

# MODELING MEDIA HISTORY

## On topic models of Swedish media politics 1945–1989

**Pelle Snickars**

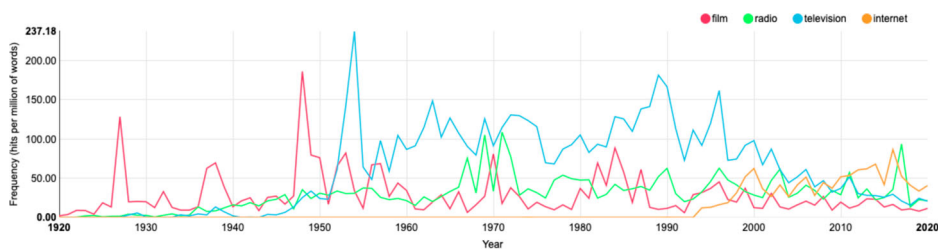
*In an explorative manner, this article uses a data-driven digital history set-up to focus on media political issues in Sweden during the second half of the twentieth century. By distant reading and topic modeling a dataset of 3100 Swedish Government Official Reports between 1945 and 1989—a corpus of some 87 million tokens—the article gives a new perspective of how the Swedish state examined and discussed media in general and media politics in particular. Topic modeling is a computational method to study latent themes or discourses in a dataset by accentuating words that tend to co-occur and together create different topics. Via a computational interrogation of the dataset in a Jupyter Lab environment a number of media topics can be detected. They include the most common words for each media topic, but also reveal temporal periodizations when media political issues were foremost discussed as well as other societal topics that media was related to.*

**KEYWORDS** digital media history; media politics; Swedish Government Official Reports; topic modeling; digital humanities

The scale of empirical material used among media historians has usually been small—some would even say tiny. A bunch of newspapers, maybe a few books or a small body of films, radio or television programs, perhaps part of a broadcast archive, have been compiled to act as empirical evidence and verification of one’s reasoning. Archival research is frequently tedious—but also anecdotal; qualitative humanistic research practices can be haphazard. Most media historians (but far from all) have consequently refrained from scaling-up the nature of their arguments towards more general assumptions—with the notable exception of non-empirical and lofty speculations about media history characteristic of so-called mediatization theory.<sup>1</sup> Yet due to digitization efforts at heritage institutions, issues of scale have during the last decade surfaced within media historical work. With computational research methods, it is now possible to examine empirical materials consisting of hundreds of millions of words or large-scale datasets of historical images, sound or film. As a result, it is today feasible to envision media historical studies that are broader in scope than just a decade ago. Digital newspaper archives are the obvious example—in a forthcoming book described as ‘an eldorado’ for media historians—with large-scale datasets used, for example, to study text reuse in millions of pages.<sup>2</sup> A number of media scholars have shown that computational methods can productively be applied to massive cultural heritage datasets to answer media historical research questions.<sup>3</sup> The Media History Digital Library has, for instance, transformed the study of film and broadcasting history, offering access to millions of pages of out-of-copyright books and magazines. Through its application Arlight computational research methods have also been integrated into the study of media history.

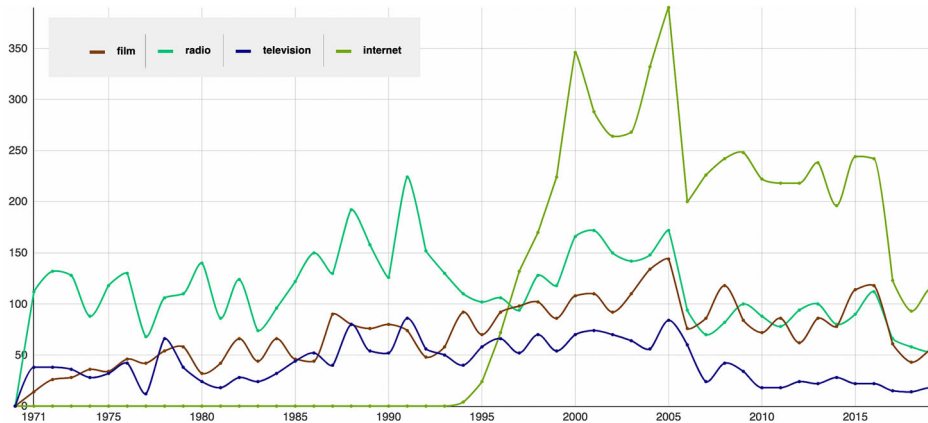
Analysing media history at scale requires access to massive datasets. Some of the newly established library labs at national libraries or major university libraries offer scholars access to such media historical datasets. Others can be found or compiled online. But in general intellectual property rights issues continue to be a problem for researching media historical matters on a more substantial scale, and particularly so for studies focusing on the twentieth century. In the online version of the digital newspaper archive at the National Library of Sweden, for example, content after 1905 cannot be accessed at all (only metadata). If it is true that the internet treats such censorship as a malfunction and routes around it, such a detour might lead the media historian to other empirical materials—such as parliamentary data. Scholars of parliamentary history have a huge advantage over media historians since their principle empirical material—parliamentary debates, motions, propositions, governmental reports etcetera—is nearly always openly available for democratic reasons. The parliamentary corpora offered within the European CLARIN infrastructure gives access to 26 datasets in different languages containing hundreds of gigabytes. Most of this data is of a more recent date, but open datasets of parliamentary content from the nineteenth and twentieth centuries can be found online as well. As Figures 1–3 illustrate, some of these datasets can also be explored in a similar manner as within Arlight.

DH-scholars are usually prone to work with such datasets with their own tools and methods. But as my simple graphs indicate, easy search interfaces are nowadays available online. The University of Huddersfield has, for example, produced an accessible web application—Hansard at Huddersfield—with all official reports from British parliamentary debates from 1803 until today. The dataset contains hundreds of millions of tokens. A similar German application, Open Discourse, gives access to all debates, *plenarprotokolle*, within the Bundestag (the German federal parliament) from 1949 until today, a corpus of some 80,000 documents with 200 million tokens. In Sweden, a comparable (but smaller) application lets researchers explore parliamentary motions (some 145,000 documents) during the last 50 years. Naturally, such massive parliamentary data contains discussions about ‘the most important topics facing societies and their citizens’—including the media, of course.<sup>4</sup> With these short examples, I simply want to stress that large-scale



**FIGURE 1**

Graphical depiction of the frequency of media keywords (film, radio, television, internet) within parliamentary discourse during different time periods—from (Figure 1) the British *Hansard* ([hansard.hud.ac.uk](http://hansard.hud.ac.uk)), from (Figure 2) from Swedish parliamentary motions ([riksdagsmotioner.dh.gu.se](http://riksdagsmotioner.dh.gu.se)), and from (Figure 3) from German plenary minutes from the Bundestag ([opendiscourse.de](http://opendiscourse.de)). The latter graph displays how often the five different political parties in Germany debated the topic of *Rundfunk*, that is broadcasting.



**FIGURE 2**

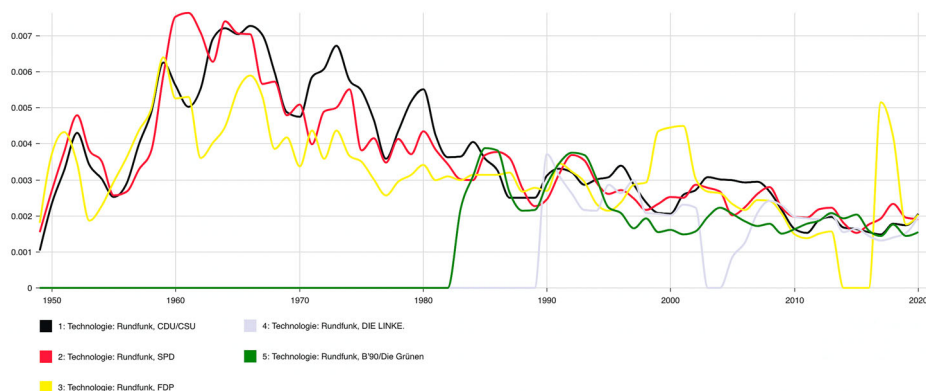
Graphical depictions of the frequency of media keywords (film, radio, television, internet) within parliamentary discourse during different time periods from Swedish parliamentary motions (riksdagsmotioner.dh.gu.se).

analysis of parliamentary datasets can potentially be media historically revealing, and especially so regarding the history of media politics.

