between actors in both the public and private sectors are seen as necessary (ibid.). Sweden has focused on social welfare policies as crime prevention (Lidskog and Persson, 2012). Sweden has its crime prevention agency Brå, the Swedish National Crime Prevention Council (Wikström and Torstensson, 1999). Brå has the purpose of increasing knowledge about crime and prevention methods, and to provide empirical evidence and a basis for decision-makers (Andersson, 2005). They also manage crime statistics and provide local crime prevention councils with knowledge (ibid.). Crime prevention often takes place on the local level, as Gustafsson (2014) argues that crimes are local. However, with fraud, this becomes muddled as it is not local in the same way as a street crime. It makes it more difficult to apply traditional crime prevention to cybercrimes (Wall, 2002).

3.1.3 Fraud and cybercrime prevention

Dorminey et al., (2012) state that there are three parts to anti-fraud measures: prevention, deterrence, and detection. Prevention aims to reduce opportunities for committing fraud, deterrence aims to deter an individual from committing fraud by catering to their fear of getting caught and/or getting punished, and detection aims to discover crimes that are happening or have already happened (ibid.). Detection can carries out via technical means, such as monitoring customer activity (Bodker et al., 2023). Fraud detection is different from

fraud prevention, which notices the fraud after it has happened and aims to mitigate its effects (Rodrigues et al, 2022). Fraud detection is based on machine learning and data analysis which learns from customers' behavioural patterns to notice deviant transactions, and from the usual modus operandi in fraud to uncover fraud as it is happening (ibid.). Abdallah et al (2016) argue that prevention is not enough to combat fraud, and that detection is just as important. Bolton and Hand (2002) claim that fraud detection must be used continuously as it is not possible to know when fraud prevention works because if it is effective, there are no visible signs, while if it fails it is visible when the crime is happening. Deterrence means that the threat of punishment or detection deters someone from committing a crime (Schneider, 2019). Schneider (2019) found that a higher clearance rate of fraud crimes had a deterring effect. The clearance rates for fraud crimes are generally low, however, which does not work in favour of fraud deterrence (Fjelkegård and Horgby, 2023).

However, not all prevention techniques are easily categorized. A common advice for romance fraud is to verify the person's identity by reverse-image searching or searching for info about the person they are in contact with (Cross, 2022). This requires victims to act to reduce their victimisation. Examples of victim-focused crime prevention efforts were found by Cross (2016), who examined different projects in Australia that identified potential victims through financial intelligence and sent them letters to inform and interrupt victimization. Cross found reduced financial losses and victimization in conjunction with the programme. This is another example of using information to prevent fraud. Leclerc and Morgenthaler (2023) outlined a crime script for fraud. A crime script details all the steps in a modus operandi and where an intervention could be suitable, which can be used for prevention purposes (Junger, Wang and Schlömer 2020; Leclerc and Morgenthaler, 2023). The script for internet fraud usually depends on other people helping the perpetrators, such as money mules or someone calling the victim (Leclerc and Morgenthaler, 2023). Leclerc and Morgenthaler (2023) claim that public awareness campaigns should be important to reduce victimisation, seeing to the fraud script. Also, information campaigns are a suggestion to reduce individuals becoming money mules by informing them about the modus and consequences (ibid.). These are examples of breaking the script, as by raising awareness, important actors like the victim and helpers are taken out, which prevents the crime.

Brenner (2004) calls for a collaborative approach to dealing with cybercrimes, as traditional law enforcement has its limitations when it comes to non-traditional crimes. With a collaborative approach, information exchange between different organisations and entities

becomes easier, and focus on training and education is important for preventing cybercrime (ibid.).

3.2 Risk management

Fraud can be seen as a modern risk within the context of the risk society. One of the key elements to risk society is the importance of experts and regulatory systems (Boudia and Jas, 2007). Risk management is a regulatory system and is a wide concept that includes, mostly in a business context, measuring and supervising potential risks (Wolke, 2017). A risk does not have a uniform definition in this context, but Wolke (2017) states that it is about potential losses or damages with no chance of potential gains. Wolke (2017) lines up a few reasons for the need for risk management, the first being a legal framework demanding it on a national and international level. Economic reasons come from the globalized market and risks coming from it, where Wolke (2017) exemplifies that the Euro was inadequately regulated and was one of the causes of a financial crisis when it was introduced. The last reason is technological advancements, where information flow is much faster, and products have shorter life cycles (ibid.). When it comes to fraud, it is a highly globalised type of crime that utilises technological advancements (Grabosky, 2001; Van Nguyen, 2022), which makes risk management relevant. Power (2013) argues that fraud prevention has become more important as a part of risk management, as crime is a risk that can bring losses. Fraud risk has changed from identifying fraudsters internally in an organisation to protecting the organisation from external threats such as cyberattacks (ibid.). Kummer et al (2014) found, in their study about non-profit organisations in New Zealand and Australia, that those organisations that had not experienced fraud mostly had ineffective measures. In contrast, when organisations got scammed, it kickstarted a process to implement risk management to prevent it from happening again (ibid.). In a sense, this underlines what Madah and Marzurki (2020) point out, that risk management should be in the culture, as the side-effect can be not seeing the danger until it appears. Most banks have transaction monitoring systems to screen transactions and manage risks, but they cannot catch all fraudulent transactions and must balance the need for protection with the need for effectiveness (Digital Fraud Committee, 2022; Wolke, 2017). Risk management is therefore connected to the risk society and fraud as it controls for threats coming from new technology. Risk management is very wide and encompassing, and it is not always clear who is the subject. In business science, the subject of risk management is the

company and its employees, but in other ways, the subject can be outside threats. In the context of fraud prevention, many different actors want to control the risk of fraud, which means that risk management is not just a concern for businesses, but for society as a whole. Understanding risk management is important to examine the management of risks within the framework of risk society and how strategies are employed to mitigate fraud risks.

4 Theories

4.1 Classic fraud theories

The fraud triangle is one of the most fundamental theories about fraud (Dorminey et al, 2012). It describes the conditions needed for a fraud crime to take place, with the parts being pressure, opportunity and rationalisation (ibid.). This theory has been a foundation for fraud research, but it has not been updated for cyber fraud, which makes it difficult to use. It is focused on the perpetrator, but a similar theory – the triangle of fraud action – describes the action itself, with the act, concealment, and conversion (Kagias et al, 2022; Mandal and Shanmugam, 2023). This triangle is more focused on the situational factors.

Even if these theories are relevant to fraud prevention, they do not serve a purpose to the aim of this thesis. This is because the aim is focused on the societal response, while the triangles are on the micro level, concerning individuals. Further, cyber fraud is complex and more focused on multiple technological, economic, and social factors, which makes it too complex to use with the classic fraud theories. Therefore, even though the fraud triangle has a long legacy of research, it is not suitable for this thesis. The theories used in this thesis are risk society, governmentality, and nodal governance. They are important to understand how the response to the rising occurrence of cyber fraud is connected to risks, how different sectors in society regulate and control behaviour connected to cyber fraud, and how these sectors and actors are governed by and govern others. These theories provide theoretical lenses which can examine the latent topics generated by the topic modelling as well as relations between documents, sectors and topics.

4.2 Risk society and security

Risk society is a concept coined by Beck (1992) and is about new risks that have been created by modern society and technology. Risk society concerns that risk is an inherent feature of contemporary societies and that modernity has created risks that humanity needs to control (Boudia and Jas, 2007). The difference between risk and older dangers such as natural disasters is that risks are connected to modernity and hazards posed by it (ibid.). Modern risks can be invisible ones, such as pollution and radioactivity, and while the concept of risk is connected to the distribution of wealth and social disadvantage, it does not always affect the lower strata of society (ibid.). Curran (2013) criticises that Beck does not take class into

account in the theory. He argues that the risk is unevenly distributed and that the risks are fundamentally different depending on the one who is experiencing it. Beck does acknowledge that risk affects people differently, but he still argues that they are fundamentally the same (Lupton, 2013). Risk and prevention are based on knowledge of the past (Adam and van Loon, 2000), and knowledge is tied to the cultural context, which Beck (2000) claims explains the difference in risk management around the world. However, knowledge, such as data, can be distorted and hidden, which gives the people who have the knowledge power over risk management (Lupton, 2013). For example, the wealthy can use their knowledge of the financial market to make more informed and less risky decisions (Curran, 2013). In a way, financial risks are less important to wealthy people as they have more resources. Boudia and Jas (2007) argue that good governance involves including citizens of different backgrounds and having an open dialogue for decision-making, echoing Curran's statements.

Risks can be exploited for economic advantage, as they can create an industry of controlling them, but new risks will always appear by the nature of modernity, meaning that risks will be infinitely produced (Boudia and Jas, 2007). Risks are nowadays seen as a political issue, as they have broader implications for society (ibid.). This means that the risk society is controlling catastrophes, which has political potential and implies that the risk society can rearrange power and authority (ibid.). Risks are politically reflexive, meaning that social actors can examine and reassess their position in society as well as how it is related to other actors (Banakar, 2015; Beck, 1992). The reflexive modernity that Beck proposes entails assessing one's position and knowledge about risk (Lupton, 2013). Globalisation is enhanced by digital technology, and because of the transnational nature of the technology, it makes risk management more difficult (ibid.). Social relations in modern society are stable and secure over time, which means that policymakers can use this to direct and shape social development by regulating individual and collective behaviour (Banakar, 2015). However, these traditional social structures that regulate behaviour are not as important in the digital age, and they shape social interaction in a different way than daily real-life situations (Lupton, 2013). The reactive law enforcement has not kept up with this development, making fighting cybercrime more difficult (Banakar, 2015; Brenner, 2004). Banakar (2015) argues that early modernity's regulatory mechanisms such as formal policy-making and informal regulation via norms are ineffective in regulating a transnational population. The nation-state was the one seen as responsible for security, but contemporary society has involved more actors in risk and

security (Zedner, 2009). Transnational organised crime poses a threat, which blurs the line between national and international security (ibid.).

Governance of security needs to be handled carefully, Zedner (2009) argues. It needs defined boundaries to not abuse power (ibid.). Further, Zedner (2009) brings up the principle of minimalism – the least number of burdens and measures in response to a question as possible – and the principle of social defence – meaning that the society has a collective responsibility to take care of security – and that they need to be balanced. Too minimalistic measures can lead to people not thinking the government is taking enough responsibility for security, which can lead to the people taking matters into their own hands which might bring dangers of vigilantism and search for vengeance. Zedner (2008) advocates that the principles of transparency, accountability and proportionality should be taken into consideration when policies get implemented and that an ethical framework should be in place to make sure that anyone with a disadvantage does not get negatively affected by the policy. Deflem (2008) argues that for a law or a policy to be accepted it must have legitimacy – being accepted by the wider society – and it needs to be administered according to pre-determined processes – i.e. having legality. This refers back to Zedner's (2009) principles of just social policies.

The main objective of crime policies can be boiled down as protection from unforeseen threats (Deflem, 2008). The traditional policing model is reactive, but crime policies have become more focused on prevention and risk management to control the future in a way, and the concept of security has become increasingly more important (Brenner, 2004; Zedner, 2009). However, security means widely different things to different actors, but the concept promises while risk threatens, which makes it a powerful political statement (Zedner, 2003). Crime prevention is closely connected to security, as crime prevention is seen as a way to combat social problems and increase perceived security (Törrönen and Korander, 2005). As society has changed, so have the ways to prevent crime. The shift is from the traditional form of reactively enforcing control, to proactively managing the risks (Banakar, 2014). Brenner (2004) raises some concerns about the risk and security concepts: where is it reasonable to put the responsibility on the individual for their own safety? Brenner (2004) argues that the Law enforcement cannot possibly take care of every single crime and that individuals should be educated and use their common sense to avoid victimisation. People's expectations can be altered slightly to make them understand that engaging in certain actions comes with certain risks, and these risks are something law enforcement cannot always help with (ibid.). Therefore, risk and security are issues for the state and other macro actors, as well as for

individuals. The problem lies in that not all individuals can be assumed to correctly estimate the risk of certain actions, which can differ for adults and children, as well as technical literacy and disabilities (ibid.). Brenner (2004) also brings up the example of websites – e.g. webshops – getting hacked and leaking credit card details, and questions how this indirect victimisation could be applied to customers. Would the webshop be responsible for the customers' victimisation or would the risk be the responsibility of the victims (ibid.)? There is no clear answer to this question, and it shows the complexity of risk and security.

4.3 Governmentality

Governmentality is a concept coined by Foucault, but like his other theoretical concepts, he refrained from giving a concrete definition (Valverde, 2017). The concept is more focused on managing risks and has governance that is less intrusive (ibid.). Power is inherent in governmentality and many of Foucault's concepts, but the concept of power does not mean a hierarchical, top-down and repressive form (Madsen, 2014). Rather, it means that power exists in the relationships between subjects, as power cannot exist without resistance (Barker, 1998). According to Foucault, knowledge and power are integrated, where power is based on knowledge but can also shape knowledge (ibid.). Knowledge can legitimize actions, and in public policies it can find ways to control the population, giving more power (Torres-Cuello and Pinzon-Salcedo, 2023). Normalisation for Foucault is a process where governance techniques become aligned with the norm and widely accepted, in other words, they become normal for the subjects they are governing (Lawlor and Nale, 2014). Disciplinary techniques need a system where it can identify the deviant actions to be able to pinpoint what the norm "should be" (ibid.). This implies that power and knowledge are interconnected and can form practices that become normalised.

Foucault came up with the concept of a "conduct of conduct", which is a way to govern subjects in a subtle, empowering way, like providing autonomy and self-realization (Madsen, 2014). Governing is not just the government telling its citizens what to do, it can be self-governance, where Valverde (2017) has the example of tracking calories. Li (2007) argues that governmentality can be seen as the will to improve the welfare of the population. This can be extended to organisations as well, where they in a way become responsible for crime prevention. Responsibilisation is a concept that arose in crime prevention during the 1980s which entails that ordinary people become more responsible for crime prevention (Valverde,

2017). This can be exemplified by employers asking their employees to keep track of visitors and unauthorised people to keep the safety of the company, rather than hiring security guards (ibid.). Governmentality can be a form of risk management, as it wants to govern future behaviour and in that sense reduce undesired behaviour (ibid.).

Foucault mentioned technologies in the sense that it is not only material technology, such as machines, but that they can be technologies of power, such as mechanisms that affect subjects' behaviour (Lemke, 2021). Technologies can produce new knowledge, such as Matthewman's (2013) example that a new technology such as a stethoscope produces more knowledge about the human body. This knowledge can also produce more power, as they are interconnected (ibid.). Other technologies are those of production, which produces or affects material things, and technologies of the self, which is connected to self-governance and affecting one's own life (Lemke, 2021). Technologies of security has its standpoint in reality as it is, rather than as it should be, and seeks to control the uncertain future (ibid.). Fraud prevention has different types of technologies used, from regulatory technologies of power to technologies of production such as transaction monitoring and surveillance.

4.4 Nodal governance

As evidenced by the above literature review, collaboration and networking are key to preventing crime. Policy networks have a more horizontal hierarchy and are inter-organisational, meaning that different stakeholders collaborate to formulate and implement policies (Marin and Mayntz, 1992). That is not to say that the network does not have any hierarchical structures at all, rather, they are dependent on power relations (ibid.). The policies do not have to be public ones, but they can be company policies and similar (ibid.). This can be seen from the perspective of decentralized social organisation and governance, which entails that society is not governed by one power, i.e. the state, but rather by multiple actors collaborating (Kenis and Schneider, 1992). Burris et al (2004:2) define governance as "the management of the course of events in a social system", and it is a complex task on all social levels. With nodal governance, the network aspect is enhanced, as Burris et al (2004) argue that governance is happening in nodes, where the actors have different sets of skills and capacities. Nodal governance can unveil how nodes are governed and govern others (ibid.).

In nodal governance, the traditional idea of the state solely overseeing governance is challenged (Holley and Shearing, 2017). Banks can be an example of a third-party actor who

is carrying out state-mandated directives (Wood and Shearing, 2006), which in this case would be fraud prevention. Third-party in this context means that the actors are not directly involved in criminal activities – neither policing nor committing – but still play a part in preventing them (ibid.). There are stronger and weaker actors with varying levels of power, but as Wood and Shearing (2006) claim, the weaker actors can mobilize their resources to gain more governance power. Shearing and Wood (2003) argue that a nodal framework does not put any actor in front and centre, rather, it is a method to examine which actors are important in a network without assuming that the state holds the most power. A nodal approach can reveal the complexity of governance at the same time as it cultivates new thinking and approaches (Holley and Shearing, 2017). Nodal governance can uncover relations and dependencies between sectors, and with topic modelling, the types of topics that are more important for some sectors can be put in relation to the other sectors more easily.

Fraud blurs the line between national and international risk management. Even within a nation-state, space is irrelevant, and the scope is so much bigger compared with traditional crimes, which makes investigation difficult. Further, within a nation-state cooperation is necessary, but the literature does not give an overview of how it works with fraud.

5 Methodology

5.1 Ontology

As evidenced by the literature review and theories, collaboration is very important in fraud prevention, and the extent of it cannot be measured by just examining one actor. Therefore, it is important to examine several actors from different sectors, and how they address the rising occurrence of cyber fraud. To pursue this aim, a larger number of documents from different actors must be analysed, which is difficult with manual document analysis. This requires an automated method. Thus, this thesis takes a mixed-methods approach with a computational method called topic modelling using Natural Language Processing (NLP). This method generates topics and can answer how cyber fraud is addressed in political documents (RQ 1) and how different sectors handle the issue (RQ 2), which will lead into a discussion about how they are connected to the principles of risk society, governmentality, and nodal governance (RQ 3). This approach can be more objective, replicable, and can reveal insights a human cannot make (Silveira et al., 2021). This method is fairly unexplored in the field of sociology of law, which also makes this a novel approach and an exploration of a new method.

5.2 Sampling

A document can have different definitions depending on the epistemological stance (Karppinen, and Moe, 2011). Lewis-Beck et al (2004) state that a document can have a wide range of ways it is produced and published, such as text, audio, or visual mediums. Karppinen and Moe (2019) mention policy and industrial documents. These kinds of documents are very broad, as Karppinen and Moe (2019) state that they encompass legal texts and reports, to press releases and opinion pieces. The authors of these documents can be anyone from the public sector or from the private sector, where the documents have the purpose of influencing the process of public decision-making (ibid.). Becker and Bryman (2012) state that policy is a vague concept that has a few defining characteristics. It describes past, present and/or future actions or practices (ibid.). It carries a commitment to these actions and practices, and it is usually focused on a specific problem (ibid.). According to Shafritz (2004), a policy is “a standing decision by an authoritative source such as a government, company or head of a family” (pg. 221). Shafritz (2004) also mentions that policies can be more general goals to be

achieved. In that sense, a policy is quite broad and does not only concern the government. In this thesis, a political document regarding fraud is defined as a document that outlines past, present and/or future actions or information within cyber fraud prevention for the organisation that has authored them. These kinds of documents are used as the data source as they can describe how different actors respond to and address cyber fraud and can be used to contextualise the problem.

