Modelling Protest Related Topics by Combining GPT-4 with State-of-the-Art Approaches

Laurens H.F. Müter[1,2], Christof van Nimwegen[2], and Remco C. Veltkamp[2]

[1] Netherlands Police, Van Deventerlaan 50, z3528 AE Utrecht, the Netherlands,
[2] Utrecht University, Nationaal Politielab AI, Leuvenlaan 4, 3584 CE Utrecht
l.h.f.muter@uu.nl

Abstract. This study explores various topic modelling techniques to analyze social media data and gain insights into a Black Lives Matter demonstration in Amsterdam during the COVID-19 pandemic. Models such as Latent Dirichlet Allocation (LDA) and Hierarchical Dirichlet Process (HDP), enriched with synonyms from the Dutch WordNet and filtered hashtags are used for this purpose. The resulting topics were fed into GPT-4 to generate both general and time-specific descriptions of the events, revealing a strong alignment between the shifts in public sentiment and key events. However, some nuanced details were missed. This research highlights the potential of combining advanced topic modelling techniques with GPT-4 for real-time event monitoring, but it also underscores the challenges of synonym-based analysis.

Keywords: Topic modelling, Open source intelligence, GPT-4

1 Introduction

During the COVID-19 pandemic in 2022, tensions in the Netherlands rose as strict government-imposed regulations led to widespread public dissatisfaction [13, 55]. When these measures were gradually eased, numerous protests emerged, with some specifically focusing on police violence and the Black Lives Matter movement [45]. These protests were sparked by the global outcry following the death of George Floyd. Although the right to protest is a fundamental freedom in the Netherlands, organising large-scale demonstrations during the pandemic posed significant challenges due to restrictions on public gatherings [32]. As the prohibition on large assemblies was still in effect, government officials had to prepare for upcoming protests in a way that balanced the public's right to demonstrate the need to uphold privacy laws and avoid violating open-source intelligence (OSINT) regulations [31].

Within law enforcement, the Open Source Intelligence (OSINT) team is responsible for gathering information that helps manage resources and monitor events leading up to and during protests, enabling timely intervention if violence occurs [37]. However, since personal data cannot be collected or analyzed, OSINT analysts must rely on publicly available information and widely discussed topics. This task is challenging: while too few details may obscure the bigger picture, too much information risks create a clutter of topics that becomes difficult to analyze effectively.

The OSINT team of the Netherlands police has turned to Twitter, now known as X, to gather valuable insights that help plan actions for large social events [56]. The dynamics of how subgroups within the larger population interact are often hidden in the vast amount of daily messages sent on the platform. Identifying these patterns is one way to understand crowd behaviour and sentiment during protests [40].

The topics discussed on social media reflect emerging trends and public sentiment, such as how different subgroups interact with one another [40]. By identifying keywords and topics, this study aims to uncover hidden patterns within the data, using state-of-the-art topic modelling techniques. The main objective of this study is to gain insights into an upcoming protest by analysing tweets over time, focusing on the evolution of topics in a time-series format. This study makes several key contributions. First, it provides an extensive topic model based on a Dutch text corpus within the context of OSINT. This topic model addresses the lack of models based on Dutch-language corpora, which are less common compared to globally spoken languages such as English or Spanish. Second, it compares different methods of topic modelling, evaluating their effectiveness in identifying trends and dynamics in social media data. Third, it analyzes how group dynamics are represented through topic usage on social networks, offering insights into how subgroups interact during protests. These contributions will enhance our understanding of group dynamics before and during large-scale social events and have the potential to strengthen the information position of the open-source intelligence team of the Netherlands police.

2 Related Work

2.1 Topics Network Structures

Twitter is a network where connected individuals discuss topics that interest them at any given time. This results in the coherence of topics spreading within specific subclusters of the social network, especially when a larger group shares in some way their opinions [18]. The dissemination of information is partially driven by interactions between mainstream media and Twitter [11], but it can also be influenced by the group dynamics within the network itself [49]. As topics emerge and circulate within the social network, journalists from mainstream media monitor trending topics on Twitter, creating a two-way interaction between mainstream media and the platform [34].

While most users receive information primarily through mainstream media, a subset of users distrust mainstream media and rely on alternative news sources [5]. These alternative sources can influence smaller clusters within the network, particularly among people who share similar thoughts and beliefs. Thus, individuals who form these clusters have to be isolated to connect with users who hold comparable viewpoints, which can be described as “Connected isolation” [17]. However, the opposite can also occur; connected users may read each other’s messages and gradually align their thoughts and beliefs. A similar phenomenon is observed in student societies, where individuals from different socioeconomic backgrounds and diverse environments find common ground to form connections, which are further strengthened through frequent interactions with processes such as integration, assimilation, separation, and marginalization [52]. This process enables people from varied backgrounds to join groups, find common ground, and reinforce shared beliefs (inspiring one another) [64].

2.2 Relevant and Non-relevant Topics

For the Netherlands police, certain topics are of particular interest, especially those with the potential to disrupt society [37]. These topics warrant close monitoring, such as discussions related to extreme right-wing groups. However, discussions often emerge within seemingly unrelated topics, attracting many users from different backgrounds. This complexity can be understood using intersectionality, as different social identities (eg. race, gender, and class) intersect to shape the experiences and perspectives of users [10]. This makes distinguishing between relevant and irrelevant messages even more challenging.

While initially less relevant, other topics still require attention due to their potential to evolve into more disruptive issues. For instance, when different topics indirectly relate to a specific event, they can galvanize people to unite, potentially leading to violent protests. An example of this is the Middle East conflict, which has led to an increase in antisemitic expressions [19].

On the other hand, many topics are not relevant to police work. These include everyday subjects like pets, hobbies, daily life, and sports. Such topics are not monitored, as government intervention in these areas would be considered an intrusion into privacy rights [30].

Despite the clear distinction in how officials prioritize different topics, there is often an interaction between relevant and non-relevant issues. For example during a church fire in Hoogmade, where a small village lost its church to a blaze [43, 58, 41, 48]. In this scenario, ordinary citizens, in a place where usually little happens, began filming the burning church. The messages, photos, and videos shared by these individuals could provide valuable insights into the situation before officials arrive, helping to assess the severity of the fire and determine if nearby residents were at risk. In such cases, a seemingly minor topic, like a church in a small rural village, suddenly becomes significant for officials.

2.3 Topic Dynamics

The way topics evolve within a network can be classified into linear [62] and viral topic evaluation [22]. Linear topic evaluation is a gradual process, where people slowly form and dissolve connections, and engage with topics through their first or second-degree connections. In contrast, viral topic evaluation occurs rapidly and often unexpectedly, typically triggered by external events.