This article is based on one such openly available parliamentary dataset, so-called Swedish Government Official Reports. In an explorative manner, my purpose is to perform a large-scale study of how national media politics evolved in Sweden during the second half of the twentieth century. The research is related to the project, Welfare State Analytics. Text Mining and Modeling Swedish Politics, Media & Culture, 1945–1989 which uses probabilistic methods and text mining models to study three massive datasets from the domains of Swedish politics, news media and literary culture.<sup>5</sup> The data used in the article is derived from 3100 Swedish Government Official Reports published between 1945 and 1989—so-called SOU-reports (*Statens offentliga utredningar*) with the terminology of this article. It is a dataset containing some 87 million tokens. By distant reading and topic modeling this dataset, the article gives a new perspective on how the Swedish state examined and discussed media in general, and media politics in particular. Via a computational interrogation of the dataset in a Jupyter Lab environment a number of media topics can be detected. They include the most common words for each media topic, but also reveal temporal periodizations when media political issues were mostly discussed, as well as other societal topics in the dataset that media was related to.

## Data and Method

Before submitting a proposal for new legislation, the Swedish government regularly examines alternatives, a task prepared by an appointed committee. Other Scandinavian countries, such as Norway, have a similar system. In general, the governmental committee



**FIGURE 3**

Graphical depictions of the frequency of media keywords from German plenary minutes from the Bundestag (opendiscourse.de). The graph displays how often the five different political parties in Germany debated the topic of *Rundfunk*, that is broadcasting.

process has been a way of accessing knowledge about various issues, and each year some 50–70 commissions deliver results in the form of SOU-reports, usually book length publications spanning some 250–350 pages. During the twentieth century governmental commissions developed into an effective instrument of parliamentarism in Sweden. Since 1922 the work of commissions has been published in a SOU-series with a distinct number; the first SOU-report devoted to film was for example SOU 1930:26. Normally, SOU-reports and work performed within governmental committees had the task of preparing the state for apt and rational decision-making. SOU-reports were hence a basis and rationale for both parliamentary debates as well as (later) practical politics. According to Rune Premfors, after 1945 ‘the range of subjects covered by governmental committees has expanded to include virtually every area of the Swedish welfare state’—the latter a term often used to designate a long postwar period of social-democratic social reforms (pensions, health insurance, child care), including a previously unprecedented economic growth. SOU-reports were not general, they explicitly addressed a particular and often thematic issue. Stressing the importance of the work executed by governmental committees, some 40% of all legislation in Sweden around 1970 was based on commission proposals.<sup>6</sup> Among historians, the SOU-series—with its external investigations usually running for a number of years—is therefore considered a valuable historical source for Swedish policy-making.<sup>7</sup> Governmental committees often included legislative suggestions; they were an arena for the exchange of factual arguments among experts (and academics), situated within a foremost social-democratic, and rationally oriented Scandinavian style of policy-making (Figure 4).

In 2015 the National Library of Sweden finished digitizing all SOU-reports from 1922 to 1999—all in all, more than 6100 publications. They are all publicly available online, OCR-



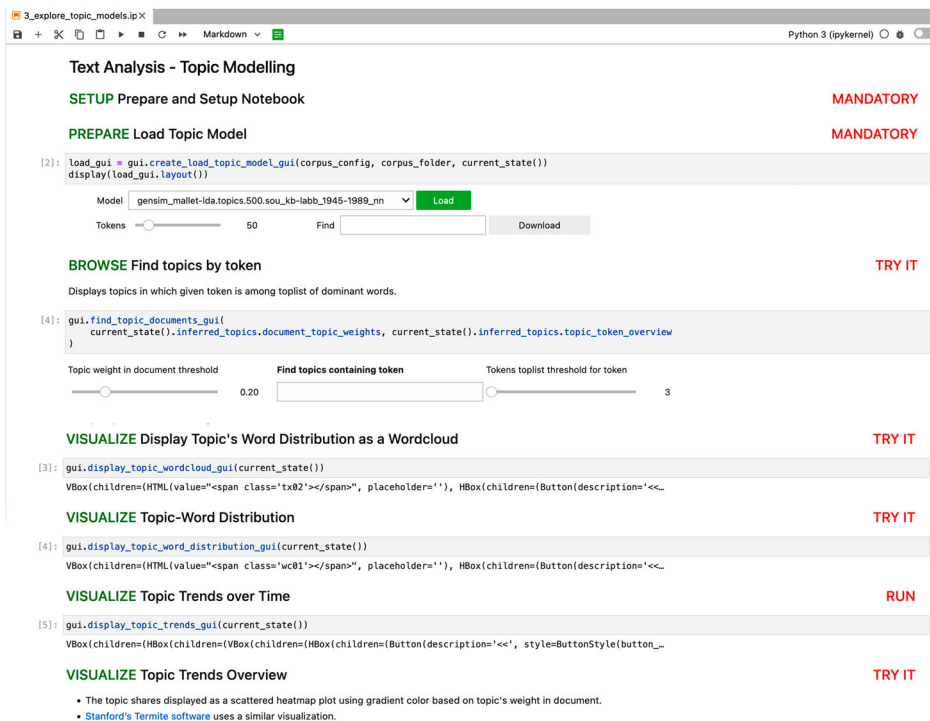


FIGURE 5

Screenshot of Jupyter Lab environment for the analyses of SOU-data—with the possibility to load different topic models with 50, 100, 200 or 500 topics. Grey areas contain executable Python code which can be run by the media historian to perform data analysis on, for example, topic weight, topic word distribution or topic trends during the time period 1945–1989.

keyword, including where this word features most prominently. Analysed within the Jupyter Lab environment, the SOU-data is an abundant empirical source (Figure 5). Naturally, topics also occur in relation to each other, and clusters or networks can be visualized at different time periods using software such as Gephi. I am aware that large-scale network data and advanced network analysis is a vibrant field within the digital humanities, often used in connection to topic modeling. Network visualizations can help to contextualize studied topics, and thus gain a more nuanced understanding of how the topics were discussed and how the overall conversation developed over time.<sup>9</sup> However, network analysis is not a distinct method used in this article—topic modeling is.

In an analysis of discussions during the nineteenth century on infrastructure in the British parliament (based on the Hansard transcripts), Jo Guldi has stated that topic modeling makes it possible to study and envision ‘the invisible categories that structure mind, language, priorities, or prejudice in a given corpus’.<sup>10</sup> Literary scholars have used topic models for more than a decade,<sup>11</sup> and the method is becoming increasingly popular among historians. In a fascinating media historical book chapter on topic modeling ‘humanism discourse’ in early nineteenth-century German press, Heidi Hakkarainen and

