

Cannabis in Danish newspapers

Esben Houborg & Oskar Enghoff

Centre for Alcohol and Drug Research, Aarhus University
eh.crf@psy.au.dk
oskar.enghoff@gmail.com

Houborg, Esben; Enghoff, Oskar (2018). 'Cannabis in Danish newspapers' i *Tidskrift for Forskning i Sygdom og Samfund*, nr. 28, 173-205.

Using quantitative methods Danish cannabis debate in national newspapers is investigated. The investigation shows that the most prevalent topics relate to law enforcement. Legalization has become an increasingly important topic in the Danish cannabis debate and the investigation shows a reframing of this debate to become increasingly related to concerns about organized crime. In this way the Danish cannabis legalization debate show the same development as the debates that have led to legalization certain states in the United States of America

Introduction

In this article, we aim to identify and quantify the dominant themes of reporting on cannabis in Danish newspapers. We do this in order to characterize the overall composition of the Danish newspaper coverage of this illicit drug and determine whether it follows any temporal and regional patterns. We begin with an introduction to the field, briefly summarizing the findings and theoretical approaches of relevant previous research and develop our aims based on this.

The topic of cannabis and illicit drugs in general is controversial and often the subject of intense discursive struggles. These struggles may take place in a num-

ber of specialized settings, e.g. scientific communities, public health administrations or parliaments, but often they also feed into and draw upon the debate in the 'public sphere' (Habermas, 1989) – that is, the collective mass of statements and interactions that are widely distributed via public platforms such as news media, television and digital social media. These platforms play an important role not only in mediating the debate, and thus influencing the dissemination of statements from different stakeholders and other participants, but also by supplying their own 'voice' via selection of relevant news events and coverage of these events.

The central role of media platforms in mediating the public debate has implications for the formation of public opinion. Although the exact causal mechanisms are difficult to determine and continually debated, much previous research has elucidated the close association between media reporting, public opinion, policy support and actual policy development. An often cited study in the drugs field was conducted by Fan (1996), who used statistical inference to demonstrate a strong correlation between American newspaper portrayals of an ongoing 'drug crisis' and a dramatic increase in general public opinion survey responses naming drug use as the United States' most important problem (from 5% to 60% in the years 1985–1994). More recently, Tieberghien (2014) analysed the coverage of drug policy in Belgian newspapers and found a narrow focus on cannabis legalization, which meant that the policy positions available to e.g. politicians who appropriated the newspaper discourse were reduced to either 'for' or 'against' cannabis legalization. The same study also examined the role of newspapers in the political construction of an 'evidence-based' drug policy, by interpreting and distributing scientific results. This research hints at the complexity of the network of actors in which 'media effects' can be theorized to take place. Other studies of newspapers have identified further relevant actors, such as medical professionals, medical cannabis patients, law enforcement and other government institutions (e.g. Hughes et al., 2011; Lewis et al., 2015).

Theorizing media effects

The impact of mass media on public opinion has been theorized in a number of ways – both in terms of how various types of media actually influence the opinions of individuals and groups, and in terms of the power structures in which this influence may originate. Discourse analysis, and especially critical discourse analysis (Fairclough, 1995), is a popular theoretical framework in this regard. This

approach highlights the social production of news discourse as embedded within an institutional and governmental ideology, in which certain exclusionary social constructions of (for example) drug use and drug user subject positions are dominant – for instance, a law enforcement discourse, in which drug use is criminal and drug users are criminals (see e.g. Månsson, 2016). Several studies of newspapers have found a preference for governmental sources in articles on drug use, such that political figures, public health professionals and especially law enforcement representatives become ‘primary definers’ (from Hall, 1978) of drug-related issues (Blood et al., 2003; Haines-Saah et al., 2014; Smith, 2015). These studies point out that the overrepresentation of official sources in newspapers has led to ‘narrow representations’ of illicit drug use, which mainly highlight the criminal, social and health-related problems associated with these substances.

While these representations of drug use are clearly of a discursive nature, another useful and popular theoretical framework is that of frame analysis. Although this approach has been criticized for an apparent lack of a clear ontology (Van Gorp, 2005), it has proven useful in a number of news media analyses because a frame lends itself more easily to empirical operationalization than a discourse. Entman (1993) offers a widely used definition of framing:

Framing essentially involves *selection* and *salience*. **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: 52; emphasis in original).

The practices of defining an issue, interpreting its causality, evaluating it morally and proposing a solution are called ‘reasoning devices’ by Van Gorp (2005), who adds that a frame also ‘consists of all indicators or framing devices by which the frame can be identified: Metaphors, catchphrases, visual images, lexical choices, selection of sources, graphics, stereotypes, dramatic characters, etc.’ (Van Gorp, 2005: 486–87). In this interpretation, frames exist mainly in the linguistic and rhetorical aspects of individual communicating texts. However, studies such as Tieberghiens’ (2014) analysis of drug policy coverage show that a similar phenomenon can also exist on a higher level of abstraction: When a broad issue (e.g. drug policy) is only covered in articles on a specific aspect of this issue (e.g. cannabis legalization), the broader issue is framed in terms of the narrower one. This is often referred to as priming: ‘Priming occurs when news content suggests to news

audiences that they ought to use specific issues as benchmarks for evaluating the performance of leaders and governments' (Scheufele and Tewksbury, 2007: 11). A very clear example of priming is found in 'event-centred' media coverage of an issue, for instance when a high-profile criminal case such as a major smuggling bust receives extended media coverage and thus has a disproportionate impact on the broader debate on drug policy (e.g. Hughes et al., 2010). Similarly, the study by Taylor (2008) shows how both media coverage and policy developed simplistic notions of drug issues at the expense of wider and more complex discussions. This is also related to the more general media effect of 'agenda setting', in which the amount of media coverage devoted to an issue affects the public and political attention given to this issue (Lancaster et al., 2011).

Research aims

In order to study framing, priming and agenda-setting in relation to cannabis and associated issues, it is necessary to collect a large and representative sample of newspaper articles, and to design the analysis in such a way that it is possible to study both the overall framing (or priming) via the range of covered events/issues, and the specific framing within individual articles. Josefine Månsson (2016) has conducted an analysis of cannabis representations in Swedish newspaper articles with a similar research design (although she employs discourse theory rather than frame analysis), collecting articles from two different years and comparing the prevalence of various cannabis discourses (e.g. medicinal and juridical) in these two periods in order to identify continuity and/or change over time. Other studies have focused on more specific cannabis-related issues, such as cannabis legalization (Tieberghien, 2014), medicinal cannabis (Sznitman and Lewis, 2015) and stigmatized cannabis users (Haines-Saah et al., 2014). Our aim is to conduct a broad longitudinal analysis similar to the one conducted by Månsson (2016), in order to quantify the prevalence of various newspaper framings of cannabis, and examine their longitudinal patterns, such as their temporal relationship to current law enforcement and policy events. Additionally, we aim to examine whether these framings vary according to the target population of different newspapers – nationwide versus regional distribution, for instance – building on previous research finding regional variations in press coverage (Månsson, 2016; Tieberghien, 2014).

Cannabis in Denmark

Drugs can be and mean many different things depending on the contexts in which they are found. The same chemical compound may act in very different ways depending on its dose, the body it interacts with, the way it is administered, the practices of use and the meanings attributed to it. Under some circumstances, it may act as a medicine; under others it may act as a poison and in still others it may act as an intoxicant. The Ancient Greek term *pharmakon*, which can refer to an array of things including remedies or medicine, poisons or magical potions, captures the multiple functions and meanings a chemical compound can have and its contingent nature. The ways drugs are framed discursively and in practice not only define the drugs themselves, but also the people who use them and the institutions that deal with them (Houborg, 2012). As medicine, drugs are part of the constitution of doctors and patients and the whole institution of medicine. As illicit intoxicants, drugs are part of the constitution of criminal offenders, law enforcement agents and the legal system. The different uses, meanings and effects of drugs mean that different actors may have different interests in how they are framed, for example believing that a chemical compound should only exist as a medicine and not as an intoxicant or as a religious sacrament, or perhaps that the compound should not exist at all within a particular social context or jurisdiction. This was the case for cannabis when the United Nations produced the Single Convention on Narcotic Drugs (1961), which regulates drugs internationally. This convention stated that cannabis should be banned from any use and that existing cultural and religious use should stop within a specific period of time after the enactment of the treaty. This kind of governance, which regulates what chemical compounds can and cannot be, can be defined as ‘political pharmacology’ (Houborg, 2012).