Linear Topic Evaluation. The linear topic evaluation reflects the slow and steady dynamics of network interactions. Group bonding and the discovery of common ground develop at a gradual pace, with new connections forming and clusters emerging over time [62].

Viral Topic Evaluation. Viral topic evaluation represents a much faster dynamic within Twitter's social network [22]. This rapid change occurs when there are significant shifts in popular topics, often driven by sudden news events or emerging trends. The current list of trending topics can change dramatically due to these events, which are influenced by several underlying factors.

2.4 External Influences on Topics

Since viral topic growth is particularly relevant for OSINT, it is important to delve deeper into the external influences that drive this rapid increase in a topic's popularity. These influences include news bursts, viral topics, and dormant topics that suddenly gain traction [35].

News Bursts. One of the most sudden and impactful dynamics is triggered by external events that occur unexpectedly and significantly affect many people, such as a terrorist attack or an aeroplane crash [38]. These events are inherently difficult to predict, as, by definition, they arise without warning.

International Sparked Topics. Another dynamic occurs when a topic rapidly gains momentum in another country and sparks interest in social clusters within the Netherlands [61]. For example, issues like police violence, gun laws, and abortion are hotly debated in the United States [39], where politicians often discuss these topics with serious implications for the population (e.g. abortion being declared illegal in certain states). Although abortion may not be on the current political agenda in the Netherlands, the intense debate in the U.S. can cause concern among Dutch citizens that similar discussions might emerge locally, potentially bringing the issue into the political spotlight [57]. As a result, protests may arise in the Netherlands driven by Dutch media picking up the issue and transforming it into a trending topic even though it initially was not [60].

Dormant Topics. Finally, there are "dormant topics", which we define as issues that have the potential to rapidly gain popularity but are not actively discussed at the moment [6]. These topics may have received significant attention in the past but currently have little influence on people's daily lives. For example, COVID-19 was a widely discussed topic in 2021 and 2022, but it has not frequently appeared in the news in 2024. However, the disruptive impact of COVID-19 is still fresh in people's memories, so if a new disease or variant emerges, the topic has the potential to quickly regain prominence.

2.5 Topic Modeling

The field of topic modelling on Twitter has seen significant advancements, with a positive trend in the sophistication and effectiveness of these models [28]. Moreover, multiple review studies have been conducted to map the various approaches to topic modelling [7, 9, 54, 12, 23]. These studies describe various approaches to model topics on Twitter. These approaches aim to uncover the hidden semantic structures within tweets, which can be challenging due to the brevity and informal nature of the text. Researchers have either developed customized topic models specifically for Twitter data or utilized pre-existing models to enhance the quality of topic discovery and analysis.

Pre-existing Topic Models. A qualitative approach was often used for coding the identified topics in studies that employed pre-existing topic models. These models generally rely on five main methodologies: linear algebra, probability, statistical distributions, neural networks, and fuzzy clustering. Linear algebra methods, such as the latent semantic analysis, use mathematical techniques to uncover latent structures in text data [14]. Probability-based approaches model the distribution of words across topics to infer underlying themes [20]. The statistical distribution methods, such as LDA, provide a robust framework for topic modelling in large datasets [8]. Neural networks offer powerful tools for modelling complex and non-linear relationships in text data [9]. Lastly, fuzzy clustering methods provide a way to handle the uncertainty and overlapping nature of topics in social media content [27, 26, 24, 25].

Latent Dirichlet Allocation. Latent Dirichlet Allocation (LDA) remains a popular method for extracting key phrases and identifying core themes in social media content. Research is conducted on LDA-based models specifically designed for extracting key phrases related to social media events, demonstrating its effectiveness in organising and summarising the vast amount of data generated on Twitter [63]. Alternatives to LDA, such as LSA (Latent Semantic Analysis) are also successfully applied on tweets [3].

Twitter-LDA. As an alternative, Twitter-LDA is a variant of the Latent Dirichlet Allocation (LDA) model tailored to handle tweets' short and noisy nature. By incorporating external resources such as WordNet and hashtags, researchers improved the quality of the topics identified. For instance, Alkhodair et al. [4] proposed a method that integrates these elements to enhance topic discovery on Twitter, further improving their work of topic detection [2].

Hierarchical Dirichlet Processes (HDP). Hierarchical Dirichlet Processes (HDP) is another pre-existing method adapted for Twitter. Srijith et al. [53] utilized HDP to detect all posts related to a specific event, which is particularly useful for tracking the spread of information during significant occurrences.

Word Co-Occurrence Analysis. Zhang et al. [65] proposed a nonparametric topic model for short text documents like tweets by integrating the recurrent Chinese restaurant process with word co-occurrence analysis. This approach allows for more flexible modelling of the diverse and dynamic content found on Twitter. In short, the Chinese restaurant process (CRP) is a clustering method that simulates how new topics are formed and how existing topics grow based on probabilistic principles [16]. As users generate new content, CRP allows for the dynamic assignment of tweets to existing topics or the creation of new topics, depending on how similar the new content is to previously observed patterns. This flexibility is especially useful for short texts like tweets, where the sparse nature of the data makes traditional topic models less effective. The combination of CRP with word co-occurrence analysis helps capture subtle relationships between words, improving topic detection and clustering accuracy.

2.6 Gaps in Current Literature

Although there is extensive literature on topic detection on social media, there is to our knowledge little known about presenting the detected topics in an automated time-based textual format. Using real-world data within the context of a well-documented event this study aims to explore a possible direction to fill this gap.

3 Method

The main objective of this study is to gain insights into events before, during, and after a large-scale protest. To achieve this, a set of labelled tweets is analyzed using various topic models (LDA, HDP, and bi-word analysis). These clustering techniques are compared using perplexity and coherence scores. Additionally, an hourly time series of words per topic is generated using the Chinese Restaurant Process (CRP) to understand how topics evolve. While the resulting topics contain descriptive terms, in the form of hashtags, the quantity and overlap of these terms make it difficult to see the bigger picture. Moreover, topic changes over time are even more difficult to grasp since more overlapping terms clutter the perspective. Therefore, a subsequent step is taken to employ a GPT-4 model to define the topics and provide a short description. Finally, the resulting descriptions are compared to actual events.

3.1 Dataset

The data is extracted from tweets posted before, during, and after a Black Lives Matter demonstration in Amsterdam during the COVID-19 pandemic [36]. The dataset contains manually labelled tweets with nouns, hashtags, and timestamps. The labelling protocol includes a description of events at the protest.