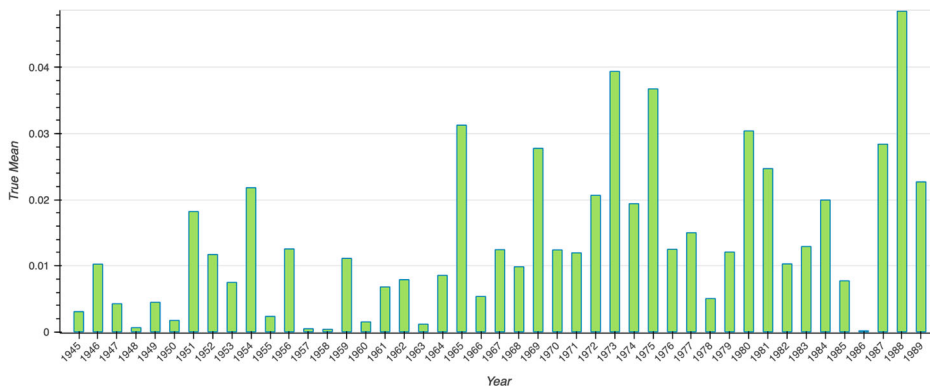


Zuhair Iftikhar state that even if the old German Fraktur typeface is challenging for OCR, topic models as a probabilistic tool 'has attracted attention among historians, because it enables detecting underlying thematic structures behind a large corpus of documents, as well as surprising connections between individual texts'.<sup>12</sup> This article focuses on topics of media in a text corpus of a more recent date and depending on how many topics a model displays different media topics can be detected. Importantly, all topic detection is concerned with *modeling data*, and there are also different forms of dynamic topic modeling or structural topic modeling.<sup>13</sup> Historical data can hence be modeled in different ways, and there are a number of ways to go about producing an almost endless number of results. It is often up to the historian to select the models most suitable. At Humlab we have used the generative probabilistic model Latent Dirichlet Allocation (LDA) and predominantly a Mallet—and occasionally a Gensim—modeling toolkit. Models can be instructed to single out  $n$  number of topics; in our case models with 50, 100, 200 and 500 topics were used. In the 50-model of the SOU-data, one single media topic could be found, whereas in the 500-model there were several dozen, with more specific traits as for example topics on film censorship or daily press subsidies. All topics are arbitrarily numbered in the Jupyter Lab environment—from 0 to 49, or from 0 to 499—and it is up to the scholar to interpret what they are about, including naming them (in this article I simply use numerals). Topic models are based on statistics and numbers, but they result in strings of words related to one another which are prone to graphic visualization. In the following, my arguments will be described as well as *shown*; the article contains 16 figures and graphical networks (with keywords from the SOU-data translated by me). Importantly, the article is not about close reading SOU-reports, even if I quote some of them. Instead, I consider all SOU-reports as one single text, as a dataset compiled by me. The empirical scale then becomes different—'the new scale changes our relationship to our object, and in fact *it changes the object itself*. The same goes for the new forms of knowledge produced by topic modeling. 'Distant reading ... is a condition of knowledge.'<sup>14</sup>

## Modeling Media

If one models the SOU-data in 50 different topics, the Jupyter Lab environment will plot results based on broad societal categories like health care, taxation, school and education or the labor market. However, the 50-model also singles out a distinct media topic, arbitrarily given number 8. This topic can be found in almost 5% of all 3100 SOU-reports. As [Figures 6](#) and [7](#) indicate, the media topic includes keywords such as film, newspaper, radio and television. It increases in strength from the mid-1960s and onwards. One historical observation is that media in general has been considered rather important by the Swedish government, partly due to the number of media-related reports or because media was discussed in other reports. Media—and particularly media politics—features in the SOU-data just like other political issues on housing, migration, or the environment. Media is hence not a topic on the margin. On the contrary, even if the SOU-data is modeled to discern relatively few topics, media is present among these.

Modeling data in only 50 topics generates more general results; more specific traits are thus often hard to detect. Regarding media topic 8, however, three things are



**FIGURE 6**  
Trends of media topic (number 8) between 1945 and 1989 within the 50-model of the SOU-data.

film	0.018171541691092	television	0.0092633681953735
newspaper	0.0167471105048147	year	0.0080452969117236
radio	0.0147620378131694	writing	0.0071687238740145
program	0.0123094595014125	publication	0.0069778866189299
sweden	0.0107544554563932	image	0.0057470319784801
advertisement	0.0104093048227952	medium	0.0052694822923115

**FIGURE 7**  
12 keywords that are most likely to appear within the media topic (number 8) between 1945 and 1989 within the 50-model of the SOU-data. Film is the medium which is most frequently addressed within the SOU-data, appearing in 1.81% of all SOU-reports.

noteworthy. The first one is obvious: film is statistically the medium that features most prominently within the media topic—an issue I will get back to. Secondly, during the welfare state years, Swedish media politics relied on governmental funding, foremost regarding public service broadcasting and film production. Nevertheless, advertisement is one of the most frequent keywords in the media topic, featuring specific words such as advertising (*annonsering* in Swedish) or tv-commercials. This is somewhat surprising, yet the issue of advertisement has a media history stretching back to the 1920s. The Swedish Newspaper Publisher Association then argued that Swedish radio should not include advertisement—not because they were benevolent towards public service but as a way to secure the commercial interests of newspapers relying on advertisement. Hence, even if commercial advertising was banned from most national media—with the daily press as



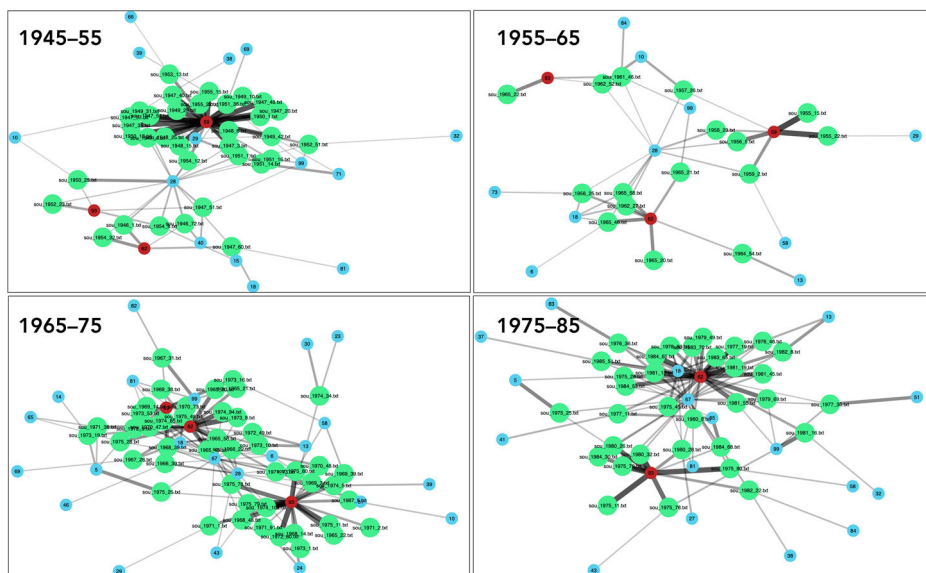
the notable exception—this contested issue was something that SOU-reports repeatedly dealt with during the welfare state period.

A third thing to note regarding media topic 8 is that among the five SOU-reports in which the topic is statistically the strongest, two publications focused on ways in which to archive audiovisual media: SOU 1974:94 on the preservation of sound and image—published by the Data preservation committee on storage techniques for broadcasting, gramophone records and celluloid film—and a later SOU 1987:51, on sound and image preservation for posterity. Governmental media politics in Sweden was hence both geared towards facilitating media *production* through state funding and subsidies, as well as *preservation* of the same media at national archives or libraries. Even though I have myself done a lot of scholarly work on the matter, this finding was completely new—and indeed displays the strength of distant reading and topic modeling.<sup>15</sup>

These somewhat surprising results from the 50-model with advertising and archiving being especially important within the media topic can be compared with previous scholarship on Swedish media history (that have often used individual SOU-reports as a singular empirical source). Classical media scholars such as Lennart Weibull and Stig Hadenius were naturally aware of the issue of advertisement, particularly regarding the introduction of Swedish television during the early 1950s when commercial broadcasting was still an option.<sup>16</sup> Yet the SOU-data reveals that advertisement continued to be a disputable issue within Swedish media politics. Within the 50-model there exists a specific enterprise and market-topic (number 35) with keywords around advertising. If one performs a search within the SOU-data where the exact word advertisement (*reklam* in Swedish) features most prominently, topic 35 is noticeable with two reports on economic planning and market regulation, SOU 1945:42 and SOU 1968:6, as well as one report on the consumption of milk, SOU 1960:29—delivered by the Milk Commission of 1954. These three market-oriented reports are statistically closely followed by one SOU-report from the late 1980s on the future of television, SOU 1989:73. In this latter report there were plenty of discussions on ‘opportunities to earn revenue from advertising in television’ (and commercials in cinemas), accentuating that the issue of advertisement was vivid in Swedish media politics during the later phase of welfare society as well.<sup>17</sup>

Modeling media-related topics can hence reveal a somewhat different historical chronology regarding the issue of advertisement than assumed by previous research. Even more striking, however, is that most Swedish media historians have ignored the fact that discussion around *preservation* was central within national media politics. Topic modeling can hence sometimes yield unexpected results. An advantage of modeling data is also that such an outcome can be further analysed. If one uses the 100-model—in which all 3100 SOU-reports are divided into 100 topics—there is one broadcast topic (number 62), which increases in strength during the 1960s. But there is also an ALM-topic (archive, library, museum) with number 39, that shares a number of keywords with the broadcasting topic—medium, information, archive, image, audience—which further attest that media politics frequently had an archival bias (Figure 8).

