Utrecht University

Coverage of the Dutch nitrogen crisis in Dutch newspapers between 2019 and 2024

Master's Thesis

Rick Marck (7094620)

Supervised by

Dr. Dennis Nguyen

July 2, 2024

Master's programme in Applied Data Science

Preface

Before you lies the master's thesis 'Coverage of the Dutch nitrogen crisis in Dutch newspapers between 2019 and 2024'. This thesis has been written to fulfill the graduation requirements for the master's degree of Applied Data Science at Utrecht University and it has been worked on between April and July of 2024.

The research question, as well as the sub-questions, were formulated together with my supervisor Dr. Dennis Nguyen and I am glad to announce that the research has been finalised successfully. I would like to express my gratitude to Dennis for his guidance in writing this thesis and for his assistance with resolving any difficulties. I would also like to thank Dr. Mirko Schaefer for wanting to assess my thesis as my second supervisor.

I wish you much reading pleasure.

Rick Marck
July 2, 2024

Abstract

The news media play an important role to assist citizens in becoming informed about crises, which allows them to evaluate the causes and responsibilities. When publishing news, news media can utilise frames to influence the perceptions and impressions of their audiences. Previous research has manually analysed the used frames regarding the coverage of the Dutch nitrogen crisis, but did so for a short period of time or did not cover the differences between the analysed outlets. Therefore, journalistic practices, regarding this crisis, have not yet been thoroughly analysed, which shows a gap in the research. The current research tries to fill this gap by analysing longitudinal shifts regarding the coverage of the Dutch nitrogen crisis. The aim of this study is to determine if computational methods can be used to determine imbalances and biases in framing practices, in order to monitor journalistic practices. To research this, 9,374 articles from June 2019 until March 2024 were extracted and analysed using Natural Language Processing techniques. Topic modelling and Named Entity Recognition were applied to perform a framing analysis on the dataset, which were visualised using knowledge graphs. The results show shifts regarding the topics and entities, with the topics being quite diverse and the entities being mostly framed politically. Lastly, the knowledge graphs showed to be successful in capturing the context of the topics and entities, while also indicating a difference between the frames that outlets applied. The results show that these methods can be used to develop a tool for analysing journalistic practices. Development of this tool should start after addressing the limitations, while the current study did not implement sentiment regarding the topics and only allowed for one topic per article. Additionally, it would be good practice to extend the case beyond the nitrogen crisis, to validate the findings of this study. The developed tool can support news organisations in monitoring imbalances and biases regarding their framing, supporting them in critically analysing their reporting styles.

Keywords - Dutch nitrogen crisis, agenda setting, news framing, journalistic quality, topic modelling, Named Entity Recognition, knowledge graphs.

Contents

1 Introduction

1.1 Contextualisation

In the latter part of 2019, thousands of farmers took to the streets to protest, occupying central parts of the Hague and leading to the largest Dutch traffic jam with 1,136 kilometers. Although these demonstrations impressed the public, they also wondered what the reasons were (Van der Ploeg, 2020).

The main reason for the protests was the Dutch high court decision in May of 2019 to suspend permits of construction projects that could pollute the atmosphere with nitrogen compounds (Stokstad, 2019). While the nitrogen oxide emissions have various sources (Stokstad, 2019), 46 percent of the total emissions in the Netherlands are caused by animal husbandry (Van der Ploeg, 2020). Furthermore, in 118 of the 162 Dutch nature reserves, nitrogen deposition exceeded 50 percent of the ecological risk thresholds on average, which attributes to a change in the ecosystems. The permit freeze thus stalled the expansion of farms, residential construction, roads and airport runways (Stokstad, 2019).

The reasoning for this decision leads back to 2015, when the Netherlands introduced a permit system which allowed for construction projects if a regional government could reduce the nitrogen emissions from other sectors. However, environmental groups were not satisfied with this system and sued the Dutch government in 2016, which eventually led to the case ending up at the Court of Justice of the European Union. This court ruled in 2018 that the permit system indeed could not successfully ensure immediate nitrogen reductions and thus ruled against the Dutch government. This verdict led to the Dutch high court decision to suspend construction permits; stating that the Dutch government should invent a better system and create a long-term plan (Stokstad, 2019).

Shortly after, politicians wanted to take immediate action, to allow the restart of new economic developments, like residential construction (Stokstad, 2019). One of the suggested measures was to reduce the highway speed limit from 130 to 100 kilometers per hour, to reduce the emissions slightly (Mendelts, 2020).

As the crisis became more prominent, media coverage of the issue also increased (Mendelts, 2020). The key function of the news media, from a democratic viewpoint, is to assist citizens in becoming informed (Strömbäck et al., 2020). Citizens gain information about a crisis based on the media coverage, which allows them to evaluate its cause and the responsibilities. Additionally, the frames that are used by the media will also have an influence on the public's perception and impressions (S.-K. An & Gower, 2009). Here, framing entails the process of selecting aspects of a perceived reality to make them more prominent, in order to promote a problem or an interpretation (Entman, 1993). Regarding the nitrogen crisis it is therefore essential to understand how the Dutch news media covered the crisis by analysing the applied frames. This can give an insight in how the the media framed the causes and responsibilities of the crisis and how this could have an effect on their audiences.

Additionally, scholars also gained interest in the Dutch nitrogen crisis. A previous study manually analysed frames used in 160 Dutch articles published in 2019 and did not find any differences between the frames, which got designated to the short time period (Visscher et al., 2022). However, they did find a difference between the mentioned actors. Another study had the objective to analyse how the farmers were framed by identifying the roles that were attributed to the farmers (Leeuwen, 2024). This study did cover a greater time period, 2019 until January 2023, though did not discuss any differences between the outlets. Lastly, other scholars studied framing within three Telegram groups that discussed the nitrogen crisis in the year 2022 (van Diemen, 2023).

One of the previous studies mentioned that differences between the outlets could not be identified, likely due to the short time period (Visscher et al., 2022). The remaining two studies did not discuss any differences between the outlets or groups. This shows that journalistic practices have not yet been thoroughly analysed, revealing a gap in this research. Therefore, the current study wants to fill the gap by analysing a greater time period and discussing any differences between the outlets. To do this, 9,374 unique articles from June 2019 until March 2024 will be analysed. Natural Language Processing (NLP) methods will be used to automatically extract and analyse information from the articles. As articles from multiple years were extracted, any changes through time can also be identified, for both frames and societal actors. This was not feasible for the earlier discussed studies, while the articles were analysed manually, which heavily impacted the number of articles that could be analysed.

To conclude, this research wants to study the yearly shifts in the coverage of the Dutch nitrogen crisis and compare any differences between newspapers. The findings can give newspapers an insight in how they frame the nitrogen crisis and how this relates to their journalistic quality. An example of this can be found in the diversity of the reporting, which implies that journalistic coverage should include different aspects and viewpoints of a problem in a neutral way (Prochazka et al., 2018), to improve the quality of the public debate (Baden & Springer, 2017). The findings of this study can also pave the way for the development of a tool using the applied methods, enabling news organisations to monitor their journalistic practices. This can lead to an improvement of journalistic quality, which will also influence the quality of the public debate (Borger et al., 2019).

1.2 Outline

In what follows, the present study will first discuss the previous studies regarding the Dutch nitrogen crisis in the theoretical framework. Then, the link between the studies and the concepts of agenda setting and framing will be discussed. These concepts, as well as their similarities, will be explained in detail, after which the link to journalistic quality will be given. Lastly, the theoretical framework will state the main research question. Next, the methodology will discuss the data extraction process and how the concepts of agenda setting and framing will be analysed. This section will highlight the concepts of topic modelling, entity recognition and knowledge graphs, based on a theoretical foundation. Using the methods, three sub-questions will be drafted, which will be used to answer the main research question. Lastly, the methodology will discuss the preprocessing steps and the used model parameters. Using the methods, the next section will discuss the results in order of the sub-questions. The results section will highlight the important results and will indicate any remarkable differences. The discussion will interpret the results in light of the earlier studies and journalistic quality, after which the limitations, implications and future work will be highlighted. To conclude, the conclusion will summarise the research steps and the important findings, after which the research question will be answered.

2 Related work

2.1 The Dutch nitrogen crisis in the media

With the actuality of the nitrogen crisis increasing, the media abundantly started to report on the issue (Mendelts, 2020). This also led to media and journalism researchers gaining interest in the topic. Generally, frame analysis is identified as a common approach when studying environmental issues (Zottola et al., 2020). When analysing these issues, frames regarding economics, social responsibility, scientific progress and public health are often mentioned. The next three paragraphs will discuss the previous findings of earlier research, regarding the Dutch nitrogen crisis.

2.1.1 Previous studies regarding the Dutch nitrogen crisis

One previous study studied the portrayal of the nitrogen debate by manually analysing 160 Dutch articles from four newspapers on frames and actors of the year 2019 (Visscher et al., 2022). Every article was checked for used frames, which could range from one to multiple. The study found that most articles discussed frames regarding the attribution of responsibility, human interests and conflicts, with morality and economic consequences receiving less attention. Here the conflict frame mostly discussed the farmer protests, while the attribution of responsibility frame mostly focused on how the government caused and should solve the problem. However, the study did not find any differences in the usage of these frames between the analysed newspapers, which could be caused by the narrow time period of four months, as proposed by the researchers. They did find a difference between the portrayed actors however, with the newspaper *de Volkskrant*, having more mentions regarding environmental experts than other newspapers and *het Financieele Dagblad* mentioning construction workers more often. Generally, the vast majority of the mentioned actors were politicians.

Another study had the objective to study the roles that were attributed to the Dutch farmers by the Dutch media, by manually analysing 176 articles published from 2019 until January 2023 (Leeuwen, 2024). The researcher identified three distinct farmer roles, namely: victim, protester and cause. The study found that farmers were mostly framed as victims and the least as protester, but did not discuss any differences between the analysed outlets.

Lastly, another study manually analysed the statements of three Telegram group chats (one Far Right news channel and two 'farmer' groups) discussing the nitrogen crisis (van Diemen, 2023). While this study did not discuss newspapers in general and only looked at the period between June and November 2022, it did identify six frames when analysing the statements. These could be divided into the cause of the nitrogen crisis, ways of responding to the nitrogen crisis and the need for immediate actions.

2.1.2 Relevance and implications of the previous findings

It can be observed that the previous studies found a variety of frames, ranging from responsibility and conflicts (Visscher et al., 2022) to farmer frames (Leeuwen, 2024) and the causes of the nitrogen crisis (van Diemen, 2023). However, it can also be observed that these studies all decided to manually analyse the documents, which prevented the use of a large dataset. Additionally, only one of the studies (Visscher et al., 2022) analysed the differences between various outlets, but only for the year 2019. While one of the studies did cover a greater time period, it did not discuss the differences between outlets (Leeuwen, 2024).

Any changes to the frames and differences between the outlets have not yet been studied, meaning that journalistic practices, regarding this subject, have not been thoroughly analysed yet, which shows a gap. One of the studies expected a difference between the outlets, but could not confirm this (Visscher et al., 2022). The researchers indicated that this could be due to the narrow time period, which resulted in articles discussing roughly the same topics. This further highlights that differences could become clear when the time period increases.

This study wants to fill this gap by analysing the yearly changes of the frames and by discussing any differences between the newspapers. Therefore, framing will be an important concept of this study. Additionally, the significance of agenda setting should not be overlooked, while some researchers argue that there is considerable overlap between the concepts (Coleman et al., 2009). The next section will discuss agenda setting, followed by news framing and the intersection of these two topics. Lastly, the link to journalistic quality will be made, after which the research question will be stated.

2.2 Agenda setting

The concept of agenda setting refers to the process by which mass media present certain issues with frequency and prominence, resulting in large segments of the public perceiving these issues as more important in comparison to other issues (Coleman et al., 2009). One of the initial agenda setting studies (McCombs & Shaw, 1972) indicates that mass media might not be very successful in telling people what to think, but that it heavily determines what the public finds important. However, the influence of agenda setting is also dependent on the category that gets discussed. For example, international conflicts and nuclear arms both have a strong agenda setting influences, in comparison to trade and politics (Wanta & Hu, 1993).

The rationale behind this phenomenon can be split up into two levels. The first level focuses on the objects that get discussed in the news (Weaver et al., 2004), which got stated as the original concept of agenda setting (Coleman et al., 2009); the idea that the amount of coverage will influence the perceived importance.

