Journal of Scandinavian Cinema
Volume 7 Number 2
© 2017 Intellect Ltd Article. English language. doi: 10.1386/jsca.7.2.155_1

feature article

FREDRIK NORÉN AND PELLE SNICKARS
Umeå University

Distant reading the history of Swedish film politics in 4500 governmental SOU reports

ABSTRACT

Using computational methods, digitized collections and archives can today be scrutinized in their entirety. By distant reading and topic modeling one particular collection – 4500 digitized Swedish Governmental Official Reports (SOU) from 1922 to 1991 – this article gives a new archival perspective of the history of Swedish film politics and policy-making. We examine different probabilistic topics related to film (and media) that the algorithm within the topic modeling software Mallet extracted from the immense text corpora of all these Official Reports. Topic modeling is a computational method to study themes in texts by accentuating words that tend to co-occur and together create different topics. Basically, it is a research tool for the discovery of hidden semantic structures, exploring a collection through the underlying topics that run through it. Hence, our article captures a number of film discourses and trends within the SOU material. In conclusion, we argue that topic modeling should be recognized as a method and research aid for gathering an overview of a major material; as a way to pose new and unforeseen research questions; and as a kind of computational support that makes it possible to apprehend major patterns more or less impossible to detect through a traditional archival investigation.

KEYWORDS

digital methods
digitized collections and archives
history of Swedish film politics
distant reading
topic modeling
Swedish Governmental Official Reports (SOU)
computational film studies

Archives are never neutral, but it is through them that bits and pieces from the past are studied. At best, scholars glean representative segments and traces. At worst, only specific samples are used, stressing the particular: 'the anecdotal nature of evidence', as digital humanist Andrew Piper has polemically characterized traditional ways of doing humanistic research (Piper 2016, para. 12). The more or less 'anecdotal nature' of humanistic inquiries and scholarly work has been due, however, to the analogue nature of archives and collections. Since all archives are (and have always been) epistemic grounds from which history is written, the structure and organization of them have determined what type of research could be executed. Then digitization happened and scholarly work could be drastically reconfigured.

The rate at which historical films and related film-historical material is digitized has increased dramatically during the past two decades. Different forms and formats of humanistic infrastructures have gradually developed from which film and media historians have benefitted (Acland and Hoyt 2016; Rieder and Röhle 2012). In Sweden, the National Library, together with the Swedish Film Institute, is increasing step by step the number of films and media-historical documents available in digitized format at filmarkivet.se and at the sister site, filmarkivforskning.se (for which co-author Snickars has been responsible). The ways that Swedish film history and cinema culture are researched and understood are bound to change as a result. If the so-called New Film History in the late 1980s and 1990s, and to some extent Media Archeology a decade later, turned out to be a reconceptualization of the history of moving images, at present scholars inclined to use digital technology and the humanistic infrastructures built around it are witnessing a similar reorganization of the practices of doing film and media-historical research.

Following film historian David Bordwell, the role models of New Film History scholars were young historians 'hunched over microfilm machines cranking through day after day of *Moving Picture World* or sitting in archives paging through studio memos' (quoted in Klenotic 1994: 46). Today, the ideal counterpart, at least as such a person is envisioned within the digital humanities, is a scholar involved in cross-disciplinary projects using digital methods and tools like data mining, Gephi visualizations, topic modeling or GIS analyses. During the last decade, digital technology has played an increasingly important role in numerous film-historical projects and platforms, ranging from early cinema programmes collected within the Siegen Cinema Databases (Garncarz and Ross 2006) to statistical data about film editing at Cinemetrics (Tsivian 2009) and the radiant Timeline of Historical Film Colors (Flueckiger 2011) to New Cinema History, with its computational focus on the circulation and consumption of films (Verhoeven 2012) – not to mention the highly elaborate platform, Lantern, used for searching, exploring and visualizing the vast collections of the Media History Digital Library (Hoyt et al. 2013). What these projects have in common are the ways in which computational analysis and quantitative research offered by databases, visualizations and data mining have allowed film-historical researchers to gather new information about film style and aesthetics as well as the history of cinema exhibition and reception. Earlier, such information would have been impossible (or too labour-intensive) to gather.

What precisely does computation allow one 'to claim that has not been seen before or that was uncertain in the world of anecdotalism?' (Piper 2016,

para. 27). Our article aims to give a new archival perspective on the history of Swedish film politics and policy-making by distant reading and topic modeling 4500 digitized Swedish Governmental Official Reports (SOU) from 1922 to 1991. We examine different probabilistic topics related to film and media that the algorithm, within our topic modeling software Mallet, extracted from the immense text corpora of these reports. Initially, we describe the work performed by governmental commissions and the SOU genre and its relation to film politics, policy-making and film scholarship. Foremost, however, we recount and analyse novel ways of how to understand and situate the history of Swedish film politics and policy through topic modeling a massive SOU corpora. The article captures a number of film discourses and trends in the joint corpus that is virtually impossible to detect through a traditional archival investigation. Topic modeling is a methodological approach to study themes in texts by accentuating words that tend to co-occur and together create topics in the form of clusters of similar words. Within topic modeling, a term or a word may be a part of several topics with different degrees of probability. The article therefore also constructs and traces a broader context of Swedish film politics over time, especially in relation to other media formats and institutional actors.

Governmental commissions and the SOU genre

Before submitting a proposal for new legislation, the Swedish government regularly examines alternatives, a task prepared by an appointed Committee or Commission of Inquiry. The governmental committee process is a way of accessing knowledge about particular issues. These might range from major policy decisions affecting Swedish society as a whole to small and technically complex issues. During the last 80 years, governmental commissions have developed into an effective instrument of majority parliamentarism. Decision-making via governmental commissions, however, has a long history in Sweden. The national legislative process providing the government, and in an earlier era, the King, with a proper basis for decision-making dates back to the seventeenth century.