Modeling the SOU-data in a hundred topics, results in a more detailed outcome, featuring increasingly specific media topics. Apart from broadcasting, there is, for example, one distinct topic on film (number 59) and one on newspapers (number 93). Since all

**FIGURE 8**

Swedish media history as a network within the 100-model of the SOU-data—at various points in time between 1945 and 1985, with a weight threshold of 10% between individual SOU-reports (in green) and the three topic nodes (marked in red) on film (topic 59), on broadcasting (topic 62) and newspapers (topic 93). Blue nodes are topics that the three media topics are related to, as for example, education (topic 5), legal issues (topic 18), or crime and prison (topic 81).

these topics are about media—number 59 on film, 62 on broadcasting, and 93 on newspapers—they share a number of keywords, and they are also related to other topics in the SOU-data. One way to illustrate these connections is to plot the three media topics (within the 100-model) in a network that changes over time during four periods: 1945–1955, 1955–1965, 1965–1975 and 1975–1985. If the weight threshold is set at 0,1—a link is then plotted if 10% (or more) of a topic occurs—it becomes apparent that the topic on film (number 59) is present in 26 SOU-reports during the first period 1945–1955, but not present at all during the last decade. Instead, the topic on broadcasting and newspapers increases in strength in the SOU-data, with a number of links between the two topics during latter decades. Especially matters around broadcasting expand, with topic 62 being present in 20 SOU-reports during the last period 1975–1985.

More interesting, however, are the linkages to other topics that can be detected in the network. Via SOU 1947:60—a report on the ordinance of the freedom of press—the broadcasting topic 62 is, for example, during the period 1945–1955, related to legal issues (topic 18) and even crime and prison (topic 81). Penal issues return again during the third period, 1965–1975; both the broadcasting and film topic were then related to punishment via SOU 1969:38, a notorious publication in Sweden on the limits of freedom of expression, violations of religious belief and morality. Even during the last period, 1975–1985, topic 81 on crime and prison is still linked to media issues. But then

religious belief is not the problem—rather personal integrity, as investigated in SOU 1980:8.

This report addresses the issues associated with invasion of privacy through the use and dissemination of privacy infringing materials. Most of these invasions of privacy occur through the dissemination of information of a private nature by the mass media, the press, radio and television.<sup>18</sup>

If one proceeds with modeling the SOU-data in even more detail, the 200-model of the Mallet toolkit (again) discerns three distinct media topics: on newspapers and daily press (number 72), on broadcasting (number 125) and on film (number 181). As is evident in Figures 9–11, these three media topics include familiar media historical keywords. The topics occur with various frequency during the welfare state period, with film (again) being the medium that the Swedish government repeatedly dealt with in different committees. When modeling the SOU-data in 200 topics, the aforementioned archival bias is once again present: the three most frequent keywords in film topic 181 are film, cinema—and archive. In fact, among the top twenty keywords of the film topic, the media archival institution ALB—a national Swedish media archive inaugurated in 1978 due to an extension of the national legal deposit law to include audiovisual media—features more prominently than the Swedish Film Institute. If one compares the most frequent keywords in each media topic in the 200-model, most of them are detached—an indication that press, broadcasting and film were examined separately with a rather media specific Swedish vocabulary. Yet as is evident in the Gephi network below they also share some word nodes as distribution, image, information, medium or content.

Finally, if one goes ahead and uses the Mallet 500-model, all 3100 SOU-reports are then divided into 500 topics. It is then possible to discern discourses and themes on a more granular level. Modeling data in such detail is usually the most interesting aspect vis-à-vis the method of topic modeling. Suddenly, smaller facets within the data become visible which are otherwise very difficult for a human reader to identify. If the 50-model had one specific media topic, there are more than 30 topics featuring media-related keywords in the 500-model—from topics on press subsidies to the future of broadcasting, as well as a topic on surveillance, and even one on community radio visible during the 1980s. Hence, all of a sudden an affluent Swedish media history appears—and not only related to media policy. There is, for instance, a topic on lovemaking (number 111), with keywords such as sexual intercourse, contraceptives, and sexual education depicted in various media; it is perhaps less surprising that the topic features most prominently during the late 1960s. Another odd theme—at least for a Swede—is a topic on warfare and media (number 63) with keywords such as war, government, resistance and occupation, as well as radio and news coverage. The topic is mostly related to issues surrounding the Cold War, and appears frequently in two reports: SOU 1953:27 on psychological defense and SOU 1984:10 on 'national non-military resistance as a complement to other total defence measures'.<sup>19</sup> Via SOU 1972:18, a slightly odd report on the Norwegian experiences during the Second World War, the warfare and media topic 63 is linked to a topic on propaganda and information (number 490), most vivid during the 1940s in SOU 1946:86, a well known publication among Swedish media historians focusing on the German

181	film	0.0964902087725414
72	newspaper	0.0862762628520339
125	radio	0.0382910405885653
125	program	0.0377953737764199
181	cinema	0.0195326549470736
125	television	0.0150819876550907
181	archive	0.0150727909563254
72	press	0.0150481853811687
125	transmission	0.0136822744560134
72	circulation	0.0132808682724995
72	daily press	0.012100924202888
72	daily newspaper	0.0111237018016238
125	medium	0.010758152004489