The second level is present in the fact that journalists can only discuss a few aspects of a news object, because of the limited capacity of the news agenda (Weaver et al., 2004). When discussing a news object (e.g. the nitrogen crisis) journalists can choose which attributes they want to discuss. For the nitrogen crisis this could be the political views for example, which could lead to other aspects being omitted or receiving less attention. In addition to the attributes that describe a topic, the second level also covers the tone in which these attributes get discussed (Coleman et al., 2009).

Due to limited time and space, news outlets will have to select certain stories, while omitting others. This leads to news selection being an essential part of agenda setting, while only the discussed news objects provide salience cues to the readers (Wanta et al., 2004), which will influence the information that the readers will receive. Additionally, agenda setting will influence how readers will develop a scheme and how they place priorities on issues (Wallington et al., 2010). To conclude, it can be stated that the selection of news objects and attributes, due to agenda setting, will have an influence on how the news outlet will present the object to their readers regarding the importance and angle.

2.3 News framing

Another important concept in media studies is framing. Just like an artist can place a frame around their painting, knowing that it will affect how viewers will interpret and react to their art, so can a journalist choose their words and the used images wisely to influence the interpretation of their audiences (Tewksbury & Scheufele, 2019). With millions of citizens turning to the news media daily, making it a corner stone of our democracies (De Vreese, 2005) it is essential to understand how this concept works.

This study will use the framing definition given by Entmann in 1993, namely: *"To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation for the item described."* (Entman, 1993). To elaborate, news media can alter public attitudes by emphasizing certain aspects of an issue more than other aspects. When reporting on poverty for example, news media could focus on unemployed individuals, but also on national policies (Kwak et al., 2020). This will result in the chosen news object becoming more salient.

News framing can be divided into two dimensions, namely emphasis framing and valence framing. Emphasis framing covers the emphasis that gets placed on a topic (Ding & Pan, 2016), while valence framing covers the positive and negative terms that are used to discuss an object (Lyu & Takikawa, 2022). These concepts will be discussed in the next two sections. However, for the scope of this research, only emphasis framing will be considered in the methodologies and analyses.

2.3.1 Emphasis framing

As discussed in the previous section, news media can lay an emphasis on certain aspects of a story. This is also called 'emphasis framing' and is caused by a simplification of reality. News media will focus on a subset of the aspects of an issue to promote a definition, interpretation or evaluation, which will make these aspects more salient (Ding & Pan, 2016). The news frames, or different presentations, can generate different reactions among the individuals that are exposed to the issue (Bullock & Vedlitz, 2017) and can affect how people will respond (Shaw et al., 2021).

However, the effectiveness of emphasis framing also depends on the audience the frame is aimed at. Previous studies have demonstrated the positive effect of framing when the frame matches the audiences' ideology. This can be designated to the phenomenon of 'motivated reasoning', which occurs when an individual will process information with the goal to arrive at a predetermined goal. An example of this is confirmation bias, which happens when a person will try to find a motivated reasoning for their beliefs and attitudes. An individual can, for example, selectively expose itself to confirming information and accept the information quickly (Luong et al., 2019).

Contrary, disconfirmation bias can also occur when an article is not inline with the beliefs of an individual (Luong et al., 2019), which can cause an opposing attitude or behaviour that may result in a boomerang effect (Byrne & Hart, 2009). When this happens a frame will be unable to change people's opinions and can even lower the support for opposing views, which could lead to even more negative attitudes (Luong et al., 2019).

By presenting an issue or object as risky or safe regarding a public domain (e.g. economics or safety), frames can reliably shift people's attitudes (Baden, 2019). This enables news outlets to orient the attention of their audience to a specific dimension, while ignoring other dimensions. These frames are established in the newsroom discourse and can be defined as newsroom frames (B. Scheufele, 2006). Newsroom frames are shared among journalists in the same organisation. Commentaries written by these journalists can thus be seen as an indicator of the frames that dominate the newsroom to which they belong. For example, a journalist may prefer to report a news object as a conflict or economic consequence, based on the subject area the journalist belongs to. A journalist belonging to the business section will likely apply economic framing more frequently, while a political journalist would prefer political framing (Brüggemann, 2014). However, it should be noted that many European public broadcasting systems have legal requirements or expectations, stating that news and information should adhere to neutral and balanced principles (Mercado-Sáez et al., 2019).

Additionally, mentioned actors also play an important role in news framing (Baden, 2019). Scholars studying the coverage of Spanish nuclear energy in newspapers found that politicians received a lot of attention, while citizens, scientists and ecologists did not receive nearly as much. Thus, the press did not connect the various interests and sensitivities of all stakeholders (Mercado-Sáez et al., 2019). Other scholars also identified differences between the prevalence of actor types, with political news dominated by politicians and labour and education news mostly referring to civil society actors (Beckers & Van Aelst, 2019). This shows that news outlets can frame certain objects by how they discuss societal actors.

2.3.2 Valence framing

Frames can affect how people will respond to certain content because of the elicited feelings (Shaw et al., 2021). Researchers determined that how an issue is framed, will have a great impact on the perception, but also on the sentiment of individuals (Balshetwar & Tugnayat, 2017). These effects will occur when an individual will alter their preferences based on almost equivalent phrases or statements (e.g. discussing 90% employment rate versus 10% unemployment rate). This phenomenon is called 'valence framing' and gets defined as the positive or negative terms that will be used to present the targeted objects (Lyu & Takikawa, 2022).

Other scholars investigated news frames and came to the same conclusions. It was found that news framing heavily depends on the emotional responses of the frame's audience and that the reactions will function as a mediator between news framing and the opinions, attitudes and behaviours of the audience (Lecheler et al., 2015). In the perspective of frame valence, one research regarding anti-China sentiment during COVID-19 stated that an increase of negative frames in the news, emphasizing negative aspects of China, can be considered as a potential cause of the expressed anti-China sentiment (in Japanese context) (Lyu & Takikawa, 2022).

Even though valence framing is an important aspect of news framing, this study will only highlight the effects of emphasis framing in the context of the Dutch nitrogen crisis. However, it is still crucial to understand that news framing can be subdivided into multiple layers.

2.4 The intersection of agenda setting and news framing

As previously indicated in section 2.2, agenda setting can be split up into two levels, where the second level covers which attributes of an object will be discussed (from now on indicated as: attribute agenda setting) as well as the tone. Furthermore, the example given in section 2.3 about poverty (when discussing framing), shows that the media are able to choose which aspects (or attributes) they want to discuss. Thus, it can be argued that there is some overlap between attribute agenda setting and the concept of framing.

Yet, there are considerable discussions about this assumption (Coleman et al., 2009). From one viewpoint, framing is seen as a vastly different concept than agenda setting, because it mainly covers the characterisation of an object and how this influences the perception of the public (D. A. Scheufele & Tewksbury, 2007). Contrary, agenda setting relies on the theory of attitude accessibility, which gets achieved by increasing the salience of issues. This will make it easier for the public to retrieve these issues from memory when making, for example, political judgements (Weaver et al., 2004).

Conversely, scholars state that attribute agenda setting and framing do show overlap, as both concepts draw attention to how objects are pictured in the news and how attributes can have an influence on this (Coleman et al., 2009). It gets added that both concepts are concerned with how issues are depicted in the media, not how often (which belongs to the first level of agenda setting) and that both concepts focus on the most salient aspects (attributes) of objects of interest (Weaver et al., 2004).

2.5 Link to journalistic quality

As stated previously, news media can focus on a subset of an issue, by promoting a definition or interpretation, which will make these parts more salient (Ding & Pan, 2016). By selectively emphasising certain aspects of a complex reality, frames will inevitably omit other aspects, which could be equally plausible or relevant. Therefore, framing also plays a crucial role in democratic debates, because it influences how audiences perceive aspects of a problem. Additionally, if an audience has the possibility to consider multiple aspects of a problem, it will influence the quality of the debate (Baden & Springer, 2017).

Regarding journalistic quality, framing is often linked to the concept of diversity, considering the variance in the discussed frames and topics inside media content (Borger et al., 2019). The concept of diversity gets described as the need for journalistic coverage to include different aspects and viewpoints of a problem. These should be considered in a neutral and unbiased way, which shapes impartiality (Prochazka et al., 2018). As framing, media diversity is deemed a crucial process for political and social discussions (Borger et al., 2019).

2.6 Research question

To consider journalistic quality regarding the nitrogen crisis in Dutch news coverage, by looking at the concepts of agenda setting and framing, yearly shifts between the outlets will be analysed. This, to determine any changes in agenda setting and framing through time, by looking at the individual sources. Additionally, it enables assessing the journalistic practices of these sources, by comparing any differences between the outlets. Therefore, this research will try to answer the following research question:

What longitudinal shifts are evident in the coverage of the nitrogen crisis within Dutch newspapers between 2019 and 2024?

3 Methodology

This section will describe the data and methods that were used for the analyses. After describing the data, a theoretical basis will be given for the methods that were used. Based on the methods, three sub-questions will be introduced, which will be used to answer the main research question. Lastly, the parameter settings of the methods will be discussed, to allow replication of this research.

3.1 Data

Using the Nexis Uni database (LexisNexis, n.d.), the data for this research was gathered. To find the relevant articles the term 'stikstofcrisis' (nitrogen crisis) got used and the output got filtered so that only newspapers would be shown, which resulted in a total of 16,292 articles. After preprocessing the data (see section 3.4), the final dataset had a size of 9,374 articles. These articles were published between the 15th of June 2019 and the 30th of March 2024.

The final dataset contains information about the title of the article, its content, source, date and author, resulting in dataset with 9,374 rows and 5 columns. For the analyses the content (body), source and dates were mostly used, to show any differences among the sources and determine changes over time.

3.2 Methods

3.2.1 Topic modelling

There are various methods which can be used to analyse agenda setting and framing computationally. For agenda setting examples of these are topic modelling (Lu et al., 2023; Baturo & Dasandi, 2017) and clustering sentence embeddings (Danner et al., 2022). Additionally, examples of computational framing analysis methods consist of various topic models, cluster analyses and neural networks (Ali & Hassan, 2022). For this study, topic modelling will be used, while it is a promising approach to determine how different topics are discussed and how they relate to a bigger issue (DiMaggio et al., 2013).

The main goal of topic modelling is to determine the characteristics that data points share. In the context of text analysis, this implies that the model will determine which events or concepts are discussed in a document, by extracting latent variables from large datasets (Vayansky & Kumar, 2020). One of the earliest and frequently utilised topic models is called 'Latent Dirichlet Allocation' (LDA). However, a study found that the topics generated by LDA were sometimes too universal and irrelevant (Egger & Yu, 2022) and it highlighted the performance of two other topic models, namely BERTopic and Top2Vec.

After applying both methods on the nitrogen crisis data, it was concluded that BERTopic would be used because of its flexibility, while parameters can be tuned. To create the topics, the model will create embeddings using Sentence-BERT, make semantically similar clusters and use a class-based version of TF-IDF to retrieve the topic representations (Grootendorst, 2022).

3.2.2 Topic model validation

Besides manually validating the topics by extracting representative and random articles as well as the corresponding headers, the topic model will also validated automatically. The two used methods are CV coherence (Borčin & Jose, 2024; Röder et al., 2015) and Normalised Pointwise Mutual Information (NPMI) coherence (Grootendorst, 2022; Y. An et al., 2023). Another study stated that NPMI should generally be considered as the most consistent measure. However it also got highlighted that the measures do not outright predict the classifier performance (Hadiat, 2022), which highlights the importance of manual validation.

3.2.3 Named Entity Recognition

To enrich the topics, Named Entity Recognition (NER) will be used. NER aims to recognize mentioned entities in a text, such as persons, locations, countries and organisations (Li et al., 2020). By implementing NER the link between societal actors and topics can be discovered.

For this task, two models were considered, namely SpaCy (Shelar et al., 2020) and Flair (Schweter & Akbik, 2020), while both methods allow for Dutch NER. After testing both methods, it was found that Flair yielded better and cleaner entities than SpaCy. Therefore, it got determined that Flair (with the model 'ner-dutch') will be used as the NER method for this study. Before this method could be used, new embeddings had to be trained, because the BERT embeddings were updated after the release of the 'ner-dutch' model.

3.2.4 Knowledge graphs

While visualising the topics and entities can be done by using generic charts, another visualisation method that can be used are knowledge graphs (KGs). The graphs can be used to precisely interlink nodes (sources, topics and entities) with edges (co-occurrence). Knowledge graphs aim to describe the semantics of real-world entities and the relations to other entities (Al-Moslmi et al., 2020) in order to convey knowledge about the real world (Schneider et al., 2022). This allows visualising how often Dutch newspapers mention certain topics and how these topics are connected via entities. Furthermore, a previous study has used knowledge graph data as a way to study agenda setting in fake news (Vargo et al., 2018).