Since 1922, the work of governmental commissions has usually been published in a series known as Statens offentliga utredningar (Swedish Government Official Reports) or SOU. All reports are published with a distinct number: the first SOU devoted to film, for example, is SOU 1930 (26). SOU reports and work performed within governmental committees have had the task of preparing the state for apt and rational decision-making. According to political scientist Rune Premfors, after 1945 'the range of subjects covered by governmental committees has expanded to include virtually every area of the Swedish welfare state' (Premfors 1983: 624). Stressing the importance of the work executed by governmental committees, Premfors states that some 40 per cent of all legislation in Sweden around 1970 was based on commission proposals (Premfors 1983). During the 1960s, certain policy issues also started to be investigated internally within governmental ministries, resulting in reports in the so-called Ds-series (Departementsserien). The SOU series, however, with its external investigations performed by a commission, usually with four to five members and often running for a number of years, is arguably a more important historical source for Swedish governmental policy-making, not least since governmental commissions were modelled

on ideas about scientific inquiries and often engaged in ambitious efforts to rewrite history.

After a governmental commission submitted an SOU report to the responsible minister, it was (and still is) also dispatched for consideration to relevant authorities, advocacy groups and the public: 'They are given an opportunity to express their views on the conclusions of an inquiry before the Government formulates a legislative proposal' (Government Offices of Sweden 2016). Governmental commissions have occasionally been accused of bias. When the government appoints a commission, it also provides a set of guidelines, instructions that might have a political tendency. Traditionally, these guidelines have specified what issue a commission should examine and what problems should be solved. As a consequence, SOU reports usually contained legislative suggestions from the committee. Policy-making thus turned into an administrative task. Already during the 1950s, governmental commissions and the administrative apparatus surrounding them were critiqued: 'One can speak of a movement from politics to administration, from principles to technique', lamented political scientist and newspaper editor Herbert Tingsten. 'As the general standard of values is so commonly accepted, the function of the state becomes so technical as to make politics appear as a kind of applied statistics' (Tingsten 1955: 147). Then again, as Jan Johansson has argued in *Det statliga kommittéväsendet* ('The governmental committee system'), the two most essential features of the system have been the focus on expert knowledge and the urge to reach some form of compromise or consensus (Johansson 1992). Governmental commissions have, in short, been an arena for the exchange of factual arguments among experts (including academics), situated within a rationally oriented Swedish style of policy-making. Ultimately they have been geared towards reaching an agreement (Veit 2009).

Regarding the medium of film, one of the first governmental commissions devoted their work to drafting a new subsidy system for the production of Swedish feature films, *Statligt stöd åt svensk filmproduktion* (SOU 1942: 36). Heading that commission was legal adviser Carl Romberg from the Swedish Department of Justice; committee members also included the head of Radio Sweden, Carl Anders Dymling, and well-known film directors Victor Sjöström and Arne Bornebusch. The mix of persons, some with hands-on experience in film, can be seen as representative of how governmental commissions on film were usually staffed. One of the most important governmental commissions on film, the 1968 Film Commission, whose work resulted in the multi-volume SOU *Samhället och filmen* ('Society and cinema') during the early 1970s, included the head of the Swedish Film Institute, Harry Schein, as well as directors Jan Troell and Kjell Grede. In addition, the director turned film historian Gösta Werner from Stockholm University contributed with a lengthy history in nineteen chapters of 'Swedish film during 75 years'. The work performed by the 1968 Film Commission in many ways testifies to the textual significance of the SOU genre. Reports were often of book length, and some were even published in several volumes. Commission work usually went on for years. In the case of the SOU 'Society and cinema', five years of investigations were wrapped up in four volumes, together spanning some 750 pages.

The SOU genre bears textual witness to and gives evidence of contemporary societal conceptions, not least since committee members often disagreed. According to law, governmental commission work had to be archived, and preserved papers from the 1968 Film Commission – in 29

(non-digitized) volumes at the Swedish National Archives (Riksarkivet) – attest, for example, to Schein's displeasure with Werner's film-historical survey: he wanted less film aesthetics and more focus on cinema audiences and film production (Schein 1973). Since governmental commissions based their work on expert knowledge, they can shed light on the role of academics and the involvement of research. Sometimes separate research anthologies were published in conjunction with SOU reports. Snickars has argued that media studies in Sweden, in fact, arose at the intersection between the media industry, the needs of media policy filtered through SOU reports and academia's new-found interest in media at both social science and humanities faculties (Hyvönen et al. 2017). This is yet another reason why the SOU genre is a relevant media-historical (and meta-academic) source. In particular, the 57 media-related SOU reports between 1960 and 1980 were an important engine for the institutionalization of media research in Sweden.

With regard to film, this is especially true during the 1960s and 1970s. Gösta Werner has been mentioned. In addition, in the mid 1960s the commission on film censorship worked in close collaboration with the Film Research Group at the Swedish Film Institute, initiated by Harry Schein in 1963. The result was the publication *Filmens inflytande på sin publik* ('The influence of film on its audience') (SOU 1967: 31), where Leif Furhammar, who would later become professor of film studies at Stockholm University, played a significant role. The first professor of film studies in Stockholm, Rune Waldekranz, also took an active part in various media-related governmental commissions during the 1970s, especially in relation to the establishment of a national film archive (Snickars 2015). After 1980, however, it became less common to include academics in governmental commission work on film. Later film professors such as Jan Olsson, Erik Hedling and Maaret Koskinen have not formally been involved in any governmental commissions on film (apart from being consulted within the general SOU remittance rounds). Media historian Mats Björkin, then at Stockholm University, was the last academic included in a governmental film commission (on the preservation of documentary film heritage; SOU 1999: 41) if one excludes film scholar turned cultural bureaucrat Jon Dunås, who served as the main secretary in the latest film commission that produced the report *Vägval för filmen* ('A crossroads for film') (SOU 2009: 73).