The tweets were collected using Twitter's open API, within a timeframe ranging from one day before the demonstration to one week after. The same queries were used during each API call to maintain consistency, with terms like "demonstration" included. Unique tweets were selected for analysis, excluding retweets. The tweets were annotated with labels such as "Relevance" (indicating the information is relevant for law enforcement), "Incident_Related" (referring to illegal activities during the protest), "Behavioural_Relevance" (expressions of discomfort related to the protest), "Event_Related" (non-illegal activities), and "High_Priority" (tweets that might require direct action from law enforcement officers).

The team of annotators consisted of six people, who labelled four different datasets over 18 months under the supervision of the Netherlands police's open-source intelligence team. The inter-annotator agreement, based on a subset of labelled tweets, averaged 0.70. The initial dataset contained 84,901 tweets, of which 55,849 were retweets. Of the non-retweeted tweets, 6,155 were labelled, representing 21.19% of the dataset. The data is available from the dedicated paper [36].

3.2 Preprocessing

The tweet data undergoes several preprocessing steps. First, Regex is used to remove URLs, punctuation, and mentions. Then, six different sets of words are created: (1) A set where hashtags are excluded, and the nouns are selected using the Spacy library [21]. The nouns are then lowercase and lemmatized (also by using Spacy). (2) A copy of set (1) without the five most occurring words (such as "demonstration"). (3) A set containing only hashtags. (4) A copy of (3) without the five most occurring hashtags. (5) A copy of (1), where for every noun, a lookup is performed in the Dutch version of WordNet [44] to include synonyms of the words. (6) A copy of the filtered preprocessed nouns from set (2) where the WordNet lookup is performed as in set (5).

3.3 Topic Model Implementation

All code is written in Python using Jupyter Notebook [29] and several standard libraries. All data is loaded into a pandas data frame [33]. The LDA [46, 15] and HDP [59] topic models are employed using the Gensim library [47]. For word co-occurrence analysis, we followed the method described by Alkhodair et al. [4]. This method combines the recurrent Chinese Restaurant Process [1] and a word co-occurrence modelling technique to derive a nonparametric dynamic topic modelling method. Gibbs sampling is used for parameter inference, a Markov Chain Monte Carlo method commonly used in topic models to estimate the posterior distribution of random variables [51].

3.4 Compare Topic Models

Several topic models are employed in this study, including LDA, Twitter-LDA, HDP, and word co-occurrence analysis. All methods are unsupervised, meaning no labels are provided to the models. The output of each model consists of a set of topics, where each topic is associated with a list of words (either hashtags or nouns). Three measures then evaluate the model's performances: perplexity, coherence u-mass, and coherence c_v scores.

Perplexity is a commonly used metric for assessing how well a probabilistic topic model predicts unseen documents. A lower perplexity score indicates better predictive performance [4].

The coherence score, another performance measure, is calculated in four steps: segmentation, probability calculation, confirmation measure, and aggregation. *Segmentation* creates word pairs used to assess topic coherence. The next step, *Probability Calculation*, defines how occurrence probabilities are calculated. The *Confirmation Measure* is calculated using the pairs and the probabilities quantifying the relationship between the word pairs of the segmentation step. Finally, the *Aggregation* step combines the confirmation measures into a single coherence score. The two coherence measures used in this study differ on the *Probability Calculation* step [50].

The u-mass coherence uses an intrinsic measure based on document co-occurrence statistics by calculating the pointwise mutual information between word pairs in a topic based on their co-occurrence in the same documents. In contrast, the c_v coherence incorporates several metrics such as cosine similarity, normalized pointwise mutual information, and a sliding window to capture word context to align better with human judgment [50].

In this study, the Python Gensim library [47] is used for perplexity and coherence score calculations. These evaluations are repeated for different topic sizes, ranging from 3 to 50.

3.5 Model Selection

After evaluating the models, the model with the highest coherence score is selected for further analysis. Additionally, the optimal number of topics is also determined. The output of the selected topic model is a list of terms (hashtags) that represent events, discussions and themes during and after the demonstration. These terms are used to present a description and timely description of the topics within the next two sections.

3.6 Topic Description

Topics identified in the previous step were fed into the pre-trained GPT-4 model ChatGPT [42], which generated summaries and descriptive titles for each topic. This process was repeated for each model, to compare the events extracted from each model with the events of the actual protest. To generate the event descriptions, the GPT-4 was prompted with a specific question, equal for every model: *"The following four lists of words belong to one topic each. Can you provide a descriptive term for each of the four topics?"*. This question is followed by the list of terms outputted by the topic model, using Python's default object representation (e.g. `["term_0", "term_1", ..., "term_n"]`). Additionally, this process was repeated to describe the topics within a 1-hour time interval for every model, using the same prompts.

The returned topic descriptions from the GPT-4 model are manually evaluated against the events detailed in the data paper [36]. These events are summarized in Table 1, which shows a chronological overview of the events before, during, and shortly after the Black Lives Matter demonstration. The events include key moments such as the preparation of the demonstration, the arrival of protesters, and the rising discontent on social media regarding violations of coronavirus measures.

The evaluation criteria assess whether the GPT-4 descriptions accurately and coherently represent the topics of interest related to the Black Lives Matter demonstration. This includes big events such as (social) media attention for the Black Lives Matter protest and (social) media attention for the overcrowdedness and non-compliance with the coronavirus measures during the demonstration. Furthermore, the discontent on (social) media regarding the actions of the local government, and specific events such as controlling the traffic towards Amsterdam was also included. Finally, the start of the speakers' program and the arrival of the mayor were also represented in the evaluation criteria.

Table 1. Chronological summary of the events during the Black Lives Matters demonstration during the COVID-19 pandemic in Amsterdam

Date Time	Event description
12:00	Briefing with the mayor of Amsterdam and the police unit commander, deciding to hold the demonstration at Dam Square instead of the larger Museum Square.
12:00 - 17:00	Demonstration organisers take measures to keep their distance (drawing crosses on the pavement where people can stand, and aisles and signs to remind protesters to comply with the coronavirus measures).
13:00 - 15:00	Announcements on Facebook attract the attention of people who would later join the demonstration.
13:00 - 18:00	Social media attracts Black Lives Matter-related messages.
17:00 - 23:00	Rising public discontent towards the Black Lives Matter protestors for violating the coronavirus measures.
16:30	Important person behind the Black Lives Matter demonstration arrives.
17:00	Speakers' program starts.
15:00 - 18:00	The number of people rises from the expected 500 to more than 10,000.
18:00	Public transport and car traffic towards Amsterdam are under control.
18:15	The mayor of Amsterdam (Femke Halsema) arrives at the demonstration wearing a Keti Koti button.
19:00 - 23:00	Rising discontent on (social) media towards the local government and the mayor of Amsterdam regarding the demonstration and non-compliance with the coronavirus measures.