This thesis has purposive sampling, which means that the researcher chooses the data they think would give the most relevant and rich information, based on theory, literature review and previous knowledge (Drisko and Maschi, 2015). The key actors are those actors that can be assumed to have a say in the matter (Spicker, 2006), in this case, fraud prevention in Sweden. The sampling of this thesis identifies relevant stakeholders regarding fraud and targets the political documents they have published. A stakeholder is affected by or participates in decisions, and in this thesis, they are one of the categories Spicker (2006) mentions, namely those who are directly involved in policymaking. According to Wilson and Laidlaw (2017), nodal governance within cybercrime can be thought as of having three sectors: governmental, business, and civil. Also, as evidenced by the literature, banks are relevant stakeholders for fraud prevention. Therefore, the sample is divided into the governmental sector, banking sector, business sector, and civil sector, as those are the overall stakeholders (Shannon et al, 2016; Wilson and Laidlaw, 2017). For specific organisations, this thesis will take inspiration from the organisations used in referrals regarding fraud. When a law is made, it goes out on referral so that relevant organisations can give their opinions on the proposal (Regeringen, 2018). The sample will also use the national organisations, as the top-down, overall perspective is of interest. Regeringskansliet (Fi2022/00489) had a referral regarding a law about tightening security with online payments, where governmental, business, and civil sectors could answer. Regeringskansliet (Fi2016/00114) had a referral regarding the implementation of EU fraud prevention rules, and the actors were mostly in the public sector in this.

For the government sector, the key actors are the Swedish government, Finansinspektionen, and Polismyndigheten as those authorities are the most important ones for crime prevention. The biggest banks in Sweden are, according to Svenska Bankföreningen (2023), Swedbank, Handelsbanken, Nordea, and SEB. Other financial institutes such as Klarna and Walley are included as well. Documents from Svårlurad, an informational campaign from Swedish banks about fraud, have been included as well. From the business sector, the national organisations

Svenskt Näringsliv and Svensk Handel have been chosen as they are gathering and organising businesses in Sweden. From the civil sector, the consumer association Sveriges Konsumenter and the retirement organisation Pensionärernas riksorganisation have been chosen as they can bring the perspectives of consumers and elderly people. Villaägarna is an association gathering homeowners in Sweden and provides another perspective from the civil sector.

The documents were taken from each organisation's website, where the search function was used with the keyword “bedrägeri” (“fraud”), and the result was screened for documents about cyber fraud and past, present and/or future actions regarding the prevention of it. The organisations were contacted as well to find out if they had some policy documents that were not available online.

5.3 Topic modelling

5.3.1 What is topic modelling?

This thesis uses Natural Language Processing (NLP) to analyse and extract meanings from the text. NLP is a form of Machine Learning and can computationally analyse natural human language (Ghavami, 2020). Jin and Mihalcea (2023) claim that NLP can be used for interpreting political decisions, where it can analyse and extract topics regarding political agendas and factors the policy responds to, which is relevant to analysing political documents regarding fraud. This type of content analysis is mixed methods, drawing from both qualitative and quantitative approaches. Text is unstructured data but can be transformed into structured data, i.e. quantitative data, using different approaches (Drisko and Maschi, 2015). The approach used in this thesis is topic modelling, which will be described below.

Topic modelling is used to find themes or subjects in texts and can reveal implicit meanings and relations between documents (Silveira et al., 2021). Topic modelling is unsupervised machine learning and uses statistical text analysis, where the words are classified and grouped depending on how and with which words they appear (Brett, 2012; Campesato, 2022). This results in topics, which are recurring terms in the text (Brett, 2012). Topic Modelling is a form of distant reading, which shows broader patterns with a larger dataset, while close reading focuses on a smaller corpus (Boyd-Graber, et al, 2017; Drakman, 2022). The goal is to gain insights from broader patterns, but with the limitations that it fails to capture nuanced meanings and contexts (Sekar, 2024). In this thesis, the aim is to find the broader themes

regarding fraud from a larger number of actors, which makes this method suitable. Topic modelling is inductive as it generates the themes from the text instead of from pre-defined categories and theories (Ignatow and Mihalcea, 2017). It can be useful for extracting data from texts, as it trains itself on the texts – called corpus – to classify entities in it (Hardeniya et al, 2016). It still requires interpretation afterwards but can be useful as the researcher's interpretation during the analysis phase might lead to important topics being rejected (Ignatow and Mihalcea, 2017). Topic modelling can reduce human bias, and Boyd-Graber et al (2017) argue that it can be used as a tool for applying grounded theory effectively. Topic modelling has several different algorithms and methods. In this thesis, Latent Dirichlet Allocation (LDA) has been chosen, as it is one the most commonly used topic modelling algorithms. It is also the most suitable for the number of documents collected, as some algorithms require a large corpus to be able to give an output, while LDA does not.

5.3.2 Latent Dirichlet Allocation

LDA is one technique for topic modelling, which sees a topic as being a mix of words, and a document as consisting of a mix of topics (Jin and Mihalcea, 2023). LDA is based on the "bag of words" approach, which entails extracting all words from the text and putting them into a matrix, which displays the words with the number of times they occur across the corpus (Dyevre, 2021; Vignoli, 2022). The bag of words focuses on the frequencies of words, and this assumes that the more frequent a word appears, the more important it is to the current topic (Asmussen and Møller, 2019). LDA calculates the probability that a certain word will belong to a certain topic and that a certain topic will belong to a certain document (Dyevre, 2021). This probability guides the output of the analysis, as it show what is deemed as important or not.

Gan and Qi (2021) state that a good topic model should have low perplexity, meaning that it should have a good predictive ability. In this case, this means that the model should be able to predict which topic belongs to which document (ibid.). The closer to 0 the perplexity score is, the higher the predictive power the model has (Blei and Lafferty, 2007; Tijare and Rani, 2020). It should have a low similarity between the topics, to avoid duplicate topics, and the model should be stable, meaning that it can be replicable (Gan and Qi, 2021). Further, to evaluate the model, coherence scores are used, which analyses the semantic similarity between words in and between topics (Pickett et al, 2020). It outputs a value between 0 and 1, and the higher the score, the more similar the words are to each other (Pickett et al, 2020; Tran, 2023). LDA requires the researcher to set the number of topics, which poses an issue as

too many topics create duplicates, overfitting the model, and too few get over-saturated, underfitting the model (Gan and Qi, 2021; Tijare and Rani, 2020). Overfitting means that the model is fitted too much to the data and captures everything, rather than the important themes, while underfitting does not manage to capture the data properly (Richert et al, 2013). There are different suggestions on how to choose the topics, as Asmussen and Møller (2019) suggest examining the perplexity for multiple numbers of topics and choosing the lowest one, while Prabhakaran (2018) suggests examining the coherence score instead. This thesis will use both approaches along with a qualitative assessment of the topic output.

5.3.3 Procedure

This thesis uses the Python programming language (version 3.11.9). There are many different packages for topic modelling in Python, and this thesis uses Gensim by Řehůřek and Sojka (2010). This choice was made as it is more flexible and is within the limits of available processing power compared to other packages (ibid.).

Before the text can be handled by Python, it must be pre-processed (Hardeniya et al, 2016). The documents are in the form of PDF files, and they are processed using the PyPDF2 module to be able to be handled by Python (Fenniak et al, 2022). Then the text is split into tokens. Tokens are the smallest part the program can handle, and here they take the form of individual words (Hardeniya et al, 2016). From these tokens, the text is cleaned in multiple steps. Punctuation and numerical characters are removed, and the words are converted to lowercase to not separate the same words that have different cases. Then the texts go through lemmatisation, meaning that each word gets reduced to its stem, to not differentiate between e.g. the words “eat” and “eaten” (ibid.). This thesis uses Simplemma by Barbaresi (2023), which has multilingual support. The next step is removing stop words. Stop words are common words that bear little to no meaning, and these words can be articles and conjunctions, among others (Hardeniya et al, 2016). They are removed to enhance the interpretability of the result and focus on the meaningful terms (ibid.). The Python package Natural Language Tool Kit (NLTK) comes already with a pre-built list of stop words for the Swedish language, with the opportunity to add more (Bird, et al, 2009). Stop words can be domain-specific and it depends on the nature of the documents what will or will not add to the result (Srinivasa-Desikan, 2018). Words like “bedrägeri” (“fraud”) were added to the stop list as it is an overarching topic and would not give any useful information to the result. Further, words that are smaller than 3 characters and only consisting of numbers were removed, as they would not provide anything to the result. The cleaning is an iterative process, where the

process must be evaluated so that only meaningful entities are left in the text, but this cannot be measured statistically and requires the researcher's judgement (ibid.).

After the pre-processing, the tokens get put into a dictionary, where each token gets a numerical value (Srinivasa-Desikan, 2018). The number is the word's word ID, which is used easier for modelling purposes (ibid.). Extreme words, i.e. words that appear very often or very seldom, were filtered from the dictionary as they would likely not add anything to the result (Asmussen and Møller, 2019). Then, the bag of words is created, which uses the dictionary to create a list of each word ID with its corresponding word count for each document in the corpus (Srinivasa-Desikan, 2018). A term frequency – inverse document frequency (TF-IDF) matrix is then constructed, which, in short, calculates how often a word appears in a document combined with how important it is across the whole corpus (Srinivasa-Desikan, 2018). TF informs how often a word appears in a document, and the IDF part gives information about how rare or common a word is across the corpus (Herdeniya et al, 2016; Srinivasa-Desikan, 2018). The TF-IDF matrix is useful when documents have varying word counts (Herdeniya et al, 2016), as the documents have in this case.

Then, the number of topics is decided upon. This is done by examining the coherence score when running the model with different numbers of topics, as inspired by Prabhakaran (2018). The code was modified to include perplexity scores and the top keywords for each topic for a qualitative assessment. When the number of topics has been decided, the LDA model for topic modelling is constructed. The LDA model itself has a few different parameters. The "passes" parameter defines how many times the algorithm goes over the corpus (Řehůřek and Sojka, 2010). Too many can lead to overfitting and too few to underfitting. The documentation of Gensim suggests testing different values of passes and examining the logs if most of the documents have converged (Řehůřek and Sojka, 2010). When the parameters have been set, the model can be run. The model trains itself on the corpus and outputs the topics. These topics are not named and consist of keywords (Dyevre, 2021). The keywords have scores showing the likelihood of belonging to the topic (Pickett et al, 2020). Further, a document-topic matrix is constructed to show which documents are more likely to have which topics (ibid.).

5.4 Reliability and validity

Computationally analysing texts does have its challenges, because a word can be interpreted in many ways and therefore can never speak for itself (Hecking and Leydesdorff, 2019). The bag of words approach has the side effect of stripping the word of its context (Vignoli, 2022), which affects the validity of the method. LDA can understand that if a word seems ambiguous as to which topic it should belong to but appears along with other words in a certain topic, it will be inferred that the word belongs to that topic (Boyd-Graber, et al, 2017). However, this is not a method that will be perfect by any means. It is used when it is not feasible to go through a large number of documents manually, and a manual document analysis has its issues with bias and subjective coding. Topic modelling has the advantage of being replicable compared with manual document analysis (Vignoli, 2022). It produces the same result with the same corpus, but especially LDA is very sensitive towards changes in the data, i.e., if documents are deleted from or added to the corpus (Hecking and Leydesdorff, 2019), which affects the replicability. The documents used are outlined in Appendix 1 and the Python code used for the analysis is in Appendix 3, to make the procedure transparent and replicable.

In this thesis, no data was generated or collected. Rather, the documents were selected, which Bowen (2009) claims make analysis more effective and faster. The documents have not been constructed for research purposes, which defies the concerns about researcher bias and reflexivity, as they are not affected by the presence of the researcher (ibid.). However, the fact that documents are not produced for research purposes can pose an issue when they cannot be adapted to the needs of the research design (ibid.). Further, not all documents are available and can be retrieved, and selection bias can be very much an issue here too (Bowen, 2009; Karppinen and Moe, 2019). This thesis has selected documents from a variety of sources, but it cannot be fully ruled out that some documents were not available on websites, and some actors did not answer the request for them. In Sweden, there is the principle of publicity (offentlighetsprincipen) that states that documents in public agencies should be publicly available (Bohlin, 2010), which infers that there would be more documents from the governmental sector. The other sectors do not have the same demands, and the banking sector has further regulations regarding secrecy. The uneven number of documents might pose an issue in the results, but the topic probabilities can also be standardised so they can be compared across sectors. Further, not all documents were exclusively about cyber fraud, which might affect the result. Nevertheless, the topic model should be able to generate relevant topics as fraud is the overarching theme of all documents. To combat overfitting and

underfitting, the number of topics is chosen using quantitative and qualitative metrics, to not choose an arbitrary number that is not connected to the dataset.

Qualitative research assumes that everyone has their subjectivities and values, and this cannot be fully removed from the research (Auerbach and Silverstein, 2003). Reflexivity is therefore important, as the researcher reflects on their own standpoints and how it may have affected the result and the research as a whole (ibid.). Because this thesis has a mixed methods approach, reflexivity is just as important here. The author of this thesis has worked with fraud prevention in a bank, which may affect the perception of fraud prevention politics. The bank perspective paints one picture, while other perspectives focus on other factors. Selection bias cannot be fully ruled out, as the author has their opinions on the subject, and full neutrality cannot be reached.

5.5 Ethics

The documents are publicly available on the internet, and in that sense, there are no ethical obstacles to using them. Confirmation bias can be an issue for the analysis, when a researcher is looking for the result they expect (Bos, 2020). To combat this, transparency and reflexivity are important (ibid.). The author's standpoint has been described in the previous part, and the method is described as transparent as possible. The documents are presented in Appendix 1, but this may pose an issue if the results can be assumed to be from certain documents, and therefore certain organisations, which may or may not reflect negatively on them depending on the nature of the result. However, the result is not going to point to any specific actor, as it will discuss the sectors on an aggregated level.

6 Results and analysis

6.1 Structure

First, some descriptive statistics such as word count and number of documents are presented. Then, the choices for the parameters of the model are outlined. The research questions are answered in the following parts, where RQ1 is answered by the topics section and RQ2 is answered by the "Sectors" section. This will form the foundation for the discussion, which will answer how the topics and sectors reflect the principles of risk society, governmentality, and nodal governance (RQ3).

6.2 Descriptive statistics of documents

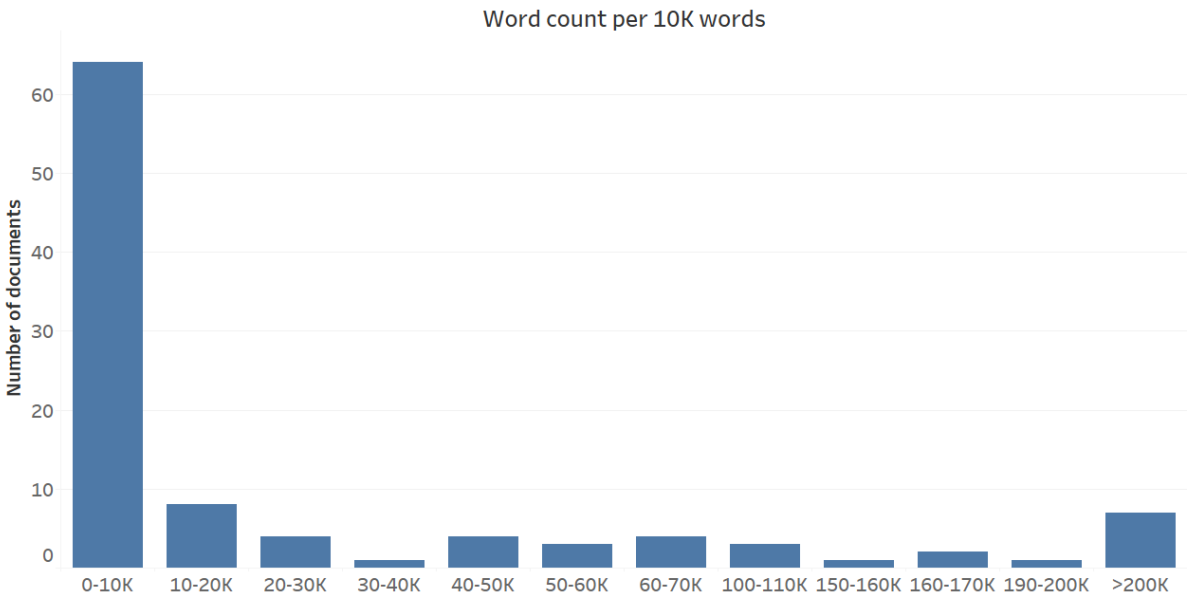

Figure 1: number of documents distributed over word count per 10,000 words.

Figure 1 shows that the majority, 63 %, have less than 10,000 words. The documents that consist of 10-20,000 words are slightly higher, 8 %, than the following categories, and the rest have similar numbers of documents. Around 7 % of the documents have more than 200,000 words.

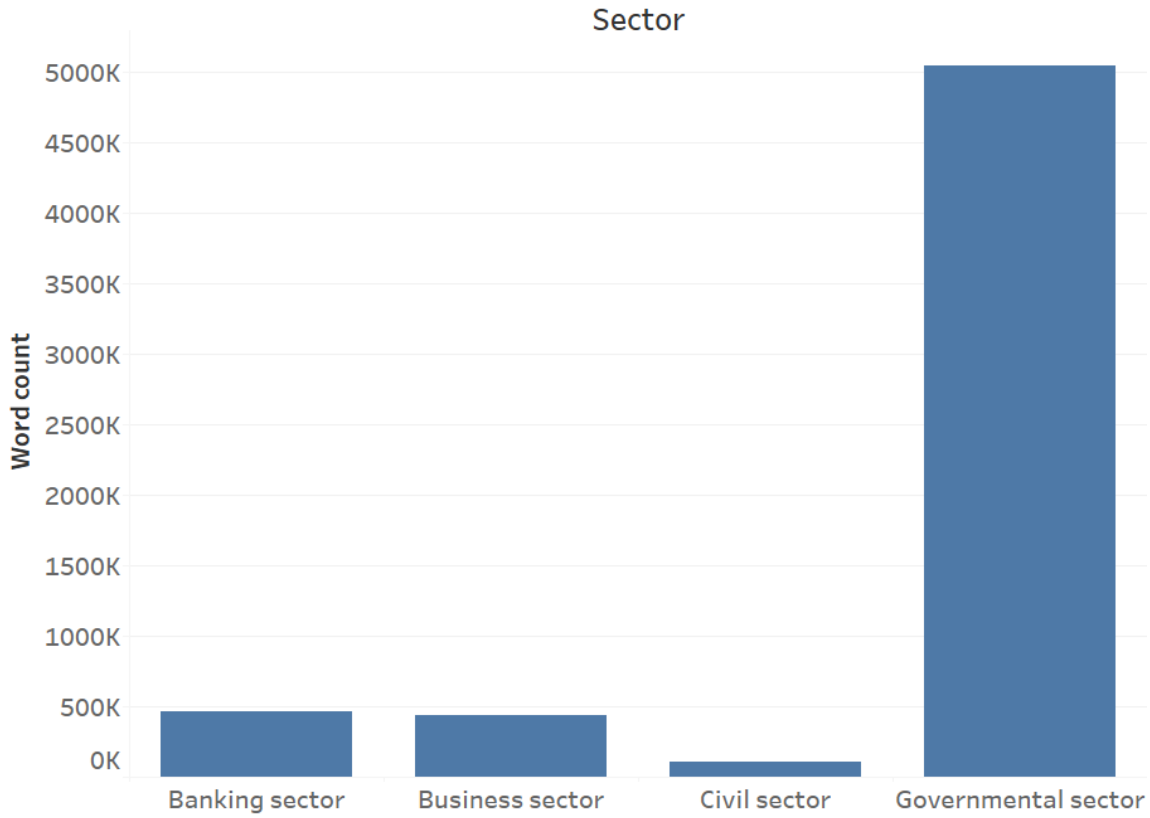

Figure 2: word count per sector

In Figure 2, the governmental sector has the absolute highest sum of word count, with the other sectors being far behind.