Cannabis was known as part of different healing compounds during the 19th century, but was not used as an intoxicant to any significant extent until the 1950s (Jepsen, 1966). In 1930 Denmark incorporated the second International Opium Convention into Danish law by amending its existing opium act. In the process, cannabis for the first time became a controlled substance as it was included among substances that could only be used for medical or scientific purposes. With no significant recreational use of cannabis, this had no practical consequences in Denmark and the regulation was only introduced as an act of international solidarity. In 1963 Denmark incorporated the 1961 UN drug convention (the ‘Single Convention’) into Danish legislation (Jepsen, 1966; Houborg, Bjerger & Frank, 2008) when

the Home Office issued a circular with a list of controlled substances. According to this document, cannabis was one of the substances prohibited in Denmark. Once again, this change in Danish cannabis control did not have any significant consequences. Only one medical product had to be discontinued and the number of cases of recreational use were few and mainly limited to a group of people who frequented a small number of night- and jazz clubs in Copenhagen (Jepsen, 1966). However, over the next couple of years, cannabis became an increasingly popular substance among young people associated with the hippie movement, and from 1965 the association between cannabis and the new youth culture became a major political issue and important topic in the news media (Houborg & Vammen, 2012). Three topics in particular dominated the public debate as cannabis use spread rapidly among the Danish youth: 1) the cultural challenge that the new youth culture posed to Danish society and the role cannabis played in this; 2) the possible risks associated with using cannabis; 3) the trade off between criminalization of drug users and cannabis use: What was the most damaging? Cannabis was mainly imported by young people travelling to the Middle East and Asia, bringing the drug home with them, and was for the most part distributed through 'social dealing'. However, there was a concern among some politicians and within the criminal justice system that cannabis trafficking and distribution would generate organized crime. This concern led the government to propose legislation that would increase sanctions for organized drug-trafficking and distribution. But another just as important concern that was articulated in the media and among politicians was that criminalization would stigmatize and alienate a large number of young people. In 1969, a compromise was made between the hard-line policy towards organized drug crime and the concerns about criminalization of drug users. The Attorney General was instructed to issue a circular ordering the criminal justice system not to prosecute violations of the drug legislation in cases of possession of illicit drugs for personal use¹. The circular made a distinction between cannabis and other controlled drugs. Whereas no legal action (or, at most, a formal caution) would be taken in cases of first-time offences related to the possession of any drug, this rule would only apply for repeat offences when they concerned cannabis. This meant a de-penalization or de facto decriminalization of possession of cannabis for personal use. This compromise came to be the basis for Danish drug policy until 2004. Drug supply was to be reduced through law enforcement and drug demand reduced through prevention and treatment. This policy ended in 2004 when the parliament amended the drug legislation and repealed the de facto decriminalization of the possession of illicit drugs for personal use. Since 2004 all possession of illicit

drugs in Denmark is punishable by at least a fine. In this way Denmark has moved towards a more prohibitive drug policy, in contrast to many other countries that have moved in the other direction. This has drastically changed the legal status of cannabis users in Denmark. It has also meant more active policing against cannabis, including the cannabis market at Christiania, leading to disruptions of the cannabis market which contributed to a gang conflict in Copenhagen.

Data

Our data consists of Danish articles published in newspapers or by news agencies that concern cannabis in some way. The articles were collected through Infomedia, a privately-owned media monitoring service that archives content from all national and local Danish newspapers and news agencies as well as a number of periodical magazines.² The aim of this data collection strategy was to be able to analyse what DiMaggio et al. (2013: 576) call the ‘environment of representations’ surrounding cannabis in newspaper articles, rather than the coverage of specific events or issues.

We collected all available articles meeting the following criteria:

- The article contains at least three appearances of the word stems ‘hash’ or ‘cannabis’
- The article is written in Danish
- The number of words in the article is at least 200
- The article has been printed in a national newspaper, a local newspaper or a periodical magazine, or has been published by a news agency.

Based on these criteria, the Infomedia database returned a wide range of article types (brief news reports, lengthy news analyses, editorials, opinion pieces, etc.) from approximately 300 distinct publications, published over a period of 27 years (1989–2016). We determined ‘hash’ and ‘cannabis’ to be the most widely used Danish terms referring to cannabis, and used a threshold of at least three instances of one or both of these terms to exclude articles in which cannabis plays a very minor role. To avoid very short and superficial articles and thus ensure a certain depth of content in our data, we collected only articles with 200 words or more. We performed the final database query on 30. September 2016, and so our data does not contain articles from the last three months of 2016. Even though Infomedia also

stores articles published solely on web platforms, we chose not to include these due to the lack of an automatized download function – including the 21,000 web articles meeting the above criteria would more than double the time required to download the articles. Even though newspapers are generally accepted as a good proxy for news reporting (see Hughes et al., 2011: 286), we propose that future research including web content is a fruitful prospect and would potentially yield other results due to the unique journalistic practices evident on the web (e.g. in terms of audience engagement and targeting).

The data collection yielded a total of 12,304 unique articles from 1989 to 2016. In total, the articles contain approximately 9.4 million words (assuming five characters per word), averaging at approximately 638 words per article. See Table 1 for an overview of articles by publication type and year.

	National dailies	Local dailies	Local weeklies	News agencies	Magazi- nes	Total unique
1989	6	0	0	0	0	6
1990	52	0	0	0	0	52
1991	120	1	0	6	0	125
1992	163	0	0	0	0	163
1993	182	0	0	2	1	185
1994	234	0	0	59	5	296
1995	195	0	0	61	3	252
1996	233	19	0	41	3	287
1997	244	27	0	46	3	313
1998	184	9	0	38	1	221
1999	142	17	0	28	1	181
2000	150	37	0	26	1	207
2001	261	82	0	49	4	370
2002	277	121	0	73	1	430
2003	444	147	0	90	12	651
2004	301	144	2	94	4	505
2005	235	242	9	82	7	529
2006	165	270	8	49	12	456
2007	128	261	20	77	6	435
2008	108	313	44	41	16	485
2009	188	400	73	62	16	698
2010	90	332	59	41	20	515
2011	139	369	85	62	25	633
2012	187	344	111	111	47	744
2013	282	463	128	201	59	1029
2014	267	504	105	258	65	1086
2015	215	343	69	162	61	779
2016	217	298	76	118	30	671
Total	5409	4743	789	1877	403	12304

Table 1: Article counts by publication type and year

Note: Some articles appear in multiple publications (e.g. syndicated content from news agencies).

Methodology

Within the broader field of print media analysis, there is a multitude of methodological approaches at various points on the qualitative-quantitative spectrum. Analysis of print media has a strong foundation within critical research, especially within the analytical framework developed by van Dijk (1988) and within Fairclough's critical discourse analysis (Fairclough, 1995; Teo, 2016: 11–13). This branch of media research typically makes use of close readings and qualitative analyses of individual texts. At the other end of the spectrum, there are various quantitative approaches (which are often theorized via frame analysis, e.g. Simon and Jerit, 2007). The most notable example is corpus linguistics (Baker, 2006), which takes its name from the fact that instead of considering individual texts, a digital (or digitized) corpus of articles or other texts become the object of analysis. The use of digital corpora allows for a quantifiable analysis of language use across a very large number of texts. Studies within corpus linguistics have mainly focused on individual terms, in terms of frequency, dispersion, collocates (common combinations with other terms), concordance (context of use within sentences) and 'keyness' (overrepresentation within certain texts) (Baker, 2006).

It is often argued that there is much to gain by integrating quantitative methods such as corpus linguistics with qualitative methods such as discourse analysis (Bednarek, 2006; Gabrielatos and Baker, 2008). Besides serving as a method of mutual validation and triangulation, quantitative media analysis makes a longitudinal analysis much more viable (Bednarek, 2006: 12). By using a fixed set of measurements to analyse a corpus of articles spanning a period of time, one can avoid the pitfalls of analysing single texts, or using a less 'strict' methodology to analyse multiple texts – which was a common critique of much early work within critical discourse analysis (O'Halloran, 2010: 564–65).