4 Result

The datasets described in Section 3.2 (nouns (1), nouns filtred (2), hashtags (3), hashtags filtred (4), nouns synonyms (5), and nouns synonyms filtered (6)) are analysed using multiple topic models (LDA and HDP). The output of these models consists of a wordlist which is presented to a GPT4 model with the question to describe the topics. Nest, a time series is generated using co-occurring bi-words in a Chinese Restaurant Process to provide a list of words for every hour of the demonstration. These wordlists are again presented to a GPT4 model to acquire an hour-to-hour description.

4.1 Topic Model Comparison

After an initial test run, the four most frequently used words (e.g. "demonstration", "people", "halsema", and "dam") were removed from the data sets. Whereafter the topic models were trained and evaluated. The LDA model with four topics achieved the highest Coherence score, though this result varied across different models. For the HDP models, the optimal number of topics was 331 for the set that includes synonyms and the most occurring words filter, as shown in Table 3. In contrast, the LDA variant performed best with four topics, although the differences between models with three, four, five and ten topics were relatively small (see Table 2). Coherence scores outside this range fluctuated, with UMass Coherence values varying between -14.42 and -1.08, and c_v scores ranging from .14 to .50 for the LDA model. Meanwhile, the HDP model's coherence scores ranged between .37 and .90.

4.2 GPT-4 Topic Descriptions

A description was generated using the topic model with the highest coherence score, which in this case was the HDP model with filtered nouns enriched with synonyms from the Dutch WordNet library. Since many topics contained only a few words, the top four topics were selected, and each topic's 100 most significant contributing words were obtained using the `num_words` parameter from the Gensim HDP model. These four lists of words, representing the key descriptors for each topic, were then presented to the pre-trained GPT-4 model of ChatGPT [42].

Although the first and last topic descriptions from the GPT-4 output were related to protests and general concerns, these descriptions were not particularly useful for gaining insights into the Black Lives Matter demonstration. Since the model only captured a topic generically describing a protest and a topic to describe "motivations of concerns". The other two topics were related

[1] Within the Gensim implementation, if a term within the topics of the trained model has a word frequency count of 0, it will output a nan-value. For more information see https://github.com/piskvorky/gensim/issues/3040.

Table 2. Comparison between different pre-processing sets and topic sizes of the LDA topic models, using perplexity, u-mass and c_v coherence scores

Pre-processing	Num Topics	Perplexity	Coherence (u-mass)	coherence c_v
nouns filter	3	-7.8636	-4.8808	.1441
nouns filter	4	-7.9642	-5.8235	.2076
nouns filter	5	-8.0396	-6.0511	.2123
nouns filter	10	-8.3440	-6.6365	.2395
nouns filter	50	-10.7724	-9.2745	.3458
nouns	3	-6.8707	-4.4980	.2040
nouns	4	-6.9668	-4.8401	.1988
nouns	5	-7.0346	-4.6870	.2074
nouns	10	-7.2764	-5.1388	.2205
nouns	50	-8.7752	-6.4710	.2607
hashtags filter	3	-5.3034	-8.9223	nan[1]
hashtags filter	4	-5.4283	-10.0962	nan[1]
hashtags filter	5	-5.5118	-9.8972	nan[1]
hashtags filter	10	-5.9106	-10.6925	nan[1]
hashtags filter	50	-11.3902	-14.4253	nan[1]
hashtags	3	-4.9138	-8.4971	nan[1]
hashtags	4	-5.0346	-8.6692	nan[1]
hashtags	5	-5.1212	-8.8447	nan[1]
hashtags	10	-5.4796	-8.6256	nan[1]
hashtags	50	-9.7637	-13.0408	nan[1]
synonyms filter	3	-8.3346	-1.4177	nan[1]
synonyms filter	4	-8.3354	-1.2525	nan[1]
synonyms filter	5	-8.3175	-1.4189	nan[1]
synonyms filter	10	-8.3014	-1.4429	nan[1]
synonyms filter	50	-9.0900	-1.1707	nan[1]
synonyms	3	-7.9876	-1.0887	.5047
synonyms	4	-7.9957	-1.0793	.4878
synonyms	5	-7.9841	-1.0829	nan[1]
synonyms	10	-8.0119	-1.1310	.4974
synonyms	50	-8.6900	-1.2615	nan[1]

Table 3. Comparison between different pre-processing sets of the HDP topic models, using perplexity and coherence. The number of topics is automatically determined.

Pre-processing	Num Topics	Perplexity	Coherence c_v
nouns filter	253	-14.1225	.3823
nouns	229	-13.7121	.3765
hashtags filter	68	-18.1934	.5037
hashtags	52	-18.1759	.5071
synonyms filter	331	-0.7056	.9019
synonyms	271	-1.1854	.7679

to instruments of measure and water control structures (since the demonstration took place at Dam Square which is a Dutch word for dame). Consequently, including synonyms in the provided word lists resulted in synonyms related to water control structures such as dams, levees, and water barriers which deviated from the original topic (the demonstration location). Therefore, a second attempt was made using the set of filtered hashtags (described in Section 3.2), which produced the following output:

- **Political Criticism and Protest Movements**: The first list seems to focus on political figures, protests, government actions, and societal issues. Keywords like "coronatijdperk", "lockdown", "Mark Rutte", and "racismeprotest" suggest a theme of political and social criticism, especially in relation to the pandemic and government policies.
- **Public Sentiment and Discontent**: The second list seems to convey public dissatisfaction and critique of the government and its handling of various crises, particularly during the pandemic. Words like "halsemaweg", "racisme", "coronahoax", and "exit" indicate a strong expression of frustration and opposition, particularly toward certain politicians.
- **Anti-Government and Anti-Establishment Sentiment**: The third list appears to reflect a more radical anti-establishment sentiment, with terms like "antilockdown", "antiantifa", "giletsjaunes", and "antifaterrorist" pointing to opposition against both governmental and societal structures, including protests and conspiracy theories.
- **Critique of Specific Politicians and Policies**: The fourth list is heavily focused on criticizing specific politicians, particularly Femke Halsema, and policies related to COVID-19. Words like "femkelockdown", "faalfemke", and "coronaboetes" suggest a theme of discontent with particular political figures and their decisions.

Since the topic model based on the synonyms preprocessing step performed best in coherence while the hashtag preprocessing resulted in the most effective GPT-4 output (given the evaluation criteria described in Section 3.6), an additional step was incorporated into the model. In this step, synonyms were used for topic modelling, while the associated hashtags were used as input for the GPT-4 model. This approach resulted in clear, well-defined topic descriptions.