Number of documents per sector	
Banking sector	30
Business sector	18
Civil sector	11
Governmental sector	43
Total	102

Table 1: number of documents per sector

Table 1 shows the number of documents in each sector. The sample is unevenly distributed, with the governmental sector having the highest number of documents while the civil sector has the lowest. as the governmental sector has 4 times as many documents as the civil sector, while the banking sector has ¾ of what the governmental sector. The number of documents is not as skewed as the word count, where the governmental sector has around 50 times more words than the civil sector, as seen in Figure 2.

6.3 Parameters for the LDA model

According to the logs, 101/102 documents were converged with 6 passes. Therefore, that parameter was chosen. Next, the number of topics was decided.

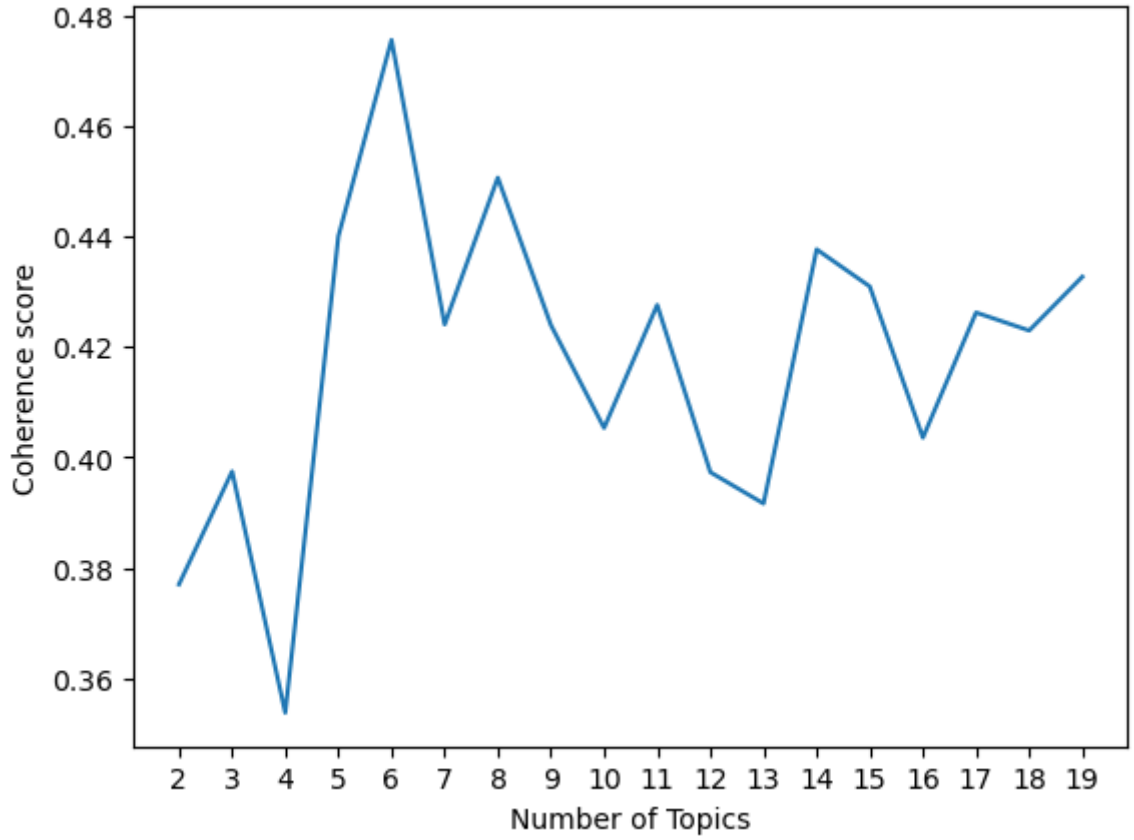

Figure 3: Coherence score for each number of topics.

Number of Topics	Coherence value
6	0,476
7	0,424
8	0,451
14	0,438

Table 2: The topics with the highest coherence scores.

The coherence value in Figure 3 is at the highest between 6 and 8 topics, or with 16 topics. The numerical values in Table 2 show that the 6-topic model has the highest coherence value, followed by 8, 14, and 7 topics.

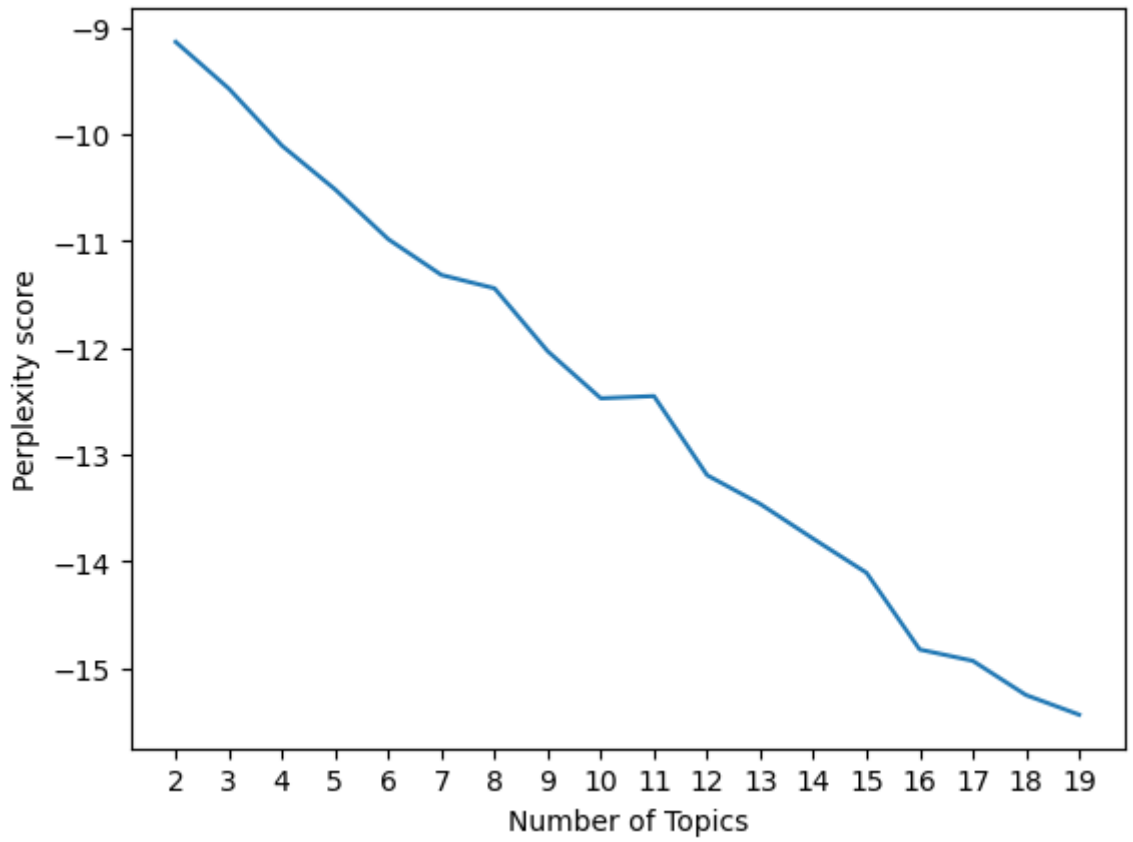

Figure 4: Perplexity score for each number of topics.

The perplexity values are not helpful, as they steadily decrease, except for two parts where there is a slight rise. The number should be as close to 0 as possible (Tijare and Rani, 2020), which the 1-topic is. However, that model cannot bring any substantial results, and the topics with over 10 topics have a higher risk of being overfitted, as seen by the coherence values and qualitative analysis.

Qualitatively assessing the topic outputs (Appendix 2) shows that the model with 6 topics has the most substantial topics. Seeing to the perplexity values in Figure 4, they indicate that higher number of topics give less predictive ability, which would also indicate overfitting. In this thesis, the broad picture is the most important, and with more topics, it becomes too specific. The 6-topic model has a balance between too broad and too narrow themes. The model will therefore be used with 6 topics and 6 passes.

6.4 Topics

6.4.1 Generated topics

This section aims to answer research question 1 – *Which topics regarding cyber fraud in Sweden can be identified in political documents?* – by outlining the topics generated by the LDA model and interpreting them. Further, this section will show how the topics are correlated and how the probability is distributed among the corpus, which will provide more understanding of the response to cyber fraud as the former can describe how similar or different the topics are, and the latter will describe which topics are seen as more important.

Topic 1 *Rules and Regulations on an EU-level*		**Topic 2** *Information-based Fraud Prevention*		**Topic 3** *Banks' Responsibilities in Fraud Prevention*	
Keyword	*%*	*Keyword*	*%*	*Keyword*	*%*
medlemsstat	0,18	bankkort	0,14	villaägare	0,21
skall	0,16	närstående	0,12	bankbedrägerier	0,19
direktiv	0,15	påstå	0,12	kontoinformation	0,13
bestämmelse	0,14	prata	0,12	betaltjänstdirektivet	0,13
betalningsinstrument	0,13	kortbedrägeri	0,12	sparbank	0,13
förordning	0,12	gammal	0,12	betalning	0,12
behörig	0,11	samtal	0,12	stoppa	0,11
byrå	0,11	lur	0,11	betaltjänstlagen	0,11
bilaga	0,1	uppmaning	0,11	falsk	0,11
kund	0,1	svårlurad	0,1	bankbedrägerierna	0,1

Topic 4 *Discovering and Reporting Fraud*		**Topic 5** *Cost of Fraud*		**Topic 6** *Online Fraud Awareness and Prevention*	
Keyword	*%*	*Keyword*	*%*	*Keyword.*	*%*
besparing	0,1	kostnad	0,22	mejl	0,24
kopia	0,09	brottslighet	0,22	faktura	0,2
logga	0,08	procent	0,22	kortuppgifter	0,18
polisanmälan	0,07	konsument	0,19	säljare	0,18
stoppa	0,07	tillsyn	0,19	aldrig	0,17
agera	0,07	företag	0,17	tipsa	0,16
bankid	0,07	finansinspektion	0,16	bestrida	0,16
genomskåda	0,07	verksamhet	0,15	annons	0,16
uppmaning	0,07	stöld	0,13	lösenord	0,15
omöjlig	0,06	drabba	0,13	bankid	0,14

Table 3: The six topics with the 10 words that have the highest probability of belonging to the topic. The probability is given in %.

The topics are outlined with their top keywords in Table 3. There are some overlapping words, such as "bankid" in topics 4 and 6, but generally, the topics are independent. However,

the words themselves are related thematically, which may lead to overlap in other ways. The probabilities for the keywords belonging to Topic 4 are especially low compared to the other topics, which could suggest that the topic is thematically weaker.

Figure 5: Word clouds for each topic, showing the words with the highest probability of appearing.

The word cloud (Figure 5) shows the most important words in each topic. In topics 1, 3 and 5, the importance of words is more evenly distributed, as quite a few words have the same size. Topics 2 and 4 have fewer words that are more important than the others. Topic 6 has two words, “faktura” (“invoice”) and “mejl” (“email”), which are clearly more important seeing to the size.

6.4.2 Analysis of topics

Here, the topics from Table 3 will be interpreted and labelled. The interpretation starts with the keywords provided for each topic and uses the context and information from the literature to explain what each topic can mean.

Topic 1 – Rules and Regulations on an EU-level

Keywords: medlemsstat, skall, direktiv, bestämmelse, betalningsinstrument, förordning, behörig, byrå, bilaga, kund

The words in this topic could be related to rules and regulations, especially on the EU level. "medlemsskap", "direktiv", "skall" and "bestämmelse" are the most important words in this topic, as seen in the word cloud (Figure 5). "Medlemsstat" ("member state") could be a reference to being a member state in the EU, while "direktiv" ("directive"), "bestämmelse" ("regulation"), and "förordning" ("ordinance") are references to regulations. The keyword "skall" ("shall") describes something coercive, which would fit with regulations. The word "betalningsinstrument" ("payment tool") indicates that this topic references the EU directive PSD2, as that is one of the things the directive aimed to regulate (Steennot, 2018). "Byrå" ("bureau") and "kund" ("customer") could describe the one who is controlling and the one who is being controlled, respectively. "Bilaga" ("appendix") seems to be an outlier that just describes the content of the document.

Topic 2 – Information-based Fraud Prevention

Keywords: bankkort, närstående, påstå, prata, kortbedrägeri, gammal, samtal, lur, uppmaning, svårlurad

This topic could be summarised as spreading information to prevent victimisation, especially for elderly people. It contains words like "närstående" ("close relatives"), "påstå" ("claim"), "prata" ("talk"), which indicates civilians preventing fraud by talking to their relatives. This is further supported by the keywords "gammal" ("old"), as most people who are scammed are older (Fjelkegård and Horgby, 2023). The topic seems to focus on phone fraud, as the words "samtal" ("call"), "lur" ("handset"), and "uppmaning" ("urging") appear. Also, "kortbedrägeri" ("card fraud") seems to indicate that cards are involved in these kinds of scams. Lastly, the word "svårlurad" appears, which is the name of the fraud prevention campaign the Swedish banks held, directed towards elderly people. The most important words in this topic are "bankkort", "närstående", "påstå", closely followed by "prata", as seen in the

word cloud (Figure 5), which supports that the essence of the topic is spreading information and awareness to one's relatives to prevent victimisation.

Topic 3 – Banks' Responsibilities in Fraud Prevention

Keywords: villaägare, bankbedrägerier, kontoinformation, betaltjänstdirektivet, sparbank, betalning, stoppa, betaltjänstlagen, falsk, bankbedrägerierna

This topic seems to be more focused on which responsibility the banks have to prevent fraud and compensate victims. In the word cloud (Figure 5), the importance of the keywords is more evenly spread out compared to previous topics, with "villaägare" being the biggest, followed by "bankbedrägerier", "kontoinformation" and "betaltjänstdirektivet". "Betaltjänstdirektivet" ("the payment services directive") and "betaltjänstlagen" ("law of payment services") are the laws and regulations that regulate how and when victims should receive compensation from the bank (SFS 2010:751; Steennot, 2018). Coupled with the keywords "sparbank" ("savings bank"), "betalning" ("payment"), and "stoppa" ("stop"), this topic could refer to the banks' responsibilities regarding fraud. The first keyword "villaägare" ("homeowner") can be a reference to the civil sector, which could mean that this is especially important for that sector.

Topic 4 – Discovering and Reporting Fraud

Keywords: besparing, kopia, logga, polisanmälan, stoppa, agera, BankID, genomskåda, uppmaning, omöjlig

This topic is a bit more fragmented and difficult to label, but to summarise, this topic is about discovering and reporting fraud. The words "besparing" ("savings") and BankID (a digital identification tool commonly used in Sweden) indicate protecting oneself against fraud when it is coupled with the words "stoppa" ("stop") and "agera" ("act"). "Genomskåda" ("see through") and polisanmälan ("police report") can be connected to discovering and acting on the fraud. "Kopia" ("copy") and "logga" ("logo") do not seem relevant at first glance, even if they are the most important words according to Figure 5, but they can be connected to spotting fabrications in emails, documents and similar, and uncovering fraud.

Topic 5 – Cost of Fraud

Keywords: kostnad, brottslighet, procent, konsument, tillsyn, företag, finansinspektion, verksamhet, stöld, drabba

This topic is about the cost of fraud. It is explicitly stated in the keyword "kostnad" ("cost"), as well as the keywords "brottslighet" ("crime"), "stöld" ("theft"), and "drabba" ("affect").

“tillsyn” (“supervision”) and “finansinspektion” (a reference to the Swedish authority Finansinspektionen) can be a reference to the measures to combat fraud, but with the connotation to companies being the object of the measures. “konsument” (“consumer”), “företag” (“company”), as well as “verksamhet” (“operations”) can refer to the victims of fraud. “procent” can be a purely descriptive word that comes from the descriptions of fraud cases. The word cloud in Figure 5 shows that “kostnad”, “brottslighet”, “procent” and “konsument” are the most important words, where the first three words clearly outline what this topic is about, while the last keyword describes who is most often victimised.

Topic 6 - Online Fraud Awareness and Prevention

Keywords: mejl, faktura, kortuppgifter, säljare, aldrig, tipsa, bestrida, annons, lösenord, BankID

This topic is also a bit fragmented but concerns spreading awareness about online fraud to prevent victimisation. “mejl” (“e-mail”), “faktura” (“invoice”), “säljare” (“seller”), and “bestrida” (“dispute”) are references to invoice fraud, as they often come via email and must be disputed. “Mejl” could also be a reference to phishing emails. The word cloud (Figure 5) clearly shows that the most important words are “faktura” and “mejl”, which could support the reference to invoice fraud and awareness of phishing emails. “kortuppgifter” (“card details”), “annons” (“advertisement”), “lösenord” (“password”) and “BankID” can be grouped with the keyword “aldrig” (“never”) because this is information one should never give out to someone else. This topic seems to be about security breaches, but it is also a bit difficult to interpret as it seems to have a part of invoice fraud in it. The keywords “BankID” and “faktura” are related to bank services, which is relevant to how banks can manage risks by keeping track of their services.

6.4.3 Topic correlations

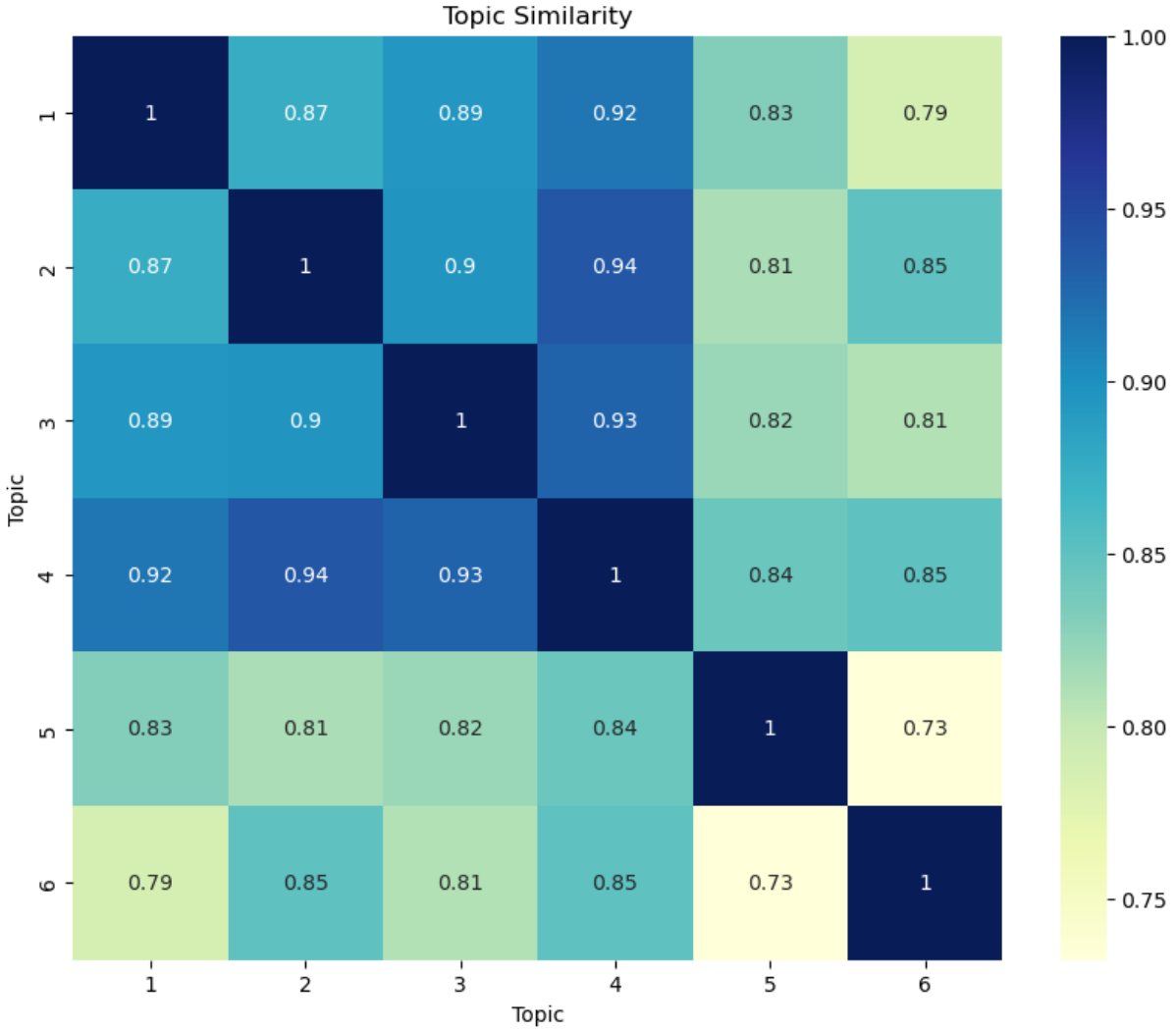

Figure 6: Heatmap showing the Cosine similarities between the topics.