Our methodological approach combines aspects of the approaches outlined above. We combine the theoretical framework of frame analysis with a corpus-based approach reminiscent of corpus linguistics. However, instead of the analytical tools associated with corpus linguistics, we use topic modelling, a text-analytical method that has been shown to have several affinities with the key aims and concepts of framing analysis (DiMaggio et al., 2013).

Via topic modelling, we can analyse the overall thematic content of the corpus, and observe patterns in the distribution of themes and frames over time and across publications. When considering individual articles – e.g. in terms of their genre, outlet, structure and script – we can 'zoom in' and conduct a more qualita-

tive framing analysis (see also Topic modelling print media discourse). Thus, our method is located somewhere between content analysis (quantitative analysis of manifest content) and discourse analysis (qualitative analysis of ‘consequential meaning’) (Richardson, 2006: 15–21).

Topic modelling

We employ digital methods to analyse the thematic structure of our data. Specifically, we use *topic modelling*, a machine learning approach to text analysis which has seen successful use in thematic analyses of text documents, for example novels (Jockers and Mimno, 2013), scientific papers (Hall et al., 2008), forum posts (Munksgaard and Demant, 2016), and news articles (DiMaggio et al., 2013; Kim and Oh, 2011; Mei and Zhai, 2005; Yang et al., 2011). The main appeal of topic modelling is the ability to automatically identify, locate and quantify ‘topics’ (i.e. themes) in very large document collections. As such, topic modelling may be superficially understood as a digital approach to automating the process of manual qualitative coding. We will now provide a brief non-technical introduction to topic models, focusing on their applications in interpretive analysis.

A topic model is a statistical model of language, designed to be able to identify the underlying thematic structure of a large collection of text documents (a corpus). The approach to language inherent in a topic model is based on relational meaning: We quantify the relations between terms, and not just the individual terms themselves (DiMaggio et al., 2013: 586–590). This sets topic modelling apart from more widespread text-analytical methods (e.g. within the field of *corpus linguistics*), which mostly take individual terms as their starting point and consider these in terms of their frequency, ‘keyness’, exclusivity, concordance and so on (see e.g. McCarthy and O’Keeffe, 2010).

A topic model takes a collection of text documents as inputs, and outputs a set number of ‘topics’. On the empirical level, a topic is essentially a group of terms which tend to occur in the same documents. On the conceptual level, topics are the thematic ‘building blocks’ of the corpus – each topic corresponds to a distinct theme, and each document consists of a unique mix of some of these themes. The assumption that a document can contain more than one topic (‘mixed membership’ modelling) is central to topic modelling, and is especially relevant in an analysis of newspaper articles as it allows us to account for the fact that these often contain several themes (e.g. drug policy and health) or even *heteroglossia* (com-

peting voices, e.g. a proponent and an opponent of a policy interviewed in the same article) (DiMaggio et al., 2013; Grimmer and Stewart, 2013). The model output represents the topics in two ways: In a topic-term matrix, which contains the probabilities of individual terms appearing within each topic; and in a document-topic matrix, which contains the proportions of each topic in each document. For example, we can say that the term 'cannabis' has a probability of 10% in a specific topic, and that this topic has a proportion of 30% in a specific document. This way of representing topics and documents is necessarily a simplification, and when combined with other simplifying documentation representation choices applied during the analysis (see Model specification and estimation), a significant amount of contextual information about the documents is lost – in terms of e.g. publication, authorship and grammar. This can be addressed by the use of rigorous qualitative inspection and validation of results, as well as the use of document metadata as covariates in the topic model (see below). However, it must be acknowledged that topic modelling and related modelling methods provide a distanced and 'macro-level' perspective of the documents analysed (Jockers 2013). This is a disadvantage in an analysis where a detailed and nuanced understanding of the content and production of individual documents is necessary, but it can alternatively be viewed as an advantage in an analysis where an inductive and explorative characterisation is sought after. Thus, topic modelling is not a replacement for more qualitative and context-sensitive readings, but should instead be viewed as a parallel perspective which can yield other insights, or as a supplementary method in a mixed-methods or 'quali-quantitative' (Venturini and Latour, 2010) research design.

In previous studies, such as the ones cited above, topics have been demonstrated to capture coherent themes in various corpora and to correspond well with human interpretations of the same corpora. An important point in this regard is that topic modelling is unsupervised – the topics are not defined a priori, but are rather derived empirically from the actual documents (Grimmer and Stewart, 2013: 280–91). This has the advantage of facilitating an inductive and explorative approach to text analysis, in which theoretically defined categories are not imposed on the data beforehand, which is a common critique of content-analytical methods (Richardson, 2006: 20). However, it also necessitates *post hoc* interpretation and validation of the model output (Grimmer and Stewart, 2013: 286). Various statistical validation measures have been developed, but several scholars argue that a qualitative, semantic validation based on expert domain knowledge is the most appropriate approach when using topic models for interpretive analysis ra-

ther than information retrieval (DiMaggio et al., 2013; Jockers, 2013; Roberts et al., 2014). This approach consists in a close reading of a topic's most probable terms and most exemplary documents (i.e. the documents with the highest proportion of the topic). The topics are considered valid if these terms and documents coherently and unambiguously correspond to a meaningful theme, and resonate with previous knowledge of the field. It should be noted that it is common for a topic model to contain a number of 'incoherent topics', which can and should simply be disregarded in the further analysis (Jockers, 2013: 128–30).

This dual process of post hoc interpretation and validation is one of the key ways in which qualitative and quantitative approaches can be combined in topic modelling. It entails close qualitative readings of individual documents, guided by quantifiable measures of topic prevalence estimated by the model. This approach may also be useful in the actual analysis by using topic proportions to identify which documents should be subjected to a 'conventional' qualitative analysis of e.g. discourse. This 'document retrieval' approach is well aligned with the original purpose of topic modelling (information retrieval in digital archives – see e.g. Blei 2012), and is another key way to combine qualitative and quantitative approaches within this methodology. However, in most cases (including our own) the ultimate goal of this process is to establish a set of valid and relevant topics which may then be used in a more quantitative analysis of the thematic structure of corpus – what Jockers (2013) has called a 'macroanalysis' or 'distant reading'. Thus, the qualitative aspects of our approach are mainly found in the validation and interpretation of the model – but this also entails the essential interpretation and theorization of more high-level model outputs such as topic correlation networks and periodical word clouds (see Findings).

Topic modelling print media discourse

Previous studies have also estimated topic models of corpora of news articles, and have demonstrated the method's analytical merits within print media analysis and frame analysis. DiMaggio et al. (2013) used a topic model of articles about governmental arts funding in order to discover how this policy domain was framed in newspapers. They discovered a number of topics corresponding to specific frames and found that some frames were highly associated with specific publications. By plotting the prevalence of two frames over time, they found that the balance between these shifted in concordance with real-world events such as public

controversies and policy developments. In two methodological papers with many similarities, Kim and Oh (2011) and Mei and Zhai (2005) estimated multiple topic models for separate 'time slices' (i.e. temporal segments spanning a fixed number of days) of a corpus of news articles and computed similarities between topics across these slices. This allowed them to identify chains or patterns of similar topics over time and, thus, to observe focus shifts in news coverage of certain issues.

These studies highlight two aspects of topic models that are especially relevant for our study: The ability to extract, quantify and compare different framings of a certain issue; and the ability to correlate these framings with document meta-data – for example, publication date or newspaper type. In this study, we choose an approach which most closely resembles that of DiMaggio et al. (2013). That is, we estimate a single model of our corpus and use article publication dates (rather than time slices) to study patterns in time and source.