The resulting model, visualised in Fig. 1, is divided into six steps: (1) Filtering the tweet text by removing punctuation, hashtags, mentions, and URLs. (2) The hashtags from the filtered text are used to create a lookup table that associates words with hashtags. Each lemmatized noun, from a tweet containing at least one hashtag, is stored in a dictionary along with the associated hashtags. (3) The nouns from step (1) are enriched with synonyms from the Dutch WordNet, using a lookup table to obtain synonyms for each word. (4) The nouns and synonyms are combined into a single list, which serves as the input for the HDP model. (5) The words related to the generated topics are translated back into associated hashtags from step (2), allowing each word to be linked with multiple hashtags. The flattened list of hashtags is then processed to find the n most frequent hashtags per topic. (6) Based on step (5) a prompt is created to generate a concise description for each topic using the GPT-4 model.

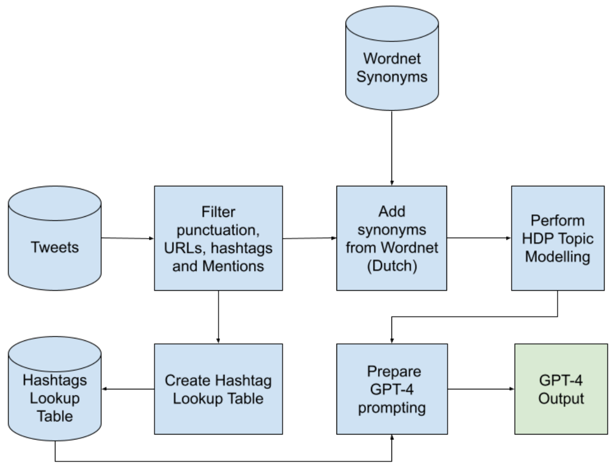

Fig. 1. Topic Model using Wordnet, HDP and GPT-4 to obtain a description of the topics from a given set of tweets

4.3 Time Series

The timestamp of each tweet is used to create a time series, allowing tracking of how a topic evolves during the protest. For this analysis, the tweets are sorted based on their timestamp, and then, for each tweet, a list of bi-words is created. These bi-words are then used in a Chinese Restaurant Process to obtain a word distribution that reflects the significance of each word concerning the topic.

The bi-words t_{bw} of a given tweet t are generated by chaining the list of nouns $w_0, w_1, w_2, w_3...w_n$ extracted from the full text of a tweet t, see Eq. 1. The noun list t_w is then divided into tuples containing pairs of consecutive nouns for each tweet (e.g. (w_0, w_1)), where the last word of tuple n becomes the first word of tuple $n+1$, as shown in Eq. 2.

$$t_w = w_0, w_1, w_2, w_3...w_{|t_w|} \tag{1}$$

$$t_{bw} = (w_0, w_1), (w_1, w_2), (w_2, w_3), ..., (w_{|t_w|-1}, w_{|t_w|}) \tag{2}$$

By summing the bi-terms t_{wb} from every tweet, a network representation of the fifty most common bi-terms can be presented, as shown in Fig. 2.

These bi-word tuples are then used to calculate word distributions $\phi^{v_j k_i}$ and topic distributions $\theta^{k_i a_b}$, using the formula from Alkhodair et al. [4] (see Eq. 3 and 4). Here, W represents the count matrix of bi-words, where $W(v_j, k_i)$ depicts

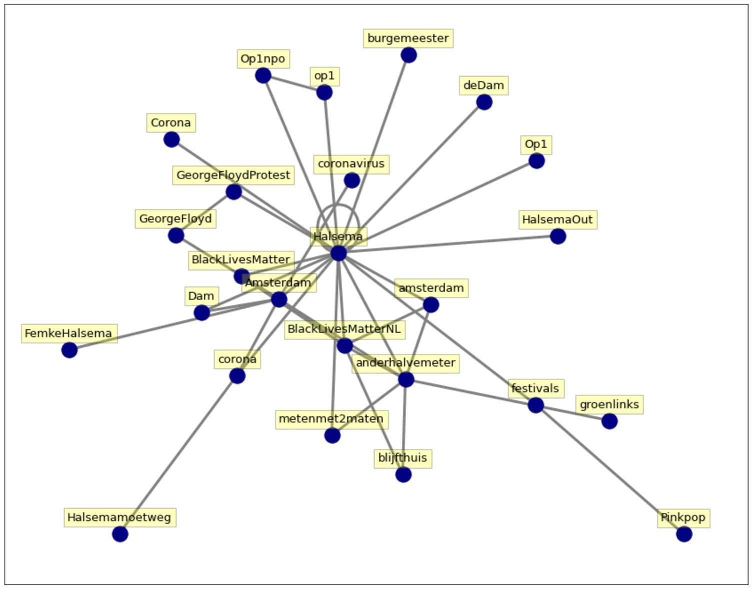

Fig. 2. Network representation of the fifty most common bi-words of the nouns extracted from the tweet texts of the entire dataset

the number of times term v_j occurs within topic k_i. For the topic distribution, T is the count matrix for every topics-author pair where $T(k_i, a_b)$ depicts the number of terms a_b used for topic k_i.

$$\phi^{v_j k_i} = \frac{W(v_j, k_i) + \beta}{\sum_{v'_j} W(v'_i, k_i) + V\beta} \tag{3}$$

$$\theta^{k_i a_b} = \frac{T(k_i, a_b) + \alpha}{\sum_{k'_i} T(k'_i, a_b) + K\alpha} \tag{4}$$

The word distributions are calculated for one-hour intervals to obtain a time series of evolving topics. The parameters are held consistent over the distribution calculations ($\alpha = 10$, $\beta = .1$) with the number of topics fixed at four. The hundred most common terms are then selected to represent a topic at time n.

Using the time series of a topic, the top most common terms can be followed, as displayed in Fig. 3, where the most common terms follow up on each other while displaying changes within the topic over time.

After gathering the list of terms per topic for each hourly epoch, a GPT-4 prompt is created, starting with the following instruction: "*The following list contains 23 lists related to Topic 1, with each list representing a one-hour timestep (starting at 00:00h). Given the words from each of these 23 lists, can you provide a timeline on how Topic 1 evolves over the 23 hours represented by these lists?*".