The heatmap in Figure 6 shows the Cosine similarities between the topics, meaning how much the topics overlap. Almost all topics have an overlap, as the closer the value is to 1, the more it is overlapping. 5 and 6 are the topics with the least overlap, while topics 1 to 4 have the most overlap. This suggests that topics 1 to 4 are the most like each other while topics 5 and 6 are the least similar. Topics 5 and 6 are quite different from each other as the first one is about the cost of fraud and the latter is about awareness and detection of fraud. The first 4 topics seem very similar but still resulted in different topics. This might also be because are very much related and some overlap is to be expected. This overlap could also explain the fragmented nature of topic 4.

6.4.4 Document-topic-distribution

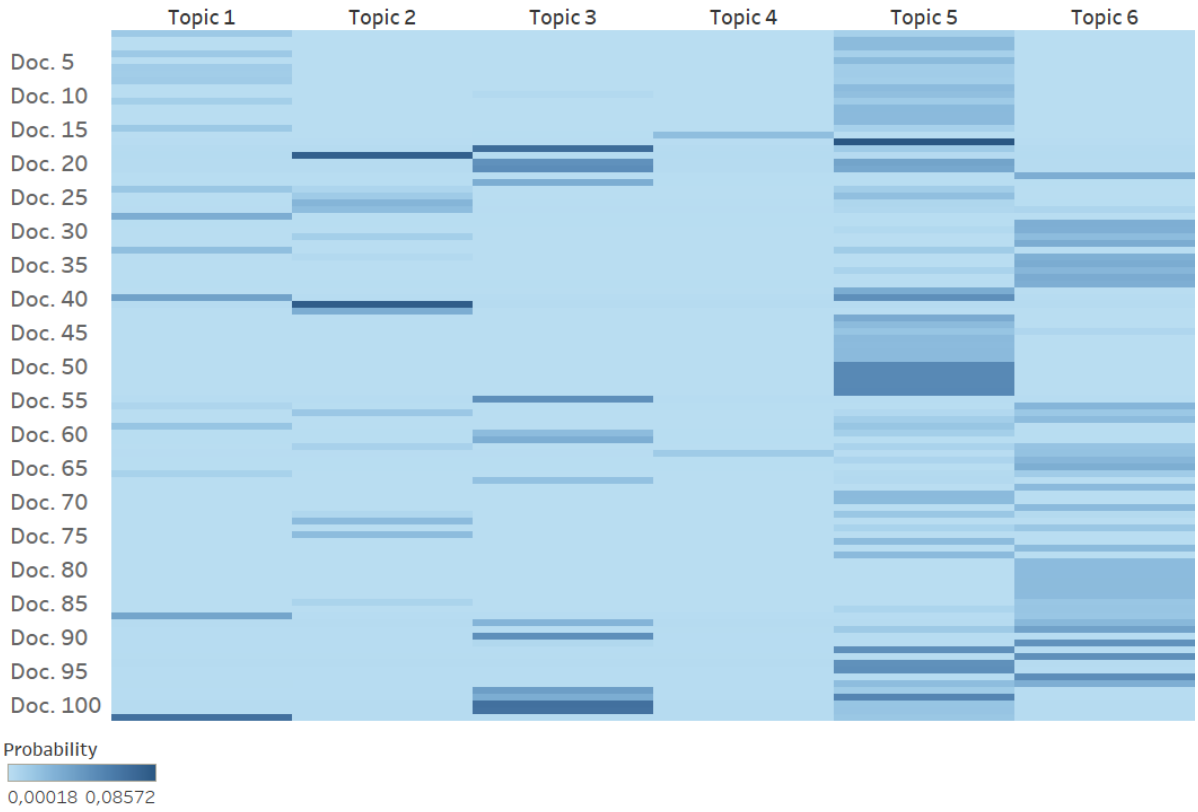

Figure 7: The probability for each modelled topic to appear in each document.

Figure 7 shows the probability of each topic appearing in each document. Topics 5 and 6 have the highest probabilities, while topic 4 has the lowest. This means that topics 5 and 6 are more widespread over the whole corpus, while topic 4 is not. This could, again, correspond with the fragmented nature of topic 4.

6.5 Sectors and topics

6.5.1 Sectors

This section outlines the relations between the banking, business, civil and governmental sectors and the topics. This will answer research question 2 – *Which similarities and/or differences can be found in the sectors in Sweden regarding topics about cyber fraud?* – and aims to explain which topics are more or less important in which sectors in relation to each other.

6.5.2 Sector-topic distribution

Figure 8: Relative importance of each topic for each sector.

Figure 8 shows the relative importance of each topic for each sector. It shows that topic 5 is dominant for business, civil and governmental sectors, while topic 6 is dominant for the banking sector. Topic 6 is the second most important topic in business and governmental sectors, but in the civil sector, topic 3 has almost as big of importance as topic 5 does, while topic 6 is one of the least important topics for that sector.

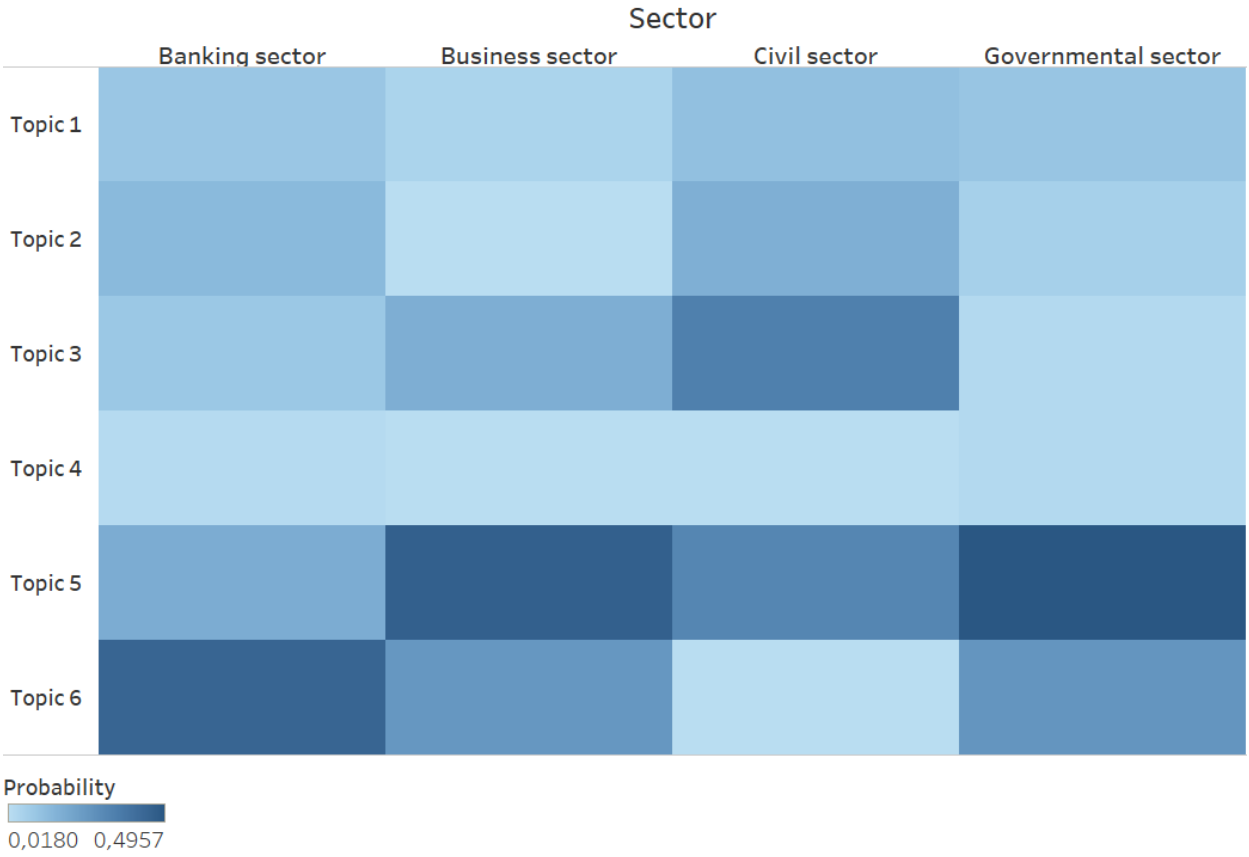

Figure 9: Probability for each modelled topic to appear in each sector.

Figure 9 shows the probability for each topic to appear in each sector. It agrees with the document-topic distribution in Figure 7 as well as the sector-topic importance in Figure 8, as topic 4 has the least probability of appearing, while topics 5 and 6 have the highest.

6.5.3 Clusters of documents and topics according to similarity

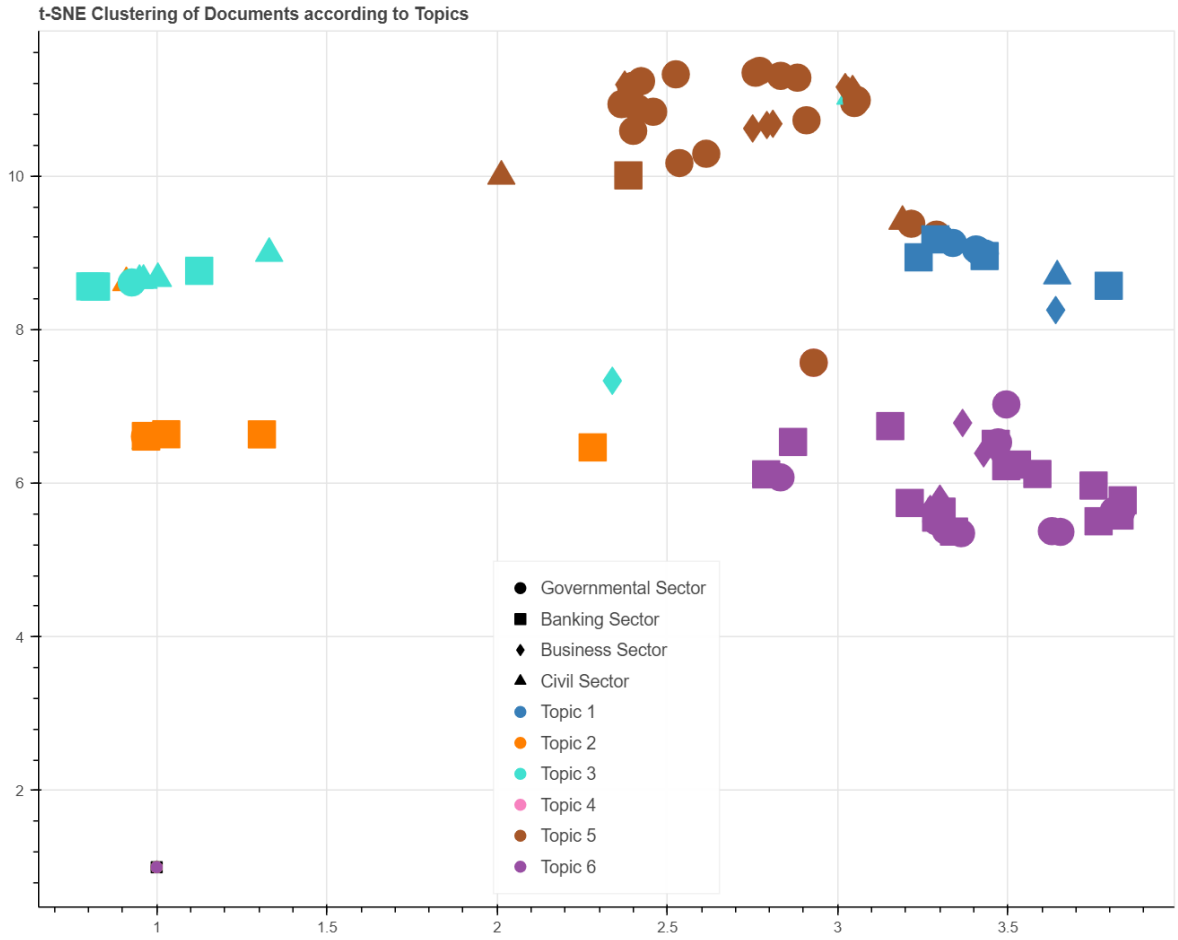

Figure 10: Cluster map of the documents clustered together according to similarities. The X and Y axes show the coordinates of where the documents are placed in the 2D plane.

T-SNE is an algorithm that can display higher dimension data on a lower scale, which makes it popular for visualisation as it can visualise how e.g. topics are clustered together in a 2D-plane (Sharma and Sharma, 2023). This cluster map (Figure 10) shows the distribution of documents and topics. Each colour is a topic, and each shape is a sector. The topics are generally clustered together, with some overlap between topics 1 and 5, as well as topics 2 and 3. Topic 6 has an outlier from the governmental sector in the lower left corner. Topics 5, 6, and 1 seem to be closer together, while topics 2 and 3 are closer. Topic 4 does not appear in the map, which corresponds to the lower likelihood of it appearing according to Figure 7 earlier. Topic 3 is dominated by the banking sector, but no inferences can be drawn from the other topics because of the skewed sampling. Topics 2, 3 and 5 seem to have the most spread, while the other topics are more grouped.

6.5.4 Analysis of sectors

Topics 5, *Cost of Fraud*, and 6, *Online Fraud Awareness and Prevention*, are the topics with the highest probability of appearing in all the documents, as indicated in Figure 7, while topic 4, *Discovering and Reporting Fraud*, is the lowest. While there has been an issue with a low rate of reports (Smith, 2007), topic 4 is seen as the less important one in the whole corpus. The probabilities for the topics appearing in the documents correspond to the probabilities of them appearing in the sectors, as exemplified by Figures 7 and 9. The exception is topic 6, which has a low probability of appearing in the civil sector. Topic 6 could be connected to the technical side, which is more relevant for the other sectors. The technologies mentioned in topic 6, such as "kortuppgifter" ("card details") and BankID are such that can be used as technologies of power because they can be monitored by the banks (Lenke, 2021). In a way, e-mails and invoices, as mentioned in the topic, can be connected to technologies of power as they request something from the receiver. The power might lay in the hands of the scammer and outside of the sectors, indicating that the sectors are not the actors in power. The cluster map (Figure 10) indicates that topics 2 - *Information-based Fraud Prevention* –, 3 – *Banks' Responsibilities in Fraud Prevention* – and 5 – *Cost of Fraud* – are more spread out. This corresponds to topic 5 being more spread out in all documents and sectors, but it does not correspond as well to topics 2 and 3. Further, topics 1 and 5 are overlapping in their clusters, which is not consistent with the topic-topic correlation (Figure 6). This indicates that there is a discrepancy between the distribution of topics for the documents and sectors, compared to the clustering of them. It could be a sign of a pattern that is not apparent from the other visualisations. One explanation could be that topics 1 and 5 both mention the cost of fraud, as well as "tillsyn" ("supervision"), which could imply that the cost is a latent theme in both topics that is only visible in the cluster map.

7 Discussion

7.1 Introduction

The previous parts outlined the topics, which consisted of rules and regulations, individuals' and banks' responsibilities to prevent fraud, prevention and detection, cost of crime, and awareness of online fraud. Topic 5, *Cost of Fraud,* has a high probability of appearing across all sectors. Topic 6, *Online Fraud Awareness and Prevention*, has a high probability as well, except for the civil sector. Topic 3, *Banks' Responsibilities in Fraud Prevention,* had the highest probability of appearing for the civil sector. This discussion will relate the findings to risk society, governmentality, and nodal governance as well as to previous research, to understand how the theoretical principles are reflected and which similarities and/or differences are in the findings compared to the literature. Therefore, it aims to answer research question 3 – *How do the findings regarding cyber fraud reflect the concepts of risk society, governmentality, and nodal governance?*

The types of scams outlined in the topics are vishing, invoice, advertisement, and card fraud, as seen in topics 2, 4 and 6. Social engineering scams such as vishing yield the most damage to its victim (Fjelkegård and Horgby, 2023). Vishing is one of the scams that has gotten the most media attention after the Supreme Court verdict T 4623-21, and it can be reflected in the topics, such as the keyword BankID in topics 4 and 6, as well as "samtal" ("call") in topic 2. Topic 6 *Online Fraud Awareness and Prevention* has a high probability for all sectors except the civil sector but the highest for the banks. This can be related to how banks are responsible for the payment services, and that other sectors hold them responsible for it. However, it is curious that the civil sector does not hold this topic in high regard, after the strong correspondence with topic 3, *Banks' Responsibilities in Fraud Prevention*. What this could mean would require more research, but it can be connected to the perception of fraud prevention, what is needed for it and who should be responsible.

7.2 Fraud prevention

7.2.1 Prevention, detection and deterrence

In general, the topics only focus on situational crime prevention. This is seen in topics 2, *Information-based Crime Prevention*, 4, *Discovering and Reporting Fraud*, as well as 6,

Online Fraud Awareness and Prevention. These topics focus on the situation, where information about fraud helps stop it, as well as ways to discover and stop fraud as it is happening. This is compliant with other literature, as fraud prevention is generally situational (Prenzler, 2016). Further, topics 1, *Rules and Regulations on an EU-level*, and 3, *Banks' Responsibilities in Fraud Prevention*, also address fraud prevention, but with different actors. Topic 1 does not indicate what kind of fraud prevention is needed, while topic 3 states that banks are responsible for both preventing fraud and compensating victims. Topic 3 also describes what happens after the event, where the victims should be compensated by the banks. This describes that different actors are responsible for managing the risks in the different topics, ranging from individuals in topic 2, banks in topic 3, as well as the government in topic 1.