We can use the topics themselves to assess which cannabis-related frames are actually present in the data, and use topic proportions to study the differing prevalence of these frames. It is also possible to calculate correlations between topic proportions and use these to identify clusters of topics that are closely related. This higher level of abstraction can be used as an overall categorization of various framings in the Danish cannabis debate. In order to study longitudinal patterns in the debate, we can plot document-topic proportions against publication date. This allows us to correlate specific real-world events with shifts in topic proportion, which not only addresses our research aims but also tests the 'predictive validity' of the model (Grimmer and Stewart, 2013). For a deeper understanding of longitudinal patterns relating to a specific topic, we can use the model's document-topic assignments to identify articles that concern this topic, and compare their language content across different time periods. Finally, we can extract the full text of articles concerning a specific topic ('exemplar documents') and subject them to close reading and qualitative analysis (Roberts et al., 2014).

We argue that the combined methodological possibilities outlined above provide a solid framework for an analysis of newspaper framings, bringing together many of the advantages of strictly quantitative approaches with those of strictly qualitative approaches. It should be noted that we are employing a complex algorithm based on advanced statistical concepts for which we have not provided a thorough mathematical explanation. Instead, we choose a heuristic and pragmatic approach, focusing not on the inner workings of topics models but on their utility in identifying and quantifying meaningful themes in our corpus (Jockers, 2013:

128–30). This approach reflects our interest in promoting the use of topic models among social scientists with non-technical backgrounds (Ramage et al., 2009).

Model specification and estimation

The articles were stored as raw text data and then refined in a number of ways in accordance with data pre-processing procedures common to the field of topic modelling. We removed several terms without thematic relevance that could possibly have biased or diluted the analysis of co-occurrence: The terms from the search string ('hash' and 'cannabis'), all publication names (DiMaggio et al., 2013: 582), all personal names, all surnames, and the term '*foto*' (photograph). We also removed a list of very common terms that carry little or no thematic significance (so-called stop words such as 'you', 'and', 'here' and 'some') (Jockers, 2013: 131). All numbers and punctuation were stripped from the texts, and the remaining terms were then reduced to word stems – for instance, 'smoke', 'smoking' and 'smoker' all become 'smok'. This step obscures any grammatical nuances but allows a single word to be traced across different conjugations and usage contexts (Jockers and Mimno, 2013: 753). Finally, any word stems consisting of fewer than three characters and/or appearing in less than 10 articles were removed. These threshold measures reduce the amount of 'noise' in the data (e.g. rare terms and spelling errors) and further serve to optimize the model estimation by reducing the corpus vocabulary. The final vocabulary consisted of 14,657 unique word stems.

We use the *Structural Topic Model*, developed and implemented in R by Roberts, Stewart and Tingley (2016). This topic model extends Blei's (2007) Correlated Topic Model in that it not only assumes correlations between topics but also enables the use of document metadata as topic prevalence covariates. This makes it especially well-suited to our corpus: News values such as recency and proximity (Bell, 1991) support the assumption that topic prevalence in newspaper articles is influenced by metadata such as publication date and publication type.

When specifying a topic model, the first and most vital step is choosing the number of topics (denoted K). The model will represent the entire corpus through the chosen number of topics, and thus different values of K will yield models with varying degrees of 'thematic granularity' and interpretability (Jockers, 2013: 127–28; Steyvers and Griffiths, 2007). When using the structural topic model, another important choice is whether to include document metadata in the model estimation, and thus introduce a degree of supervision in an otherwise unsupervised

model. Metadata (e.g. publication year and publication type) can be included as covariates which inform either document-topic proportions (topic prevalence) or topic-term probabilities (topic content) (Roberts et al., 2014: 5). Topic models with document metadata covariates have been shown to produce more coherent and domain-specific topics (Yang et al., 2015), and to perform better in terms of statistical quantities of interest, e.g. calculated covariate relationships with uncertainty estimates (Roberts et al., 2014). In our study, inclusion of prevalence covariates will allow the model to take into account that certain topics are more prevalent in articles from certain time periods or media outlets.

We estimated several models with a value of K between 10 and 100 with and without covariates. Through experimentation and inspection of results, we decided to use a model with 40 topics and with both publication date and publication type as topic prevalence covariates (respectively represented as number of days since 1. January 1970 and dummy variables for each publication type) as this model specification resulted in a high number of coherent and readily interpretable topics with a thematic granularity suitable for our analysis. We decided not to include topic content covariates as this resulted in fewer interpretable topics. This heuristic process of trial and error in the search for a reasonable K is a common approach in applications of topic modelling within social sciences and humanities, where the interpretability and analytic usefulness of the model is valued higher than various statistical measures of topic quality (which may correlate negatively with human evaluations of topic quality, see e.g. Chang et al. 2009). See the appendix for a complete specification of the modelling procedure.

All the tested models were very similar in the way they represented the corpus, varying mainly in granularity. Models with a lower K tended to produce very broad topics, aggregating themes that were more meaningful to describe separately, while models with a greater K tended to produce overly specific topics that were often incoherent. The model we chose provided a sound compromise and enabled us to isolate the themes that were relevant to our analysis (Jockers, 2013: 134–35).

Table 2 provides 10 examples of topics from the model. Each row describes one topic, displaying the 10 most probable terms within the topic (Danish word stems and English-translated full words), a label based on our qualitative interpretation of the topic's term probabilities and exemplary texts, and finally the mean proportion of the topic across all articles.

Ten most probable terms (Danish stems and full English words)	Label	Mean
Politi, anhold, kilo, København, sigt, person, aktion, sag, beslaglag, mænd <i>Police, arrest, kilo, Copenhagen, charge, person, action, case, confiscate, men</i>	Police raids	5.35%
Legalisering, København, brug, forbud, legalis, kriminel, stof, lov, politik, gør <i>Legalization, Copenhagen, use, prohibition, legalise, criminal, drug, law, policy, do</i>	Legalization debate	5.14%
Politi, mand, bil, sigt, klok, indbrud, stjal, anhold, kørt, nat <i>Police, man, car, charge, time, break-in, stole, arrest, drove, night</i>	Crime reports	4.08%
Stof, ung, procent, prøv, ryg, undersøg, alkohol, bland, vis, men <i>Drug, young, percent, try, smoke, study, alcohol, mix, certain, but</i>	Youth drug use	3.81%
Politi, Danmark, ton, dansk, indsmugling, sag, anhold, København, spani, kilo <i>Police, Denmark, tonne, Danish, smuggling, case, arrest, Copenhagen, Spain, kilo</i>	Smugglers and distributors	3.76%
Christiania, push, politi, street, fristad, christianit, København, bod, pusherstre, dag <i>Christiania, pusher, police, street, free-town, Christiania resident, Copenhagen, stall, Pusher Street, day</i>	Christiania	3.60%
Misbrug, ung, behandling, problem, ryg, hashmisbrug, hjælp, fler, stofmisbrug, dag <i>Abuse, youth, treatment, problem, smoke, cannabis abuse, help, more, drug abuse, day</i>	Addiction and treatment	2.95%
Patient, medicinsk, medicin, læg, brug, smert, behandling, sygdom, syg, effek <i>Patient, medicinal, medicine, doctor, use(r), pain, treatment, illness, ill, effect</i>	Medical cannabis	2.04%
Bebo, område, politi, hus, sted, problem, nabo, gad, handl, park <i>Resident, area, police, house, place, problem, neighbour, street, trading, park</i>	Dealing and local areas	1.59%
Kør, bil, bilist, kørekort, blod, politi, stof, påvirk, tag, bød <i>Drive, car, driver, driver's license, blood, police, drug, affected, apprehended, fine</i>	High drivers	1.49%

Table 2: Ten coherent topics from the model.

Most probable terms, labels, categories and mean proportions

13 out of 40 topics were deemed to be incoherent and are thus excluded from the analysis (not shown above). We also exclude two topics which are purely linguistic and thus carry no thematic significance ('Conversational language' and 'Biographical language' respectively). The fairly high number of incoherent topics may be an indicator of thematic noise in the corpus (DiMaggio et al., 2013: 582) – for example, the fact that a certain proportion of the corpus consists of a large number of diverse themes which are each too minor to merit more than a few articles, and that this sparse diversity cannot be represented by 40 topics (this explanation resonates well with the genre of news articles). This is an example of the compromises involved in the selection of topic models. However, the incoherent topics can be discarded without consequences for the remaining topics (Jockers, 2013: 128–30).