To create the knowledge graphs in Python, the library NetworkX will be used, which allows to explore and analyse networks and network algorithms. It also has the functionality to utilise PyGraphviz, which is a drawing package based on Graphviz (Hagberg et al., 2008). After comparing the standard drawing techniques against the Graphviz techniques (Ellson et al., 2002), it was found that Graphviz returned better graphs, especially the 'Neato' layout. Therefore Graphviz in combination with the Neato layout will be used to visualise the knowledge graphs in NetworkX.

3.3 Sub-questions

To answer the main research question (*What longitudinal shifts are evident in the coverage of the nitrogen crisis within Dutch newspapers between 2019 and 2024?*), three sub-questions have been drafted. In order to look at the concepts of agenda setting and framing, the first sub-question will be used to discuss the important topics and will answer the following question: *What are the predominant topics discussed in the media regarding the nitrogen crisis?*

Additionally, to determine any changes in mentioned societal actors and how these are related to the topics, entity recognition will be used. Therefore, the second question was drawn up as: *In what ways are societal actors associated with the topics?*

Lastly, to look at the yearly shifts and determine any differences between the outlets (using the knowledge graphs), the last question was drawn up as follows: *What are the primary differences in coverage among various media outlets?*

3.4 Preprocessing and parameter settings

3.4.1 Dataset preprocessing

As mentioned in section 3.1, the dataset got reduced from 16,292 to 9,374 articles. When loading the original data, there were various non-Dutch articles and articles without content, which resulted in twenty articles being dropped.

Next it was found that the newspaper *Algemeen Dagblad* has seven regional newspapers, which closely follow the publications of the main newspaper. It got decided that every regional newspaper would be changed to *Algemeen Dagblad* making this outlet very substantial. Removing the duplicate articles without combining the regional outlets would result in the articles being spread across these outlets and thus decreasing the importance of the main outlet.

Lastly, all articles were checked for duplication based on their titles, which showed that there were a total of 6,898 duplicate articles. Dropping these articles resulted in a total dataset size of 9,374 unique articles.

3.4.2 Training the BERTopic model

To train the BERTopic model, first the article texts were extracted from the preprocessed dataset and embedded using the *"paraphrase-multilingual-MiniLM-L12-v2"* embedding model (Reimers & Gurevych, 2019).

Next, the parameter settings for the topic model could be defined. The number of important n-grams per topic which should be extracted was set on 10, with an n-gram range of 1-5, which allows the model to extract bi-grams such as 'maximum speed' or 'nitrogen crisis'. Additionally a Dutch stop words list was initialised using the NLTK library and the word 'we' got added, while it occurred very frequently and it did not contain any semantic information. Using the stop words and n-grams, a count vectorizer from scikit-learn got initialised.

After initialising the count vectorizer, the HDBSCAN model that would be used got set up. Here, the important parameter is the minimum cluster size, which has a significant impact on the number of clusters (and thus topics) that will be generated. After some testing, the parameter was set on 25, which ensures that every topic will have at least 25 articles.

Using the embedding, vectorizer and HDBSCAN models, as well as the most important words and the n-gram range, the BERTopic model could be trained. The language was set to multilingual to support extraction of the Dutch topics. After training the model, the topics and their probabilities were estimated and stored inside their corresponding variables.

Lastly, the model identified a substantial amount of outliers, denoted by articles which were difficult to assign to a class and could belong to multiple. The model indicated 5,216 articles (55.6%) as outliers, which were automatically assigned to their most probable class using built-in functions.

3.4.3 Validation and topic labelling

Calculating the coherence scores using the library Gensim resulted in a CV score of 0.86 and an NPMI score of 0.19. This shows that, solely based on the CV score (range 0 - 1), the topic model would perform exceptionally well, however the NPMI score shows that it performs a bit better than average (while NPMI ranges from -1 to 1). Like discussed in the methodology, these metrics do not outright predict the performance of the classifier, which shows the need for manual validation.

The topics were manually validated by extracting the most representative article and a random article per topic, as well as the ten keywords and ten random headlines. Every topic was checked based on the extracted information to determine any inconsistencies. Additionally, it was found that there was enough distinction between the topics for them to be manually labelled. Labelling the topics was done by using the same information as the validation process to provide enough context. The labelled topics along with the keywords and the occurrences can be found in table 3 of appendix A.

To understand the changes of topics throughout time, meta topics were also created. These meta topics contain multiple normal topics, which allows them to cover one big problem or topic, instead of multiple nuanced ones. The topics were manually assigned to the meta topics and the division of these topics across the meta topics can be seen in table 4 of appendix B.

3.4.4 Entity preprocessing

Flair automatically extracts the entities related to persons, locations, organisations and 'miscellaneous'. These entities were extracted per article and then combined in one column, to store all entities per article. Next, the entity occurrences were counted and the top 200 entities were extracted. These entities were checked for duplication and merged were necessary. After merging the entities, the original dataframe got updated. Even though only the top 200 entities were checked for duplication, it does not mean that the other entities were dropped, but these were just deemed as less significant.

One last alteration that was made to the entities was dropping the entity 'Nederland' (the Netherlands), because it was found that it had a negative influence on the visualisations of the knowledge graphs. This while a significant part of the articles discussed the Netherlands in general, resulting in very cluttered, interlinked visualisations. However, the entity was only dropped when creating the knowledge graphs, so it can still be identified in the other graphs relating to the entities.

3.4.5 Knowledge graphs

Lastly the knowledge graphs were created. To prevent the graphs from becoming disorganised the number of sources per graph was set to seven, with the number of topics that would be extracted per source being set to three. Lastly, the number of entities per topic was set to two. The knowledge graphs can be found in figures 6 to 11 of appendix C, containing information about the sources (the blue nodes), topics (red), entities (beige) and the co-occurrences (edge labels). Lastly, the node size gives a visual representation of the importance of the node, with bigger nodes being mentioned more often.

4 Results

This section will discuss the results to answer the research question: *What longitudinal shifts are evident in the coverage of the nitrogen crisis within Dutch newspapers between 2019 and 2024?* This will be done by dividing the main question into three sub-questions; discussing the predominant topics, the associations with societal actors and primary differences between the outlets. However, before discussing the sub-questions, the data will be explored briefly.

4.1 Data exploration

To get an insight in the data, a plot was created showing the number of publications of the seven most important sources, which can be seen in figure 1. Here two major peaks can be seen regarding the nitrogen crisis, namely at the beginning of 2020 and from 2022 until 2023. This plot can be used as a basis for the sub-questions (e.g. when sudden changes in topic occurrences are found).

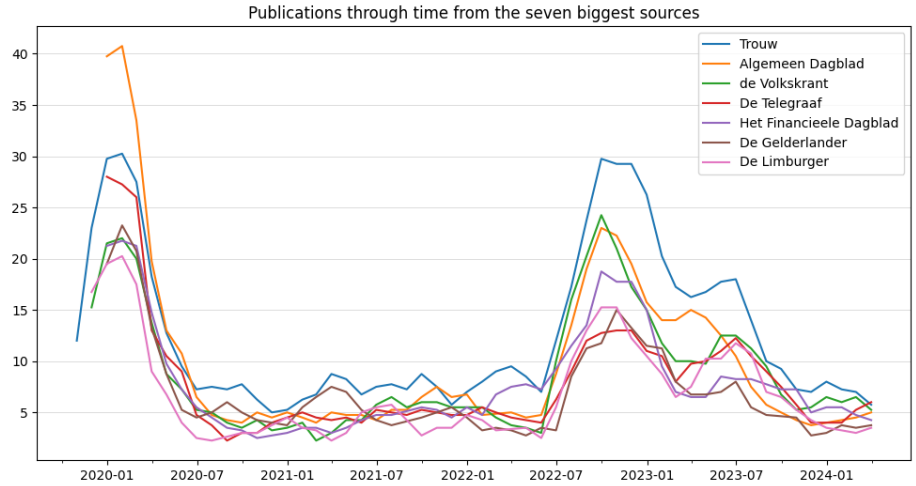

Figure 1: Number of publications from the seven biggest sources.

Additionally, a table containing the days on which the articles were published got created, which can be seen in table 1. Here it is interesting to see that that most articles are published on a Saturday, instead of a regular day during the working week.

Table 1: Published articles per day

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
82	991	1480	1428	1623	1595	2175

4.2 Predominant topics

The first sub-question that will be discussed is: *What are the predominant topics discussed in the media regarding the nitrogen crisis?* To find the predominant topics, the results of the BERTopic model will be used. Running the model on the dataset resulted in 46 distinct topics. The top ten most mentioned topics can be found in figure 2 and an overview of all topics in table 3 of appendix A.

Figure 2 shows the evolving topic importances throughout time. In the beginning of the nitrogen crisis, the most covered topic was the traffic, which was triggered by the decision to reduce the speed limit on Dutch highways from 130 to 100 kilometers per hour (Mendelts, 2020). This decision was closely related to the stoppage to all construction projects (Stokstad, 2019). However it is interesting to see that the topic of housing did not get discussed as much as the traffic. Lastly, with the increase of coverage in 2022, topics like the agricultural impact on the environment and political responses also received more attention, while the topic of traffic almost diminished.

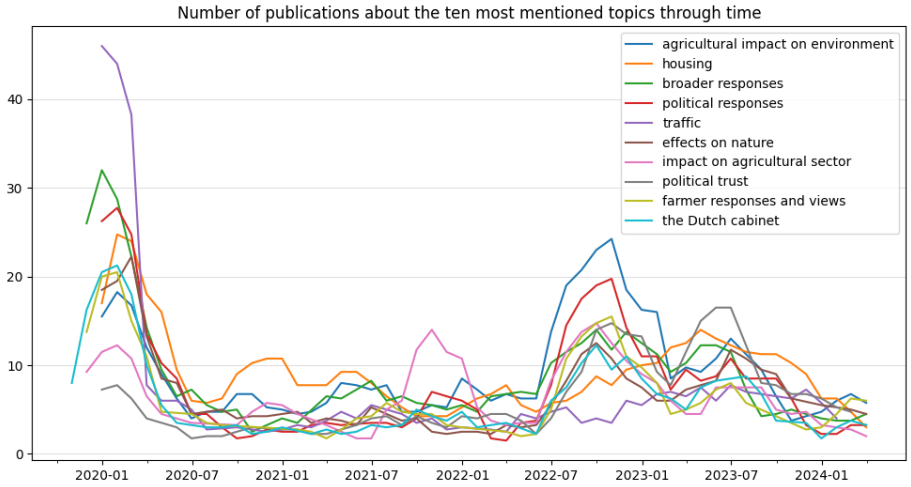

Figure 2: Publications about the ten most important topics through time.

To get an insight in all topic importances throughout time, the meta topics can be used. The graph in figure 3 shows that, instead of the traffic (infrastructure and development), the agricultural issues and tensions received the most coverage in the beginning of the nitrogen crisis, which is also logical because of the farmer protests in the latter part of 2019 (Van der Ploeg, 2020). This figure also shows the general trends of the newspapers, closely following the Dutch elections in 2021 and 2023, which can be seen by the increase of the topic about political responses and governance. Additionally, the graph shows that agricultural issues, infrastructural consequences and political responses received the most coverage, with the other seven meta topics receiving less attention. Based on the meta topics, no main, salient topic could be identified.

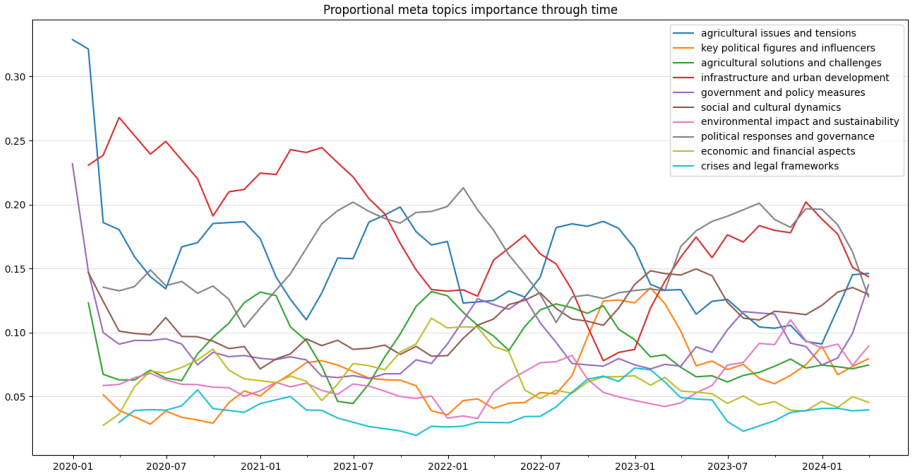

Figure 3: Proportions of the meta topics through time.