The topics address prevention and detection in Dorminey's et al (2012) suggestion to combat fraud, but they do not address deterrence. This could also be coupled with the spatial dimension, as it makes it more difficult to pin down a perpetrator, which is the object of deterrence (Van Nguyen, 2022). The proactive and reactive angles differ between the topics. Topic 2 has a proactive angle, and topics 4 and 6 are more reactive, as they focus on detecting and stopping fraud. It is similar to traditional law enforcement, which Brenner (2004) says is based on reacting to crimes as they happen. Topics 2 and 3 are different levels of prevention, with topic 2 being on the individual level and topic 3 focusing on the banks. In a sense, fraud is an unforeseen threat that could happen at any moment, and one must protect oneself towards it, implying that there is no way to eliminate the perpetrator. A higher clearance rate could have a deterring effect, as Schneider (2019) argues. The clearance rate could be connected to the number of police reports. A big issue has been that fraud most often does not get reported to the police, but it does get more often reported to banks (Smith, 2007). It could be connected to topic 2, as keywords like "genomskåda" make the individual responsible for uncovering the fraud. Topic 4, *Discovering and Reporting Fraud*, has the keyword "polisanmälan" ("police report"), and the topic itself references discovering and reporting crime. This could indicate that the low numbers of reports are seen as an issue and that with topic 5, *Cost of Fraud*, it could infer that the monetary loss seen in statistics is not as high as the actual number. Card fraud is generally more often reported, with social engineering and advertisement frauds having half as high of a crime rate as card frauds (Fjelkegård and Horgby, 2023). Topic 4 could agree with Abdallah et al (2016) and Bolton and Hard (2002) when they claim that fraud detection is just as important as prevention. Fraud attempts can

slip through prevention measures, which makes it as important to discover fraud as it is happening (Bolton and Hand, 2002). It is not prevention, but topic 4 suggests that this is equally important.

7.2.2 Technologies of power

There are different types of technologies visible in the topics. Technologies of power such as surveillance, can be connected to the banks and government, which is visible in topics 1 and 3, EU regulations and banks' responsibilities. This is a form of power that seeks to control the future, and by extension, the population through banks using monitoring systems and governments using legislative means (Matthewman, 2013; Wolke, 2017). In topic 3, this power is instead used by other sectors to make the banks responsible, which can indicate that an actor's power can be used by other actors. Power only exists when there are relations (Valverde, 2017), and in a way, the civil sector lends power to the banking sector. Topic 2, *Information-based Fraud Prevention*, can be connected to technologies of the self, as they have the purpose of informing people to escape their victimisation (Lenke, 2021). In a normalisation process, these technologies are being seen as the norm (Lawlor and Nale, 2014), and recurring themes throughout the topics are information, awareness and discovering fraud. The focus is on the victim, as the victim seems to be the only one who can stop the crime, while the perpetrator cannot be caught, which is corresponding to research and statistics (Cross, 2016; Fjelkegård and Horgby, 2023).

7.2.3 Responsibility to handle prevention, detection, and compensation for fraud

The civil sector was an outlier as it has the highest probabilities for topics 2, *Information-based Fraud Prevention,* and 3, *Banks' Responsibilities in Fraud Prevention*, while the lowest for topic 6, *Online Fraud Awareness and Prevention*. Topic 2 is about the individual preventing fraud, which corresponds to the civil sector as it focuses on responsibilisation. Topic 3 is about banks taking responsibility for fraud and is the most probable topic in the civil sector. In a sense, topic 3 wants the banks to have more coercive methods to govern security, which infers that security is not only the government's issue. The government does give banks more responsibility, but it is in a broader sense about anti-money laundering and financial crime prevention (prop 2016/17:173). The notion of authorised and unauthorised transactions makes it more difficult for banks to compensate victims, as most fraud cases have been signed by the customer (Maher, 2021). However, the Supreme Court case (verdict T 4623-21) has raised questions about what can and cannot be seen as unauthorised

transactions. The victim was deceived convincingly by the scammer, and in that case, the court ruled that the bank was obligated to compensate their customer. This ruling could be one of the implicit foundations of topic 3. Topic 3 is the third most probable topic for the business sector and one of the least important for the governmental sector. This could infer that the government does not hold the banks responsible or does not deem them as important as the other sectors. The business sector does lose more monetary value than the civil sector but does not seem to want to hold the banks responsible. The Supreme Court ruling does not apply to companies (Krantz, 2024), which could explain the low probability of the topic in that sector.

On the other hand, responsibilisation of the individual is strong in topic 2, *Information-based Fraud Prevention*. It is up to the individual to take action and prevent their own victimisation, and it is also up to their relatives to inform them about fraud modus. This can be connected to Brenner's (2004) claim that the individual has been given more responsibility to manage their risks, and it harks back to the question: when is it reasonable to put the responsibility on the individual? This topic suggests that is the way to prevent fraud, which is in line with Leclerc and Morgenthaler's (2023) claim that information campaigns are the most effective way to prevent fraud. This kind of prevention makes the crime script more apparent, as it outlines how the scam is perpetrated and how to stop it (ibid.). This knowledge could be an empowering approach to governance, the conduct of conduct as Foucault puts it (Madsen, 2014), which enables individual's proactive and informed decisions. Further, information campaigns foster collaborative approaches, which seem to be the key to fraud prevention (Brenner, 2004; Dupont, 2004; Wall, 2002). This accommodates transparency and accountability, which is a foundation for good governance (Boudia and Jas, 2007).

In the media, there has been a lot of discourse about which responsibility the banks have (SVT, 2022). Examining the sector-topic correlations in Figure 8, this theme is the strongest in the civil sector, which corresponds to the media picture. This once again begs the question: who is responsible for risk management? Within nodal governance, every actor has their own set of skills (Burris et al, 2004), which could entail that banks are seen as having more power to prevent fraud. It can also reflect a broader discussion about societal structures, where the banking sector is seen as an actor with a lot of power – as they have technologies of power in the form of monitoring systems, for example (Lemke, 2021) – and the assumption is that they could take more responsibility for fraud victims. This could also be seen as risk being unevenly distributed, where the victims often are civilians who do not have as much power as

the banks. The state is no longer the one who has the most power, rather, it is the bank that can stop fraud. Further, this topic states that because the bank is the one who is supposed to stop fraud, they should compensate the victim if they fail to do so. Responsibility seems to be scattered depending on the perspective. On a higher level, risk can be a political question and could rearrange the power structures (Boudia and Jas, 2007). The topics seem to indicate who should be responsible for risk management from different viewpoints, rather than who is. This is not the straightforward governing power that is usually seen in the literature. Rather, it is the power to govern other parties by assigning them tasks to mitigate the risks of other sectors.

7.3 The risk society and fraud

7.3.1 Risk culture and normalisation

Risk society highlights the importance of public awareness for managing risks (Boudia and Jas, 2007), which can be seen in topic 2, *Information-based Fraud Prevention*. Madah and Marzurki's (2020) notion that risk management should be in the culture of an organisation aligns with the normalisation practices in governmentality. If risks are seen as inherent and inescapable, the measures to handle risks are seen as necessary and therefore a normalisation process starts. The normalisation of risk management strategies can be connected to topics 4, *Discovering and Reporting Fraud*, and 6, *Online Fraud Awareness and Prevention*, because they both promote a culture of being aware of cybersecurity threats and acting on suspicions. However, the minimalism principles that Zedner (2009) discusses can be applied here. The principle of minimalism entails that the least burden shall come from an intervention, and this needs to be balanced with the need for effective operations (Wolke, 2017; Zedner, 2009). If every transaction had to be screened, that would only halt the operations and potentially lead to losses. In that case, Banakar's (2015) claim about policymaking being ineffective would ring true, especially when he claims that formally regulating the transnational population is not effective. The topics are ambiguous if they entail monitoring systems from the banks, except for topic 3 *Banks' Responsibilities in Fraud Prevention.* Topic 1 suggests that EU regulations are important in fraud prevention. The legislation serves as one of the foundations of risk management (Wolke, 2017), and topic 1 has the highest probability for the civil sector (Figure 9), which could indicate that the civil sector trusts the legislation to help their causes. This topic could also be connected to the Supreme Court case, as it is highly relevant for the

victims of fraud in the civil sector. This could imply a form of pressure on the banks to focus on preventing fraud, or they might go with financial losses by monetarily compensating fraud victims.

7.3.2 Managing the risk of fraud

Risk management, as outlined by Wolke (2017), exists because a legal framework demands it or for economic or technological reasons. The legal framework reasoning is visible in topic 1, with the EU regulations. Economic reasons are visible in topic 5, *Cost of Crime*, but also a bit in topic 4, with the word "besparingar" ("savings"). Technological reasons are visible in topic 6, but also in topic 4 with its reference to BankID but is a bit more implicit in all the other topics, as they reference the way to make payments. Technology is a big reason the fraud crime rates have risen (Junger, Wang and Schlömer, 2020), so this reason being implicit in all the topics is not unexpected. This shows that risk management is relevant and needed for fraud, and that it must continue to evolve and adapt itself to the current fraud climate, especially when knowledge in this field is lacking (McCord, et. al., 2022).

Topic 5, *Cost of Fraud*, has a high probability of appearing in all the sectors but is especially of essence for business and government (Figure 9). This can be because the business sector has the highest monetary loss (Levi et al., 2017), which might lead to pressure on the government to act. The nodal governance is not as clear in the topics, as they seem to focus on different actors that have the power. Topic 1 focuses on EU regulations, topic 3 focuses on the banks. This implies that the pressure of fraud prevention is especially on those two actors.

7.3.3 Unequal distribution of risk

The unequal distribution of risk can be inferred from topic 2, which is about crime prevention. It has "gammal" ("old") as a keyword, which could mean that older people more often get scammed. This is compliant with statistics (Fjelkegård and Horgby, 2023). and it can show that some groups are at more risk than others. Risks are very different depending on who is experiencing them (Curran, 2013), and this is implicit in the topics. Topic 5 mentions actors such as consumers and companies, and topic 3 mentions "villaägare" ("house owners"). This could also infer a disparity in power between the actors, as the actors with more power – banks in this case – experience less risk compared to individuals (Levi et al., 2017). This could be connected to knowledge, even if it is not explicit in the theme. The knowledge from the past is used to predict the future and those people with knowledge of risks can afford to take more risks than others (Adam and van Loon, 2000; Curran, 2013). Topics 2, *Information-*

based Fraud Prevention, and 6, *Online Fraud Awareness and Prevention*, are connected to knowledge via spreading awareness. This knowledge, in a way, gives power to potential victims by letting them know how to not fall for the scammer's tricks, effectively taking power away from the scammer. However, knowledge and risk are unequally distributed, and sometimes knowledge and risk management are not seen as important until they get victimised (Kummer et al, 2014). This could infer the importance of spreading awareness. Risk society calls for an open dialogue between the government and its citizens to shape risk management (Boudia and Jas, 2007), which can be applied to the topics. The topics are on different levels of the populations, but the intersection seems to only happen in topic 5, *Cost of Fraud*, as it has a high probability of appearing in all sectors (see Figure 9). This could infer that all sectors agree that fraud is a big problem, but they do not agree on who is responsible for preventing it or what should be done.

7.4 Power relations and governance between the sectors

Crime prevention strategies are dependent on collaboration and networking, especially in the Nordics, and policy networks are dependent on power relations (Aromaa and Takala, 2005; Marin and Mayntz, 1992). The network is visible in the topics, and from the importance the sectors have in the network, power relations can be inferred. The banking, civil, and governmental sectors seem to be the most important nodes, seeing how the topics relate to the sectors, with the business sector taking a backseat. The topics have references to those sectors, and especially the banking sector is seen as a central node as it is constantly referenced throughout the topics. Banks are dependent on the other sectors for their survival, but they still hold a lot of power over them. Banks have technologies to monitor their customers, effectively the other sectors, which gives them more power, according to Foucault's technologies of power (Matthewman, 2013). Further, the EU has a place as another powerful node, from topic 1, *rules and regulations on an EU-level*, but according to Figure 8, this is not as important for the sectors as the other topics. This is a higher level, where the demands on the banks from a state level show the coercive power the state has. The banking sector is getting demands from both the civil and governmental sectors, but the business sector does not seem to make the same demands. Businesses do still rely on banks for their financial services, but the fraud prevention demands on the banks being less important can be explained by organisations having their own risk management strategies, and that the Supreme Court

verdict does not apply to companies and organisations (Krantz, 2024; Wolke, 2017). The power relations are not clear-cut and hierarchical, but they are all interdependent (Shearing and Wood, 2003). The result does agree with the core ideas of nodal governance: that it is not only the government that has the governance tasks (Holley and Shearing, 2017). The civil sector can be seen as a weaker actor that mobilises its resources to gain more power over governance (Wood and Shearing, 2006), and here they collaborate to keep the banking sector responsible for preventing and compensating victims of scams. It could be said that the banking sector governs and is governed by the civil sector, while the government governs the banking sector.

7.5 The relevance of space and scope in fraud

Space and scope are recurring, implicit themes. Topic 1 concerns crime over nation borders and topic 6 is about online means of committing fraud, implying that space is irrelevant when it comes to cyber fraud (Van Nguyen, 2022). Topic 5 details the scope, as the costs of crime are high and can affect anyone, as fraud can reach a vast number of victims (ibid.). Topic 2 combines space and scope, by describing modus and volume as something individuals must be aware of. In a way, anyone, anywhere can be victimised (Wall, 2007). This can also be connected to politics. Boudin and Jas (2007) claim that risk is being seen as more of a political issue. The European Union is a supranational organisation, and it is very relevant as fraud does not have the same spatial component as other crimes. Different countries can be used to laundry the money, depending on the legislation (Grabosky, 2001), and EU regulations could be a way to prevent it. This could once again be connected to nodal governance: multiple actors are involved and have the power to act upon fraud, but it can be difficult concerning space. Space concerning different jurisdictions can give varying responses to fraud prevention (Beck, 2000), but even within the same jurisdiction, space has a big part to play. The perpetrator is more difficult to catch when they are not in the same physical space as the victim. Globalisation and technology have brought innovations in how crime is carried out, while the criminal justice system is largely unchanged (Banakar, 2015; Törrönen and Korander, 2005). Crime prevention has changed but is not compatible with the current legal system, as it cannot handle the scope of the crimes (Törrönen and Korander, 2005; Wall, 2007). Therefore, the focus of the topics on different prevention measures (topics 2, 3 and 6) as well as the transnational angle with the scope (topics 1 and 5), requires a new

kind of crime prevention that would be able to handle crime regardless of space and scope. This can also be connected to risk society, as space and scope are modern risks that appear alongside technology.

8 Conclusion

8.1 Cyber fraud as a societal issue

This thesis aimed to examine how cyber fraud is addressed in political documents by different sectors in Sweden, as there is a research gap regarding the organisation and implementation of measures against cyber fraud. It also explored how the issue can be related to risk society and governmentality. Topic modelling was performed on a corpus of 102 documents from 4 different sectors in society: civil, business, banking and governmental sectors. The main findings consisted of topics generated by topic modelling. There were six topics identified: Topic 1 described EU regulations regarding fraud, topic 2 described crime prevention in the way that the individual should be informed about the risks and how to avoid them, and topic 3 described the banking sector's responsibility of preventing fraud and compensating fraud victims. Further, topic 4 was about detecting and reporting fraud, topic 5 described the cost and damages caused by fraud, and finally, topic 6 described awareness and prevention of online fraud. Topic 5 was the most important overall in all the documents and the most important for the business and governmental sectors. In contrast, topic 4 was the least important in all documents and sectors. Topic 6 was the most important in the banking sector and topic 3 was the most important in the civil sector.

The topics generally focused on situational crime prevention, and it was especially seen as a responsibilisation of the individual. Further, especially the civil sector wanted to put the responsibility on the banking sector. This implies that the perpetrator cannot be dealt with and that the victims must prevent their own victimisation, or that an actor in power, such as the banks, is the one responsible for prevention. This implies further that the perpetrator and space are irrelevant and with the larger scope, a new type of crime prevention is necessary. When it comes to prevention, the banking sector governs and is governed by the civil sector, while the governmental sector governs the banking sector. This implies that the banking sector is a central actor that has demands from different actors and an expectation to handle the fraud risk. This also implies the power relations, seeing to technologies of power that the powerful actors in banking and governmental sectors have. Their resources and power can be the reason they are seen as the ones who should be responsible. The risk of fraud is seen as inescapable and fosters a culture of risk mitigation. This can be connected to normalisation because prevention strategies such as fraud detection and awareness are more widespread and accepted forms of fraud prevention, and therefore normalised. The normalisation of the fraud

crime prevention practises can be seen in topics 4 and 6, as they are focused on detecting fraud and being aware of risks.

This thesis's strengths lie in the computational method that is novel in the field of sociology of law which can bring a more objective and different perspective. This strength is also the biggest weakness, as while it can get rid of human biases, it can also lose context. That is why reflexivity and placing the topic output into context is especially important. It has been a challenge to use topic modelling. While there is a lot written regarding how to use topic modelling in practice, the theory behind it is a bit vaguer. It required a lot of search for literature to understand how to use the concepts of perplexity and coherence, as well as how to decide the number of topics. Further, the approach is not common in social sciences, as articles covering or using topic modelling were mostly from computer science, which shows that there is a gap to be bridged between these disciplines.

Theoretical implications are that the risk society, governmentality, and nodal governance make up a suitable framework to analyse modern-day risks such as fraud. In fraud prevention, there is an unequal distribution of risk, which is connected to power. This implies that power is a central concept in fraud prevention and must be considered. Further, knowledge is as much important and can give more power to the weaker actors. This can be connected to practical implications, as it focuses on prevention by informing individuals and holding the banking sector responsible. This can also be a foundation of further research, as it is not clear that the responsibilisation of the individual and banking sector are the most effective ways to prevent fraud.

Cyber fraud is a complex issue that is different from traditional types of crime. It requires a new way of law enforcement and a new way of prevention. Technology can cause damage, but it can also prevent and save. Therefore, it is important to understand and examine the issue of cyber fraud and how different actors and societal actors can collaborate to overcome it. This thesis showcased the societal interplay between different actors to handle the issue of fraud. The results imply that victimisation is being regulated by policies, and that power is very much at play here. This infers that power must be examined to understand the extent and efficiency of fraud prevention measures, as all the different actors have different resources, knowledge, and power that must interact. The policies and social norms are interplaying, as the norms affect which policies are made and the policies, in a way, affect the norms, as in the case of the normalisation of fraud prevention measures. Banakar (2014) has noticed a shift to

proactive risk management, which corresponds to the findings in this thesis. A proactive plan to deal with the threats of cyber fraud is needed.

The contribution to the sociology of law from this thesis consists of an exploration of a novel approach in topic modelling, as well as a discussion about the responsibilisation of different actors and the interplay of fraud prevention measures and their normalisation. This can lay the foundation for further research regarding these themes, and serve as a basis for creating a new fraud prevention strategy. It is most likely that cyber fraud will continue to be a problem in the future, and that it will follow the advancement of technology, which means that Law Enforcement and prevention strategies must keep up to be able to combat it.

8.2 Limitations of the study

The statistical model can have the issue of overfitting and underfitting, where overfitting gets attached to the characteristics of the data and cannot generalise to unseen data, meaning that the model is specific to the data it has been provided with (Rongpeng, 2020). Underfitting is when the model is too simple to make predictions about unseen data (ibid.). In Table 2 the coherence value for this topic model was 0,476. The coherence value ranges from 0 to 1 (Tran, 2023), which implies that the model is moderately coherent. As topic modelling is a relatively novel approach in the social sciences, it is a method that should be developed further. The perplexity did not give any guidance for choosing the right model, and it did almost steadily go down with the rising number of topics. This implies that the model has a low degree of predictive power (Blei and Lafferty, 2007). This could also indicate that the model is overfitted, considering that the data size is on the lower end. The literature did not guide this either, as some sources stated a minimum of a thousand documents (Brett, 2012), while the other sources left the number of documents vague. Originally, a vector-based module called BERT would have been used because it considers the context rather than the bag-of-words approach used in LDA (Jeet Rawat et al, 2022), but it did not yield any results during the test run because the corpus was too small. This could be mitigated by having a larger corpus, but as mentioned, the ideal size is vague and must be explored.