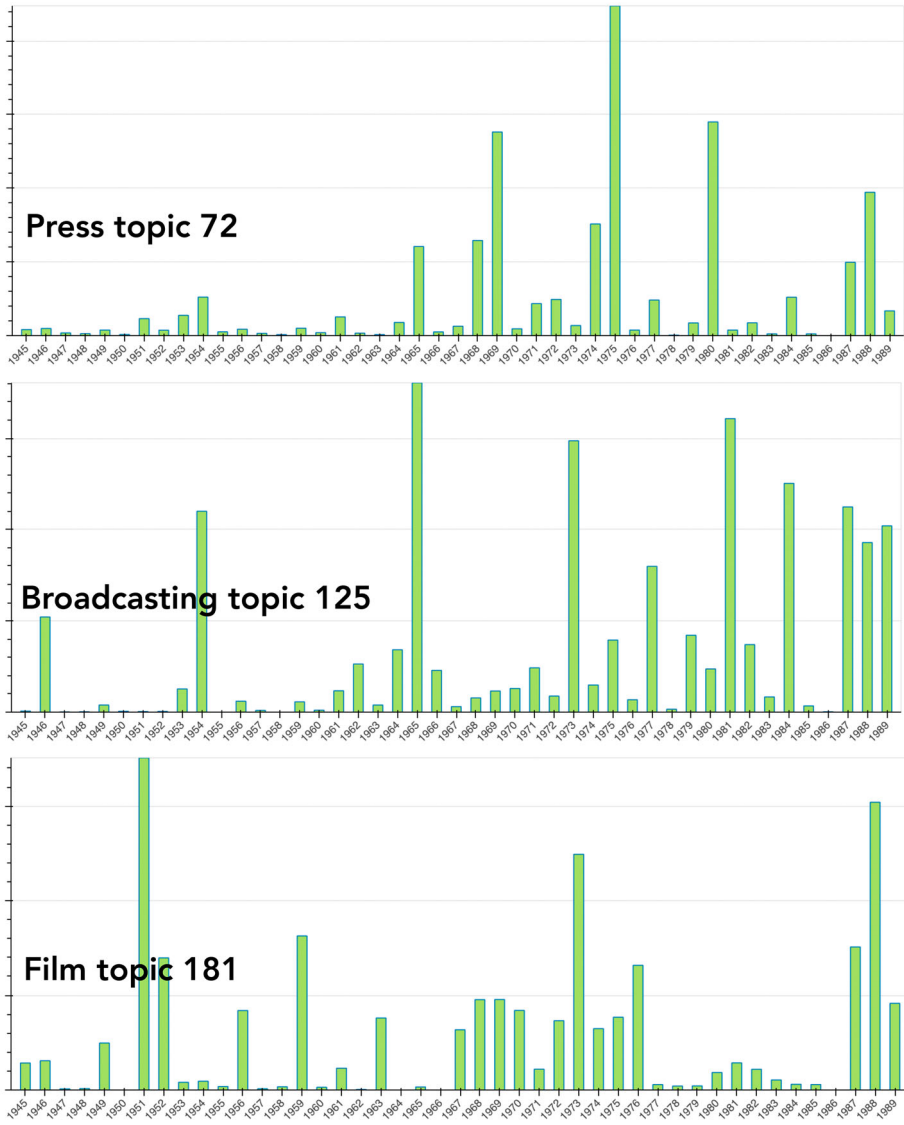
**FIGURE 9**

The 13 most frequent keywords within three distinct media topics of the 200-model of the SOU-data—on press (number 72), on broadcasting (number 125) and on film (number 181)—as well as time periods when they occur most frequently. The network in Gephi (with the algorithm Force Atlas) includes some 250 keywords from the three media topics.

propaganda in Sweden between 1939 and 1945. The two topics also share keywords such as war, German, power, newspaper, police and press.

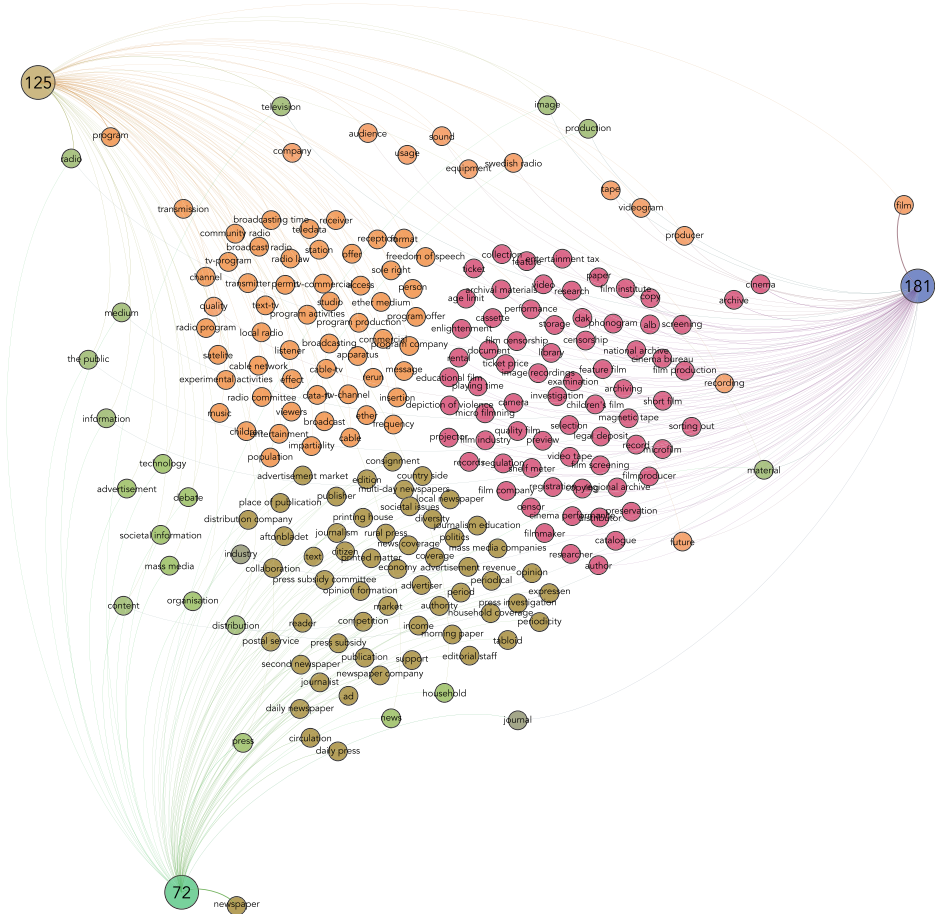
Within the 500-model there is even a meta-topic on media—with keywords such as information, journalist, mass media research and journalism education. This meta-media topic (number 492) increases in strength after 1968 and is particularly strong in two SOU-reports from 1975 and 1977. The first of these, SOU 1975:25 dealt with the need to establish a more general journalism education at Swedish universities, and the other SOU 1977:11, was a report that explicitly addressed Swedish media studies, *Forskning om massmedier* [Research about Mass Media]. As I have argued previously in this journal, the formation of media research in Sweden can partly be seen as an effect of politicians and the media industry wanting to be better informed on issues such as media influence, media ownership and the habits and composition of media audiences.<sup>20</sup> The meta-media topic 492 clearly indicates that such a need became increasingly important for the Swedish government during the 1970s and 1980s (Figures 12 and 13).

Given that there are more than 30 media-related topics in the 500-model, there are numerous ways to model and explore the data. As might be expected, media topics are related to one another, and sometimes share keywords. In general, broadcasting topics and topics concerning earlier forms of automatic data processing appear from the 1960s and onwards, whereas topics on film are more evenly spread, with peaks in the



**FIGURE 10**  
 The 13 most frequent keywords within three distinct media topics of the 200-model of the SOU-data—on press (number 72), on broadcasting (number 125) and on film (number 181)—as well as time periods when they occur most frequently.

early 1950s, around 1970 and towards the late 1980s. Again, it needs to be stressed that the Swedish government and the inquiries they commissioned put a lot of emphasis on the medium of film—both prior to the so-called film reform of 1963 (and the inauguration of the Swedish Film Institute) as well as afterwards during the 1970s. Previous film

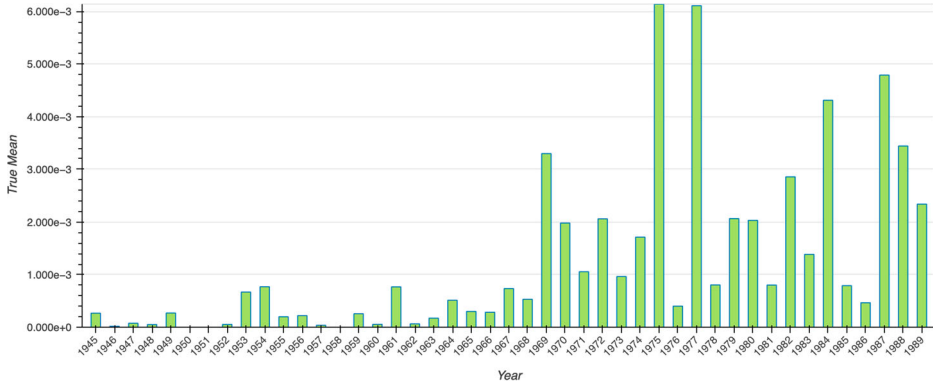


**FIGURE 11**

The network in Gephi (with the algorithm Force Atlas) includes some 250 keywords from the three media topics: on press (number 72), on broadcasting (number 125) and on film (number 181).

historical scholarship in Sweden have been unaware of the range of attention that the government put on the medium of film. In his dissertation, *Staten och och filmen* [The State and Film] Roger Blomberg, for example, used SOU-reports on film, but only a dozen or so. The same goes for film historian Leif Furhammar.<sup>21</sup> Topic modeling all SOU-reports as one dataset however reveals that film was the most prominent medium within the 50-model, and hence occurs frequently within the 500-model as well, so my remaining analyses will focus on moving images. Within the 500-model there is one distinct cinema topic (number 181) with keywords such as film, cinema, screening, producer, film production, television, production, board of film censors and film institute. Most of the SOU-reports where this topic is statistically most frequent were delivered by the 1968 Film





**FIGURE 12** Trends of meta-media topic (number 492) between 1945 and 1989 within the 500-model of the SOU-data.

Commission and its four volume series, *Samhället och filmen I-IV [Society and film]* published during the early 1970s. However, the cinema topic also shares a number of keywords with a topic on authorship, copyright and jurisprudence (number 155) as well as a topic on archival matters (number 343)—again stressing that media and preservation were issues that the government frequently addressed. The linkages between these three topics also suggest that the medium of film was located in quite different cultural policy contexts during the welfare state years (Figure 14).