4.3 Link between topics and societal actors

This section will consider the question: *In what ways are societal actors associated with the topics?* First the top twenty entities were extracted. Figure 4 shows that political parties (e.g. VVD, CDA, D66, BBB, ChristenUnie and Groenlinks) form a big part of the societal actors. Additionally political figures such as Mark Rutte, Johan Remkes, Carola Schouten and Christianne van der Wal are also mentioned often, which shows that the top entities are mostly political. It becomes clear that the most important entity is Nederland (the Netherlands) which makes sense because of the context of this research.

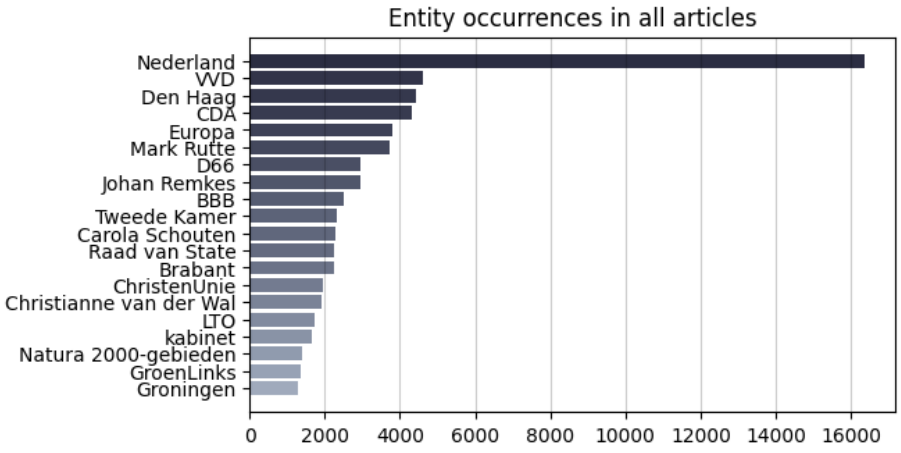

Figure 4: Entity occurrences in all articles.

Additionally the entity mentions through time were plotted in figure 5. This graph mostly follows the same trends as figure 1, though it is interesting to see actors suddenly getting mentioned more often, namely politician Johan Remkes and the political party BBB. The peak regarding Johan Remkes can be linked back to a report that he published in October of 2022 (Remkes, 2022), stating the discussion contents about a sustainable future of the agricultural sector. The peak regarding the political party BBB, which advocates for the agricultural sector, can be designated to their political victory in the Dutch senate elections.

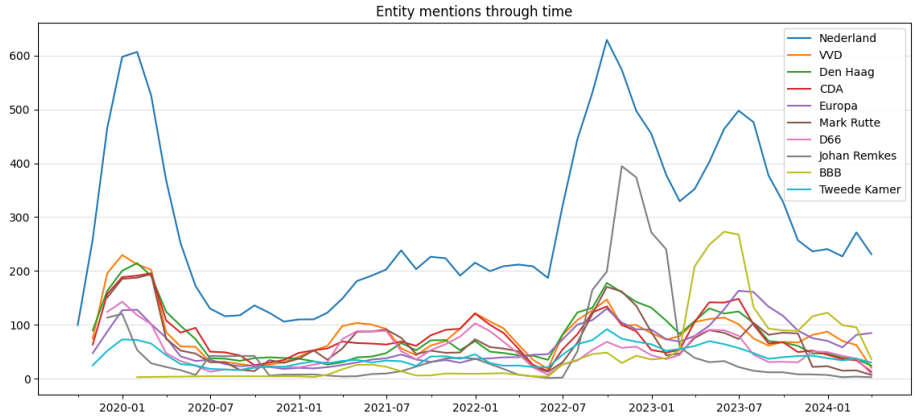

Figure 5: Entity mentions through time.

To find the associations between the societal actors and the topics, the knowledge graphs will be used. Based on figures 6 and 7 it can be determined that Carola Schouten is one of the important societal actors in the years 2019 and 2020. There are links between her and various agricultural and political topics, which makes sense because she was the Dutch minister of agriculture, nature and food quality in these years. Den Haag, the Dutch political centre, also plays an important role in both years. In 2019 the entity often co-occurs with farmer tensions and traffic, while in 2020 it co-occurs more often with economical and social topics. Additionally, Raad van State (the State Council) occurs with two topics in 2019 regarding construction projects.

The year 2021, figure 8, shows more political related entities, with VVD and D66 as political parties, Mark Rutte as the Dutch prime minister and Den Haag. VVD, D66 and Mark Rutte all co-occur quite frequently with economical topics (housing and economy), as well as political topics (political shifts and coalition parties). Den Haag again co-occurs with broader agricultural and economical topics (such as construction project delays). Lastly PBL, a government agency related to the living environment, co-occurs with two agricultural topics. This political focus can be traced back to the Dutch elections of that year.

2022, figure 9, had one very important entity, namely Johan Remkes. This can be linked back to the report that he published in October of that year. One consideration here is that the data also contained a topic called 'Johan Remkes' which led to the Python program automatically combining the topic and entity. This entity co-occurs very frequently with political and agricultural topics, as well as broader responses, which is inline with the contents of his report (Remkes, 2022). The new minister of nature and nitrogen, Christianne van der Wal, also co-occurs with Johan Remkes and environmental topics. Europa (Europe) is also quite significant in the year 2022. This could be due to the proposed European Nature Restoration Act (from June 2022), while the knowledge graphs shows co-occurrences between environmental subjects and Europe.

Again, 2023 (seen in figure 10) has a political focus, just as 2021. BBB and VVD, two political parties, co-occur frequently with social and political topics, while the political leader of another political party, Frans Timmermans, occurs frequently with prime minister Mark Rutte and european laws. This political focus can be referred back to the Dutch elections of 2023.

Lastly, the year 2024 (figure 11) shows that Europe is again an important entity in the nitrogen crisis. It co-occurs with agricultural tensions, views and impacts, but also with european laws. Additionally, Léon Faassen, a deputy of the province of Limburg, also co-occurs with the Dutch cabinet and with broader responses. However, while this study has been conducted in 2024, not much data regarding this year could be gathered.

4.4 Primary differences between outlets

This section will consider the question: *What are the primary differences in coverage among various media outlets?* To answer the question, the yearly differences between the outlets will be discussed, by using the knowledge graphs.

The seven (biggest) outlets that will be compared can be found in table 2. Here, *de Gelderlander* and *de Limburger* are regional newspapers (of the Dutch provinces Gelderland and Limburg), while the others are national newspapers. For the yearly analyses, the top three most discussed topics per source will be used to compare the main focus of every source.

Table 2: Number of articles per analysed source

Source	**N**
Trouw	701
Algemeen Dagblad	582
de Volkskrant	485
de Telegraaf	445
het Financieele Dagblad	437
de Gelderlander	379
de Limburger	375

Figure 6 shows the connections between the sources, topics and entities for 2019. Nine topics can be determined, with the topics about traffic and broader responses being the most discussed. From the seven sources, only *de Gelderlander* did not discuss the broader responses. The same goes for *het Financieele Dagblad* not discussing traffic. Besides the main two topics, *Trouw* considered the effects of the crisis on nature, *Algemeen Dagblad* opted for economical impacts and *de Gelderlander* discussed the agricultural impacts and responses. Lastly the remaining four sources, *de Volkskrant, de Limburger, de Telegraaf* and *het Financieele Dagblad*, discussed the political side of the nitrogen crisis.

The year 2020 (figure 7) shows a shift, with the number of important topics increasing to thirteen, showing that the outlets chose to discuss more varying topics. Again, *Trouw* discussed the effects of the nitrogen crisis on nature, while also discussing the impact from farmers and *Algemeen Dagblad* focused on the economical impacts. The source *de Limburger* slightly shifted to economical views by discussing the construction projects, but also continued to discuss the political side of the crisis. Here the State Council topic can also be linked to the construction projects, while the decisions made by the State Council heavily influenced the progress of these projects. Three other shifts can be seen, with *de Telegraaf* and *het Financieele Dagblad* considering the economical impacts, *de Volkskrant* considering the natural aspects and agricultural impacts, instead of the political side. Lastly, *de Gelderlander* focused on different aspects by discussing festivals, traffic and also nature.

In 2021 (figure 8), eleven main topics can be determined. *De Volkskrant, het Financieele Dagblad, Algemeen Dagblad* and *de Limburger* all discussed housing and the Dutch coalition parties, because of the Dutch elections regarding the House of Representatives. Here *de Volkskrant* shifted from agricultural impacts and effects in 2020, to more political and economical aspects in 2021. The other three sources mainly substituted an economical aspect in 2020 for a political aspect in 2021. Again *Trouw* discussed the effects on nature, but also shifted its attention to the coalition and political shifts. Lastly, the sources *de Telegraaf* and *de Gelderlander* both did not discuss any political aspects or even housing, which is vastly different from the other five sources. While *de Telegraaf* mostly focused on economical aspects and farmers' impact on the crisis, *de Gelderlander* mainly laid its attention on the agricultural sector and traffic.

The knowledge graph of 2022 (figure 9) shows the twelve most important topics of that year. Just like previous years, *Trouw* focused on the effects on nature, but now also laid the focus on social interactions and responses. *De Volkskrant* also reported roughly the same topics, but exchanged one economical topic (housing) in 2021 for a more political topic (the Dutch cabinet) in 2022. On the other hand, *het Financieele Dagblad* shifted from political aspects in 2021 to a more agricultural and natural focus in 2022. Another shift can be seen in *Algemeen Dagblad* which shifted from economical views in 2021, to more diverse reporting in 2022, considering the agricultural sector, housing and political responses. *De Gelderlander* reported roughly the same topics as *Algemeen Dagblad*, thus keeping to their diverse way of reporting. The last two sources, *de Limburger* and *de Telegraaf* both addressed the report from Johan Remkes. Jo-

han Remkes is displayed as an entity in the knowledge graph, which is due to the fact that the topic and entity regarding Johan Remkes were automatically merged by the Python program. Between *de Limburger* and *de Telegraaf*, *de Telegraaf* dedicated more attention to responses of the public and politics, while *de Limburger* decided to report general crises and impacts on the agricultural sector. Thus, both sources shifted from economical impacts to societal responses (*de Telegraaf*) and agricultural impacts (*de Limburger*).

For the year 2023, thirteen important topics can be determined (figure 10). *Trouw, Algemeen Dagblad* and *het Financieele Dagblad* discussed varying topics, with all sources discussing political and social topics. Beside these topics, *Trouw* again focused on nature, while *Algemeen Dagblad* and *het Financieele Dagblad* focused on housing. In comparison to 2022, *Trouw* and *Algemeen Dagblad* did not change their publishing styles, by staying quite diverse, while *het Financieele Dagblad* shifted to more socially focused aspects. *De Limburger* also did not change their reporting that much, exchanging a socially focused topic (crises) in 2022 for a more economical topic (housing) in 2023. Additionally, the sources *de Volkskrant* and *de Telegraaf* focused on political aspects in 2023, exchanging a topic regarding the agricultural sector (*de Volkskrant*) and social aspects (*de Telegraaf*). Lastly, *de Gelderlander* again followed a quite diverse reporting style, only exchanging one topic about the agricultural sector in 2022 to a more social topic (festivals) in 2023.

Lastly, for the year 2024 (figure 11) sixteen topics can be determined. This diversity is due to a lack of data regarding this year. The shifts regarding this year will be discussed more concisely. Both *Trouw* and *Algemeen Dagblad* considered varying topics regarding politics, social and agricultural aspects. It is notable that *Trouw* did not discuss any nature related topics up until March of 2024. The sources *de Limburger* and *de Telegraaf* mainly focused on political and social aspects by considering important public figures, European laws and responses, which denotes a big shift from *de Limburger* in 2024. *De Volkskrant* shifted from political aspects in 2023 to more social aspects, while *het Financieele Dagblad* and *de Gelderlander* shifted to more environmental and economical topics respectively.