Other types of packages could have been used, with their pros and cons. The choice of Gensim was because it was deemed the most suitable for analysing the Swedish language, at the same time as it did not use too many computer resources (Řehůřek and Sojka, 2010). The lemmatiser did not seem to work in practice and documentation was lacking. There were other

issues with the lemmatiser. It changed the word “skall” (“shall”) to the word “skalle” (“skull”), which changed the context. This implies that NLP is not as developed for languages other than English. The technical tools might need to be further developed to be suitable for more languages, but that is outside of the scope for this thesis. One can only speculate how and if the results could have been more substantial if the model was better, but this method needs more development and research.

The documents were unevenly distributed over the sectors, as seen in Table 1. The uneven distribution can be attributed to the governmental and banking sectors being more document-based and bureaucratic, while the civil sector does not have the same demands in the law. This choice was made to contribute to the quantity of documents, but it might have affected and skewed the data towards the sectors with more documents. Further, the documents were of varying lengths, as seen by the word count in Figure 2, which also could have affected it in the same way. The literature did not give any guidance on how and if the varying document lengths would affect the result, which establishes that more exploration of the method is needed to understand its limits and capabilities.

Topics 1-4 overlap according to Figure 6, which could indicate that the model is underfitted and fails to capture nuances in the data set (Tijare and Rani, 2020). With the qualitative assessment, topic 4 was more fragmented and difficult to set a label on, which could also indicate underfitting (ibid.). However, topics 1-3 were more distinct and easier to interpret, which implies that the topic model’s fit needs a balance between quantitative and qualitative measures. As Sekar (2024) claims, distant reading identifies broader themes and can fail to capture context and nuances. With a manual document analysis, the data could have been cross-verified, but that would have been out of the scope of this thesis. Further, a manual document analysis would have made it more difficult to analyse all the documents and infer the relations between sectors, documents, and topics in the same way that this approach did.

Bibliography

Abdallah, A., Maarof, M.A., and Zainal, A.B. (2016). Fraud detection system: A survey. *J. Netw. Comput. Appl.*, 68, 90-113.

Adam, B. and van Loon, J. (2000). *Introduction: Repositioning Theory Risk; the Challenge for Social Theory.* In: Adam, B., Beck, U., and Van, L. J. (eds.). *The Risk Society and Beyond: Critical Issues for Social Theory.* (pp. 10-39). London: SAGE Publications.

Alvesalo, A. and Tombs, S. (2004). Economic Crime Control in Finland. *Sociology*, 38(1), 165–174.

Andersson, J. (2005). The Swedish National Council for Crime Prevention: A Short Presentation. *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 6(1), 74–88.

Arner, D., Zetzsche, D., Buckley, R. and Barberis, J. (2018). The Identity Challenge in Finance: From Analogue Identity to Digitized Identification to Digital KYC Utilities. *SSRN Electronic Journal*, 10.2139/ssrn.3224115.

Aromaa, K. and Takala, J-P. (2005). Recent Developments in Crime Prevention and Safety Policies in Finland. *Canadian Journal of Criminology and Criminal Justice*, 47(2), 389–406.

Asmussen, C.B. and Møller, C. (2019). Smart literature review: A practical topic modelling approach to exploratory literature review. *Journal of Big Data*, 6(1), 93.

Atkins, B. and Huang, W. (2013). A Study of Social Engineering in Online Frauds. *Open Journal of Social Sciences*, 1, 23-32.

Auerbach, C. and Silverstein, L.B. (2003). *Qualitative Data: An Introduction to Coding and Analysis.* New York: NYU Press.

Banakar, R. (2014). *Normativity in legal sociology: Methodological reflections on law and regulation in late modernity*. New York: Springer International Publishing AG.

Barbaresi A. (2023). *Simplemma: a simple multilingual lemmatizer for Python [Computer software] (*Version 0.9.1). Berlin: Berlin-Brandenburg Academy of Sciences. Available from https://github.com/adbar/simplemma

Barker, P. (1998). *Michel Foucault: An Introduction.* Edinburgh: Edinburgh University Press.

Bird, S., Klein, E. and Loper, E. (2009). Natural language processing with Python: Analyzing text with the natural language toolkit. Newton: O'Reilly Media, Inc.

Beck, U. (1992). *Risk society: Towards a new modernity.* London: Sage Publications.

Beck, U. (2000). *Risk Society Revisited: Theory, Politics and Research Programmes.* In: Adam, B., Beck, U. and Van, L.J. (Eds.). *The Risk Society and Beyond: Critical Issues for Social Theory.* (Pp. 208-225.) SAGE Publications, Limited.

Becker, S. and Bryman, A. (2012). *Understanding Research for Social Policy and Social Work. Themes, Methods and Approaches* (Second edition). Bristol: Policy Press.

Blei, D.M. and Lafferty, J.D. (2007). A correlated topic model of Science. *The Annals of Applied Statistics*, 1(1).

Bodker, A., Connolly, P., Sing, O., Hutchins, B., Townsley, M. and Drew, J. (2023). Card-not-present fraud: Using crime scripts to inform crime prevention initiatives. *Security Journal,* 36(4), 693–711.

Bohlin, A. (2010). *Offentlighetsprincipen* (Eighth edition). Stockholm: Norstedts juridik.

Bolton, R.J. and Hand, D.J. (2002). Statistical Fraud Detection: A Review. *Statistical Science,* 17(3), 235–249.

Bos, J. (2020). *Research ethics for students in the social sciences* (First edition). Berlin: Springer Nature.

Boudia, S. and Jas, N. (2007). Introduction: Risk and 'Risk Society' in Historical Perspective, *History and Technology*, 23:4, 317-331.

Bowen, G.A. (2009). Document Analysis as a Qualitative Research Method. *Qualitative Research Journal (RMIT Training Pty Ltd Trading as RMIT Publishing)*, 9(2), 27–40.

Boyd-Graber, J., Hu, Y. and Mimno, D. (2017). Applications of Topic Models. *Foundations and Trends in Information Retrieval:* Vol. 11: No, 2-3, pp 143-296.

Brenner, S.W. (2004). Toward criminal law for cyberspace: new model of law enforcement. *Rutgers Computer & Technology Law Journal*, 30(1), 1-104.

Brett, M.R. (2012). Topic Modeling: A Basic Introduction. *Journal of Digital Humanities,* 2(1), Winter.

Broadhurst, R. (2021). *Cybercrime: Thieves, Swindlers, Bandits, and Privateers in Cyberspace.* In: Cornish, P. (ed.). *The Oxford Handbook of Cyber Security.* Oxford: Oxford Academic.

Brottsofferjouren (n.d.). *Bedrägerier.* Retrieved 2024-03-11 from https://www.brottsofferjouren.se/brottsofferstod/olika-brottstyper/bedragerier/

Brown, R. and Smith, R.G. (2018). Exploring the relationship between organised crime and volume crime. *Trends and Issues in Crime and Criminal Justice,* 565.

Burgard, A. and Schlembach, C. (2013). Frames of Fraud: A Qualitative Analysis of the Structure and Process of Victimization on the Internet. *International Journal of Cyber Criminology,* 7, 112.

Burris, S.C., Drahos, P. and Shearing, C.D. (2005). Nodal Governance. *Australian Journal of Legal Philosophy*, 30, 30-58.

Button, M and Cross, C. (2017). *Cyber Frauds, Scams and their Victims* (First edition). London: Routledge.

Campesato, O. (2022). *Natural Language Processing Using R: Pocket Primer* (First edition). Herndon: Mercury Learning and Information.

Choi, K., Lee, J. and Chun, Y. (2017). Voice phishing fraud and its modus operandi. *Security Journal*. 30.

Cross, C. (2016). Using financial intelligence to target online fraud victimisation: Applying a tertiary prevention perspective. *Criminal Justice Studies*, 29(2), 125–142.

Cross, C. (2022). Using artificial intelligence (AI) and deepfakes to deceive victims: The need to rethink current romance fraud prevention messaging. *Crime Prevention and Community Safety*, 24(1), 30–41.

Curran, D. (2013). Risk Society and the Distribution of Bads: Theorizing Class in the Risk Society. *The British journal of sociology.* 64. 44-62.

Curran, D. (2016). *Risk, risk society, risk behavior and social problems*. In: Ritzer, G. (ed.). *The Blackwell Encyclopedia of Sociology*. Oxford: Blackwell.

Deflem, M. (2008). *Sociology of Law: Visions of a Scholarly Tradition.* Cambridge: Cambridge University Press.

Digital Fraud Committee (2022). *Fighting Fraud: Breaking the Chain.* Committee Office, House of Lords, London. Retrieved 2024-03-08 from https://committees.parliament.uk/publications/31584/documents/177260/default/

Di Ronco, A. (2016). Inspecting the European crime prevention strategy towards incivilities. *Crime Prevention and Community Safety*, 18(2), 141–160.

Drisko, J.W. and Maschi, T. (2015). *Content Analysis*. New York: Oxford Academic

Dorminey J., Fleming, A.S., Kranacher, M-J. and Riley jr., R. (2012). The evolution of fraud theory. *Issues in Accounting Education, 27*, 555–579.

Drakman, A. and Gelfgren, S. (2022). How to combine close and distant reading within the history of science and ideas: Two examples from ongoing research. *Lychnos*, 85–108.

Dupont, B. (2004). Security in the age of networks. *Policing and Society*, 14:1, 76-91.

Dyevre, A. (2021). Text-mining for Lawyers: How Machine Learning Techniques Can Advance our Understanding of Legal Discourse. *Erasmus Law Review*, 14(1).

Fenniak, M., Stamy, M., pubpub-zz, Thoma, M., Peveler, M., Exiledkingcc, and ypdf Contributors. (2022). The pypdf library. Retrieved from https://pypi.org/project/pypdf/

Fjelkegård, L. and Horgby, A. (2023). *Bedrägerier mot privatpersoner. De förebyggande åtgärdernas träffsäkerhet. Rapport 2023:11.* Stockholm: Brottsförebyggande rådet.

Gan, J. and Qi, Y. (2021). Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example. *Entropy*, 23(10), 1301.

Ghavami, P. (2020). *Big Data Analytics Methods: Analytics Techniques in Data Mining, Deep Learning and Natural Language Processing*. Berlin, Boston: De Gruyter.

Global Anti-Scam Alliance, GASA (2023). *The Global State of Scams – 2023.* Retrieved 2024-05-16 from https://www.gasa.org/files/ugd/b63e7d_92ac212a168843219668d5a28510ce16.pdf

Graham, J. (1993). Crime prevention policies in Europe. *European Journal of Crime, Criminal Law and Criminal Justice*, 1(2), 126-142.

Grabosky, P.N. (2001). Virtual Criminality: Old Wine in New Bottles? *Social & Legal Studies*, 10(2), 243–249.

Grant, H. (2015). *The "Social" of Crime Prevention: Meaning and Implications of Social Crime Prevention for Police in the Developing World.* In: Grant, H. (ed.). *Social Crime Prevention in the Developing World* (First edition). Cham: Springer International Publishing.

Gustafsson, C. (2014). *Crime prevention in Sweden.* Kraków: Wydawnictwo Jak.

Hardeniya, N. Perkins, N., Chopra, D., Joshi, N. and Mathur, I. (2016). *Natural Language Processing: Python and NLTK*. Birmingham: Packt Publishing.

Hecking, T. and Leydesdorff, L. (2019). Can topic models be used in research evaluations? Reproducibility, validity, and reliability when compared with semantic maps. *Research Evaluation. 28*. 10.1093/reseval/rvz015.

Holst, E., Kamra Kregert, K., Viberg, J., Wallin, S., and Westerberg, S. (2023). *Nationella trygghetsundersökningen 2023: Om utsatthet, otrygghet och förtroende. Rapport 2023:9.* Stockholm: Brottsförebyggande rådet.

Holley, C. and Shearing, C. (2017). *A nodal perspective of governance: Advances in nodal governance thinking.* In: Drahos, P. (ed.). *Regulatory Theory* (First edition, pp. 163–180). Canberra: ANU Press.

Högsta domstolen verdict 2022-06-21 in Case nr. T 4623–21.

Ignatow, G. and Mihalcea, R. (2017). *Text Mining: A Guidebook for the Social Sciences.* Los Angeles: SAGE Publications, Inc.

Jeet Rawat, A., Ghildiyal, S. and Dixit, A.K. (2022). Topic modelling of legal documents using NLP and bidirectional encoder representations from transformers. *Indonesian Journal of Electrical Engineering and Computer Science*, 28(3), 1749.

Jin, Z. and Mihalcea, R. (2023). *Natural Language Processing for Policymaking.* In: Bertoni, E., Fontana, M., Gabrielli, L., Signorelli, S. and Vespe, M. (eds.). *Handbook of Computational Social Science for Policy* (pp. 141-162). Cham: Springer International Publishing.

Junger M., Wang V. and Schlömer M. (2020). Fraud against businesses both online and offline: crime scripts, business characteristics, efforts, and benefits. *Crime Science*, 9(1), 1–15.

Kagias, P., Cheliatsidou, A., Garefalakis, A., Azibi, J. and Sariannidis, N. (2022), The fraud triangle – an alternative approach. *Journal of Financial Crime,* Vol. 29 No. 3, pp. 908-924

Karppinen, K. and Moe, H. (2011). *What we talk about when we talk about document analysis*. In: Just, N. and Puppis M. (eds.). *Trends in Communication Policy Research: New Theories, Methods and Subjects* (pp. 177-194). Bristol: Intellect.

Karppinen, K. and Moe, H. (2019). *Texts as Data I: Document Analysis.* In: Van den Bulck, H., Puppis, M., Donders, K. and Van Audenhove, L. (eds.). *The Palgrave Handbook of Methods for Media Policy Research* (pp. 249-262). London: Palgrave Macmillan.

Kenis, P. and Schneider, V. (1992). *Policy Networks and Policy Analysis: Scrutinizing a New Analytical Toolbox.* In: Marin, B. and Mayntz, R. (eds.). *Policy Networks: Empirical Evidence and Theoretical Considerations* (pp. 25-58). Frankfurt: Campus Verlag.

Kenis, P. and Schneider, V. (2019). *Analyzing Policy-Making II: Policy Network Analysis*. In: Van den Bulck, H., Puppis, M., Donders, K. and Van Audenhove, L. (eds.). *The Palgrave Handbook of Methods for Media Policy Research* (pp. 471–491). London: Palgrave Macmillan.

Kemp, S., Miró-Llinares, F. and Moneva, A. (2020). The Dark Figure and the Cyber Fraud Rise in Europe: Evidence from Spain. *European Journal on Criminal Policy and Research*, 26(3), 293–312.

Krantz, M. (22 February 2024). *Företagare extra utsatta vid bankbedrägerier*. Retrieved 2024-04-16 from https://www.foretagarna.se/nyheter/riks/2024/februari/foretagare-extra-utsatta-vid-bankbedragerier/

Kummer, T-F., Singh, K. and Best, P. (2014). The effect of fraud on risk management in not-for-profit organizations. *Corporate Ownership and Control.* 12. 641-655.

Lappi-Seppälä, T. and Tonry, M. (2011). Crime, Criminal Justice, and Criminology in the Nordic Countries. *Crime and Justice*, 40(1), 1–32.

Lawlor, L. and Nale, J. (2014). *The Cambridge Foucault Lexicon.* Cambridge: Cambridge University Press.

Lemke, T. (2021). *The Government of Things: Foucault and the New Materialisms.* New York: New York University Press.

Leppänen, A. and Kankaanranta, T. (2017). Cybercrime investigation in Finland. *Journal of Scandinavian Studies in Criminology and Crime Prevention, 18*(2), 157–175.

Levi, M. (2008). Organized fraud and organizing frauds: Unpacking research on networks and organization. *Criminology and Criminal Justice, 8*, 389–419.

Levi, M. and Doig, A. (2020). Exploring the 'Shadows' in the Implementation Processes for National Anti-fraud Strategies at the Local Level: Aims, Ownership, and Impact. *European Journal on Criminal Policy and Research*, 26(3), 313–333.

Levi, M. and Maguire, M. (2004). Reducing and preventing organised crime: An evidence-based critique. *Crime, Law and Social Change,* 41(5), 397–469.

Levi, M., Doig, A., Gundur, R., Wall, D. and Williams, M. (2017). Cyberfraud and the implications for effective risk-based responses: Themes from UK research. *Crime, Law and Social Change*, 67(1), 77–96.

Lewis-Beck, M. S., Bryman, A. and Futing Liao, T. (2004). Content analysis. In: Lewis-Beck, M. S., Bryman, A. and Futing Liao, T. (eds.). *The SAGE Encyclopedia of Social Science Research Methods* (pp. 187-190). Los Angeles: Sage Publications, Inc.

Li, T. M. (2007). Governmentality. *Anthropologica*, *49*(2), 275–281.

Lidskog, R. and Persson, M. (2012). Community Safety Policies in Sweden. A Policy Change in Crime Control Strategies? *International Journal of Public Administration,* 35:5, 293-302.

Lupton, D. (2013). *Risk: Second edition.* London: Taylor & Francis Group.

Madah Marzuki, M., Nik Abdul Majid, W.Z., Azis, N.K., Rosman, R. and Haji Abdulatiff, N.K. (2020). Fraud Risk Management Model: A Content Analysis Approach. *The Journal of Asian Finance, Economics and Business*, 7(10), 717–728.

Madsen, O.J. (2014). *Governmentality*. In: Teo, T. (ed.) *Encyclopedia of Critical Psychology*. New York: Springer.

Mandal, A. and Shanmugam, A. (2023). Fathoming fraud: unveiling theories, investigating pathways and combating fraud. *Journal of Financial Crime.*

Maher, R. (2021). Critical Analysis of Recent Efforts in the United Kingdom to Tackle Authorised Push Payment Scams and the Impact on the Bank-Customer Relationship. *Trinity College Law Review, 24*, 134-145.

Matthewman, S. (2013). Michel Foucault, Technology, and Actor-Network Theory. *Techné: Research in Philosophy and Technology.* 17. 274-292.

Marin, B. and Mayntz, R. (1992). *Introduction: Studying Policy Networks.* In: Marin, B. and Mayntz, R. (eds.). *Policy Networks: Empirical Evidence and Theoretical Considerations* (pp. 11-23). Frankfurt: Campus Verlag.

Moretti, F. (2013). *Distant reading*. Brooklyn: Verso.

Näsi, M., Danielsson, P. and Kaakinen, M. (2023). Cybercrime Victimisation and Polyvictimisation in Finland—Prevalence and Risk Factors. *European Journal on Criminal Policy and Research,* 29(2), 283–301.

Pickett, M., Cline, D. and Ryan, J. (2020). *Exploring Coherence Metrics for Optimizing Topic Models of Humpback Song.* Massachusetts: Massachusetts Institute of Technology.

Power, M. (2013). The apparatus of fraud risk. *Accounting, Organizations and Society,* 38(6–7), 525–543.

Prabhakaran, S. (2018). *Topic Modeling with Gensim (Python).* Retrieved 2024-04-04 from https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/

Prenzler, T. (2016). Welfare fraud prevention in Australia: a follow-up study. *Crime Prevention and Community Safety*, Vol. 18 No. 3, pp. 187-203.

Prenzler, T. (2017). *Fraud Victimisation and Prevention.* In: Deckert, A. and Sarre, R. (eds.). *The Palgrave Handbook of Australian and New Zealand Criminology, Crime and Justice* (pp. 269–283). Cham: Springer International Publishing.

Prenzler, T. (2020). What works in fraud prevention: A review of real-world intervention projects. *Journal of Criminological Research, Policy and Practice,* 6(1), 83–96.