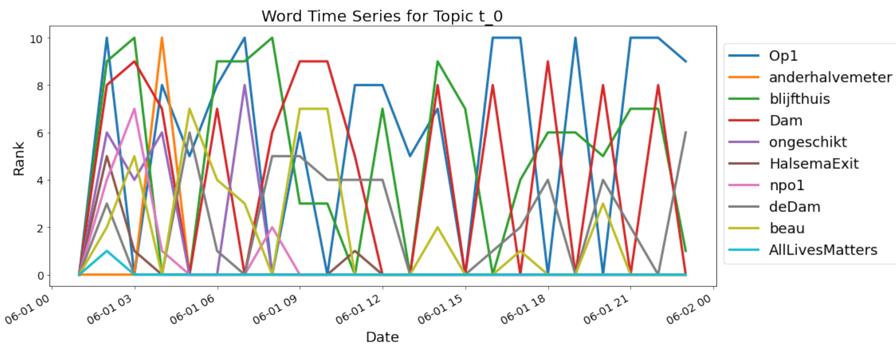

Fig. 3. Evaluation of the top most common terms per hour during the demonstration for topic 0

The output provided a detailed description of how public sentiment towards the demonstration and the government shifted over time. The timeline was divided into periods of two to five hours, with each time chunk given a descriptive title.

- Hours 0-5: Public Health Measures and Protests Begin to Gain Attention
- Hours 6-9: Heightened Political Discontent and Calls for Resignations
- Hours 10-12: Protests at Dam Square Intensify
- Hours 13-16: Tensions Rise Between Health Concerns and Social Movements
- Hours 17-19: Escalating Political Fallout
- Hours 20-23: Consolidation of Key Themes – Pandemic and Protests

Each title also includes a brief description of key highlights derived from the hashtags. For example, during the 13:00-16:00 time chunk:

- A mix of COVID-19 concerns ("lockdown", "COVIDIOTS") with social justice movements (Black Lives Matter) and counter-movements ("AllLivesMatter").
- Mentions of George Floyd and references to global racial justice protests highlight the international context that is fueling local demonstrations.
- Focus on government and leadership failures, with increased mentions of Rutte (Dutch Prime Minister) and "HalsemaExit".

Finally, the GPT-4 model also generated a summary of the overall trends observed throughout the protest.

> Over the 23-hour period, we observe an increasing intertwining of COVID-19 pandemic management and social justice protests (especially Black Lives Matter), along with political controversies surrounding the handling of both issues. The timeline reflects growing public discontent with leadership (particularly, Femke Halsema) and the conflicting pressures between enforcing health measures and allowing public protests.

The output generated by GPT-4 aligned with the events described in Table 1 regarding the demonstration. All big events were accurately reflected in the models' output, whereas additional details regarding the content of the protests were included. For example, the catalyst for the protest (the death of George Floyd, which sparked a global outcry against police violence) was captured in the description. The increasing public criticism directed at the Mayor of Amsterdam, which intensified during the demonstration, was also correctly depicted. Additionally, a noticeable topic shift occurred around 16:00 when more people than expected came to attend the protest, leading to a shift in public discourse around 17:00 from topics related to Black Lives Matter to concerns about COVID-19 and its impact on daily life. These concerns regarding COVID-19 eventually resulted in criticism towards the government from 17:00 onwards.

However, the model did not capture every specific event. For instance, before the demonstration began, government officials decided to relocate it from the larger Museum Square to the smaller Dam Square, a decision that contributed to the overcrowding. This critical detail was not reflected in the model's output. Additionally, the arrival of the important person at 16:30 and the mayor at 18:15 wearing a Keti Koti button was also missing from the final descriptions.

5 Discussion

The described shifts in group sentiment were consistent with the public interest and sentiment during the demonstration. Clear criticism of the government and the mayor of Amsterdam was evident in both the data and the model's output. Furthermore, the progression of topics, shifting from those related to Black Lives Matter to those focused on COVID-19 and the associated restrictions, was distinctly visible in the GPT-4 output. The topics remained distinguishable over time, with minimal overlap or interference.

While the dataset using synonyms performed poorly compared to the hashtag-based set, it raises questions about why this occurred. We hypothesize that the introduction of bias through synonyms played a significant role. For example, the word "Dam", referring to the demonstration's location, has several synonyms related to water management (e.g. dikes, locks, dams, reservoirs), which introduced irrelevant terms that were unrelated to the protest. This made it more difficult for GPT-4 to accurately interpret the topics within the protest's context, as it struggled to differentiate between relevant protest-related terms and unrelated synonyms.

5.1 Limitations

This study did not test several other topic models, such as BERTtopic. Since transformer-based models require significant computational resources to train and are difficult to fully understand, we chose not to include them in this analysis. Additionally, this research focused on a single protest. Although the method is unsupervised, we anticipate similar performance on other protest-related tweets, but this needs to be verified in future work. It is unclear how the model would perform if, for example, two demonstrations in different cities on the same or similar topics occurred simultaneously. Future research should explore how the models handle such scenarios. Another area of interest is examining how topics are described when a demonstration and a counter-demonstration occur in the same city.

5.2 Future Work

This study utilized a relatively limited dataset of labelled tweets, especially when analysing smaller time windows, such as those shorter than the one-hour intervals used in this research. Future studies could benefit from using a larger dataset to determine whether the quality of the topic models and the results improve.

Additionally, integrating more sophisticated context-aware models or incorporating external knowledge sources could help bridge the gap between general event summaries and detailed event tracking. Despite these limitations, the combination of topic modelling and advanced language models like GPT-4 shows great potential for enhancing the analysis of social movements and large-scale public demonstrations.

Another promising direction for future research is the analysis of different protests. A mixed dataset containing tweets from multiple protests could present new challenges and raise interesting research questions.

One such challenge might involve distinguishing between two simultaneous protests. In this case, topic models would need to separate the topics of each protest, which could be difficult if the protests occur in the same city and involve related, yet opposing, themes. This situation could arise, for example, when a protest triggers a counter-protest, leading to shared topics but opposing contexts.

Finally, tweets are not limited to text alone; they include metadata, URLs, images, videos, and other media objects. These additional dimensions could enhance topic models by, for example, identifying the most descriptive images or extracting more contextual information from URLs. Incorporating multimedia elements could also strengthen authorities' ability to gather information by learning from past events.

6 Concluding Remarks

In this study, we compared various topic modelling techniques to gain insights into a Black Lives Matter demonstration in Amsterdam during the COVID-19 pandemic. The resulting topics and associated words were used to generate related hashtags, which were then fed into a GPT-4 model. From this model, both a general overview and a time-specific description of the events were obtained. When comparing the model's output with the actual events, we concluded that this approach shows promising results. However, although the big events of the demonstrations were depicted accurately, more specific events were missing in the final description. These specific events are difficult to spot and were, in this case, detected only after an in-depth analysis of the dataset. While the method demonstrates the potential for capturing broader themes and key moments, it highlights the challenge of identifying granular details through automated approaches. Overall, the described techniques show that it is possible to gain insights into an upcoming protest by analysing tweets over time, on a global scale without granular details and within a relatively short time window.