SOU reports on film, to date 23 in number, have frequently been cited in film-historical research in Sweden. The political scientist Roger Blomgren referred to numerous SOU publications in his dissertation on film and the Swedish state, *Staten och filmen: svensk filmpolitik 1909–1993* (1998), and in Furhammar's classic study, *Filmen i Sverige* ('Film in Sweden') (1991), some fifteen SOUs on film are discussed and referenced. Our use of the SOU genre differs from previous film-historical research since we include all publications. There is, after all, a methodological difference between a close reading of a dozen or so publications and a distant reading of more than 4000. We acknowledge, however, that our article does not alter the perception of the history of Swedish film politics in a radical way. Our analyses largely confirm previous findings, though via different computation-driven methodological approaches, but they also suggest areas and intersections of importance and relevance for further research. Or as Andrew Piper has put it: 'computation forces us to rethink current disciplinary practices from the ground up. What counts as evidence?' (Piper 2016, para. 3).

SOU reports – from a distance

During the last decade, the literary historian Franco Moretti has developed into the principal proponent for what he terms 'distant reading' (Moretti 2000: 56–57). In an age of huge numbers of digitized books, understanding literature can and should, according to Moretti, move from studying only particular texts within a canon to the aggregation and analyses of massive quantities of textual data. Since Moretti coined the term, distant reading has become a popular way of broadly describing the analyses of major textual corpora, usually in the form of loosely tied algorithmic text-mining approaches. Moretti has asserted that 'If you want to look beyond the canon […] close reading will not do it', and continues:

> It's not designed to do it, it's designed to do the opposite. At bottom, it's a theological exercise – very solemn treatment of very few texts taken very seriously – whereas what we really need is a little pact with the devil: we know how to read texts, now let's learn how not to read them. Distant reading: where distance, let me repeat it, is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes – or genres and systems.
>
> (2013: 46)

Distant reading is hence envisioned as a 'condition of knowledge'. Moretti has also stressed that 'to understand the system in its entirety', scholars must be prepared to 'accept losing something' (2013: 46). As a literary scholar, Moretti has foremost done work on fiction, but in later years he has started to analyse other forms of textual accounts in the Literary Lab he established (and ran for a number of years) at Stanford University. In an article co-authored with science historian Dominique Pestre, he has, for example, used the mode of distant reading to analyse economic vocabulary in the World Bank's yearly reports (Moretti and Pestre 2015). In some ways, these reports are similar to the Swedish SOU genre, but they also differ. There is, for instance, no equivalent to a uniform World Bank 'management discourse' in the SOU material as a whole, since the latter genre is much more heterogeneous. Moretti and Pestre are primarily interested in grammatical patterns and the semantic transformations of 'Bankspeak', and they do not use topic modeling. Nevertheless, our article is inspired by their study, particularly with regard to their analyses of textual data in its entirety and over a long time period.

Concerning material and method, our study departs from the fact that all SOU reports have been digitized by the National Library of Sweden during the last five years, in all, 6129 SOU publications between 1922 and 1999. After 1999, all published SOU reports can be found in original PDF versions (and other formats) at the Swedish Government website (https://data.riksdagen.se). Our study, however, does not utilize all digitized reports. We have analysed 4522 documents published between 1922 and 1991. The reason is simple: at the time of our analyses, Språkbanken (The Swedish Language Bank) at Gothenburg University had XML translated all SOUs up to the early 1990s. Språkbanken's presentation of corpora and linguistic data is mostly in the form of concordances (accessed online), but to do proper topic modeling, text material needs to be XML translated. XML (Extensible Markup Language) defines a set of rules for encoding textual documents in a format that is machine readable, which makes it possible to structure text in ways suitable for the analysis, for example extracting all nouns (independent of lemmatizing).

From a film-historical perspective, the selection of material for our study is somewhat arbitrary, not least regarding periodization and the year our analysed material ends (1991). Still, in some ways the time span coincides neatly with a traditional media history of 'classic' mass media – that is radio, television and film – prior to the introduction of the Internet, which during the last two decades has influenced discussions of film politics. In addition, and encouraged by work at Stanford Literary Lab, our article emphasizes the need for an exploratory research approach, where the methodological process rather than results themselves are foregrounded. Still, we also reveal a series of findings, mostly presented in the form of Gephi illustrations from the SOU corpora. Yet at present, methods and tools within the digital humanities lack standardization (Clement 2016), so 'trial and error' often tend to serve as guiding principles within the computationally driven research process.

There are different ways to algorithmically study themes and contexts in a massive text collection. With well-known tools for studying word frequency trends such as Google Ngram Viewer, a researcher might discern connections between changes in language and changes in society. In the Ngram Viewer, the term 'gender', for instance, is almost never mentioned before 1960, but thereafter increases rapidly, especially after 1980 as a result (one might argue) of the feminist movement. A word or term may nevertheless have multiple meanings in various contexts over time (Mimno 2012). Naturally, a word is not a discourse; they have different ontologies. Instead, topic modeling is a methodological approach to study themes in texts by looking at words that tend to occur in clusters and together create a topic. Topics are more or less translatable into a common discourse or a theme in a large textual corpus. Topic modeling, in turn, is a computer-generated and automated method for organizing, managing and delivering results, where the latter depend on the algorithm being used. Themes are thus automatically discovered – inductively – from analysing original texts in their entirety.