In total, the incoherent topics account for 20.6% of the corpus. This means that on the level of individual words, approximately one fifth of the corpus is not coded with a coherent topic. However, 89.48% of the individual articles have a proportion of at least 10% of at least one coherent topic. If this is accepted as an indicator that an article contains at least one coherent topic, only 10.52% of the corpus is not coded on the article level.

Findings

In order to assess the overall thematic structure of corpus, we constructed a network graph that maps out positive correlations between the 25 thematically coherent topics. The spatial layout of the network was generated with a force-directed graph drawing algorithm³, which means that positively correlated topics are drawn together, while negatively correlated or uncorrelated topics are pushed apart. Thus, the network layout is a simple indicator of thematic similarities and differences in the corpus, and can aid us in identifying some broader thematic categories. We have established three categories, using the network of relations as a guide in our qualitative interpretations of the terms and documents associated with each topic (see Data). These categories are: *Law enforcement* (eleven topics), *social and health problems* (eight topics) and *international* (two topics). The remaining four topics are considered separately, as they hold special interest for the analysis and/or do not fit into a category.

Figure 1 visualizes this network with the categorisation of topics represented as seven distinct node colours. Each topic is shown as a network node, the size of which corresponds to the topic's mean proportion in the corpus. Positive correlations between topics are represented as edges connecting the nodes, and the width of each edge corresponds linearly to the magnitude of the correlation coefficient (Pearson's r), which spans from approximately 0.01 to 0.3.

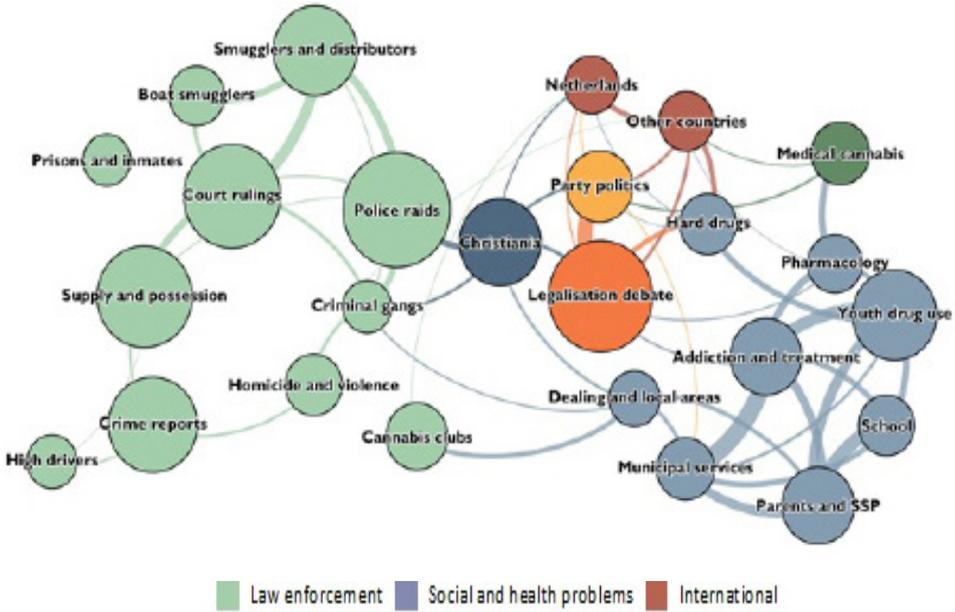


Figure 1: Correlation network of coherent topics

When considering the spatial distribution of categories in the network, it becomes clear that the corpus is thematically polarized. The law enforcement topics are gathered on the left-hand side of the network with few or no connections to the right-hand side, which is almost entirely occupied by topics covering political, social and health-related issues. The topic of Christiania lies in the middle of the network with connections to major topics on both sides. This indicates that the specific case of Christiania often occasions the discussion of more general topics such as cannabis legalization.

Category	Mean proportion
Law enforcement	32.8%
Social and health issues	19.0%
Legalization	5.1%
Christiania	3.6%
International	3.6%
Party politics	2.6%
Medical cannabis	2.0%

Table 3: Corpus-wide proportions of topic categories

In Table 3 we show the corpus-wide proportions of the various topic categories. Law enforcement is the most prevalent category accounting for 32.76% of the corpus. This finding echoes numerous previous quantitative studies, which have found law enforcement or criminal justice to be the most common frame in newspaper coverage of illicit drug issues (e.g. Hughes et al., 2010; Noto et al., 2006; Paimre, 2015). However, these studies have typically estimated the actual proportion of this frame to be between 50% and 60%. The lower proportion in our data may indicate a difference in Danish newspaper coverage, but these corpus-wide proportions are subject to some uncertainty due to our data's uneven coverage of each publication type (see Data). Thus, we provide publication-specific proportions of the law enforcement category in Table 4.

Publication type	Mean proportion of 'law enforcement'
National publications	27.0%
News agencies	44.5%
Local weeklies	37.2%
Regional publications	37.6%
Magazines	3.9%

Table 4: 'Law enforcement' prevalence vs. publication type

None of the publication types contain the 50-60% identified in previous studies, which strengthens the assumption of a difference in Danish coverage. However, it should be noted that the cited studies used articles as the unit of analysis, whereas

we use a mixed-membership model in which one article is allowed to deal with multiple topics. If we simply count the number of articles with a proportion of law enforcement topics of at least 10%, these articles account for 56% of the total corpus⁴.

Magazines and national publications have the lowest mean proportion of this category, while news agencies as well as local and regional publications have a significantly larger proportion. The high proportion in articles from news agencies may be an indication of a close relationship between law enforcement agents and news agencies, which often rely on governmental institutions as a stable source of routinely generated news events (Blood et al., 2003; Chermak, 1997). The difference between national and local/regional publications highlights the importance of regionality and the 'proximity' news value (Price and Tewksbury, 1997): Criminal activity in specific regions seems to be a more salient issue for local residents than for the population in general.

The above findings aggregate all articles from the entire time period and, thus, obscure any temporal patterns. In order to gain an overview of longitudinal thematic developments, we can plot the distribution of categories over time. In Figure 2, we show relative categorical proportions for each year in the time period, calculated as the sum of proportions of each topic in the category and normalized such that the categories sum to 100% for each year. We provide separate diagrams for four different publication types due to the fact that an aggregate diagram would necessarily be biased by the temporal variation in coverage of each publication type (see Data), and since the prevalence of topics may vary according to publication type (e.g. Table 4) this could lead to false impressions of the temporal variation of category proportions.

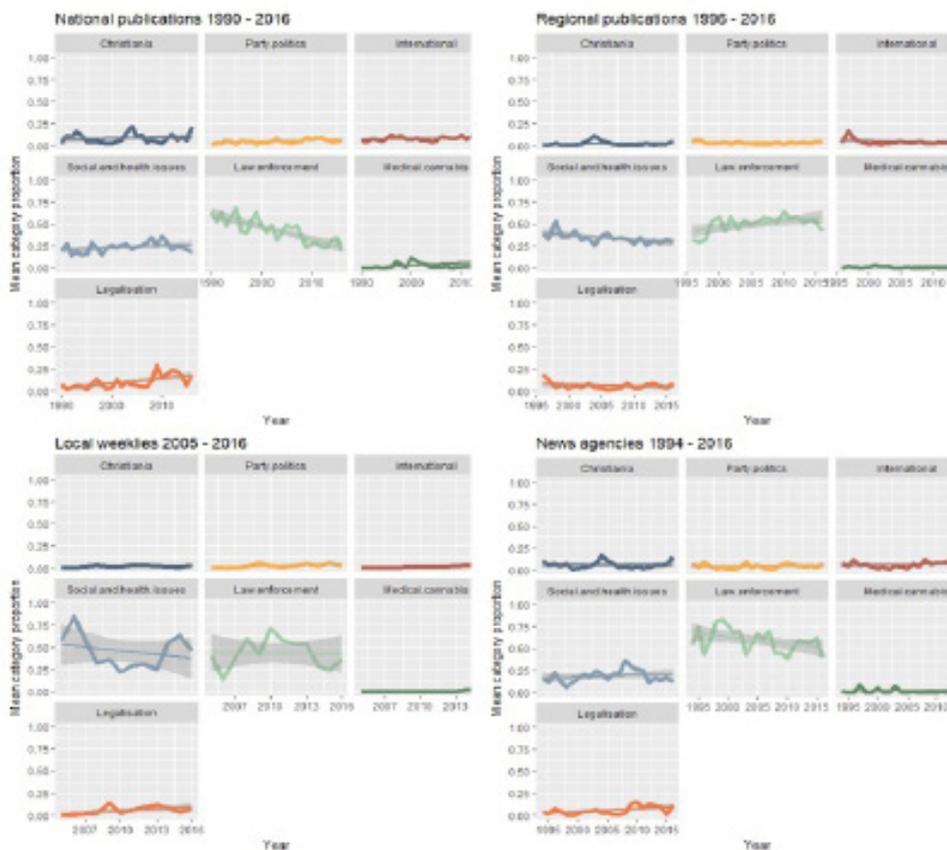


Figure 2: Category proportions per year with trendlines and confidence intervals (linear model)