5 Discussion

5.1 Interpretation

This section will discuss the interpretation of the results per sub-question, by discussing the expectations, possible explanations and new insights. This will be done by linking the results back to previous research regarding the nitrogen crisis, as well as the general findings of the theoretical framework.

The first sub-question states: *What are the predominant topics discussed in the media regarding the nitrogen crisis?* As stated in the results, shifts regarding the individual topics were found (figure 2) and the meta topics (figure 3) showed that the media tend to focus on agricultural tensions, infrastructure and political responses. The meta topics are mostly in line with previous research, which specified attribution of responsibility (regarding the government) and conflicts (farmer protests) as two important frames (Visscher et al., 2022). These roughly overlap with the political responses and agricultural tensions meta topics. Figure 2 also shows that the topics 'agricultural impact on environment', 'impact on agricultural sector' and 'farmer responses and views' are quite relevant. These also roughly overlap with the findings of previous research, stating that farmers were mostly framed as the main cause, victims and protesters regarding the nitrogen crisis (Leeuwen, 2024).

These findings can be linked back to the concepts of agenda setting and news framing. As seen in figure 2 and discussed in the results, traffic was seen as an important topic in the beginning of the nitrogen crisis, after fading away almost completely. This is a sign of agenda setting, which discusses that the amount of coverage about a topic will influence the perceived importance of the topic (Coleman et al., 2009). Additionally, with agricultural tensions, infrastructure and political responses forming an important part of the meta topics, it shows that the media tend to emphasise these aspects of the nitrogen crisis the most, which is a sign of emphasis framing (Ding & Pan, 2016). However, the meta topics do not show one specific, salient aspect, with the top three topics being almost equally important. Based on the topics, no sole frame could be indicated.

The second sub-question was: *In what ways are societal actors associated with these topics?*. First, the top twenty entities were plotted and it was found that most entities were related to the political sphere. This finding is in line with previous research which found that politicians formed the vast majority of all mentioned actors (Visscher et al., 2022). Previous research found that actors play an important role in framing the news (Baden, 2019), with politicians being mentioned most often in political news (Mercado-Sáez et al., 2019). When only considering the societal actors it could indicate that the media want to frame the nitrogen crisis as a political issue. Additionally, the results show that the main shifts related to the entities were due to the number of published articles. However, two peaks regarding Johan Remkes and BBB were also found, which were due to social developments.

When considering the associations between the topics and the societal actors no standouts could be identified. Political actors are mentioned most often

(with two or more associations to the topics), with agricultural subjects relating to ministers of agriculture for example. However, it was striking to see that the knowledge graph of 2022 (figure 9) was able to capture the essence of the report published by Johan Remkes. As mentioned previously, Johan Remkes published a report containing the contents of various discussions between the Dutch cabinet, the agricultural sector and other stakeholders (Remkes, 2022). The edges in the knowledge graph between Johan Remkes and these various stakeholders (the topics and entities), show that this method is successful in capturing this context and its associations.

The last sub-question stated: *What are the primary differences in coverage among various media outlets?* Previous research regarding the Dutch nitrogen crisis did not find any differences between the usage of frames in the analysed newspapers (Visscher et al., 2022). However, the current study did find small differences between the outlets. Two examples here are *Algemeen Dagblad* and *Trouw*, with the first focusing slightly more on the economical aspects and the latter discussing the effects on nature quite often. The reason for this finding could be the bigger time period of roughly four years, in comparison to previous research, which only analysed articles from several months (Visscher et al., 2022).

This is an interesting finding, while it is expected that European news media will adhere to neutral and balanced principles based on legal requirements (Mercado-Sáez et al., 2019). The discovered viewpoints of *Algemeen Dagblad* and *Trouw* could provide an insight in the newsroom frames of these outlets (B. Scheufele, 2006). For example, for *Trouw* this could imply that their newsroom has a tendency to focus on environmental aspects in their reporting, while *Algemeen Dagblad* focuses more on economical aspects. These frames could interfere with the journalistic concept of diversity, which describes the need to include multiple aspects and viewpoints in journalistic coverage (Prochazka et al., 2018). If these sources actually follow a certain newsroom frame, it could hinder the diversity of their reporting, which, in turn, could have an influence on political and societal discussions (Borger et al., 2019).

5.2 Limitations

The results show that the knowledge graphs gave a good understanding of the associations between sources, topics and entities. Differences between the used frames among sources were found and discussed in the previous section. However, there is one drawback regarding the knowledge graphs in there current state. That is that the graphs do not show the sentiment regarding the sources and topics, which makes it impossible to indicate the tone in which sources discuss the topics. For example, it was found that *Trouw* discusses the environmental side frequently, by discussing the effects on nature. However, it is not clear in what way this topic gets discussed, so it is impossible to say if *Trouw* has a positive or negative viewpoint regarding this topic. To solve this, sentiment could be added to the graphs by colouring the associations, with lighter colours meaning a positive relation and darker colours being more negative (or critical).

Another limitation of this study can be found in the topic assignments. Previous research regarding the nitrogen crisis, using manual analyses, allowed for multiple topics per article (Visscher et al., 2022; Leeuwen, 2024), however the current study only assigned one topic per article. While automatically assigning multiple topics to an article is possible using BERTopic (using the topic distributions), the importance of the topics per article should also be considered. While the topic model could indicate that an article belongs to a topic with a likelihood of 80%, the other topics consisting of the last 20% should not be as prominent in the final analyses as the main topic. Therefore, due to this difficulty and time constraints, it got decided that only one topic would be assigned per article. However, using the functionality of BERTopic and a different implementation of the analyses, by taking these topic importances into account (e.g. using weights), this limitation can be solved.

Although differences were found between the used frames regarding the nitrogen crisis, it is difficult to determine if the outcome of this study can be generalised. The current study focused on one case, the nitrogen crisis, which, as indicated by the mentioned actors, is quite a political subject. Therefore it is difficult to indicate if *Algemeen Dagblad* and *Trouw* actually tend to focus more on the economical and environmental side respectively, while the focus of this study only lied on one case. This can be solved by gathering all articles from these sources, over one or multiple years. Then the analyses can be performed again to determine the frames. This could give an insight in the newsroom frames of these sources and it enables checking their reporting diversity.

While the described methods provide good insights in the used frames regarding the Dutch nitrogen crisis, the results should be approached with caution, while the limitations could have an influence on the final results. However, the provided solutions provide good opportunities to address the limitations.

5.3 Implications

The results of this study pave the way for new developments. As indicated previously, the knowledge graphs were capable of capturing the context regarding topics and entities (for example regarding the report of Johan Remkes in 2022). Additionally, the knowledge graphs indicated differences between certain sources in how they discussed the nitrogen crisis. These findings indicate that the combination of topic modelling, named entity recognition and knowledge graphs is a successful combination for determining news frames. Addressing the mentioned limitations could improve the performance even further.

Using these methods, a tool can be developed for news organisations to check their reporting for frames. This tool can contain the described implementations of Named Entity Recognition and topic modelling, to automatically output the required knowledge graphs. Ideally, the tool should be very flexible, providing support for tuning the topic model and adding additional parameters to the knowledge graphs (e.g. sentiment support). A news organisation could hire someone internally to use the developed tool to analyse all the published articles of the organisation. This employee will have the responsibility

to check the articles for frames and determine any tendencies regarding the reporting, by extracting the articles and visualising the outputs using the tool. If a news organisation has the tendency to primarily report on economical aspects of a news object for example, the responsible employee should inform the organisation about this tendency and encourage them to implement some changes. The tool will thus enable organisations to get an overview of their articles and how they frame them, enabling them to make changes if necessary.

This tool can therefore be used to monitor journalistic practices, by determining how sources frame certain news objects. If news organisations will change their ways of reporting based on the results of the suggested tool, it could have a positive influence on journalistic quality. Preventing a source from mainly discussing one frame by introducing additional viewpoints and other frames, the diversity of their reporting will improve. This will allow their audience to consider multiple viewpoints and will have an influence on the quality of the public debate (Baden & Springer, 2017).

Additionally, the discussed methods can provide journalists with valuable insights about their own reporting style. A different tool, for example a dashboard with information about the journalist, their articles and used frames, could allow journalists to view their reporting style or allow managers to view the styles of their journalists. This could open a discussion between the manager and the journalist about their reporting styles and enable a manager to steer a journalist individually if necessary. Additionally, a journalist can view their own dashboard and perhaps encourage themselves to cover multiple views.

5.4 Future work

A future study could focus on the first two mentioned limitations, regarding the sentiment and the topic assignments. This study could research if an addition of sentiment to the knowledge graphs, combined with the capability to allow for multiple topics per article, would have an influence on the discovered frames. These two limitations can easily be combined, while only the topic assignments has an influence on the frames. Adding sentiment to the knowledge graph will create extra context, providing valuable insights about the associations between sources and topics, or even topics and entities.

Another study could address the third limitation by performing this research again, but for all articles regarding one or multiple sources. As mentioned in the limitations, analysis of all articles of a news organisation will enable that organisation to get an insight in their reporting and enables them to check for reporting diversity. This future study should be performed after addressing the first two limitations, allowing for more (and preciser) results.

Lastly, future work could try to develop a tool using the discussed methods of this study. This could either be the generic tool to check for any frames regarding a whole news organisation, or an implementation of the proposed dashboard, providing insights to the individual journalists and managers. For practicality, it would be a good approach to develop this tool after addressing the mentioned limitations, to check the effectiveness of the methods.

6 Conclusion

This study researched the news reporting regarding the Dutch nitrogen crisis and aimed to answer the following research question: *What longitudinal shifts are evident in the coverage of the nitrogen crisis within Dutch newspapers between 2019 and 2024?* Although previous studies also tried to analyse this subject, the studies manually analysed the articles and did not discuss any differences between the analysed outlets over a longer period of time. This indicated that journalistic practices, regarding this subject, were not yet thoroughly researched. The current study aimed to fill this gap by applying Natural Language Processing over 9,374 articles between June 2019 and March 2024, to enable comparisons through multiple years and to identify any shifts.

To perform this research, first the findings of previous research regarding the Dutch nitrogen crisis were introduced and the link to the concepts of agenda setting, news framing and journalistic quality were discussed. Next, the methods to perform a framing analysis on the extracted data were introduced. The current study made use of Flair and BERTopic to extract the entities and the topics respectively. Alternatives to both methods were tried, but these two methods performed the best for this study. Knowledge graphs were used to visualise the semantic relationships between the sources, topics and entities.

The results of these methods were introduced in order of the drawn subquestions. The first sub-question regarded the predominant topics and no specific salient topic was determined, with agricultural issues, political responses and infrastructural consequences all receiving attention. The results showed that the topic importances also shifted through time. The second sub-question related to the important entities and the associations to the topics. It was found that the most important entities were mostly politically related. The knowledge graphs showed no obvious standouts in the topic and entity associations, but it did show that it was successful in capturing essential information and context regarding the entities. Lastly, the third sub-question focused on the primary differences among outlets. Differences between the outlets were identified, with *Algmeen Dagblad* focusing slightly more on economical aspects and *Trouw* on environmental impacts.

To conclude, longitudinal shifts between 2019 and 2024 regarding the coverage of the nitrogen crisis are evident in both the extracted topics and entities. Additionally, the knowledge graphs supported this finding and showed shifts in the discussed frames between sources. It became evident that the knowledge graphs were successful in capturing the essential context regarding the topics and entities. The current study suggests the development of a tool to combine topic modelling, entity recognition and knowledge graphs. This tool can be used to monitor journalistic practices and could help an organisation to critically analyse their reporting or support an individual employee in their writing. The suggested tool should be developed after carefully considering the limitations, mentioned in the discussion (section 5).

References

Ali, M., & Hassan, N. (2022). A survey of computational framing analysis approaches. In *Proceedings of the 2022 conference on empirical methods in natural language processing* (pp. 9335–9348).

Al-Moslmi, T., Ocaña, M. G., Opdahl, A. L., & Veres, C. (2020). Named entity extraction for knowledge graphs: A literature overview. *IEEE Access*, *8*, 32862–32881.

An, S.-K., & Gower, K. K. (2009). How do the news media frame crises? a content analysis of crisis news coverage. *Public relations review*, *35*(2), 107–112.

An, Y., Oh, H., & Lee, J. (2023). Marketing insights from reviews using topic modeling with bertopic and deep clustering network. *Applied Sciences*, *13*(16), 9443.

Baden, C. (2019). Framing the news. *The handbook of journalism studies*, 229–245.