Topic modeling, and especially unsupervised topic modeling, has won increasing popularity in the digital humanities because of low technical thresholds. It nevertheless needs to be envisioned within a strict mathematical framework of statistics. In our study, we have used the popular topic modelling tool Mallet, based on the so-called latent Dirichlet allocation algorithm (LDA) that clusters words into topics and also bundles underlying documents into different mixtures of topics (Blei et al. 2003). Without going into technical details, LDA uses a Bayesian inference model that associates each document with a probability distribution of topics, and where topics are the probability distribution of words. The model, in short, provides results in which each document is represented by a random mixture of latent topics, and each topic is characterized by a distribution of words generated according to a probability for respective topics. A word may thus be part of several topics with different degrees of probability (Blei et al. 2012).

SOU analytics

Using contextual clues, in our case, prior knowledge about Swedish film history, topic models can connect words with similar meanings and also distinguish between uses of words with multiple meanings. Collaborating with programmer Roger Mähler at Humlab, the digital humanities hub at Umeå University, scripts in the programming language Python were developed to prepare the SOU material. By utilizing the tagged XML versions of

the reports, we extracted the word lemmatize of all nouns as a way to study themes in all texts (e.g. the two words 'film' and 'films' count as two hits of the lemmatized noun 'film'). To generate better probabilistic topics, the reports were chopped into smaller chunks of 1000 nouns each (Jockers 2013). Mallet was then instructed to return 500 latent topics for all SOU report 'chunks' from 1922 to 1991. In our results, one topic, arbitrarily numbered by the software as '158', was identified as a distinct film topic. The number '0.00383' was the probabilistic weight of that topic in the total SOU corpora. The twenty most likely words within this distinct topic are presented below (the English translations are the authors'):

> 158 0,00383 film (film), biograf (cinema), visning (screening), producent (producer), produktion (production), nöjesskatt (entertainment tax), filmproduktion (film production), kortfilm (short film), videogram (video), biografbyrå (censor agency), år (year), stöd (subsidy), barnfilm (children's film), föreställning (performance/screening), censur (censorship), filminstitut (film institute), framställning (representation), spelfilm (feature film), filmbransch (film industry), publik (audience).

The film topic revealed a preassumed idea that one general theme of Swedish film politics had to do with economically connotated keywords. However, by utilizing the 'word count' output data from Mallet, it was possible to reveal the textual nature of the SOU film discourse in its entirety. A Python script expanded the topic into a spreadsheet with its 500 most likely nouns. The data were imported into Gephi, a widely used open-source network visualization tool. In Figure 1, nouns closer to the centre are more likely to occur in the topic.

In essence, the illustration displays a governmental (and to some extent national) film discourse, with words and terms graphically mapped out in proportions of probability. As is evident, among words most likely to appear are 'film', '*biograf*' ('cinema') and '*visning*' ('screening'). Moving outwards within the concentric circles of the word cloud, in the middle (with a slightly lower probability to occur in the topic), terms like '*filmvetenskap*' ('film studies'), '*visningskopia*' ('screening copy') and '*barncensur*' ('child censorship') appear. Probability and leverage of ascendancy of terms are hence the most interesting features in the illustration.

Studying this concentric frequency list of the most likely nouns in the film topic makes it possible for film scholars to find entries and approaches on how to research Swedish film politics in novel ways. For example, with a simple search on '*barn*' ('child/children'), we find terms such as '*barnfilm*' ('children's film') displayed in the centre of the word cloud, '*barnfilmsklubbar*' ('children's film clubs') in the middle and '*barnfilmsjury*' ('children's film jury') in the periphery (i.e. less likely to occur). The term '*barnfilm*' is also much more likely to feature in the topic than, for example, '*forskning*' ('research') or '*långfilmsproduktion*' ('feature film production'), an indication that the governmental film policy considered the children's film genre to be of importance. At the periphery of the SOU film discourse, it is also surprising that words at the outer end of the topic were still strongly associated with film, such as '*krigsfilm*' ('war movie'), '*underhållningsvåld*' ('entertainment violence'), '*instruktionsfilm*' ('instructional film') or '*filmkritik*' ('film criticism'). The latter was perceived to be of less importance within the national SOU film discourse than, for example, '*pornografi*' ('pornography') – displayed almost in the centre.

Figure 1: The topic 'Film (158)' – arbitrarily numbered by the topic modeling software Mallet – taken from the entire SOU corpora 1922–91, and displayed in a word cloud with the 500 most likely nouns related to film (with terms in the centre most likely to occur). The graph was modulated in Gephi by the Force Atlas algorithm with Repulsion strength set to 200,000.

In the periphery, keywords regarding non-traditional screening venues and film formats were also quite frequent: '*folketshusföreningar*' ('People's Parks and Community Centres'), '*studentfilmstudio*' ('student film studio') and '*smalfilms-distribution*' ('small gauge film distribution'), an indication that neglected parts of Swedish film history were considered significant within the SOU material.

It is also noteworthy that the peculiar Swedish term '*kvalitetsfilm*' ('quality film') was likely to occur in the topic, located just outside the innermost circle. As is well known, '*kvalitetsfilm*' was a benchmark for the so-called Swedish Film Reform of 1963 designed by Harry Schein. It thus comes as no surprise

that '*kvalitetsfilm*', '*kvalitet*' ('quality') and '*kvalitetsbidrag*' ('quality subsidy') are all featured within the film topic with relative high probability. Hence, the SOU film discourse also gives a kind of normative evidence of the ways in which government policy sought to support production and preservation of films of high quality, however difficult it might be to define the concept. In a 1965 BBC interview, Schein sidesteps this issue, stating that 'the entire matter of who is to decide whether a film is good or bad seems to me an irrelevant question because good films are films that people who make professional judgements about film consider to be good' (Snickars 2010: 163).