Several notable trends emerge from these diagrams. Law enforcement and social/health problems are consistently the most dominant themes. This is equivalent to the findings of Månsson (2016) in her qualitative analysis of Swedish newspaper coverage of cannabis in two distinct time periods, where juridical and ‘social problem’ discourses were consistently the most prominent in both periods. In our data, however, national publications show a marked decrease in content related to law enforcement, with a simultaneous increase in content related to legalization and medical cannabis. This finding indicates an actual shift in the coverage of cannabis-related issues. The same trend is evident in articles from news agencies, although to a lesser degree. Finally, regional publications show the opposite ten-

gency – a slight increase in law enforcement content. These findings highlight the importance of differentiating between various types of news media.

The topic of Christiania is characterized by two major peaks around 1993 (national publications) and 2004 (national publications and news agencies), corresponding to periods of increased police activity – most notably the major ‘crackdown’ in 2004. This indication of ‘cyclic’ or ‘event-centred’ press coverage (Paimre, 2015; Tieberghien, 2014) can also be seen with the topic of medical cannabis, which generally has a very low relative proportion that only increases periodically around the turn of the century and in the past three years. The law enforcement topic also has notable peaks, which correspond to ‘celebrated cases’ such as the disruption of major smuggling operations, while the legalization topic has several peaks coinciding with a number of policy proposals, to which we shall return later. In order to account for these peaks, we show the longitudinal trends of the prevalence of four topic categories in Figure 3 (all newspaper types), along with highlighted news events that feature prominently in articles from periods with increased coverage. This figure visualizes and explains some of the longitudinal trends identified via the model, but also serves as a demonstration of the model’s predictive validity, i.e. the correlation between topic prevalence peaks and real-world events.



Figure 3: Longitudinal trends of four topic categories (all newspaper types).

Figure 3 aggregates all newspaper types and thus is subject to some bias from the time-variant coverage of each newspaper type. Furthermore, we have previously shown that both overall topic proportions and longitudinal trends may vary

among newspaper types (see Table 4 and Figure 2). This means, that considering topic proportions within a single newspaper type yields the most clear and interpretable results. One example of this is the legalization topic. This topic is most likely to appear in national newspapers (see Figure 4), which indicates that the legalization debate is a national issue more than it is a local concern. To further our understanding of the legalization debate as a national issue, we created a graph showing topic category proportions over time (excluding incoherent topics) in articles published in national newspapers – see Figure 5.

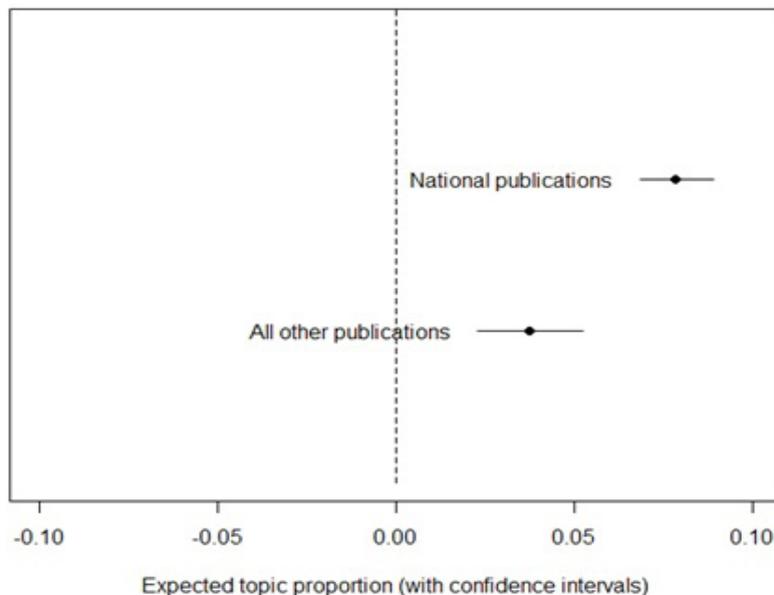


Figure 4: Legalization topic in national vs. other publications.

Between 1989 and 2008, the proportion of the legalization topic fluctuates between 2% and 15%. However, from 2009 onwards, another pattern develops, with a sharp peak at 35.4% in 2009 and a proportion fluctuating between 20% and 27.5% in the years 2010–2014. In other words, the coverage of the legalization debate shows a lasting increase in national newspapers in the years following the sharp peak of 2009. It is notable that this increase is not evident when considering the category proportions in all newspaper types (Figure 3).

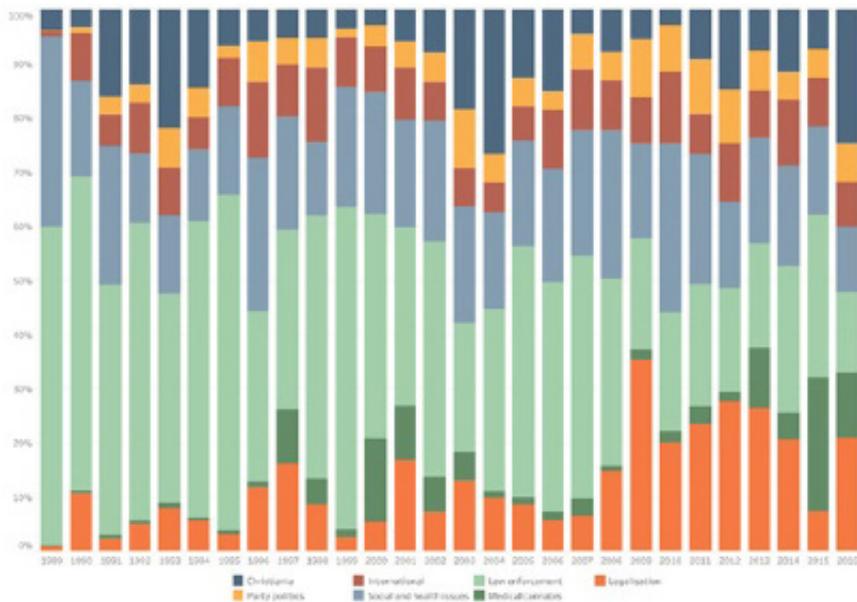


Figure 5 Category proportions per year: National publications.

As is reflected in Figure 3, 2009, the year marking the beginning of a period with more coverage of the legalization issue, saw severable notable events related to cannabis policy both in Denmark and internationally. In order to further our understanding of this ‘turning point’, we can analyse the *content* of the articles on legalization published before and after 2009, and, thus, determine whether the shift in magnitude is accompanied by a shift in content.