I have previously addressed the archival bias of a number of SOU-reports on media, and in a similar way a more fine-grained analysis and modeling of the data can reveal how the medium of film belonged to an extensive judicial context. Since a number of SOU-

information	0.2211080092317305	citizen	0.011932717212743
mass media	0.0265361861665138	radio	0.0112574705461435
the public	0.0214264837789625	material	0.0103101095512129
journalist	0.0190883162170061	medium	0.010169013232819
contact	0.0172842990032552	public information	0.0096449411930701
content	0.014835270048275	information activities	0.0086975801981395
press	0.0147546435806214	news	0.0086673452727694
magazine	0.0135855597996432	audience	0.0081331949245638

**FIGURE 13** 16 keywords that are most likely to appear within the meta-media topic (number 492) between 1945 and 1989 within the 500-model of the SOU-data.

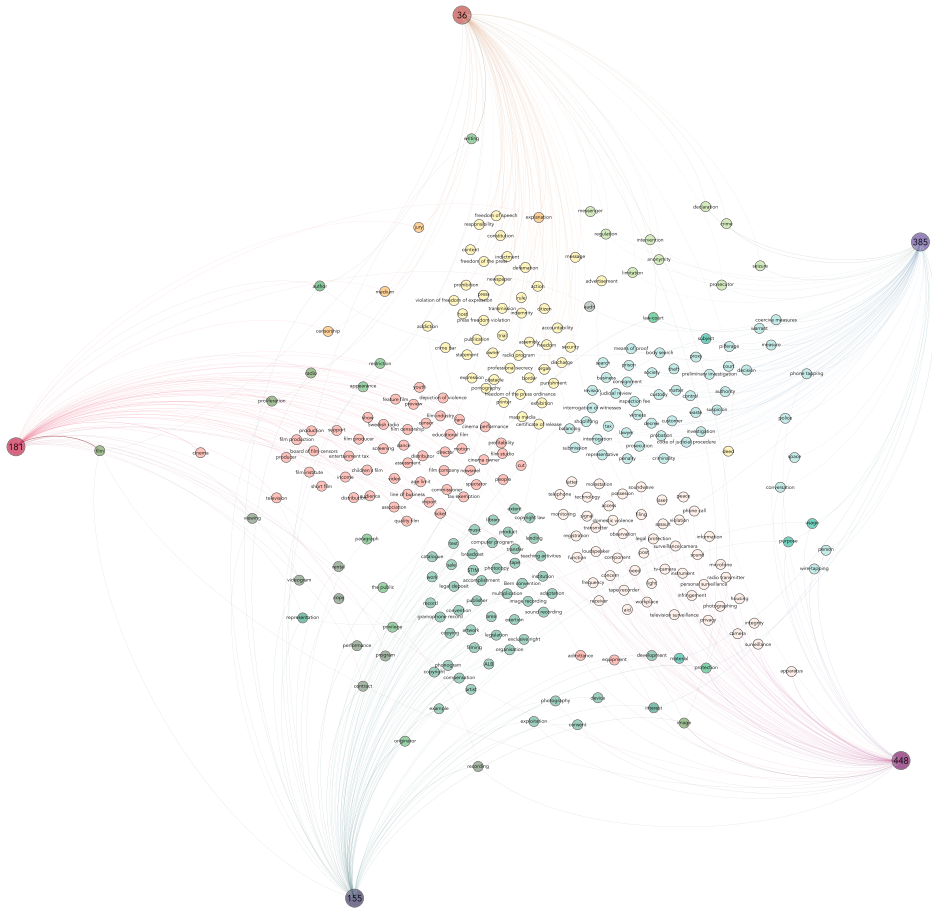
Topic number	Probability	Words in order of prominence
		AUTORSHIP TOPIC
155	0.006873	work of art, originator, copyright protection, copy, recording, contract, artist, copyright, compensation, phonogram, the public, usage, image, judicial, ALB, legal right, privilege, radio, performance, recording, material, format, utilisation, year, reproduction, interest, artwork, videogram, film
		CINEMA TOPIC
181	0.013874	film, cinema, screening, producer, film production, television, production, board of film censors, film institute, short film, support, earning, film industry, videogram, entertainment tax, censorship, film producer, fiction film, distribution, children's film, audience, feature film, show
		ARCHIVAL TOPIC
343	0.047464	archive, film, material, document, national archive, tape, paper, micro film, recording, government authority, archiving, magnetic tape, copy, weeding, preservation, information, image, DAK, cost, regional state archive, usage, conservation, research, record, institution, archival material, future, medium

FIGURE 14

Within the 500-model of the SOU-data details become visible—such as the relation between media and copyright issues or between media and archival matters. The keyword *film* is a frequent term appearing in all three topics.

reports were geared towards legislative issues this is not surprising. Yet, if one plots keywords within the authorship topic (155) and the topic on cinema (181), and compares them with three other topics—on freedom of speech (36), on surveillance (385) and on personal integrity (448)—an unusual context for the medium of film appears. Swedish film history is here linked to discourses on defamation, on regulation, on prosecution, and even police and criminality. Visualized in a Gephi network the five topics include a number of nodes with the most frequent keywords for each topic, but they also share words such as author, medium, regulation, audit, regulation and deed. Examination and censorship of film, furthermore—a contested issue in Sweden ever since a Board of Film Censors was established in 1911—is within the network connected to both copyright issues, pornography, wire tapping and prison, a somewhat sinister milieu for a medium of entertainment. Modeling the data and visualizing relations between keywords as nodes in a network thus situates the medium of film in a different socio-cultural context, accentuating legal matters around both copyright and freedom of speech, as well as the (claimed) inherent psychological dangers of moving images (Figure 15).