Another way to graphically present the same 'Film (158)' topic is to look at its weight distribution over a longer time period, as displayed in Figure 2. Around 1970, for example, a number of dots are clustered, largely due to work performed by one of the most important governmental commissions on film, the aforementioned 'Society and cinema'. Not surprisingly, the publications of such specific SOU film reports, seventeen in all during the period under consideration, coincide with the peaks of the film topic. The three blue dots with most weight (high up in the illustration) refer to *Filmstöd och biografnöjesskatt* ('Film subsidies and cinema entertainment tax') (SOU 1959: 2), *Samhället och filmen* ('Society and cinema') (SOU 1973: 16) and *Statens ansvar för visning och spridning av värdefull film* ('State responsibility for the screening and distribution of quality film) (SOU 1988: 37).

By using LDA topic modeling and analysing the entire SOU corpora, rather than examining a few individual reports, as has been the usual practice, it is indeed possible to apprehend Swedish film politics as a whole from a distance as well as over time. Topic modeling should be regarded as a research tool that not only gives a general overview but also helps scholars select SOUs of greatest importance, that is, with the highest probabilistic weight of film-related topics. Topic modeling the SOU corpus over time can also help scholars to accurately distinguish periods where a film-related topic was intensely discussed and debated. The illustration above reveals a cluster of dots from approximately 1945 to 1955, a period during which the Swedish government planned and prepared for the Film Reform. The matter was discussed in various SOU publications for more than two decades, beginning with SOU (1942: 36), before the Film Reform was implemented. This historical perspective differs from the present-day official version, 'History of the Film Institute', presented at the Swedish Film Institute website (SFI 2016).

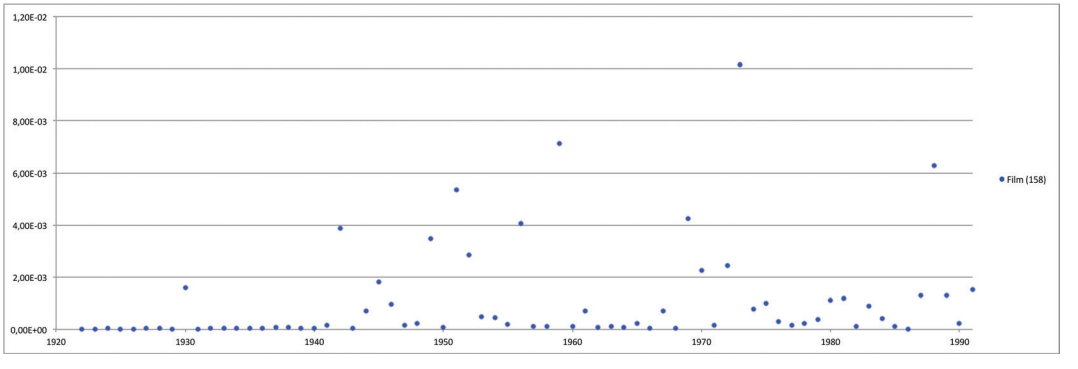

Figure 2: 'Film (158)' topic of the entire SOU corpora with its probabilistic weight distribution between 1922 and 1991.

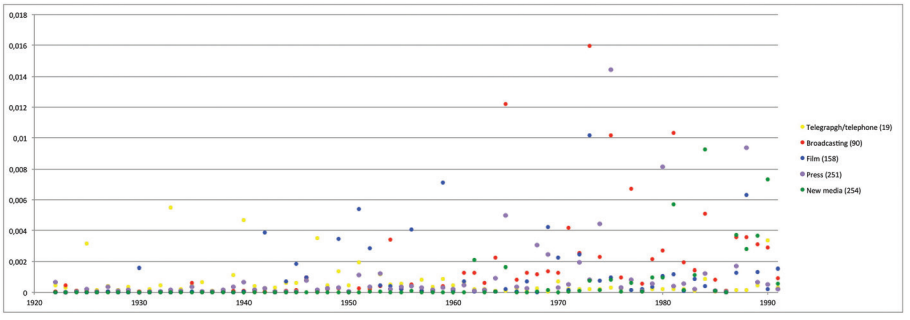

Figure 3: Five different media topics within the SOU corpora (with arbitrary numbers): 'Telegraph/Telephone (19)'; 'Broadcasting (90)'; 'Film (158)'; 'Press (251)'; and 'New media (254)' displayed with probabilistic topic weight through time.

As previously mentioned, the time span of the analysed SOU corpora coincides with a traditional history of classic mass media. Hence, one interesting research question is how the relationship between film and other media was presented in the SOU reports. Using the output dataset from Mallet, we were able to detect other distinct media topics, such as 'Telegraph/Telephone (19)', 'Broadcasting (90)' – due to the state broadcasting monopoly, SOUs in general treated radio and television as interlinked media, and they were also merged by the LDA algorithm – 'Press (251)' and 'New media (254)' (including telecommunications and satellite/cable television). Adding these topics to the illustration below, it is possible to observe how Sweden's governmental interest in mass media increased after World War II, especially since the 1960s (Figure 3).