Baden, C., & Springer, N. (2017). Conceptualizing viewpoint diversity in news discourse. *Journalism*, *18*(2), 176–194.

Balshetwar, S., & Tugnayat, R. (2017). Framing and sentiment: Cumulative effect. In *2017 international conference on energy, communication, data analytics and soft computing (icecds)* (pp. 2873–2877).

Baturo, A., & Dasandi, N. (2017). What drives the international development agenda? an nlp analysis of the united nations general debate 1970–2016. In *2017 international conference on the frontiers and advances in data science (fads)* (pp. 171–176).

Beckers, K., & Van Aelst, P. (2019). Look who's talking: An analysis of actors in television news (2003–2016). *Journalism Studies*, *20*(6), 872–890.

Borčin, M., & Jose, J. M. (2024). Optimizing bertopic: Analysis and reproducibility study of parameter influences on topic modeling. In *European conference on information retrieval* (pp. 147–160).

Borger, M., van Hoof, A., & Sanders, J. (2019). Exploring participatory journalistic content: Objectivity and diversity in five examples of participatory journalism. *Journalism*, *20*(3), 444–466.

Brüggemann, M. (2014). Between frame setting and frame sending: How journalists contribute to news frames. *Communication Theory*, *24*(1), 61–82.

Bullock, J. B., & Vedlitz, A. (2017). Emphasis framing and the role of perceived knowledge: A survey experiment. *Review of Policy Research*, *34*(4), 485–503.

Byrne, S., & Hart, P. S. (2009). The boomerang effect a synthesis of findings and a preliminary theoretical framework. *Annals of the International Communication Association*, *33*(1), 3–37.

Coleman, R., McCombs, M., Shaw, D., & Weaver, D. (2009). Agenda setting. In *The handbook of journalism studies* (pp. 167–180). Routledge.

Danner, H., Hagerer, G., Pan, Y., & Groh, G. (2022). The news media and its audience: Agenda setting on organic food in the united states and germany. *Journal of Cleaner Production*, *354*, 131503.

De Vreese, C. H. (2005). News framing: Theory and typology. *Information design journal+ document design*, *13*(1), 51–62.

DiMaggio, P., Nag, M., & Blei, D. (2013). Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of us government arts funding. *Poetics*, *41*(6), 570–606.

Ding, T., & Pan, S. (2016). Personalized emphasis framing for persuasive message generation. *arXiv preprint arXiv:1607.08898*.

Egger, R., & Yu, J. (2022). A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. *Frontiers in sociology*, *7*, 886498.

Ellson, J., Gansner, E., Koutsofios, L., North, S. C., & Woodhull, G. (2002). Graphviz—open source graph drawing tools. In *Graph drawing: 9th international symposium, gd 2001 vienna, austria, september 23–26, 2001 revised papers 9* (pp. 483–484).

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of communication*, *43*(4), 51–58.

Grootendorst, M. (2022). Bertopic: Neural topic modeling with a class-based tf-idf procedure. *arXiv preprint arXiv:2203.05794*.

Hadiat, A. R. (2022). *Topic modeling evaluations: The relationship between coherency and accuracy* (Unpublished doctoral dissertation).

Hagberg, A., Swart, P. J., & Schult, D. A. (2008). *Exploring network structure, dynamics, and function using networkx* (Tech. Rep.). Los Alamos National Laboratory (LANL), Los Alamos, NM (United States).

Kwak, H., An, J., & Ahn, Y.-Y. (2020). A systematic media frame analysis of 1.5 million new york times articles from 2000 to 2017. In *Proceedings of the 12th acm conference on web science* (pp. 305–314).

Lecheler, S., Bos, L., & Vliegenthart, R. (2015). The mediating role of emotions: News framing effects on opinions about immigration. *Journalism & Mass Communication Quarterly*, *92*(4), 812–838.

Leeuwen, A. v. (2024). *The dutch farmer as boundary object: Media frames and roles of the dutch farmer pertaining to the nitrogen crisis* (Unpublished master's thesis).

LexisNexis. (n.d.). *Nexis uni — academic research tool for universities & libraries.* `lexisnexis.com/en-us/professional/academic/nexis-uni.page`.

Li, J., Sun, A., Han, J., & Li, C. (2020). A survey on deep learning for named entity recognition. *IEEE transactions on knowledge and data engineering*, *34*(1), 50–70.

Lu, G., Businger, M., Dollfus, C., Wozniak, T., Fleck, M., Heroth, T., … Lipenkova, J. (2023). Agenda-setting for covid-19: A study of large-scale economic news coverage using natural language processing. *International Journal of Data Science and Analytics*, *15*(3), 291–312.

Luong, K. T., Garrett, R. K., & Slater, M. D. (2019). Promoting persuasion with ideologically tailored science messages: A novel approach to research on emphasis framing. *Science Communication*, *41*(4), 488–515.

Lyu, Z., & Takikawa, H. (2022). Media framing and expression of anti-china sentiment in covid-19-related news discourse: An analysis using deep learning methods. *Heliyon*, *8*(8).

McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *Public opinion quarterly*, *36*(2), 176–187.

Mendelts, P. (2020). Uit de stikstofcrisis? *NBR Tijdschrift Natuurbeschermingsrecht*, *3*(1), 11–16.

Mercado-Sáez, M.-T., Marco-Crespo, E., & Álvarez-Villa, À. (2019). Exploring news frames, sources and editorial lines on newspaper coverage of nuclear energy in spain. *Environmental Communication*, *13*(4), 546–559.

Prochazka, F., Weber, P., & Schweiger, W. (2018). Effects of civility and reasoning in user comments on perceived journalistic quality. *Journalism studies*, *19*(1), 62–78.

Reimers, N., & Gurevych, I. (2019, 11). Sentence-bert: Sentence embeddings using siamese bert-networks. In *Proceedings of the 2019 conference on empirical methods in natural language processing*. Association for Computational Linguistics. Retrieved from `http://arxiv.org/abs/1908.10084`

Remkes, J. W. (2022, 10). *Wat wel kan. uit de impasse en een aanzet voor perspectief.* `rijksoverheid.nl/documenten/rapporten/2022/10/05/wat-wel-kan`. (rijksoverheid.nl)

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the eighth acm international conference on web search and data mining* (pp. 399–408).

Scheufele, B. (2006). Frames, schemata, and news reporting.

Scheufele, D. A., & Tewksbury, D. (2007). Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of communication*, *57*(1), 9–20.

Schneider, P., Schopf, T., Vladika, J., Galkin, M., Simperl, E., & Matthes, F. (2022). A decade of knowledge graphs in natural language processing: A survey. *arXiv preprint arXiv:2210.00105*.

Schweter, S., & Akbik, A. (2020). *Flert: Document-level features for named entity recognition.*

Shaw, B., Campbell, T., & Radler, B. T. (2021). Testing emphasis message frames and metaphors on social media to engage boaters to learn about preventing the spread of zebra mussels. *Environmental Management*, *68*(6), 824–834.

Shelar, H., Kaur, G., Heda, N., & Agrawal, P. (2020). Named entity recognition approaches and their comparison for custom ner model. *Science & Technology Libraries*, *39*(3), 324–337.

Stokstad, E. (2019). *Nitrogen crisis threatens dutch environment—and economy.* American Association for the Advancement of Science.

Strömbäck, J., Tsfati, Y., Boomgaarden, H., Damstra, A., Lindgren, E., Vliegenthart, R., & Lindholm, T. (2020). News media trust and its impact on media use: Toward a framework for future research. *Annals of the International Communication Association*, *44*(2), 139–156.

Tewksbury, D., & Scheufele, D. A. (2019). News framing theory and research. In *Media effects* (pp. 51–68). Routledge.

Van der Ploeg, J. D. (2020). Farmers' upheaval, climate crisis and populism. *The Journal of Peasant Studies*, *47*(3), 589–605.

van Diemen, C. (2023). Farmers, framing, and the far right: A content analysis of far right framing in the case of the dutch'nitrogen crisis'.

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New media & society*, *20*(5), 2028–2049.

Vayansky, I., & Kumar, S. A. (2020). A review of topic modeling methods. *Information Systems*, *94*, 101582.

Visscher, M., Cucurachi, S., & Smeets, I. (2022). Portrayal of the nitrogen debate in dutch newspapers.

Wallington, S. F., Blake, K., Taylor-Clark, K., & Viswanath, K. (2010). Antecedents to agenda setting and framing in health news: An examination of priority, angle, source, and resource usage from a national survey of us health reporters and editors. *Journal of health communication*, *15*(1), 76–94.

Wanta, W., Golan, G., & Lee, C. (2004). Agenda setting and international news: Media influence on public perceptions of foreign nations. *Journalism & Mass Communication Quarterly*, *81*(2), 364–377.

Wanta, W., & Hu, Y.-W. (1993). The agenda-setting effects of international news coverage: An examination of differing news frames. *International Journal of Public Opinion Research*, *5*(3), 250–264.

Weaver, D., McCombs, M., & Shaw, D. L. (2004). Agenda-setting research: Issues, attributes, and influences. *Handbook of political communication research*, 275–300.

Zottola, A., Atanasova, D., Cardwell, E., Forrester, J., & Stevens, C. (2020). Framing nitrogen pollution in the british press: 1984–2018. *Discourse & Communication*, *14*(1), 84–103.

A Topics

Table 3: Topics

Topic Labels	Translated representation	N
agricultural impact on environment	['farmers', 'agriculture', 'netherlands', 'nature', 'year', 'farmer', 'must', 'well', 'less', 'says']	509
housing	['homes', 'year', '000', 'build', 'new', 'houses', 'construction', 'well', 'percent', 'municipality']	505
broader responses	['nitrogen', 'nature', 'cabinet', 'farmers', 'netherlands', 'must', 'nitrogen crisis', 'well', 'year', 'goes']	496
political responses	['farmers', 'agriculture', 'cabinet', 'lto', 'must', 'well', 'nitrogen crisis', 'nature', 'minister', 'netherlands']	426
traffic	['kilometres', 'hour', 'car', '100', 'per', 'the', 'year', 'per hour', 'kilometres per', 'maximum speed']	388
effects on nature	['nature', 'netherlands', 'should', 'says', 'does', 'year', 'nitrogen', 'areas', 'goes', 'farmers']	363
impact on agricultural sector	['farmers', 'nitrogen', 'nature', 'agriculture', 'should', 'well', 'cabinet', 'companies', 'nitrogen emissions', 'nitrogen crisis']	331
political trust	['people', 'politics', 'well', 'netherlands', 'cabinet', 'den', 'trust', 'year', 'must', 'goes']	326
farmer responses and views	['farmers', 'farmer', 'well', 'year', 'says', 'true', 'go', 'should', 'goes', 'people']	317
the Dutch cabinet	['cabinet', 'cda', 'vvd', 'year', 'nitrogen', 'goes', 'well', 'rutte', 'nitrogen crisis', 'farmers']	315
farmers' tension towards government measures	['farmers', 'den', 'nederland', 'nitrogen crisis', 'well', 'den haag', 'haag', 'says', 'agriculture', 'goes']	314
prime minister Mark Rutte	['rutte', 'vvd', 'well', 'prime minister', 'cabinet', 'year', 'politics', 'people', 'kamer', 'true']	299

Table 3: Topics

Topic Labels	Translated representation	N
political shifts	['bbb', 'party', 'vvd', 'parties', 'well', 'cda', 'people', 'elections', 'year', 'kamer']	293
coalition parties	['cabinet', 'rutte', 'd66', 'vvd', 'cda', 'parties', 'party', 'kamer', 'kaag', 'christenunie']	290
the economy	['year', 'economy', 'netherlands', 'companies', 'growth', 'money', 'cabinet', 'euro', 'percent', 'economic']	224
construction projects	['year', 'construction', 'projects', 'well', 'euro', 'should', 'percent', 'big', 'people', 'new']	209
financial aspects	['farmers', 'cabinet', 'billion', 'should', 'percent', 'emissions', 'euro', 'nitrogen', 'year', 'netherlands']	206
farmers' impact on nitrogen crisis	['nitrogen', 'nature', 'netherlands', 'emissions', 'farmers', 'agriculture', 'ammonia', 'percent', 'should', 'nitrogen emissions']	206
construction projects delay	['projects', 'new', 'construction', 'year', 'nitrogen crisis', 'according', 'according', 'delay', 'project', 'goes', 'extra']	200
social interactions and responses	['people', 'well', 'year', 'where', 'goes', 'says', 'book', 'our', 'we', 'go']	188
the State Council	['county', 'nitrogen', 'council', 'permit', 'state', 'state council', 'ruling', 'according', 'nature', 'construction']	187
cow farmers and their livelihood	['farmers', 'cows', 'company', 'says', 'well', 'year', 'farmer', 'netherlands', 'goes', 'going']	180
Johan Remkes	['remkes', 'cabinet', 'farmers', 'johan', 'johan remkes', 'lto', 'netherlands', 'conversation', 'must', 'wal']	174
pig farmers	['lely', 'farmers', 'meat', 'pigs', 'year', 'says', 'according', 'company', 'sphere', 'netherlands']	167
general crises	['crisis', 'people', 'crises', 'year', 'our', 'well', 'world', 'always', 'word', 'netherlands']	161