To do this, we analysed term frequencies in articles about legalization (topic proportion of at least 10%) from before and after 2009. Figure 6 displays word clouds of the 100 most frequent terms from each period, with each term scaled by frequency. The word clouds demonstrate some remarkable differences between the two periods, most notably in the term ‘København’ (Copenhagen), which goes from being relatively infrequent to being by far the most frequent term.

tential threat to vulnerable young people. In this way these articles articulate one of the most important representations of this drug since the 1960s: Cannabis as an alien entity that has the potential to corrupt young people and spread among them as an epidemic (Houborg 2008; Houborg & Vammen 2012). A recurrent type of news story is, thus, about outbreaks of cannabis use in a group of young people in a particular location. In the terminology of the *pharmakon*, it could be said that cannabis is framed as a poison or a contagion. In Figure 1, the law enforcement topics and the social and health issues topic were almost unrelated. However, there were some links between the two topics, namely between 'dealing in local areas' on the social and health issues side and 'cannabis clubs' and 'criminal gangs' on the law enforcement side. Again, these are mainly local issues relating to particular problems: Drug dealing in local areas can be both a law enforcement and a social issue. Christiania was situated between law enforcement and social and health issues, linking the two topics. Christiania, or rather the cannabis market at Christiania, appears to be a topic that has come to symbolize the cannabis issue in Denmark. Both proponents and opponents of the liberal Danish cannabis policy, which de-penalized possession of cannabis for personal use from 1969 to 2004, have viewed the Christiania market as a manifestation of this policy. This may be why 'Christiania' appears as an important term in the legalization topic, particularly during the period 1989–2009. Furthermore, the open cannabis market has been a challenge to a cannabis policy that decriminalized only the possession of cannabis for personal use, not its distribution, particularly when the cannabis market at Christiania seemed to come under the control of organized crime gangs.

The cannabis legalization debate has changed significantly in recent years. In the current public debate about cannabis in Denmark, the main issue concerns the legal status of cannabis, particularly whether or not the drug should be legalized, that is, allowed to exist as a legal intoxicant. During the late 1960s and early 1970s, the legalization of cannabis was seriously debated in Denmark, for instance during a 1969 parliamentary debate about which policy would be best in relation to the drug culture that had developed since the early 1960s (with cannabis as the main ingredient). However, due to insecurity and lack of knowledge about the risks associated with cannabis use and the effects of legalization on cannabis consumption and the cannabis market in Denmark, a majority decided to continue prohibition for two more years until 1971 in the hope that more knowledge would be available by then. This decision was taken on the same day that the parliament decided to introduce an increased penalty for professional drug trafficking and drug dealing and simultaneously to de-penalize possession of cannabis for per-

sonal use (Houborg & Vammen 2012). The two major concerns were the possible effects of cannabis policy on public health and public order. Would legalization increase the number of people who used cannabis, and which public health effects would this have? Would prohibition prevent cannabis-related organized crime or would it be a driver for such crime? Since 1969, one of these questions has been answered. Prohibition did not prevent cannabis-related organized crime: Rather, the illegal cannabis market is a major source of revenue for organized crime. It is difficult to say whether decriminalization of cannabis for personal use meant that more or fewer people have used cannabis, but after a zero-tolerance policy towards all possession of cannabis was introduced in 2004, there is no indication that cannabis consumption has decreased. The year 2009 marks a turning point in the cannabis legalization debate in Denmark. The word cloud suggests that the proposal of the Copenhagen Municipality to legalize cannabis was an important event in the reframing of the cannabis legalization debate. The proposal framed cannabis as a criminogenic substance that had acted as a driver of organized crime and of a wave of violence that erupted when the police disturbed the cannabis market at Christiania. This represented the opposite approach to that taken in the 1969 debate, in which increased criminalization was introduced to combat organized crime. In the more recent context, the main arguments against cannabis legalization are public health concerns about increased consumption of cannabis and as a consequence higher incidence of cannabis dependence and mental illness associated with cannabis use. Less often heard are public health arguments in favor of legalization. These could concern the current lack of regulation of cannabis products and the impossibility of introducing a harm-reduction policy when the drug is illegal, which may increase the incidence of dependence.

To conclude, we would like to observe that there seem to be different ways in which cannabis is articulated and framed in Danish newspapers that are not necessarily related. The most important topics in terms of the weight they have in our corpus are the local, regional and news agency stories that report on day-to-day cannabis concerns and cannabis regulation in Denmark under the current policy. They concern cannabis-related crime and social and health problems. On the other hand, there are the topics concerning which kind of cannabis policy should be adopted in Denmark. Within this topic, our data show an increased salience of the cannabis legalization issue and also what could be seen as a reframing of the debate to focus more on the issue of organized crime. This reflects the cannabis debates in other countries. For example, Room (2013) has noticed that it has been the criminological rather than the public health arguments that have led to can-

nabis legalization in several states in the USA and in Uruguay. The question is whether we will see a similar change of cannabis policy in Denmark and a reframing of cannabis as a legal intoxicant. The research presented here shows the long term development of the cannabis issue in Danish newspapers. This approach can be a useful addition to case studies of specific events where media frame particular drugs and/or drug related issues and how this affects policy, such as the study by Hughes et al. about mephedrone (Hughes et al. 2010). It could be of benefit to our understanding of the relationship between media framing and drug policy to put such events into a longitudinal perspective. During the period covered in the research we present here, we have not noticed events or moral panics like the one reported by Hughes et al. This may be because our research concerns cannabis which has been part of Danish society since the mid 1960s. If we had reported on XTC instead, we would have come across such moral panics during the early 2000s in Denmark. The research presented here may also add to the existing literature about the relationship between media and drug policy and more generally how drugs are framed. Our research shows that national newspapers on the one hand and regional and local newspapers on the other tend to report on cannabis and cannabis related issues in different ways. Where the national media tend to report on cannabis in relation to particular events, such as large drug seizures or political proposals and to be more directly concerned with policy issues, the local and regional media tend to report on the day-to-day concerns and regulation of cannabis and cannabis related issues. The local and regional newspapers do report on policy issues like new legislation etc. like the national newspapers, but they also report on local cannabis policy and regulation that more rarely find their way into the national newspapers. The research presented here therefore point to the relevance of investigating how different types of media frame drugs and the different media effects this may produce.

Notes

- 1: This was the rule, but in certain cases prosecution could take place - for instance when this was necessary to build cases against drug dealers.
- 2: All data processing was conducted via the programming language R (R Core Team, 2016), using a combination of publicly available and custom-built scripts and algorithms.
- 3: Using the Force Atlas 2 algorithm from the Gephi software package (Bastian et al., 2009).

4: In order to make a meaningful comparison between these figures, it would be necessary to replicate our method with the same corpora as these previous studies.

References

- Baker, Paul (2006) Using Corpora in Discourse Analysis. Continuum.
- Bastian, Mathieu, Heymann, Sebastien and Jacomy, Mathieu (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Third International AAAI Conference on Weblogs and Social Media.*: 361–362.
- Bednarek, Monika (2006) *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*. Continuum.
- Bell, Allan (1991) *The language of news media*. Blackwell.
- Blei, David M (2012) Probabilistic topic models. *Communications of the ACM*. 55(4): 77–84. <https://doi.org/10.1145/2133806.2133826>.
- Blei, David M and Lafferty, John D (2007) A correlated topic model of Science. *The Annals of Applied Statistics*. 1(1): 17–35. <https://doi.org/10.1214/07-AOAS114>.
- Blood, R Warwick, Williams, Jordan and Mccailum, Kerry (2003) Representations of Public Risk : Illegal Drugs in the Australian Press. *Media Interntional Australia incorporating Culture and Policy*. (108): 82–100. <https://doi.org/10.1177/1329878X0310800110>.
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., and Blei, D. M. (2009) Reading tea leaves: How humans interpret topic models. *NIPS '09 Proceedings of the 22nd International Conference on Neural Information Processing Systems*: 288–296.
- Chermak, Steven (1997) The presentation of drugs in the news media: The news sources involved in the construction of social problems. *Justice Quarterly*. 14(4): 687–718. <https://doi.org/10.1080/07418829700093551>.
- DiMaggio, Paul, Nag, Manish and Blei, David (2013) Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*. 41(6): 570–606. <https://doi.org/10.1016/j.poetic.2013.08.004>.
- Entman, Robert M (1993) Framing: Toward Clarification of a Fractured Paradigm. *Journal of Communication*. 43(4): 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>.
- Fairclough, Norman (1995) *Media discourse*. E. Arnold.
- Fan, David P (1996) News Media Framing Sets Public Opinion That Drugs Is the Country' s Most Important Problem. *Substance Use & Misuse*. 31(10): 1413–1421. <https://doi.org/10.3109/10826089609063984>.
- Gabrielatos, Costas and Baker, Paul (2008) Fleeing, Sneaking, Flooding. *Journal of English Linguistics*. 36(1): 5–38. <https://doi.org/10.1177/0075424207311247>.
- Grimmer, Justin and Stewart, Brandon M (2013) Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*. 21(3): 267–297. <https://doi.org/10.1093/pan/mps028>.
- Haines-Saah, Rebecca J, Johnson, Joy L, Repta, Robin, et al. (2014) The privileged normalization of marijuana use - an analysis of Canadian newspaper reporting, 1997-2007. *Critical public health*. 24(1): 47–61. <https://doi.org/10.1080/09581596.2013.771812>.