By utilizing the word-count dataset from Mallet's output, and with the same Python script used previously, each of the five topics' spreadsheets were imported to Gephi. The data were also modulated with the same settings, although we decided to decrease the number of nouns from 500 to 200 to get a more readable and less messy result. The illustration above displays a traditional perspective on twentieth-century media history: prior to about 1940, governmental commissions focused primarily on media forms such as the telegraph and telephone; later, film was scrutinized; after 1960, work was increasingly devoted to broadcasting formats such as television and radio; and since the mid-1970s, different forms of new media have drawn attention.

Figure 3 seems to suggest that film politics gradually became integrated into a broader scope of mass media politics. However, examining specific words and terms within the five topics reveals that the opposite is true. Figure 4 displays the terms and words that were featured within the five distinct media topics in a manner similar to the concentric word cloud around the topic 'Film (158)' above. The terms 'station', '*förbindelse*' ('connection') and 'telegram' were featured within the topic 'Telegraph/Telephone (19)'; '*tittare*' ('viewer/s'), '*riksprogram*' ('national program') and '*samhällsinformation*' ('societal information') within the topic 'Broadcasting (90)'; the terms 'journalist', '*annons*' ('advertisement') and '*upplaga*' ('circulation number') in the 'Press (251)' topic; and terms like '*sattelitsänding*' ('satellite transmission'), '*nätägare*' ('net owner') and 'teledata' in the topic 'New media (254)'.

The illustration vividly displays which of the most likely 200 nouns in each of the five media topics co-occurs with words in other media topics. The purpose of the illustration is to visually present word clusters closest to each media topic (containing unique nouns not shared with any other topic) as well as smaller,

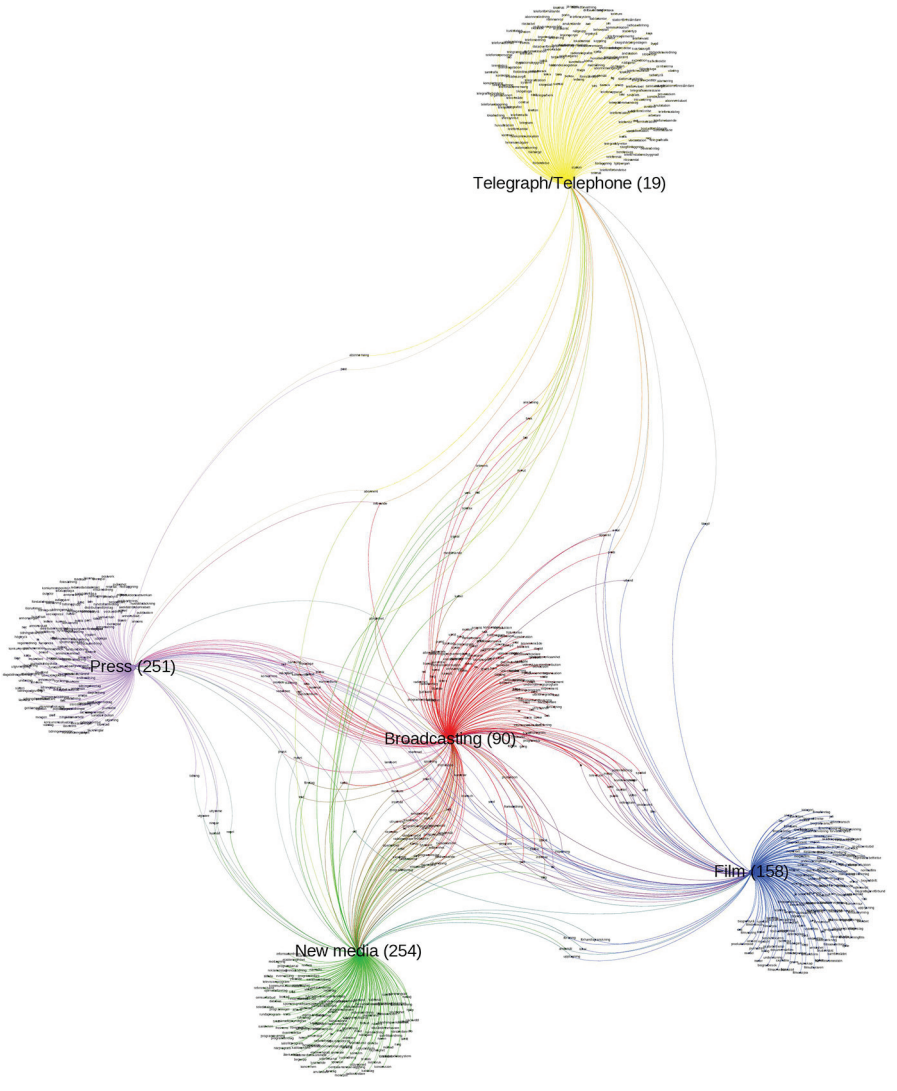

Figure 4: Five distinct, and to some extent connected, media topics within the SOU corpora between 1922 and 1991: 'Telegraph/Telephone (19)'; 'Broadcasting (90)'; 'Film (158)'; 'Press (251)'; and 'New media (254)'. The graph was modulated in Gephi by the Force Atlas algorithm with Repulsion strength set to 200,000 and Closeness-Centrality for nodes.

in-between clusters farther away from each core topic where nouns are shared in two or more media topics. Distinct media topics (the five major clusters) can easily be identified within the SOU corpora. Yet surprisingly, each topic, understood as a media discourse, is detached and relatively separate, as if parallel worlds exist in the SOU textual universe regarding media, as if the state dealt solely with distinct, isolated media forms. The 'Telegraph/Telephone (19)' topic does not, in fact, share a single common noun with any of the other four topics.