7 Ethical Statement

Ethical issues might arise related to this study, the most prominent issues revolve around privacy. To protect the privacy of the Twitter users the data used in this study is carefully anonymised. Thus, user IDs or any other personal data from Twitter users are not stored during this study. The analysis is mainly conducted on a published dataset that was masked and preprocessed to exclude any personal information. For more information about the dataset, contact the data protection officer[3].

Another concern might arise from the possibilities this study poses, for example, enriching governmental organisations with possibly sensitive information might open the door to surveillance of citizens and unwanted monitoring of online activity. However, we believe that the Police should be given the best tools possible within the extent and constraints of the law. Additional rules and restrictions might arise with technological advances, but these applications can also, for example, serve to accommodate non-violent protests.

References

1. Amr Ahmed and Eric Xing. Dynamic non-parametric mixture models and the recurrent chinese restaurant process: with applications to evolutionary clustering. In *Proceedings of the 2008 SIAM international conference on data mining*, pages 219–230. SIAM, 2008.
2. Hayder M Alash and Ghaidaa A Al-Sultany. Improve topic modeling algorithms based on twitter hashtags. In *Journal of Physics: Conference Series*, volume 1660, page 012100. IOP Publishing, 2020.
3. Rubayyi Alghamdi and Khalid Alfalqi. A survey of topic modeling in text mining. *Int. J. Adv. Comput. Sci. Appl.(IJACSA)*, 6(1), 2015.

[3] https://www.uu.nl/en/organisation/practical-matters/privacy/data-protection-officer

4. Sarah A Alkhodair, Benjamin CM Fung, Osmud Rahman, and Patrick CK Hung. Improving interpretations of topic modeling in microblogs. *Journal of the Association for Information Science and Technology*, 69(4):528–540, 2018.
5. Kim Andersen, Adam Shehata, and Dennis Andersson. Alternative news orientation and trust in mainstream media: A longitudinal audience perspective. *Digital Journalism*, 11(5):833–852, 2023.
6. Lachlan Birdsey, Claudia Szabo, and Yong Meng Teo. Twitter knows: understanding the emergence of topics in social networks. In *2015 Winter Simulation Conference (WSC)*, pages 4009–4020. IEEE, 2015.
7. David M Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.
8. David M Blei, Andrew Y Ng, and Michael I Jordan. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022, 2003.
9. Jordan Boyd-Graber, Yuening Hu, David Mimno, et al. Applications of topic models. *Foundations and Trends® in Information Retrieval*, 11(2-3):143–296, 2017.
10. Pete Burnap and Matthew L Williams. Us and them: identifying cyber hate on twitter across multiple protected characteristics. *EPJ Data science*, 5:1–15, 2016.
11. Julia Cagé, Nicolas Hervé, and Béatrice Mazoyer. Social media influence mainstream media: Evidence from two billion tweets. *Available at SSRN 3663899*, 2020.
12. Steven P Crain, Ke Zhou, Shuang-Hong Yang, and Hongyuan Zha. Dimensionality reduction and topic modeling: From latent semantic indexing to latent dirichlet allocation and beyond. *Mining text data*, pages 129–161, 2012.
13. Pablo De Pedraza, Martin Guzi, and Kea Tijdens. Life dissatisfaction and anxiety in covid-19 pandemic. Technical report, MUNI ECON Working Paper, 2020.
14. Scott Deerwester, Susan T Dumais, George W Furnas, Thomas K Landauer, and Richard Harshman. Indexing by latent semantic analysis. *Journal of the American society for information science*, 41(6):391–407, 1990.
15. Daniel Falush, Matthew Stephens, and Jonathan K Pritchard. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164(4):1567–1587, 2003.
16. Thomas Griffiths, Michael Jordan, Joshua Tenenbaum, and David Blei. Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 16, 2003.
17. Stephen Francis Groening. *Connected isolation: Screens, mobility, and globalized media culture*. University of Minnesota, 2008.
18. Yosh Halberstam and Brian Knight. Homophily, group size, and the diffusion of political information in social networks: Evidence from twitter. *Journal of public economics*, 143:73–88, 2016.
19. Ottar Hellevik. Antisemitism and islamophobia in norway: A survey analysis of prevalence, trends and possible causes of negative attitudes towards jews and muslims. In *The shifting boundaries of prejudice: Antisemitism and Islamophobia in contemporary Norway*, pages 108–154. Scandinavian University Press (Universitetsforlaget AS), 2020.
20. Thomas Hofmann. Unsupervised learning by probabilistic latent semantic analysis. *Machine learning*, 42:177–196, 2001.
21. Matthew Honnibal and Ines Montani. spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. To appear, 2017.

22. José Luis Iribarren and Esteban Moro. Branching dynamics of viral information spreading. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 84(4):046116, 2011.
23. Amir Karami. *Fuzzy topic modeling for medical corpora*. University of Maryland, Baltimore County, 2015.
24. Amir Karami. Taming wild high dimensional text data with a fuzzy lash. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 518–522. IEEE, 2017.
25. Amir Karami. Application of fuzzy clustering for text data dimensionality reduction. *International Journal of Knowledge Engineering and Data Mining*, 6(3):289–306, 2019.
26. Amir Karami, Aryya Gangopadhyay, Bin Zhou, and Hadi Kharrazi. Flatm: A fuzzy logic approach topic model for medical documents. In *2015 Annual Conference of the North American Fuzzy Information Processing Society (NAFIPS) held jointly with 2015 5th World Conference on Soft Computing (WConSC)*, pages 1–6. IEEE, 2015.
27. Amir Karami, Aryya Gangopadhyay, Bin Zhou, and Hadi Kharrazi. Fuzzy approach topic discovery in health and medical corpora. *International Journal of Fuzzy Systems*, 20:1334–1345, 2018.
28. Amir Karami, Morgan Lundy, Frank Webb, and Yogesh K Dwivedi. Twitter and research: A systematic literature review through text mining. *IEEE access*, 8:67698–67717, 2020.
29. Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, and Carol Willing. Jupyter notebooks – a publishing format for reproducible computational workflows. In F. Loizides and B. Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press, 2016.
30. Bert-Jaap Koops. The shifting" balance'between criminal investigation and privacy a case study of communications interception law in the netherlands. *Information Communication & Society*, 6(3):380–403, 2003.
31. Bert-Jaap Koops, Jaap-Henk Hoepman, and Ronald Leenes. Open-source intelligence and privacy by design. *Computer Law & Security Review*, 29(6):676–688, 2013.
32. Tom Massart, Thijs Vos, Clara Egger, Claire Dupuy, Constance Morel-Jean, Raul Magni-Berton, and Sébastian Roché. The resilience of democracy in the midst of the covid-19 pandemic: Democratic compensators in belgium, the netherlands and france. *Politics Low Countries*, 3:113, 2021.
33. Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.
34. Logan Molyneux and Rachel R Mourão. Political journalists' normalization of twitter: Interaction and new affordances. *Journalism Studies*, 20(2):248–266, 2019.
35. Marçal Mora-Cantallops, Salvador Sánchez-Alonso, and Anna Visvizi. The influence of external political events on social networks: The case of the brexit twitter network. *Journal of Ambient Intelligence and Humanized Computing*, 12(4):4363–4375, 2021.
36. Laurens HF Müter, Christof van Nimwegen, and Remco C Veltkamp. Dutch tweets before, during, and after a black lives matter demonstration in amsterdam: Expert annotated data, protocol, and labelling tool. *Data in Brief*, 55:110527, 2024.