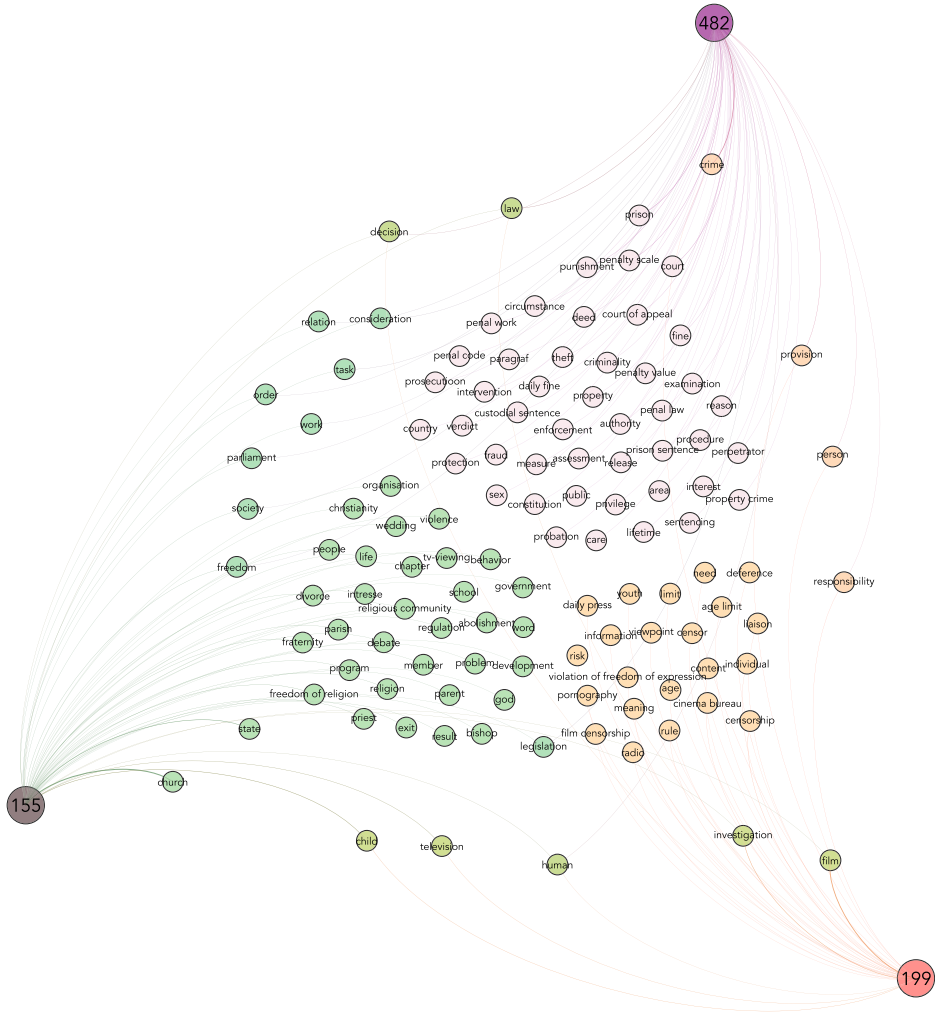
If one, finally, models the SOU-data in 500 topics with a Gensim modeling toolkit (rather than Mallet) yet another surprising cinematic context appears—this time with moral and religious connotations. The difference between Mallet and Gensim is, in short, that the latter uses a so-called Variational Bayes sampling method, which is considered faster (but less precise) than Mallet's Gibbs Sampling. The point to be made, however, is that media historical data can be *modeled* in different ways, producing slightly



**FIGURE 15**

Network in Gephi (Force Atlas) from the 500-model of the SOU-data, with some 300 keywords from five different topics—36 (on freedom of speech), 155 (on authorship), 181 (on cinema), 385 (on surveillance) and 448 (on personal integrity).

different results and interpretations. The Gensim toolkit, for example, detects a specific topic on film censorship (number 199) which Mallet does not, with links to topics on Christianity, law and order. Moreover, the censorship topic contains the keyword *filmforskningsgruppen*, a film research group set-up by the Swedish Film Institute in 1962 to investigate issues of censorship and cinematic influences. The final report from the group, SOU 1967:31 *Filmens inflytande på sin publik* [Cinema's influence on its audience] was unequivocal: it was not possible to demonstrate any psychological effects whatsoever of film viewing. Within the SOU-data, however, there are a number of linkages between this censorship topic 199 and a topic on religion and church (number 155), as well as a topic on crime and punishment (number 452)—producing yet another eerie context for the medium of film. The three topics share keywords such as responsibility, legislation and law. The topic on religion and church appears most frequent in the SOU-data during the 1960s, foremost related to SOU 1964:13 on freedom of religion, and SOU 1964:16



**FIGURE 16**  
 Network in Gephi (Force Atlas) from the 500-model of the SOU-data using a Gensim modeling toolkit, with some 150 keywords from three different topics—155 (on religion och church), 199 (on film censorship), and 482 (on crime and punishment). Only a few keywords are shared among all topics, among them *law*.

on the meaning of the notion of the church (both reports were delivered by a 1958 commission investigating the relation between state and church). The topic on religion and church, with keywords such as Christianity, priest, bishop—and even God—also shares the keyword child with the film censorship topic, suggesting that moral issues related to childhood and youth were part of national media policy at the time. During the heyday of the Social-democratic welfare state Christian values were in decline (some argued), hence the Christian-democratic party in Sweden was founded in 1964—at least to some extent (some argue) as a conservative reaction towards Vilgot Sjöman’s feature film 491 (which dealt with homosexuality and juvenile delinquency) (Figure 16).

## Conclusion

More than a hundred Swedish Government Official Reports between 1945 and 1989 dealt with media issues. If one defines the notion of medium in a broad sense there are even more media reports. It is hence not surprising that media topics occur within the SOU-data. Since some SOU-reports addressed particular issues, such as film censorship, it is furthermore hardly unexpected that modeling the data will reveal and discern keywords associated with such themes. Topic modeling is indeed a digital method that can uncover invisible categories and discourses in a massive dataset. Yet if this dataset is compiled and based on specific publications—such as particular SOU-reports—they will become visible and surface when modeling the data (in one way or another). The concern and critique of digital methods such as topic modeling has consequently often been related to a critical assessment of the data at hand. The curation of a dataset needs to be transparent but a topic model will always be based on a particular dataset, hence neglecting other potentially relevant historical sources. Then again, as an openly available parliamentary dataset, SOU-reports are handy to explore, model and examine for a media historian. But more heterogeneous parliamentary discussions as the Hansard, the German *plenarprotokolle* or Swedish parliamentary debates are actually a more appropriate empirical source for detecting unexpected findings. In such datasets discussions on media and media politics would naturally also be prevalent (somewhere), but they would not feature in a specific setting or form—apart from when media politics is explicitly debated. Consequently, within our research project Welfare State Analytics we are currently preparing and curating the debates from Swedish parliament for large-scale analyses.

Then again, in an explorative manner this article has modeled Swedish media history in different ways. By distant reading a dataset of 3100 SOU-reports some findings have been made which previous media historical scholarship have neglected—or rather, have been unable to detect due to traditional, small-scale examinations of only a few SOU-reports. Film was, for example, the medium which the Swedish government dealt most explicitly and extensively with. New media such as television—and later telecommunications—were hence not predominant within national media policy, but rather the old medium of moving images. Another finding that previous research has disregarded was that media politics in Sweden was geared towards both media production and preservation. All models with 50, 100, 200 or 500 topics generate results which confirm that media politics repeatedly had an archival bias.

Under the headline ‘a one medium history’ in the research anthology, *A History of Swedish Broadcasting: Communicative Ethos, Genres and Institutional Change*, editors Monika Djerf-Pierre and Mats Ekström state that the

arguments for studying broadcasting separately are not only based on the centrality of broadcasting in 20th century social, cultural and political life, but also on the fact that radio and television share unique communicative properties and ways of organizing public communication.<sup>22</sup>

Such claims make little sense when modeling the SOU-data. On the contrary, the most important finding of this article is that when modeling media history interrelations

between media are everywhere. They can also be statistically measured, even if topic modeling is based on probability (that can differ when rerunning a model). While it is true that individual SOU-reports often dealt with one medium at a time, the SOU-data repeatedly displays interconnections between different media topics, as well as other societal topics that media were related to. Modeling media history through computational interrogations can thus both confirm previous historical assertions, but more importantly reveal a number of divergent and previously neglected histories.

## Funding

This work was supported by the Swedish Research Council [grant number 2018-06063].

## Disclosure Statement

No potential conflict of interest was reported by the author.