Still, if one examines the shared terms among the five media topics, it immediately becomes apparent that the 'Broadcasting (90)' topic has a strong

entanglement with the other media discourses. It is hence no coincidence that it is centred in the middle of the illustration. Among the shared terms between the broadcasting and the film topics are words connected to '*underhållning*' ('entertainment'), 'video', '*publik*' ('audience') and '*våld*' ('violence'). Interestingly, '*kvalitet*' ('quality') was also a shared term, hinting at the ways the Swedish government sought to steer and implement a certain output of content, whether in the form of 'good' films, radio or television programmes. In comparison, terms shared among the broadcasting, new media and film topics were more associated with technical matters, such as '*inspelning*' ('recording') and 'studio'. Broadcasting and new media shared the most terms, evident from the cluster of nouns between these topics and hinting at the similarities between various electronic media forms and formats. In addition, shared terms between the broadcasting topic and the press topic were more geared towards actual content: words like '*redaktion*' ('editorial office'), 'reportage', '*debatt*', and '*händelse*' ('event') stressed similarities in the way these media were discussed in the SOU corpus. Finally, between the 'Telegraph/Telephone (19)' topic in the outermost periphery, terms were (not surprisingly) mostly associated with infrastructure: 'signal', '*apparat*', '*kabel*' and '*abonnemang*' ('subscription').

In general, the illustration makes it apparent that it is indeed possible to map out the discursive densities of media topics within the SOU corpora in its entirety, both regarding shared terms and the degree of media specificity. On the one hand, a scholar can thus zoom in on terms within a specific media topic – like our discussion above of the 'Film (158)' topic – and on the other, zoom out and analyse broader relations between the displayed media topics. In spite of that, congruent and colourful topic model illustrations can also be deceiving, especially if one does not fully understand the algorithms or the software settings behind them. In our case, the Mallet LDA algorithm found five different and very distinct media topics that at least on the surface looked self-centred. Some terms were shared, but the five distinguishable media topics should foremost be perceived as a proof of the strength of one generative statistical model (LDA). Regarding the medium of film, in a collection of 4522 SOU reports, the LDA algorithm found a solid latent topic that we could immediately associate with a film discourse. Interestingly, however, the algorithm automatically also separated the 'Film (158)' topic from other media-related topics.

Hence, in more than one way the Gephi graph above is the perfect illustration of the discursive divide within Swedish media politics during the post-war establishment of the welfare state. As Snickars and others have argued elsewhere, the problem of separation was primarily organizational, since the different governmental commissions dealing with media naturally mirrored the prevailing ministerial structure within the Swedish state apparatus. It was thus the system itself that caused problems; a kind of silo thinking affected the very nature of media inquiries commissioned, dictated as they were by government offices (Hyvönen et al. 2015). Media were, in short, always investigated separately, a point stressed – and criticized – in a 1971 memo from Harry Schein, then serving as the Ministry of Education's media expert, to Ingvar Carlsson, at the time Minister of Education:

> There is a film commission, a literature commission, a mass media commission, a press subsidies commission, etcetera. And I am also currently dealing with cable television. Is it right to isolate the issues from each other in this way? Should there not be a coherent information and communications policy?
>
> (Schein 1971)

Regarding SOU commission work and the reliance on expert and academic knowledge, it should be stressed that at least some media researchers during the 1960s and 70s had an ambition to take a collective approach to all mass media. In fact, the most important SOU on media research, *Forskning om massmedier* ('Research about mass media') (SOU 1977: 11), argued 'for research that addresses all media and not just the press, or radio and TV'. Artificial boundaries 'within the field (often) prevent the necessary comparison between media' (SOU 1977: 11: 18), the report further stated, bemoaning that national media research 'was far too concentrated on individual "channels" such as TV, radio or the press, despite it being generally most advantageous to study the mass media situation as a whole' (SOU 1977: 11: 11). The desire to include several media forms and formats in comparative studies had in fact been a recurring feature in several media-related SOU reports, from the radio inquiry (SOU 1965: 20) to the press inquiry (SOU 1974: 102; Hyvönen et al. 2017).

Yet another way to graphically study the broader context of film politics is to utilize the output dataset from Mallet as a way to create network graphs of co-occurring topics, that is, to position the distinct film topic among other topics found in the entire SOU material. The 'Film (158)' topic does not necessarily have a lot in common with these other topics, but there is always some form of connection (algorithmically defined). In the following two illustrations (Illustrations 5 and 6), topics are displayed together in linked clusters depending on the probabilistic weights between topics (source) and SOU reports (source target). The computed average topic weights for each report (based on weights in each text chunk of 1000 nouns) were used to visualize a network of SOU reports and their most dominant topics (with some close or distant relationship to the 'Film (158)' topic). The weakest 'SOU-report-to-topic' links were filtered out based on a configurable threshold (with a probability weight less than one per cent). The script repeated the process for all reports between 1922 and 1991, and for all the 500 topics, turning the result into a spreadsheet. As stated, topic modeling is not a static computational process (since it is based on probability).

The Gephi graph was modulated by the so-called Closeness-Centrality and Force Atlas as well as a Modularity Class algorithm to sort topics and reports into thematic clusters. In order to classify a topic, all SOU reports and topics were also manually examined as a way to identify the common theme of a topic cluster. As is apparent, once again the 'Broadcasting (90)' topic and its technical sister topic 'Transmission (254)' were situated in the centre of the illustration, hinting at the centrality of radio and television (not the least from a regulatory perspective) for the Swedish state apparatus.