Table 3: Topics

Topic Labels	Translated representation	N
political figures	['schouten', 'farmers', 'well', 'says', 'minister', 'adema', 'agriculture', 'plas', 'people', 'should']	153
festivals and events	['00', 'year', 'well', 'circuit', 'festival', 'says', '30', 'again', 'goes', 'bike']	151
agricultural challenges and permits	['farmers', 'county', 'licence', 'cows', 'nitrogen', 'nature', 'animals', 'council', 'mob', 'stable']	140
water quality	['water', 'water quality', 'waters', 'netherlands', '2027', 'water boards', 'should', 'says', 'european', 'substances']	135
european law	['european', 'law', 'member states', 'nature restoration law', 'netherlands', 'eu', 'timmermans', 'parliament', 'brussels', 'commission']	131
coalition in the province of Brabant	['cda', 'vvd', 'brabant', 'coalition', 'party', 'forum', 'brabant', 'hoekstra', 'd66', 'parties']	130
reflections and social analysis	['people', 'well', 'year', 'netherlands', 'book', 'where', 'goes', 'jaspers', 'nature', 'go']	128
cabinet money schemes regarding farmers	['farmers', 'euros', 'cabinet', 'money', 'million', 'billion', 'agriculture', 'scheme', 'nature', 'million euros']	125
potential agricultural solutions	['nitrogen', 'farmers', 'manure', 'ammonia', 'emissions', 'nature', 'percent', 'nitrogen space', 'less', 'netherlands']	119
cabinet climate policy	['cabinet', 'climate', '2030', 'emissions', 'co2', 'netherlands', 'must', 'year', 'new', 'pbl']	114
COVID-19	['virus', 'people', 'well', 'corona', 'our', 'year', 'should', 'goes', 'again', 'netherlands']	107
harbours and shipping	[['port', 'year', 'rotterdam', 'netherlands', 'says', 'gas', 'canal', 'according', 'geerlings', 'water']]	101
airports	['lelystad', 'schiphol', 'airport', 'lelystad airport', 'flights', 'airport', 'opening', 'airport', 'aviation', 'adegeest']	100

Table 3: Topics

Topic Labels	Translated representation	N
spirituality	['00', 'people', 'staphorst', 'church', 'well', 'community', 'year', 'god', 'goes', 'going']	90
Rabobank	['bank', 'rabobank', 'farmers', 'loans', 'draijer', 'banks', 'decraene', 'rabo', 'agriculture', 'well']	82
trains	['trains', 'track', 'prorail', 'transport', 'ov', 'drive', 'train', 'public transport', 'public', 'station']	70
biobased construction	['materials', 'wood', 'biobased', 'building', 'well', 'says', 'co2', 'duine', 'emissions', 'year']	57
asylum seekers	['asylum seekers', 'refugees', 'reception', 'people', 'netherlands', 'well', 'cabinet', 'municipalities', 'apel', 'ter apel']	56
minister Henk Staghouwer	['staghouwer', 'minister', 'henk', 'farmers', 'cabinet', 'kamer', 'christenunie', 'agriculture', 'second', 'henk staghouwer']	41
sustainable energy	['energy', 'lohuis', 'year', 'pure energy', 'solar panels', 'pure', 'efteling', 'heat pumps', 'netherlands', 'park']	38
substance PFAS	['pfas', 'substances', 'standard', '3m', 'soil', 'microgram', 'chemours', 'veldhoven', 'chemical', 'dredgers']	32

B Meta topics

Table 4: Meta topics

Meta topic	Topics	N
infrastructure and urban development	housing traffic biobased construction construction projects delay construction projects trains harbours and shipping airports	1630
agricultural issues and tensions	farmers' tension towards government measures agricultural impact on environment farmers' impact on nitrogen crisis farmer responses and views pig farmers	1513
political responses and governance	political responses political trust coalition parties coalition in the province of Brabant political shifts	1465
social and cultural dynamics	broader responses reflections and social analysis social interactions and responses spirituality festivals and events	1053
government and policy measures	the Dutch cabinet the State Council cabinet money schemes regarding farmers cabinet climate policy European law	872
agricultural solutions and challenges	impact on agricultural sector agricultural challenges and permits potential agricultural solutions cow farmers and their livelihood	770
key political figures and influencers	prime minister Mark Rutte informant Johan Remkes minister Henk Staghouwer political figures	667

Table 4: Meta topics

Meta topic	Topics	N
environmental impact and sustainability	effects on nature water quality substance PFAS sustainable energy	568
economic and financial aspects	the economy financial aspects Rabobank	512
crises and legal frameworks	COVID-19 asylum seekers general crises	324

C Knowledge graphs

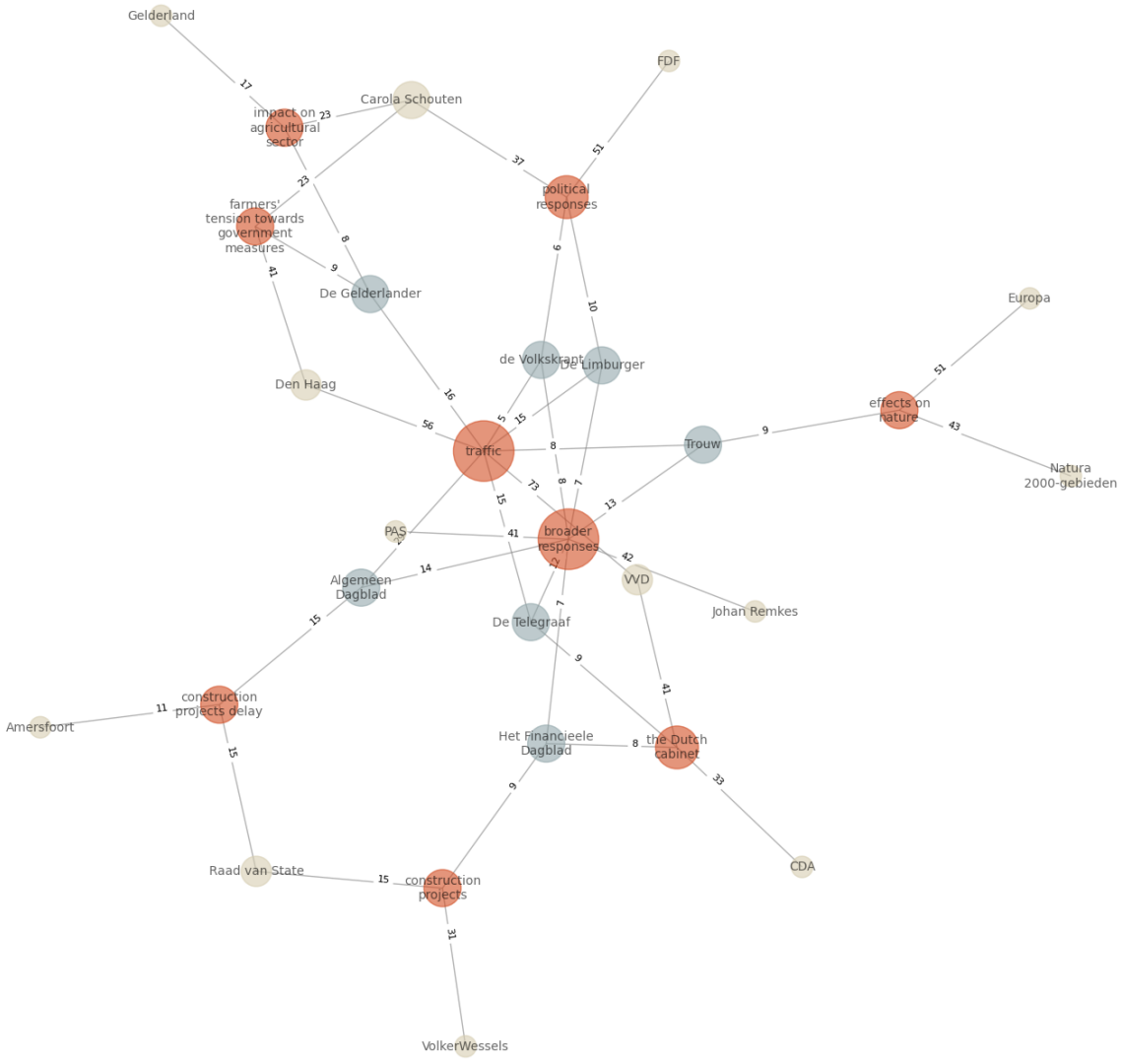

Figure 6: KG with the top sources, topics and entities for 2019.

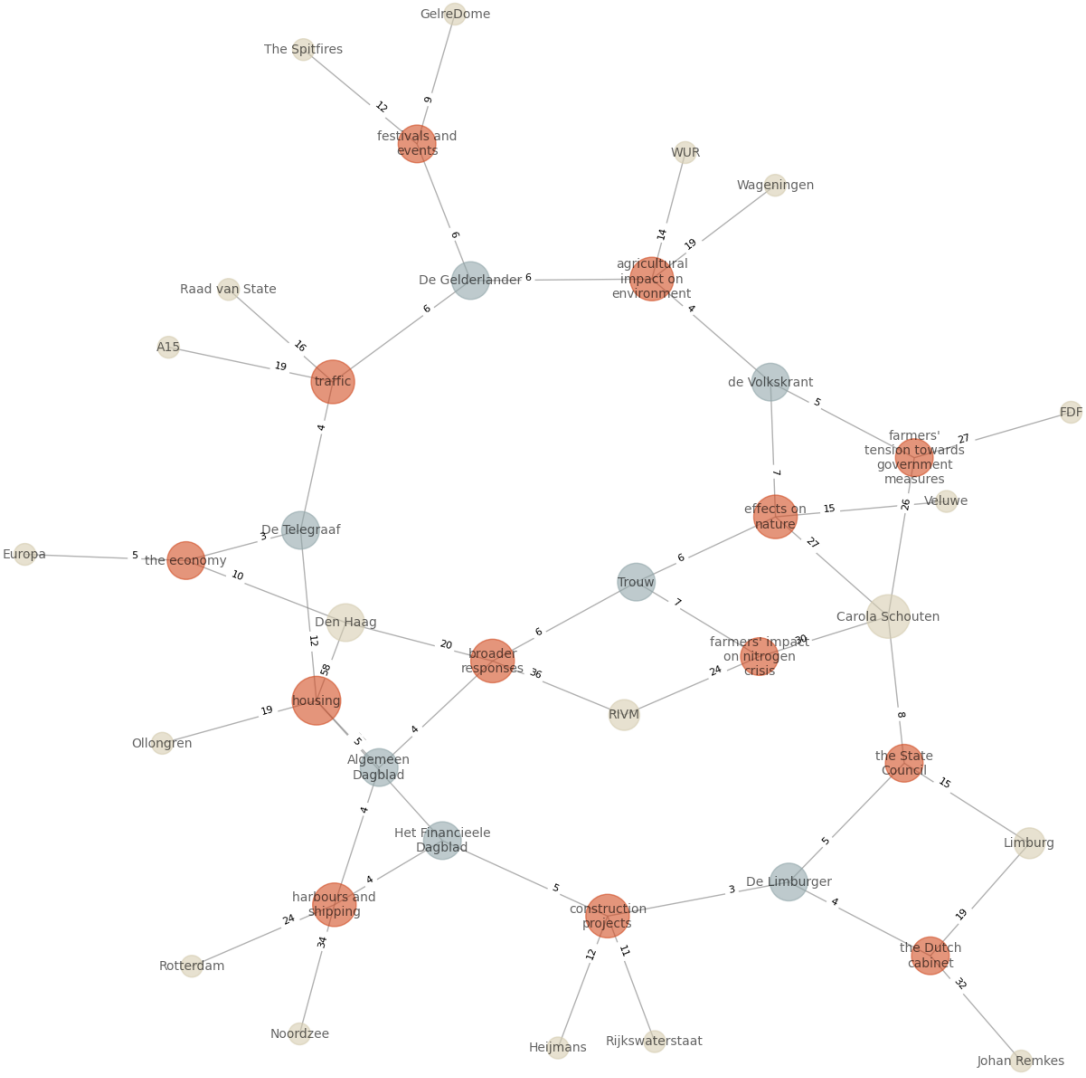

Figure 7: KG with the top sources, topics and entities for 2020.