Prop. 2016/17:173. *Ytterligare åtgärder mot penningtvätt och finansiering av terrorism.* https://www.riksdagen.se/sv/dokument-och-lagar/dokument/proposition/ytterligare-atgarder-mot-penningtvatt-och_h403173/html/

Raschka, S. and Mirjalili, V. (2017). *Python Machine Learning - Second Edition: Unlock Modern Machine Learning and Deep Learning Techniques with Python by Using the Latest Cutting-edge Open Source Python Libraries.* Birmingham: Packt Publishing.

Rege, A. (2009). What's Love Got to Do with It? Exploring Online Dating Scams and Identity Fraud. *International Journal of Cyber Criminology (IJCC).* 3. 974-2891.

Řehůřek, R., & Sojka, P. (2010). *Software Framework for Topic Modelling with Large Corpora. Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* (pp. 45–50). Valletta: ELRA.

Regeringen (7 September 2023). *Bedrägerier mot äldre i fokus på rundabordssamtal.* Retrieved 2024-03-11 from https://www.regeringen.se/artiklar/2023/09/bedragerier-mot-aldre-i-fokus-pa-rundabordssamtal/

Regeringen (3 April 2018). *Swedish legislation - how laws are made*. Retrieved 2024-03-18 from https://www.government.se/how-sweden-is-governed/swedish-legislation---how-laws-are-made/

Regeringskansliet (Fi2022/00489). *Remiss: Promemorian Ett stärkt skydd mot bedrägerier vid betalningar online.* https://www.regeringen.se/contentassets/22bb150ea23c40769f49092791b4b2aa/remissinstanser-ett-starkt-skydd-mot-bedragerier-vid-betalningar-online.pdf.pdf

Regeringskansliet (Fi2016/00114). *Remiss: Departementspromemorian Kontroller och inspektioner i Sverige av Europeiska byrån för bedrägeribekämpning (Ds 2016:1).* Retrieved from: https://www.regeringen.se/contentassets/7e24caf8d6bc44cda2891048536a3fce/remissinstanser-kontroller-och-inspektioner-i-sverige-av-europeiska-byran-forbedrageribekampning.pdf

Richert, W., Coelho, L. P., Chaffer, J. and Swedberg, K. (2013). *Building machine learning systems with Python.* Birmingham: Packt Publishing.

Rodrigues, V. F., Policarpo, L. M., da Silveira, D. E., da Rosa Righi, R., da Costa, C. A., Barbosa, J. L. V., Antunes, R. S., Scorsatto, R., and Arcot, T. (2022). Fraud detection and prevention in e-commerce: A systematic literature review. *Electronic Commerce Research and Applications*, 56(Complete).

Rongpeng L. (2020). *Essential Statistics for Non-STEM Data Analysts: Get to Grips with the Statistics and Math Knowledge Needed to Enter the World of Data Science with Python.* Birmingham: Packt Publishing.

Salehijam, M. (2018). The Value of Systematic Content Analysis in Legal Research. *Tilburg Law Review, 23*(1-2), 34–42.

Savolainen, J. (2005). Think Nationally, Act Locally: The Municipal-Level Effects of the National Crime Prevention Program in Finland. *European Journal on Criminal Policy and Research*, 11(2), 175–192.

Sekar, J.J. (2024). How Distant is 'Distant Reading'? A Paradigm Shift in Pedagogy. *Asian Journal of Language, Literature and Culture Studies*, 7(1), 84-99.

Schneider, A. (2019). Deterrence Theory in Paraguay: Exploring Fraud and Violation of Trust Cases. *Social Sciences*, 8(1), 23.

SFS (1962:700). Brottsbalk.

SFS (2017:630). Lag om åtgärder mot penningtvätt och finansiering av terrorism

SFS (2010:751) Lag om betaltjänster

Shannon, D., Hradilova, S.K., Skinnari, J. and Hörnqvist, L. (2016). *Bedrägeribrottsligheten i Sverige – Kartläggning och åtgärdsförslag. Rapport 2016:9.* Stockholm: Brottsförebyggande rådet.

Shannon, D. (2022). *Brottsutvecklingen till och med 2021. Kortanalys 5/2022.* Stockholm: Brottsförebyggande rådet.

Shafritz, J. (2004). *The Dictionary Of Public Policy And Administration.* London: Routledge.

Sharma, N. and Sharma, S. (2023). *Optimization of t-SNE by Tuning Perplexity for Dimensionality Reduction in NLP.* In: Kumar, S., Hiranwal, S., Purohit, S. and Prasad, M. (eds). *Proceedings of International Conference on Communication and Computational Technologies. ICCCT 2023. Algorithms for Intelligent Systems.* Singapore: Springer.

Shearing, C., and Wood, J. (2003). Nodal Governance, Democracy, and the New 'Denizens'. *Journal of Law and Society,* 30(3), 400–419.

Silveira, R., Fernandes, C.G.O., Neto, J. A. M., Furtado, V. and Filho, E. P. (2021). Topic Modelling of Legal Documents via LEGAL-BERT. *CEUR Workshop Proceedings,* 2896, 64–72

Snortum, J.R. (1983). Police practice and crime prevention: Swedish perspectives and U.S. problems. *Police Journal*, 56(3), 224-240.

Spicker, P. (2006). *Policy analysis for practice: Applying social policy*. Bristol, UK: Policy Press.

Steennot, R. (2018). Reduced payer's liability for unauthorized payment transactions under the second Payment Services Directive (PSD2). *Computer Law & Security Review*, 34, 954-964.

SVT (25 december 2023) *100-tals lurade får rätt mot bankerna – efter bedrägerierna.* Retrieved 2024-03-27 from *https*://www.svt.se/nyheter/lokalt/stockholm/100-tals-lurade-far-ratt-mot-bankerna-efter-bedragerierna.

Srinivasa-Desikan, B. (2018). *Natural Language Processing and Computational Linguistics: A Practical Guide to Text Analysis with Python, Gensim, SpaCy, and Keras.* Birmingham: Packt Publishing.

Svenska Bankföreningen (24 March 2023). *De stora bankkoncernerna.* Retrieved 2024-03-20 from https://www.swedishbankers.se/fakta-och-rapporter/svensk-bankmarknad/de-stora-bankkoncernerna/.

Taylor, J.L. and Galica, T. (2020). A new code to protect victims in the UK from authorised push payments fraud. *Banking & Finance Law Review,* 35(2), 327-332.

Tijare, P. and Rani, P. (2020). Exploring popular topic models. *Journal of Physics: Conference Series.* 1706. 012171.

Torres-Cuello, M.A. and Pinzon-Salcedo, L.A. (2023). A Look at Power Issues in Collaborative Program Evaluations Under Michel Foucault´s Conception of Power-Knowledge. *American Journal of Evaluation,* 44(3), 424-446.

Tran, T. (6 November 2023). *Topic Modelling: Crafting an LDA Model with Python for Analyzing Dialogue in the 'Friends' Sitcom.* Retrieved 2024-05-07 from https://medium.com/@trantphuongthao99/topic-modelling-crafting-an-lda-model-with-python-for-analyzing-dialogue-in-the-friends-sitcom-9527bd3fd442

Törrönen, J., and Korander, T. (2005). Preventive Policing and Security Plans: The Reception of New Crime Prevention Strategies in Three Finnish Cities. *Journal of Scandinavian Studies in Criminology & Crime Prevention,* 6(2), 106–127.

Valverde, M. (2017). *Michel Foucault.* London: Routledge.

Van Nguyen, T. (2022). The modus operandi of transnational computer fraud: A crime script analysis in Vietnam. *Trends in Organized Crime,* 25(2), 226–247.

Vignoli, V. (2022). *Text as Data.* In: Mello, P.A. and Ostermann, F. (eds.). *Routledge Handbook of Foreign Policy Analysis Methods*. London: Routledge.

Wall, D.S. (2002). *Insecurity and the policing of cyberspace.* In: Crawford, A. (ed.). *Crime and insecurity* (pp. 186–209). Cullompton: Willan.

Wall, D.S. (2007). Policing Cybercrimes: Situating the Public Police in Networks of Security within Cyberspace. *Police Practice and Research*, 8(2), 183–205.

Weisburd, D., Waring, E. and Chayet E. (2002). *White-Collar Crime and Criminal Careers*. Cambridge: Cambridge University Press.

Welsh, B.C. and Farrington, D.P. (2012). *The Oxford handbook of crime prevention.* Oxford: University Press.

Wikström, P-O. and Torstensson, M. (1999). Local crime prevention and its national support: Organization and direction. *European Journal of Criminal Policy and Research,* 7(4), 459–481.

Wilson, C. and Laidlaw, G. (2017). Applying Nodal Governance to Combat Cybercrime: A Novel Approach. *Journal of The Colloquium for Information System Security Education (CISSE), 5*(1).

Wood, J. and Shearing, C. (2006). *Security and nodal governance.* Paper prepared for seminar at the Temple University Beasley School of Law, 25 October 2006. Philadelphia, PA.

Wolke, T. (2017). *Risk management*. Berlin: De Gruyter Oldenbourg.

Zedner, L. (2003), The concept of security: an agenda for comparative analysis. *Legal Studies*, 23: 153-175.

Zedner, L. (2009). *Security.* Abingdon: Routledge.

APPENDIX 1

Documents used for analysis.

Regeringen

1. Lagrådsremiss Kontroller och inspektioner i Sverige av Europeiska byrån för bedrägeribekämpning
2. Fi2022/00489 Ett stärkt skydd mot bedrägerier vid betalningar online
3. Lagrådsremiss Stark kundautentisering vid fakturabetalningar online
4. Regeringens proposition 2016/17:56 Kontroller och inspektioner i Sverige av Europeiska byrån för bedrägeribekämpning
5. Regeringens proposition 2022/23:9 Stark kundautentisering vid fakturabetalningar online
6. Regeringens proposition 2020/21:73 En ny straffbestämmelse som skyddar betalningsverktyg
7. Ett nytt brott om olovlig befattning med betalningsinstrument Genomförande av non-cash-direktivet Ds 2020:1
8. Lagrådsremiss En ny straffbestämmelse som skyddar betalningsverktyg
9. Kommittédirektiv Säker och tillgänglig digital identitet
10. SOU 2019:14 Ett säkert statligt ID-kort – med e-legitimation
11. promemorian Brott och brottsutredning i IT-miljö; Europarådets konvention om IT-relaterad brottslighet med tilläggsprotokoll.
12. SOU 2015:77 Fakturabedrägerier
13. Lagrådsremiss Grovt fordringsbedrägeri och andra förmögenhetsbrott
14. Kommittédirektiv: åtgärder mot fakturabedrägerier
15. Regeringens proposition 2000/01:133

Pensionärernas riksorganisation

16. "Agera – stoppa bedrägerierna!"
17. "HANDLINGSPROGRAM Gäller 2022-2026"
18. Krav för att få stopp på bedrägerierna
19. PRO bjöd in bankerna och pratade bedrägerier
20. PRO uppmanar finansmarknadsministern till krafttag mot bedrägerier

Svenska Bankföreningen

21. Vad innebär den nya betaltjänstlagen och PSD2?
22. Säkrare betalningar
23. PSD2
24. Kampanj ska rusta kunderna mot bedrägeriförsök
25. Staten måste ta ansvar för id-handlingarna
26. Felaktig bild om kortbedrägerierna
27. Bli Svårlurad!

Swedbank

28. ”Policy mot penningtvätt och terroristfinansiering”
29. Anmäla bedrägeri
30. Bedrägeri – skydda dig mot bedrägerier
31. Ligg steget före – skydda dig mot bedrägerier
32. Undvik bedragare – ligg steget före

Handelsbanken

33. ”Nya regler om betaltjänster”
34. Tillsammans mot bedrägerier
35. Undvik bedrägerier under semestertider
36. Så skyddar du företaget mot cyberbrott
37. De vanligaste bedrägerierna

Klarna

38. Bedrägeripolicy för Sverige

Sveriges konsumenter

39. “Banker – gör det enda rätta!”
40. Tuffare krav på bankerna i gemensamt utspel

Svårlurad

41. Varning för falska SMS
42. Det är lätt att bli svårlurad

Finansinspektionen

43. Årsredovisning 2023
44. Användningen av öppna finansiella tjänster i Sverige
45. Policy om intern styrning och kontroll
46. Kontobedrägerier
47. Penningtvätt – upplägg med osanna fakturor
48. Konsumentskyddrapport 2023

Svenskt Näringsliv

49. Brottslighetens kostnader för svenska företag
50. Brottslighetens kostnader 2020
51. Brottslighetens kostnader 2022
52. FÖRETAGEN OCH IT-SÄKERHETEN – hotbilder, motåtgärder och behov
53. Brottslighetens kostnader 2023
54. Träffsäkra åtgärder krävs för att bekämpa brottsligheten mot företag

Nordea

55. Anmäl misstänkt bedrägeri
56. Kortreklamation – reklamera kortköp
57. Minska risken att bli utsatt för bedrägeri
58. Skydda företaget och undvik bedrägerier

SEB

59. Integritetspolicy privat
60. Betaltjänstdirektivet PSD2
61. Informationssäkerhet
62. Tips för att undvika bedrägerier
63. Skydda dig mot bedrägerier

Walley

64. Bedrägeripolicy B2B

Polismyndigheten

65. Försök inte lura mig
66. Anmälningarna av bedrägerier ökade under 2023 – det här gör polisen
67. Bedrägeri genom hotellbokningar
68. Bedrägerier – det spelar roll vad du gör
69. Bedrägerier och penningtvätt
70. De dödliga bedrägerierna
71. Det spelar roll vad du gör – bedrägerier
72. Försök inte lura mig – mötesmanual
73. Försök inte lura mig i hemmet
74. Försök inte lura mig på nätet
75. Försök inte lura mig på stan
76. Lägesbild för december 2023
77. Bedrägeri vid näthandel och vilseledande annonser
78. Organiserade bedrägerierna
79. Fakturabedrägeri
80. Identitetsintrång, ID-kapning
81. Kortbedrägeri
82. Nätfiske, phishing
83. Näthandel och vilseledande annonser
84. Romansbedrägeri
85. Telefonbedrägeri, vishing

Svensk Handel

86. Bedrägeri i flera steg lurar företag in i dyra telefonabonnemang
87. Bedrägerier i coronavirusets spår ökar

88. Falska inkassokrav
89. Hantera bluffaktura
90. Information från Polisens bedrägerisektion region Stockholm
91. Kontokortbedrägerier (CNP) – utan fysiskt kort
92. Kreditbedrägeri – beställningar mot faktura som inte betalas
93. Minska risken för kortbedrägerier
94. Utbetalningar till felaktiga konton
95. VD-bedrägeri (business e-mail compromise)
96. VD-bedrägerierna ökar
97. Ökade bedrägerier gör Trygg E-handel ännu viktigare

Villaägarna

98. Bankbedrägerier Finansierar Organiserad Brottslighet
99. Bankerna kan göra mer för att minska bedrägerier
100. Dessa banker skyddar dig bäst mot bedragare
101. Regeringen Kan och Måste Stoppa Bankbedrägerier
102. Större möjligheter att få tillbaka pengarna vid bankbedrägerier

APPENDIX 2

The models here are displayed with the five words that have the highest probability of appearing.

Model with 6 topics:

Topic 1: "medlemsstat", "skalle", "direktiv", "bestämmelse", "betalningsinstrument"
Topic 2: "bankkort", "närstående", "påstå", "prata", "kortbedrägeri"
Topic 3: "villaägare", "bankbedrägerier", "kontoinformation", "betaltjänstdirektivet", "sparbank"
Topic 4: "besparing", "kopia", "logga", "polisanmälan", "stoppa"
Topic 5: "kostnad", "brottslighet", "procent", "konsument", "tillsyn"
Topic 6: "mejl", "faktura", "kortuppgifter", "säljare", "aldrig"

Coherence Score: 0.475561019244068
Perplexity Score: -10.983555659194987

Model with 7 topics:

Topic 1: "bestämmelse", "avse", "konsument", "förordning", "skalle"
Topic 2: "mejl", "faktura", "aldrig", "tipsa", "falsk"
Topic 3: "ehandel", "villaägare", "konsument", "film", "klandervärd"
Topic 4: "beställning", "kopia", "dörr", "adress", "kreditbedrägeri"
Topic 5: "osann", "stega", "utbetalning", "blankett", "inkassokrav"
Topic 6: "kostnad", "brottslighet", "procent", "stöld", "miljard"
Topic 7: "tillsyn", "kundautentisering", "konsumentkreditlagen", "handlare", "konsumentverk"

Coherence Score: 0.42403253955191117
Perplexity Score: -11.318749859456936

Model with 8 topics:

Topic 1: "tillsyn", "bestämmelse", "bankbedrägerier", "förslag", "regering"
Topic 2: "annons", "vilseleda", "skriftlig", "polisanmälan", "betalningsinstrument"
Topic 3: "styrning", "byrå", "hotell", "intern", "förordning"
Topic 4: "oväntad", "business", "kontoinformation", "svårlurad", "behandla"
Topic 5: "utredare", "kortköp", "beställning", "reklamera", "utbetalning"
Topic 6: "företag", "brottslighet", "procent", "bankid", "kostnad"
Topic 7: "inkassokrav", "lägesbild", "bilaga", "genomskåda", "surfplatta"
Topic 8: "villaägare", "film", "blankett", "bankkund", "oreglerad"

Coherence Score: 0.45062032418576004
Perplexity Score: -11.443880256883583

Model with 14 topics:

Topic 1: "konsument", "villaägare", "bestämmelse", "förordning", "utredning"
Topic 2: "annons", "skriftlig", "ortuppgifter", "film", "gärningsperson"
Topic 3: "styrning", "betaltjänstdirektivet", "hotell", "svårlurad", "kontoinformation"
Topic 4: "kont", "bankkont", "falsk", "gammal", "gneta"
Topic 5: "utredare", "bluffakturor", "inkasso", "telefonförsäljning", "fordring"
Topic 6: "mejl", "skalle", "nät", "undvika", "sparbank"
Topic 7: "marknadsföring", "säljare", "telefonabonnemang", "giltig", "avtal"
Topic 8: "tillsyn", "verksamhet", "konsumentkreditlagen", "kundautentisering", "konsumentverk"
Topic 9: "kampanj", "moms", "hane", "cyberbrott", "koppling"
Topic 10: "fakturabedrägeri", "reklamera", "blankett", "kortköp", "stega"
Topic 11: "direktiv", "olovlig", "betalningsinstrument", "grov", "ehandel"
Topic 12: "utbetalning", "träf", "december", "plånbok", "lägesbild"
Topic 13: "företag", "brottslighet", "kostnad", "procent", "bankid"
Topic 14: "handlare", "kortuppgifter", "business", "rutin", "kontouppgifter"