## Notes

1. For a media historical critique of mediatisation theory, see Snickars, "Media and Mediatization."
2. Bunout, Ehrmann and Clavert, *Digitised Newspapers*; Salmi et al., "The Reuse of Texts in Finnish Newspapers and Journals, 1771–1920."
3. Williams, "Networking Moving Image History"; Fickers, Snickars, and Williams, "Audiovisual Data in Digital Humanities"; Lingold, Mueller, and Trettien, *Digital Sound Studies*; Jofre et al., "What's in a Face?"
4. Abercrombie and Batista-Navarro, "Sentiment and Position-Taking Analysis of Parliamentary Debates." Hansard at Huddersfield can be found at <https://hansard.hud.ac.uk/site/site.php>. Swedish parliamentary motions 1971–2021 at <https://riksdagsmotioner.dh.gu.se/>, and German plenary minutes from the Bundestag 1949–2021 at <https://opendiscourse.de/>.
5. For more information about the research project Welfare State Analytics, see <https://www.westac.se/en/>. Datasets and scripts used by the project are available on Github, <https://github.com/welfare-state-analytics>.
6. Premfors, "Governmental Commissions in Sweden," 624.
7. Åmark, *Hundra år av välfärdspolitik*; Östberg and Andersson, *Sveriges historia 1965–2012*; Norén and Snickars, "Distant Reading the History of Swedish Film Politics."
8. Grimmer and Stewart, "Text as Data."
9. For a discussion, see Ahnert et al., *The Network Turn*.
10. Guldi, "Parliament's Debates About Infrastructure."
11. Jockers, *Macroanalyses*.
12. Hakkarainen and Iftikhar, "The Many Themes of Humanism," 261.
13. See for example, Isoaho, Gritsenko, and Mäkelä, "Topic Modeling and Text Analysis for Qualitative Policy Research," "Patterns and Interpretation"; Moretti, *Distant Reading*, 46.
14. Moretti, "Patterns and Interpretation"; Moretti, *Distant Reading*, 46.
15. Snickars, *Kulturarvets mediehistoria*.



16. See for example, Hadenius, *Kampen om monopolet* and Weibull, "New Media Between Technology and Content."
17. SOU 1989:73, 77.
18. SOU 1980:8, 9.
19. SOU 1984:10, 3.
20. Hyvönen, Snickars, and Vesterlund, "The Formation of Swedish Media Studies."
21. Roger Blomgren, *Staten och filmen* and Leif Furhammar, *Filmen i Sverige*.
22. Djerf-Pierre and Ekström, "Approaching Broadcast History," 17.

## Bibliography

- SOU 1980:8 *Privatlivets fred*.
- SOU 1984:10 *Kompletterande motståndsformer*.
- SOU 1989:73 *TV-politiken*.
- Abercrombie, Gavin, and Riza Batista-Navarro. "Sentiment and Position-Taking Analysis of Parliamentary Debates: A Systematic Literature Review." *Journal of Computational Social Science* 3 (2020): 245–270.
- Ahnert, Ruth, Sebastian E. Ahnert, Catherine Nicole Coleman, and Scott B. Weingart. *The Network Turn. Changing Perspectives in the Humanities*. Cambridge: Cambridge University Press, 2020.
- Åmark, Klas. *Hundra år av välfärdspolitik*. Umeå: Boréa, 2005.
- Blomgren, Roger. *Staten och filmen: svensk filmpolitik 1909–1993*. Stockholm: Gidlund, 1998.
- Bunout, Estelle, Maud Ehrmann, and Frédéric Clavert, eds. *Digitised Newspapers – A New Eldorado for Historians? Tools, Methodology, Epistemology, and the Changing Practices of Writing History in the Context of Historical Newspapers Mass Digitization*. Berlin: de Gruyter, 2022.
- Djerf-Pierre, Monika, and Mats Ekström. "Approaching Broadcast History. An Introduction." In *A History of Swedish Broadcasting: Communicative Ethos, Genres and Institutional Change*, edited by Monika Djerf-Pierre and Mats Ekström, 11–27. Göteborg: Nordicom, 2013.
- Fickers, Andreas, Snickars Pelle, and Mark Williams. "Editorial Special Issue: Audiovisual Data in Digital Humanities." *VIEW. Journal of European Television History and Culture* 14 (2018): 1–4.
- Furhammar, Leif. *Filmen i Sverige*. Höganäs: Wiken, 1991.
- Grimmer, Justin, and Brandon M. Stewart. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21 (2013): 267–297.
- Guldi, Jo. "Parliament's Debates About Infrastructure: An Exercise in Using Dynamic Topic Models to Synthesize Historical Change." *Technology and Culture* 1 (2019): 1–33.
- Hadenius, Stig. *Kampen om monopolet. Sveriges Radio och TV under 1900-talet*. Stockholm: Prisma, 1998.
- Hakkarainen, Heidi, and Zuhair Iftikhar. "The Many Themes of Humanism. Topic Modelling Humanism Discourse in Early 19th-Century German-Language Press." In *Digital Histories. Emergent Approaches Within the New Digital History*, edited by Mats Fridlund, Mila Oiva, and Petri Paju, 259–277. Helsinki: Helsinki University Press.
- Hyvönen, Mats, Pelle Snickars, and Per Vesterlund. "The Formation of Swedish Media Studies." *Media History* 1 (2018): 86–98.

- Isoaho, Karoliina, Dario Gritsenko, and Eetu Mäkelä. "Topic Modeling and Text Analysis for Qualitative Policy Research." *Policy Studies Journal History* (19 June 2019). doi://doi.org/10.1111/psj.12343.
- Jockers, Matthew. *Macroanalyses. Digital Methods & Literary History*. Urbana: University of Illinois Press, 2013.
- Jofre, Ana, Josh Cole, Vincent Berardi, Carl Bennet, and Michael Reale. "What's in a Face? Gender Representation of Faces in Time, 1940s–1990s." *Journal of Cultural Analytics* (19 March 2020). doi:10.22148/001c.12266.
- Lingold, Mary Caton, Darren Mueller, and Whitney Trettien, eds. *Digital Sound Studies*. Durham, NC: Duke University Press, 2018.
- Moretti, Franco. "Patterns and Interpretation." *Pamphlets of the Stanford Literary Lab*, no. 15 (2017). Accessed October 1, 2021. <https://litlab.stanford.edu/LiteraryLabPamphlet15.pdf>.
- Moretti, Franco. *Distant Reading*. London: Verso, 2013.
- Norén, Fredrik, and Pelle Snickars. "Distant Reading the History of Swedish Film Politics—In 4,500 Governmental SOU Reports." *Journal of Scandinavian Cinema* 2 (2017): 155–175.
- Östberg, Kjell, and Jenny Andersson. *Sveriges historia 1965–2012*. Stockholm: Norstedt, 2013.
- Premfors, Rune. "Governmental Commissions in Sweden." *American Behavioral Scientist* (1 May 1983). doi:10.1177/000276483026005008.
- Salmi, Hannu, Petri Paju, Heli Rantala, Asko Nivala, Alekski Vesanto, and Filip Ginter. "The Reuse of Texts in Finnish Newspapers and Journals, 1771–1920: A Digital Humanities Perspective." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 1 (2021): 14–28.
- Snickars, Pelle. *Kulturarvets mediehistoria: dokumentation och representation 1750–1950*. Lund: Mediehistoriskt arkiv, 2020.
- Snickars, Pelle. "Media and Mediatization." In *Routledge Companion to Cultural History in the Western World*, edited by Alessandro Arcangeli, Jörg Rogge, and Hannu Salmi, 488–506. London: Routledge, 2020.
- Weibull, Lennart. "New Media Between Technology and Content. The Introduction of Radio and Television in Sweden." In *A History of Swedish Broadcasting: Communicative Ethos, Genres and Institutional Change*, edited by Monika Djerf-Pierre and Mats Ekström, 31–54. Göteborg: Nordicom, 2013.
- Williams, Mark. "Networking Moving Image History: Archives, Scholars and the Media Ecology Project." In *The Arclight Guidebook to Media History and the Digital Humanities*, edited by Charles Acland and Eric Hoyt, 335–346. Falmer: REFRAFRAME Books, 2016.

**Pelle Snickars**, Department of Arts & Cultural Sciences, Division of ALM & Digital Cultures, Lund University, Lund, Sweden. E-mail: pelle.snickars@kultur.lu.se