Topic and entity knowledge graph for the year 2021

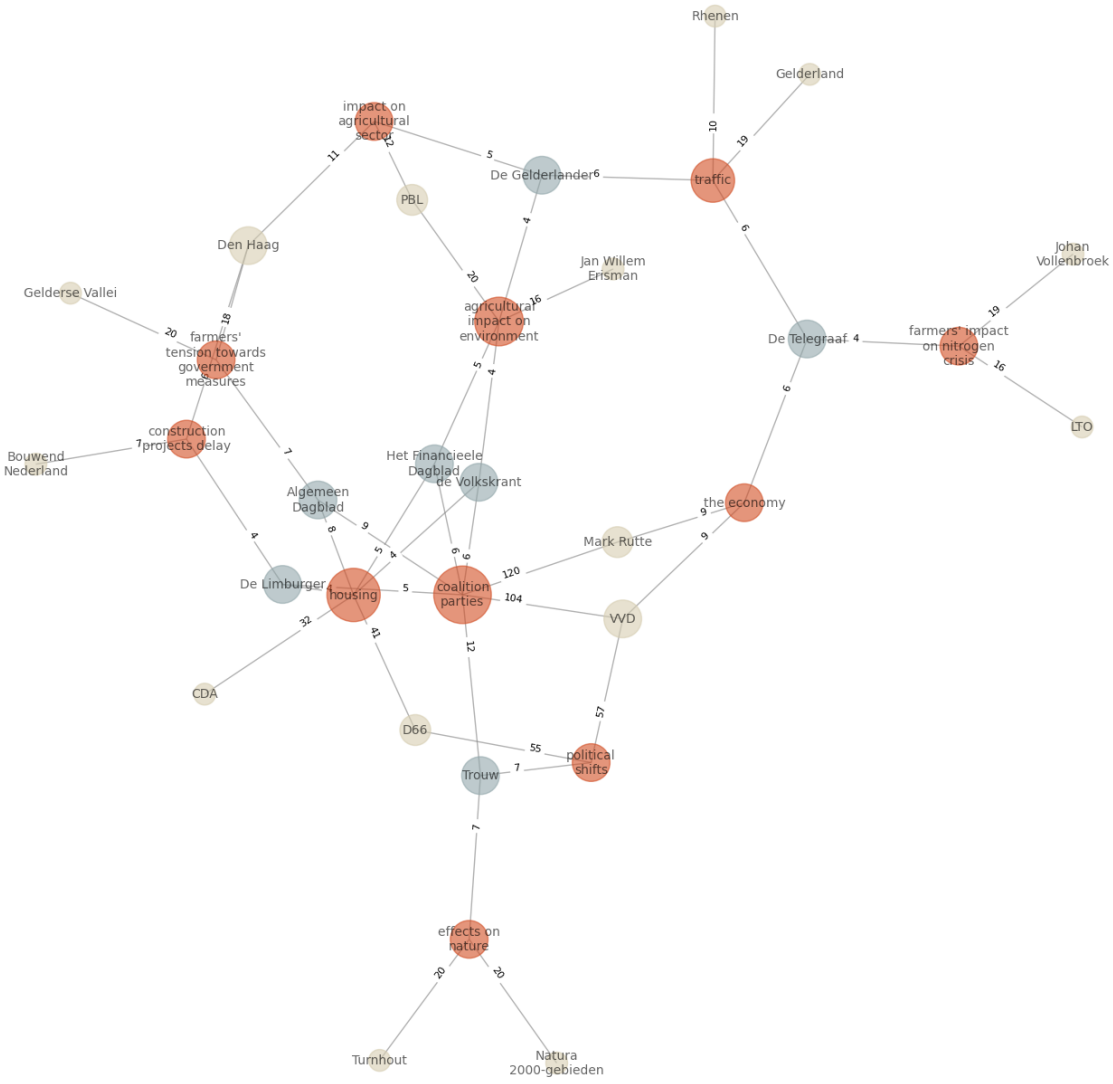

Figure 8: KG with the top sources, topics and entities for 2021.

Topic and entity knowledge graph for the year 2022

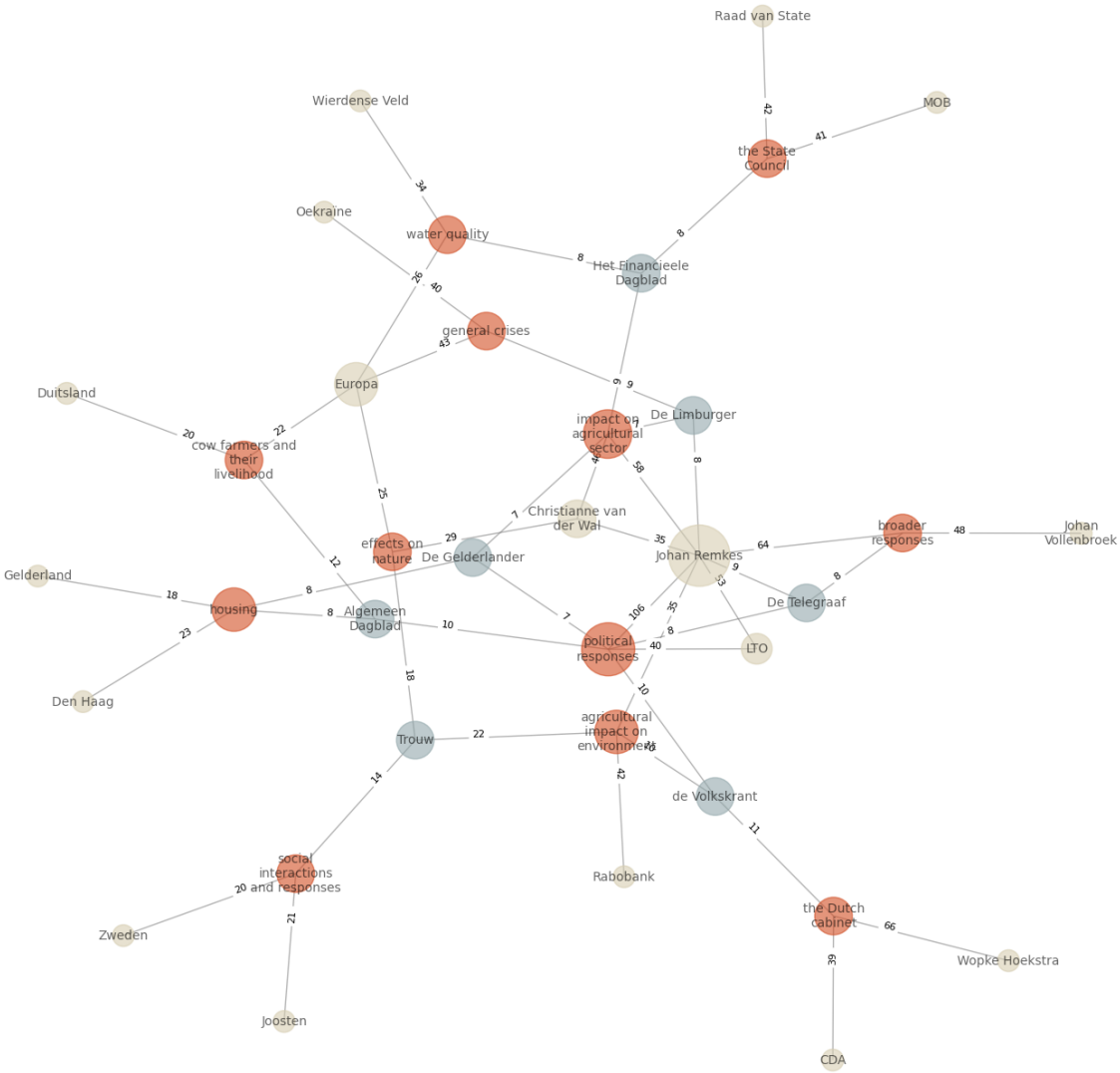

Figure 9: KG with the top sources, topics and entities for 2022.

Topic and entity knowledge graph for the year 2023

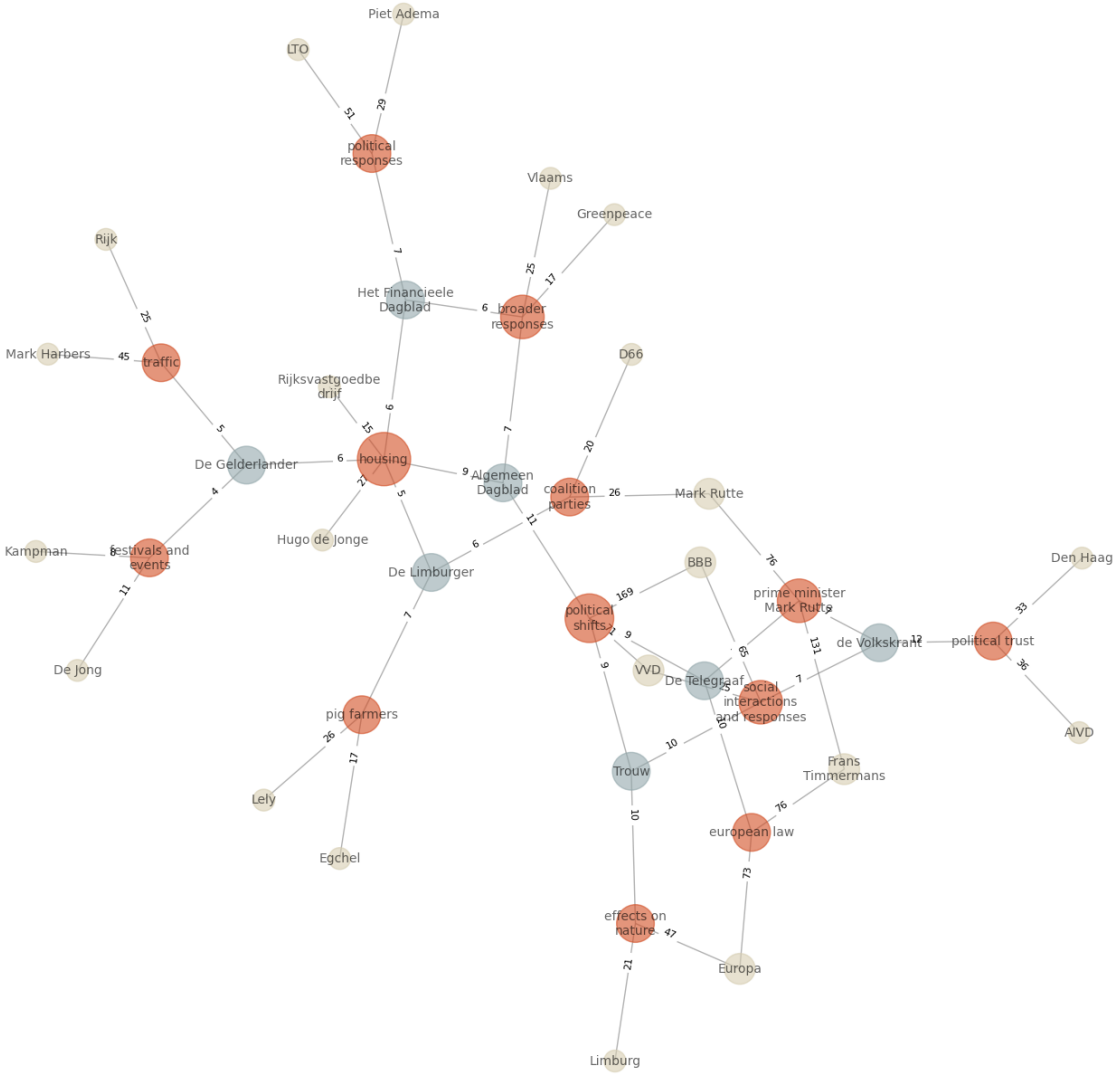

Figure 10: KG with the top sources, topics and entities for 2023.

Topic and entity knowledge graph for the year 2024

Figure 11: KG with the top sources, topics and entities for 2024.