Coherence Score: 0.437632926498301
Perplexity Score: -13.791193040847396

APPENDIX 3

Source code used for the analysis in Python.

```
import re
from nltk.corpus import stopwords
import PyPDF2
from gensim import corpora, models
from gensim.models import CoherenceModel
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.pyplot as plt
import pandas as pd
import seaborn as sns
from sklearn.manifold import TSNE
from bokeh.plotting import figure, show
from bokeh.models import Legend
from bokeh.io import export_png
import simplemma
import logging
from sklearn.metrics.pairwise import cosine_similarity
from wordcloud import WordCloud

# initialising the list with Swedish stop words
stop_words = stopwords.words('swedish')

# making a new list with stop words identified during the analysis
new_stopwords = ["skall", "dock",
"artikel","brott","bedragare","enligt","bedrägeri",
                 "bedr","ägeri","bedr","agare", "ägeribr", "aga", "ägerier",
"du","ed","oller","ontr","pros",
                 "veriering","istället","säkr","försk","etage","samor","ebygga
nde","ottet","andr","identier",
                 "empel","etag","xempel","ända","and","ddar","sky"]

# adding the list of new stop words to the already existing list
stop_words.extend(new_stopwords)

def read_pdf(file):
    '''This function processes a PDF file into a readable format'''
    reader = PyPDF2.PdfReader(file)
    text = ""
    for page in reader.pages:
        text += page.extract_text()
    return text

def preprocess_text(text):
```

```
    ''' This function preprocesses the text and prepares it for analysis with
the topic model. The reason this is done in
    many steps is that it was difficult for the programme to handle if
everything was in the same command'''

    # Defining regular expression pattern to match hyphenated words split by
line breaks
    hyphenated_pattern = r'(\w+)-\n(\w+)'

    # Removing punctuation and handle hyphenated words
    text = re.sub(hyphenated_pattern, r'\1\2', text)

    # Removing non-alphabetical characters and keep line breaks
    text = re.sub(r"[^A-ZÅÄÖa-zåäö ]", "", text)

    # Tokenising the text in different steps. First by splitting the text into
individual words and removing leading or trailing whitespace
    # Then, by removing words that are numeric, shorter than 3 characters, and
words that match the number:number pattern
    tokens = [word.lower().strip() for word in text.lower().split()]
    tokens = [word.lower() for word in tokens if not word.isnumeric()
                                            and len(word) > 3
                                            and not re.match(r".*\d+:\d+.*",
word)]

    # The tokens are lemmatised to their base form
    tokens = [simplemma.lemmatize(word.lower(), lang='sv') for word in tokens]

    # Now stop words are removed
    tokens = [word.lower() for word in tokens if word.lower() not in
stop_words]

    return tokens


def compute_coherence_values(dictionary, corpus, texts, limit, start, step):
    """This function calculates and returns lists of coherence and perplexity
values for different numbers of topics"""

    # Lists to store values are intialised
    coherence_values = []
    perplexity_values = []
    model_list = []
    top_words_list = []
    topic_labels_list = []

    # Now the model are created, iterating through different numbers of
topics. Random state is used to make the result replicable.
    for num_topics in range(start, limit, step):
```

```
        lda_model = models.LdaModel(corpus=corpus, id2word=dictionary,
        num_topics=num_topics, passes=6, random_state=8366)
        # The model is added to the list
        model_list.append(lda_model)

        # Computing coherence score and adding to list
        coherencemodel = CoherenceModel(model=lda_model, texts=texts,
        dictionary=dictionary, coherence='c_v')
        coherence_values.append(coherencemodel.get_coherence())

        # Perplexity is computed and added to the list
        perplexity = lda_model.log_perplexity(corpus)
        perplexity_values.append(perplexity)

        # Adding top words for each model for qualitative assessment
        top_words = lda_model.print_topics(num_words=5)
        top_words_list.append(top_words)

        # Making labels for each topic
        topic_labels = [f"Topic {i+1}" for i in range(num_topics)]
        topic_labels_list.append(topic_labels)

    return model_list, coherence_values, perplexity_values, top_words_list,
           topic_labels_list

def visualise_coh_and_perp(coherence_values, perplexity_values,
top_words_list, topic_labels_list):
    '''This function visualises the coherence and perplexity values for
       different numbers of topics, based on the result from
       compute_coherence_values, as well as top words for qualitative
       assessment'''

    # Printing the top words for each topic
    for i, (top_words, topic_labels) in enumerate(zip(top_words_list,
topic_labels_list)):
        print(f"Model with {i+2} topics:")
        for topic, label in zip(top_words, topic_labels):
            print(f"{label}: {topic}")

    # Visualising the Coherence scores
    start=2; limit=20; step=1;
    x = range(start, limit, step)
    plt.plot(x, coherence_values)
    plt.xlabel("Number of Topics")
    plt.ylabel("Coherence score")
    plt.xticks(range(start, limit))
    plt.show()

    # Listing the coherence score for each number of topic
```

```
    for m, cv in zip(x, coherence_values):
        print("Num Topics =", m, " has Coherence Value of", round(cv, 4))

    # Visualising the perplexity scores
    start=2; limit=20; step=1;
    x = range(start, limit, step)
    plt.plot(x, perplexity_values)
    plt.xlabel("Number of Topics")
    plt.ylabel("Perplexity score")
    plt.xticks(range(start, limit))
    plt.show()

    # Listing the perplexity score for each number of topic
    for m, p in zip(x, perplexity_values):
        print("Num Topics =", m, " has Perplexity Value of", round(p, 4))


def get_document_topic_df(lda_model, corpus):
    '''This function makes a dataframe that shows the probability of each
topic to appear in each document'''

    document_topic_distribution = lda_model.get_document_topics(corpus,
minimum_probability=0)

    # Making a list of the probabilities of each topic appearing, iterating
through all documents
    doc_topic_data = []
    for i, topic_probs in enumerate(document_topic_distribution):
        doc_topic_data.append({f'Topic {topic_id+1}': prob for topic_id, prob
in topic_probs})

    # Creating a dataframe
    document_topic_df = pd.DataFrame(doc_topic_data, index=[f'doc {i+1}' for i
in range(len(corpus))])

    return document_topic_df


def long_format(doc_topic_df):
    '''This function makes the document-topic DataFrame into a long format for
easier visualisation'''

    # Resetting index to make 'Document' a regular column
    doc_topic_df.reset_index(inplace=True)

    # Melting the DataFrame to transform it into long format and renaming the
columns
    doc_topic_long_df = pd.melt(doc_topic_df, id_vars=['index'],
var_name='Topic', value_name='Value')
```

```
    doc_topic_long_df.rename(columns={'index': 'Document'}, inplace=True)

    # Sorting the DataFrame by document and then by topic
    doc_topic_long_df.sort_values(by=['Document', 'Topic'], inplace=True)

    # Setting index to Document
    doc_topic_long_df.set_index('Document', inplace=True)

    return doc_topic_long_df


def topic_df(lda_model):
    '''This function makes a DataFrame for the topics with their top words,
for the
    purpose of making a word cloud'''
    # Displaying the topics in a DataFrame
    topic_terms = lda_model.show_topics(num_topics=-1, num_words=10,
formatted=False)

    # Initialisining an empty DataFrame
    topic_terms_df = pd.DataFrame(columns=['Topic', 'Top Terms'])

    # Iterate over each topic and its associated terms
    for topic_id, terms in topic_terms:
        top_terms = ", ".join([term for term, _ in terms])

        # Creating a DataFrame with the current topic and its top terms
        current_topic_df = pd.DataFrame({'Topic': [topic_id+1], 'Top Terms':
[top_terms]})

        # Concatenating the DataFrame with the existing DataFrame
        topic_terms_df = pd.concat([topic_terms_df, current_topic_df],
ignore_index=True)

    return topic_terms_df


def generate_word_cloud(topic_number, terms):
    '''This function generates word clouds for each topic'''

    # Replacing words to make them correct
    terms = terms.replace("lösenor", "lösenord")
    terms = terms.replace("skalle", "skall")

    # Making the word cloud and visualising it
    wordcloud = WordCloud(width=800, height=400,
background_color='white').generate(terms)
    plt.figure(figsize=(10, 5))
    plt.imshow(wordcloud, interpolation='bilinear')
```

```
    plt.title(f"Word Cloud for Topic {topic_number}")
    plt.axis("off")
    plt.show()


def word_count_dist(corpus):
    '''This function creates a DataFrame, detailing the word count for each
document'''

    # Getting the word count for each document
    doc_lens = [len(d) for d in corpus]

    #Getting the labels for each document
    ind_doc = []
    for i in range(1, 103):
        ind = f"Document {i}"
        ind_doc.append(ind)

    # Adding both lists to a DataFrame
    df = pd.DataFrame(doc_lens, columns=['Word count'], index=ind_doc)

    return df


def visualize_sector_topic_matrix(df):
    '''This function creates a heatmap where it shows the probability for each
topic to belong to each
       document'''
    df.index = range(len(df))
# Create a heatmap using Seaborn
    plt.figure(figsize=(10, 8))
    sns.heatmap(df, cmap='viridis')
    plt.title('Sector-Topic Distribution')
    plt.xlabel('Topic')
    plt.ylabel('Sector')
    plt.show()

def topic_similarity(lda_model):
    '''This function calculates the similarities between topics and visualises
it in a heatmap'''
    topic_vectors = lda_model.get_topics()

# Calculating pairwise cosine similarity between topics
    similarity_matrix = cosine_similarity(topic_vectors)

# Plotting heatmap
    plt.figure(figsize=(10, 8))
    sns.heatmap(similarity_matrix, annot=True, cmap="YlGnBu",
xticklabels=range(1,7), yticklabels=range(1,7))
```

```
    plt.title('Topic Similarity')
    plt.xlabel('Topic')
    plt.ylabel('Topic')
    plt.show()

def TSNE_clustering(lda_model, corpus_tfidf):
    '''This function clusters documents together, and is the foundation for
visualisation to uncover hidden patterns not shown
    by the other functions'''

    # Getting topic weights
    topic_weights = []
    for doc_topics in lda_model[corpus_tfidf]:
        topic_weights.append([weight for _, weight in doc_topics])

    # Making an array of topic weights
    arr = pd.DataFrame(topic_weights).fillna(0).values

    # Keeping the well-separated points
    arr = arr[np.amax(arr, axis=1) > 0.35]

    # Finding the dominant topic number in each doc
    topic_num = np.argmax(arr, axis=1)

    # Doing tSNE Dimension Reduction
    tsne_model = TSNE(n_components=2, verbose=1, random_state=55, angle=.99,
init='pca', perplexity=50, n_iter=1000)
    tsne_lda = tsne_model.fit_transform(arr)
    return topic_num, tsne_lda

def visualize_cluster(document_numbers, topic_num, sector_shapes,
topic_colors, tsne_lda):
    '''This function visualises the t-SNE clustering'''
    # Plotting the Topic Clusters
    plot = figure(title="t-SNE Clustering of Documents according to Topics",
                  width=900, height=700)


    # Adding scatter plots with different shapes and colors for each document
    for doc_num, topic in zip(document_numbers, topic_num):
        sector = find_sector(doc_num)  # Finding the sector of the document
        color = topic_colors[topic]  # Getting color based on topic
        shape = sector_shapes[sector]  # Getting shape based on sector
        plot.scatter(x=tsne_lda[doc_num-1, 0], y=tsne_lda[doc_num-1, 1],
color=color, size=10, marker=shape)

    # Initialising empty lists for the legends
    sector_legend_items = []
    topic_legend_items = []
```

```
    # Creating a legend for each shape and sector
    for sector, shape in sector_shapes.items():
        renderer = plot.scatter(x=[1], y=[1], size=20, color="black",
marker=shape)
        sector_legend_items.append((sector, [renderer]))

    # Creating a legend for each colour and topic
    for topic, color in topic_colors.items():
        renderer = plot.scatter(x=[1], y=[1], size=20, color=color,
marker="circle")
        topic_legend_items.append((f"Topic {topic+1}", [renderer]))

    # Adding the legend to the cluster
    legend_items = sector_legend_items + topic_legend_items
    legend = Legend(items=legend_items, location="bottom_center")
    plot.add_layout(legend)

    # Showing the plot
    show(plot)

    # Saving the plot as a PNG
    export_png(plot, filename="tsne_cluster.png")


def find_sector(doc_num):
    '''Helper function to visualise_cluster, which finds the right sector
depending on the document number'''
    if doc_num in range(1, 17) or doc_num in range(43, 49) or doc_num in
range(65, 86):
        return "Governmental Sector"
    elif doc_num in range(21, 39) or doc_num in range(41, 43) or doc_num in
range(55, 65):
        return "Banking Sector"
    elif doc_num in range(49, 55) or doc_num in range(86, 98):
        return "Business Sector"
    elif doc_num in range(16, 21) or doc_num in range(39, 41) or doc_num in
range(98, 103):
        return "Civil Sector"


def main():
    '''This part executes all the functions in practice'''
    # Initialising an empty list to store the corpus when it has been read
into the programme
    corpus = []

    # Making a list of all documents used for the analysis
```

```
    pdf_files =
["ORG1DOC1.pdf","ORG1DOC2.pdf","ORG1DOC3.pdf","ORG1DOC4.pdf","ORG1DOC5.pdf",
"ORG1DOC6.pdf","ORG1DOC7.pdf","ORG1DOC8.pdf","ORG1DOC9.pdf","ORG1DOC10.pdf",
"ORG1DOC11.pdf","ORG1DOC12.pdf","ORG1DOC13.pdf","ORG1DOC14.pdf",
"ORG1DOC15.pdf",
"ORG2DOC1.pdf", "ORG2DOC2.pdf","ORG2DOC3.pdf","ORG2DOC4.pdf","ORG2DOC5.pdf",
"ORG3DOC2.pdf","ORG3DOC3.pdf","ORG3DOC4.pdf","ORG3DOC5.pdf","ORG3DOC6.pdf",
"ORG3DOC7.pdf","ORG3DOC8.pdf",
"ORG4DOC1.pdf","ORG4DOC2.pdf","ORG4DOC3.pdf","ORG4DOC4.pdf","ORG4DOC5.pdf",
"ORG5DOC1.pdf","ORG5DOC2.pdf","ORG5DOC3.pdf","ORG5DOC4.pdf","ORG5DOC5.pdf",
"ORG6DOC1.pdf",
"ORG7DOC1.pdf", "ORG7DOC2.pdf",
"ORG8DOC1.pdf","ORG8DOC2.pdf",
"ORG9DOC1.pdf","ORG9DOC2.pdf","ORG9DOC3.pdf","ORG9DOC4.pdf","ORG9DOC5.pdf",
"ORG9DOC6.pdf",
"ORG10DOC1.pdf", "ORG10DOC2.pdf", "ORG10DOC3.pdf", "ORG10DOC4.pdf",
"ORG10DOC5.pdf", "ORG10DOC6.pdf",
"ORG11DOC1.pdf", "ORG11DOC2.pdf", "ORG11DOC3.pdf", "ORG11DOC4.pdf",
"ORG12DOC1.pdf","ORG12DOC2.pdf","ORG12DOC3.pdf","ORG12DOC4.pdf",
"ORG12DOC5.pdf",
"ORG13DOC1.pdf",
"ORG14DOC1.pdf",
"ORG14DOC2.pdf","ORG14DOC3.pdf","ORG14DOC4.pdf","ORG14DOC5.pdf",
"ORG14DOC6.pdf","ORG14DOC7.pdf","ORG14DOC8.pdf","ORG14DOC9.pdf",
"ORG14DOC10.pdf","ORG14DOC11.pdf","ORG14DOC12.pdf","ORG14DOC13.pdf",
"ORG14DOC14.pdf","ORG14DOC15.pdf","ORG14DOC16.pdf","ORG14DOC17.pdf",
"ORG14DOC18.pdf","ORG14DOC19.pdf","ORG14DOC20.pdf","ORG14DOC21.pdf",
"ORG15DOC1.pdf","ORG15DOC2.pdf","ORG15DOC3.pdf","ORG15DOC4.pdf",
"ORG15DOC5.pdf","ORG15DOC6.pdf","ORG15DOC7.pdf","ORG15DOC8.pdf",
"ORG15DOC9.pdf","ORG15DOC10.pdf","ORG15DOC11.pdf","ORG15DOC12.pdf",
"ORG16DOC1.pdf","ORG16DOC2.pdf","ORG16DOC3.pdf","ORG16DOC4.pdf",
"ORG16DOC5.pdf"]

    # Making the PDF files readable and adding them to the corpus list
    for pdf_file in pdf_files:
        text = read_pdf(pdf_file)
        corpus.append(text)

    # Preprocessing the corpus and making a new list
    processed_corpus = [preprocess_text(doc) for doc in corpus]

    # Making a dictionary of the corpus, i.e. assigning a numerical value to
each word
    dictionary = corpora.Dictionary(processed_corpus)

    # Filtering out the 4 lowest occurring words, and the 2 most occuring
words, to not skew the result
    dictionary.filter_extremes(no_below=4)
    dictionary.filter_n_most_frequent(2)
```

```
    # Creating a bag of words model of the corpus, and then making it into a
TFIDF model
    corpus_bow = [dictionary.doc2bow(doc) for doc in processed_corpus]
    tfidf = models.TfidfModel(corpus_bow)
    corpus_tfidf = tfidf[corpus_bow]

    # Activating the logging for the purpose of finding the right number of
passes
    for handler in logging.root.handlers[:]:
        logging.root.removeHandler(handler)
        logging.basicConfig(format='%(asctime)s : %(levelname)s :
%(message)s', level=logging.DEBUG,filename='logs.log')

    # Setting the parameters of the model
    model_list, coherence_values, perplexity_values, top_words_list,
topic_labels_list = compute_coherence_values(dictionary=dictionary,
corpus=corpus_tfidf, texts=processed_corpus, start=2, limit=20, step=1)
    visualise_coh_and_perp(coherence_values, perplexity_values,
top_words_list, topic_labels_list)

    # The previous function returns a list of lda_models. This prints out the
parameters to make sure
    # that the right one is chosen
    for idx, lda_model in enumerate(model_list):
        print(f"Parameters for LDA model {idx + 1}:")
        print("Number of topics:", lda_model.num_topics)
        print(f"index is {idx}")

    # Choosing the LDA model based on quantitative and qualitative assessment
    lda_model = model_list[4]


    # Creating a document topic dataframe do see the probabilites of each
topic for each document
    document_topic_df = get_document_topic_df(lda_model, corpus_tfidf)

    # Making it into a long format for easier visualisation, and then saving
it as a CSV file
    document_topic_df_long = long_format(document_topic_df)
    #document_topic_df_long.to_csv('document_topic_matrix_final.csv',
header=True)

    # Visualising the topic similarities
    topic_similarity(lda_model)

    # Creating a word count DataFrame and exporting it to a CSV file
    word_count_df = word_count_dist(corpus)
    #word_count_df.to_csv('word_count_dist.csv', header=True, index=True)
```

```
    # Getting a DataFRame for the top words for each topic and the visualising
it in word clouds.
    topic_terms_df = topic_df(lda_model)
    for _, row in topic_terms_df.iterrows():
        generate_word_cloud(row['Topic'], row['Top Terms'])


    # This section is for visualisation with t-SNE, i.e. clustering. First,
the document numbers,
    # sector shapes and topic colours are defined.

    document_numbers = np.concatenate([
        np.arange(1, 17), np.arange(16,21), np.arange(22, 39),
        np.arange(39,41),
        np.arange(41,43), np.arange(43, 49),
        np.arange(49, 55), np.arange(55, 65), np.arange(65, 86),
        np.arange(86, 98), np.arange(98, 103)
    ])

    sector_shapes = {
        'Governmental Sector': 'circle',
        'Banking Sector': 'square',
        'Business Sector': 'diamond',
        'Civil Sector': 'triangle'
        }

    topic_colors = {
        0: ‘#377eb8’,
        1: '#ff7f08',
        2: '#40e0d0',
        3: '#f781bf',
        4: '#a65628',
        5: '#984ea3'
    }

    # Creating the t-SNE model.
    topic_num, tsne_lda = TSNE_clustering(lda_model, corpus_tfidf)

    # Visualising the t-SNE clustering.
    visualize_cluster(document_numbers, topic_num, sector_shapes,
topic_colors, tsne_lda)



if __name__ == "__main__":
    main()
```