- Hall, David, Jurafsky, Daniel and Manning, Christopher D (2008) Studying the history of ideas using topic models. *EMNLP '08 Proceedings of the Conference on Empirical Methods in Natural Language Processing*. 363–371. <https://doi.org/10.3115/1613715.1613763>.
- Hall, Stuart (1978) *Policing the crisis : mugging, the state, and law and order*. Macmillan. <https://doi.org/10.1007/978-1-349-15881-2>.
- Hughes, Caitlin Elizabeth, Spicer, Bridget, Lancaster, Kari, et al. (2010) *Media reporting on illicit drugs in Australia: Trends and impacts on youth attitudes to illicit drug use*. Sydney: National Drug and Alcohol Research Centre.
- Hughes, Caitlin Elizabeth, Lancaster, Kari and Spicer, Bridget (2011) How do Australian news media depict illicit drug issues? An analysis of print media reporting across and between illicit drugs, 2003–2008. *International Journal of Drug Policy*. 22(4): 285–291. <https://doi.org/10.1016/j.drugpo.2011.05.008>.
- Jockers, Matthew L (2013) *Macroanalysis: Digital Methods and Literary History*. Urbana: University of Illinois Press.
- Jockers, Matthew L and Mimno, David (2013) Significant themes in 19th-century literature. *Poetics*. 41(6): 750–769. <https://doi.org/10.1016/j.poetic.2013.08.005>.
- Kim, Dongwoo and Oh, Alice (2011) Topic chains for understanding a news corpus. In: *International Conference on Intelligent Text Processing and Computational Linguistics*. https://doi.org/10.1007/978-3-642-19437-5_13.
- Lancaster, Kari, Hughes, Caitlin Elizabeth, Spicer, Bridget, et al. (2011) Illicit drugs and the media: Models of media effects for use in drug policy research. *Drug and Alcohol Review*. 30(4): 397–402. <https://doi.org/10.1111/j.1465-3362.2010.00239.x>.
- Lewis, Nehama, Broitman, Doron and Sznitman, Sharon R (2015) Medical Cannabis. *Science Communication*. 37(6): 675–702. <https://doi.org/10.1177/1075547015608507>.
- McCarthy, Michael and O’Keeffe, Anne (2010) *The Routledge Handbook of Corpus Linguistics*. Routledge.
- Mei, Qiaozhu and Zhai, ChengXiang (2005) Discovering evolutionary theme patterns from text. *Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining - KDD '05*. <https://doi.org/10.1145/1081870.1081895>.
- Munksgaard, Rasmus and Demant, Jakob (2016) Mixing politics and crime? The prevalence and decline of political discourse on the cryptomarket. *International Journal of Drug Policy*. 35: 77–83. <https://doi.org/10.1016/j.drugpo.2016.04.021>.
- Månsson, Josefin (2016) The same old story? Continuity and change in Swedish print media constructions of cannabis. *NAD Nordic Studies on Alcohol and Drugs*. 33(3): 267–285. <https://doi.org/10.1515/nsad-2016-0021>.
- Noto, Ana Regina, Pinsky, Ilana and Mastroianni, Fábio de Carvalho (2006) Drugs in the Brazilian Print Media: An Exploratory Survey of Newspaper and Magazine Stories in the Year 2000. *Substance Use & Misuse*. 41(9): 1263–1276. <https://doi.org/10.1080/10826080600754868>.
- O’Halloran, Kieran (2010) How to use corpus linguistics in the study of media discourse. In: O’Keeffe, Anne and McCarthy, Michael (eds) *The Routledge Handbook of Corpus Linguistics*. Routledge. <https://doi.org/10.4324/9780203856949.ch40>.
- Paimre, Marianne (2015) *Mapping Trajectories of Attention to Drug Related Issues in Estonian Main Dailies*. (June): 39–48. <https://doi.org/10.4236/ajc.2015.32005>.

- Price, Vincent and Tewksbury, David (1997) News values and public opinion: A theoretical account of media priming and framing. In: Barnett, GA and Boster, FJ (eds) *Progress in the communication sciences* (Vol. 13). Ablex.
- Ramage, Daniel, Rosen, Evan, Chuang, Jason, et al. (2009) Topic Modeling for the Social Sciences. *Artificial Intelligence*.: 2–5.
- Richardson, John E (2006) *Analysing Newspapers: An Approach from Critical Discourse Analysis*. Palgrave Macmillan.
- Roberts, Margaret E, Stewart, Brandon M, Tingley, Dustin, et al. (2014) Structural topic models for open-ended survey responses. *American Journal of Political Science*. 58(4): 1064–1082. <https://doi.org/10.1111/ajps.12103>.
- Roberts, Margaret E, Stewart, Brandon M and Tingley, Dustin (2016) *stm: R Package for Structural Topic Models*.
- Scheufele, Dietram A and Tewksbury, David (2007) Framing, agenda setting, and priming: The evolution of three media effects models. *Journal of Communication*. 57(1): 9–20. <https://doi.org/10.1111/j.0021-9916.2007.00326.x>.
- Simon, Adam F and Jerit, Jennifer (2007) Toward a theory relating political discourse, media, and public opinion. *Journal of Communication*. 57(2): 254–271. <https://doi.org/10.1111/j.1460-2466.2007.00342.x>.
- Smith, Alison (2015) Mind-Altering Discourse: Experts and Sources in the News on the UK Drug Policy Reform Debate. *Journal of Promotional Communications*. 3(2): 261–267.
- Steyvers, M and Griffiths, T (2007) Probabilistic topic models. In: Landauer, T, McNamara, S Dennis, and Kintsch, W (eds) *Latent Semantic Analysis: A Road to Meaning*. Laurence Erlbaum.
- Sznitman, Sharon R and Lewis, Nehama (2015) Is cannabis an illicit drug or a medicine? A quantitative framing analysis of Israeli newspaper coverage. *International Journal of Drug Policy*. 26(5): 446–452. <https://doi.org/10.1016/j.drugpo.2015.01.010>.
- Taylor, S. (2008) Outside the outsiders: Media representations of drug use. *Probation Journal*. 55(4): 369–387. <https://doi.org/10.1177/0264550508096493>.
- Teo, Peter (2000) Racism in the News: A Critical Discourse Analysis of News Reporting in Two Australian Newspapers. *Discourse & Society*. 11(1): 7–49. <https://doi.org/10.1177/0957926500011001002>.
- Tieberghien, Julie (2014) The role of the media in the science-policy nexus. Some critical reflections based on an analysis of the Belgian drug policy debate (1996–2003). *International Journal of Drug Policy*. 25(2): 276–281. <https://doi.org/10.1016/j.drugpo.2013.05.014>
- van Dijk, Teun A (1988) *News analysis: Case studies of international and national news in the press*. Lawrence Erlbaum.
- Van Gorp, Baldwin (2005) Where is the Frame? *European Journal of Communication*. 20(4): 484–507. <https://doi.org/10.1177/0267323105058253>.
- Venturini, Tomasso, & Latour, Bruno (2010). The social fabric: Digital traces and qualitative methods. *Proceedings of Futur en Seine* 2009.
- Yang, Tze-I, Torget, Andrew J and Mihalcea, Rada (2011) Topic modeling on historical newspapers. *LaTeCH '11 Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. (June): 96–104.
- Yang, Yi, Downey, Doug, Boyd-graber, Jordan, et al. (2015) Efficient Methods for Incorporating Knowledge into Topic Models. In: *Conference on Empirical Methods in Natural Language Processing*. <https://doi.org/10.18653/v1/D15-1037>.