37. Laurens HF Müter and Remco C Veltkamp. Analysing protest-related tweets: An evaluation of techniques by the open source intelligence team. In *Future of Information and Communication Conference*, pages 41–60. Springer, 2024.
38. Seth A Myers, Chenguang Zhu, and Jure Leskovec. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 33–41, 2012.
39. RYAN Nall, Lauren Solberg, ROBERT L Cook, and Brenda Rogers. A framework for discussing controversial topics in medical education. *Contemporary Challenges in Medical Education: From Theory to Practice*, 2019.
40. Debashis Naskar, Sanasam Ranbir Singh, Durgesh Kumar, Sukumar Nandi, and Eva Onaindia de la Rivaherrera. Emotion dynamics of public opinions on twitter. *ACM Transactions on Information Systems (TOIS)*, 38(2):1–24, 2020.
41. nos. Torenspits in Hoogmade ingestort bij grote kerkbrand — nos.nl. https://nos.nl/artikel/2309029-torenspits-in-hoogmade-ingestort-bij-grote-kerkbrand, 2019. [Accessed 18-11-2024].
42. OpenAI. Introducing chatgpt. https://openai.com/index/chatgpt/, 2022. [Accessed 05-09-2024].
43. Kees-Jan Oskam. Hoogmade daags na de kerkbrand: kijken, praten en foto's nemen. https://www.nd.nl/nieuws/nederland/515725/hoogmade-daags-na-de-kerkbrand-kijken-praten-en-foto-s-nemen, 2019. [Accessed 18-11-2024].
44. Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vossen. Open Dutch WordNet. In *Proceedings of the Eight Global Wordnet Conference*, Bucharest, Romania, 2016.
45. Maximilian A Primbs, Rob W Holland, Freek Oude Maatman, Tessa AM Lansu, Ruddy Faure, and Gijsbert Bijlstra. The effects of the 2020 blm protests on racial bias in the united states. *Personality and Social Psychology Bulletin*, page 01461672241269841, 2022.
46. Jonathan K Pritchard, Matthew Stephens, and Peter Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–959, 2000.
47. R Rehurek and Sojka P. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2), 2011.
48. Tara Riem. DPG Media Privacy Gate — ad.nl. https://www.ad.nl/alphen/zo-reageert-nederland-op-de-kerkbrand-in-hoogmade~a28d8337/, 2019. [Accessed 18-11-2024].
49. Fabián Riquelme and Pablo González-Cantergiani. Measuring user influence on twitter: A survey. *Information processing & management*, 52(5):949–975, 2016.
50. Michael Röder, Andreas Both, and Alexander Hinneburg. Exploring the space of topic coherence measures. In *Proceedings of the eighth ACM international conference on Web search and data mining*, pages 399–408, 2015.
51. Michal Rosen-Zvi, Chaitanya Chemudugunta, Thomas Griffiths, Padhraic Smyth, and Mark Steyvers. Learning author-topic models from text corpora. *ACM Transactions on Information Systems (TOIS)*, 28(1):1–38, 2010.
52. David L Sam and John W Berry. Acculturation: When individuals and groups of different cultural backgrounds meet. *Perspectives on psychological science*, 5(4):472–481, 2010.
53. PK Srijith, Mark Hepple, Kalina Bontcheva, and Daniel Preotiuc-Pietro. Sub-story detection in twitter with hierarchical dirichlet processes. *Information Processing & Management*, 53(4):989–1003, 2017.

54. Mark Steyvers and Tom Griffiths. Probabilistic topic models. In *Handbook of latent semantic analysis*, pages 439–460. Psychology Press, 2007.
55. Jan Terpstra, Jacques de Maillard, Renze Salet, and Sebastian Roché. Policing the corona crisis: A comparison between france and the netherlands. *International journal of police science & management*, 23(2):168–181, 2021.
56. Daniel Trottier. Open source intelligence, social media and law enforcement: Visions, constraints and critiques. *European Journal of Cultural Studies*, 18(4-5):530–547, 2015.
57. Annick C van Brouwershaven, Sophie H Bolt, and Jeroen GF Jonkman. The ongoing and contentious coverage of abortion in a progressive context: a long-term cross-outlet assessment of dutch abortion news (2000–2022). *Culture, Health & Sexuality*, pages 1–18, 2024.
58. Chris van Dijk. Kerk Hoogmade volledig verwoest na grote brand: 'Dit was het hart van Hoogmade' — omroepwest.nl. https://www.omroepwest.nl/nieuws/3960685/kerk-hoogmade-volledig-verwoest-na-grote-brand-dit-was-het-hart-van-hoogmade, 2019. [Accessed 18-11-2024].
59. Chong Wang, John Paisley, and David M Blei. Online variational inference for the hierarchical dirichlet process. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 752–760. JMLR Workshop and Conference Proceedings, 2011.
60. Ruud Wouters and Jonas Lefevere. Making their mark? how protest sparks, surfs, and sustains media issue attention. *Political Communication*, 40(5):615–632, 2023.
61. Jinghong Xu, Difan Guo, Shaoqiang Liu, et al. How social media expedites the crisis spillover effect: A case study of tesla's recall event. *Public Relations Review*, 50(1):102432, 2024.
62. Jaewon Yang and Jure Leskovec. Modeling information diffusion in implicit networks. In *2010 IEEE international conference on data mining*, pages 599–608. IEEE, 2010.
63. Min Yang, Yuzhi Liang, Wei Zhao, Wei Xu, Jia Zhu, and Qiang Qu. Task-oriented keyphrase extraction from social media. *Multimedia Tools and Applications*, 77:3171–3187, 2018.
64. Amy Yau, Ben Marder, and Stephanie O'Donohoe. The role of social media in negotiating identity during the process of acculturation. *Information Technology & People*, 33(2):554–575, 2020.
65. Yuhao Zhang, Wenji Mao, and Junjie Lin. Modeling topic evolution in social media short texts. In *2017 IEEE International Conference on Big Knowledge (ICBK)*, pages 315–319. IEEE, 2017.