In the graphs, topics do not co-occur with each other directly, but through SOU reports. For example, the publication *Via satellit och kabel* ('Via satellite and cable') (SOU 1984: 65) connected topic 'Film (158)' with topic 'Broadcasting (90)', but also, following one of the edges, the small, thin red lines transgressing '1984065', with the peripheral topic 'Telegraph/Telephone (19)'. Evidently the film topic was linked via various SOU reports to other media topics like 'Broadcasting (90)' or 'New media (254)', but also more surprisingly to the topic 'Archive (295)' with (non-displayed) terms like 'handling' (document), 'material', '*gallring*' ('culling'), '*papper*' ('paper'), '*kopia*' ('copy') and 'film'. Another interesting aspect of the illustration is that it visualizes the relationship between distinct media topics in the SOU material and copyright issues. The 'Copyright (347)' topic was linked to other media topics through a number of SOU reports; the publication *Översyn av upphovsrättslagstiftningen* ('Evaluation of the Copyright Law') (SOU 1983: 65), for example, connected the film and

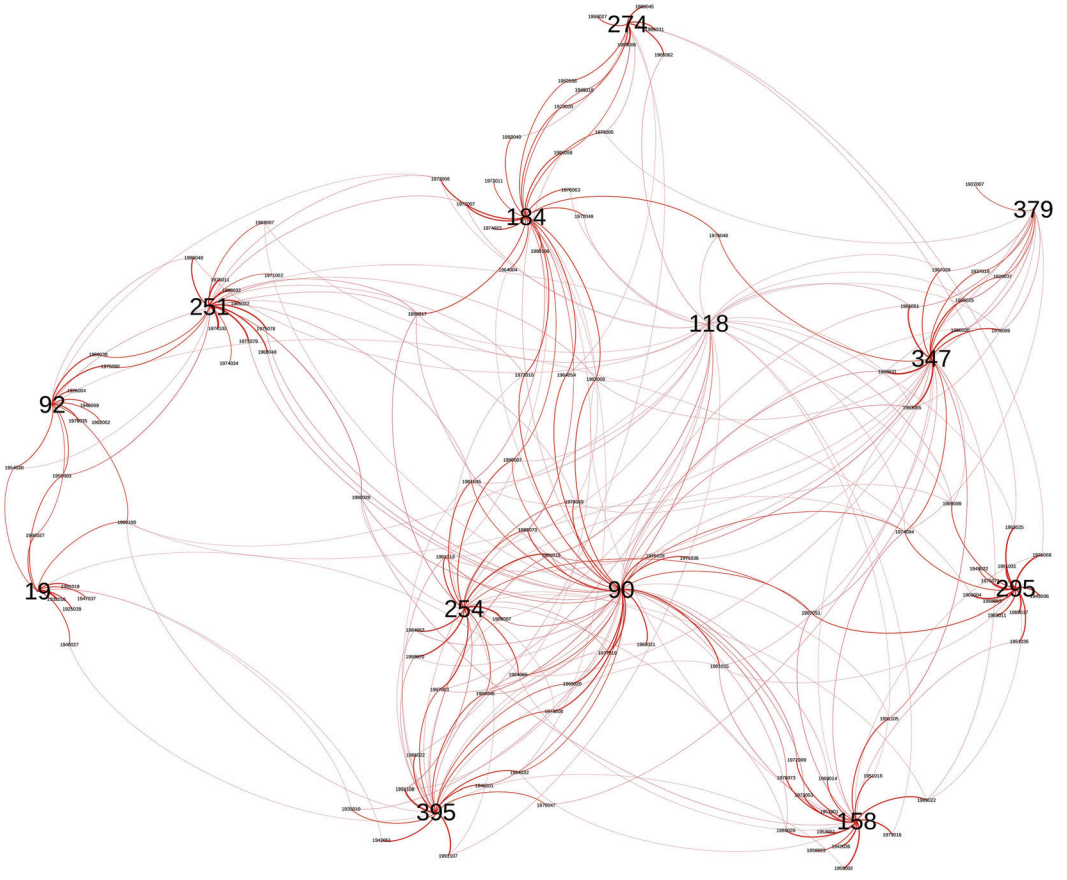

Figure 5: SOU topics related to the 'Film (158)' topic (below to the right). To reduce the risk of making the Gephi graph too messy, we chose to keep the original (arbitrary) topic numbers. Bold figures in the graph are topic numbers, and smaller numbers are SOU reports connected to topics – that is, '1959002', where 1959 is the year and 53 the report serial number, referring to the publication Filmstöd och biografnöjesskatt *('Film subsidies and cinema entertainment tax') (SOU 1959: 2).*

copyright topics. Since the copyright topic was also linked to 'Advertisement (184)' in the SOU material, the illustration vividly displays various camouflaged relations between discourses within work performed by film-related governmental commissions. If the topic 'Film (158)' was primarily perceived as a distinct and separate topic (given our discussion above), it can also be empirically interlinked to a number of varied and comprehensive clusters of a broadly defined media discourse, mostly hidden among the semantic structures, but which the LDA algorithm uncovered.

One last illustration can serve as a case in point. Finding a distinct film topic within a mass media meta-topic was not surprising. However, Gephi's Modularity Class algorithm (like Mallet) is based on a probabilistic model that alters the result of network connections each time it is run, sometimes marginally, sometimes a bit more. When a topic becomes situated in different clusters, it is interlinked to different and altered SOU contexts. The topic 'Film (158)', for example, often appeared in a similar media cluster as above when

running various executions of Gephi. However, sometimes the same film topic appeared in totally different clusters, occasionally with negative media connotation (i.e 'the danger of film') invoking a silent cinema reformist discourse on children and film.

Figure 6 displays one of these Gephi executions where the film topic was linked to a number of topics concerning youth and children. 'Quality films' was not the common denominator, however, but rather the psychological dangers of the medium and the negative influence of film. The film topic was even

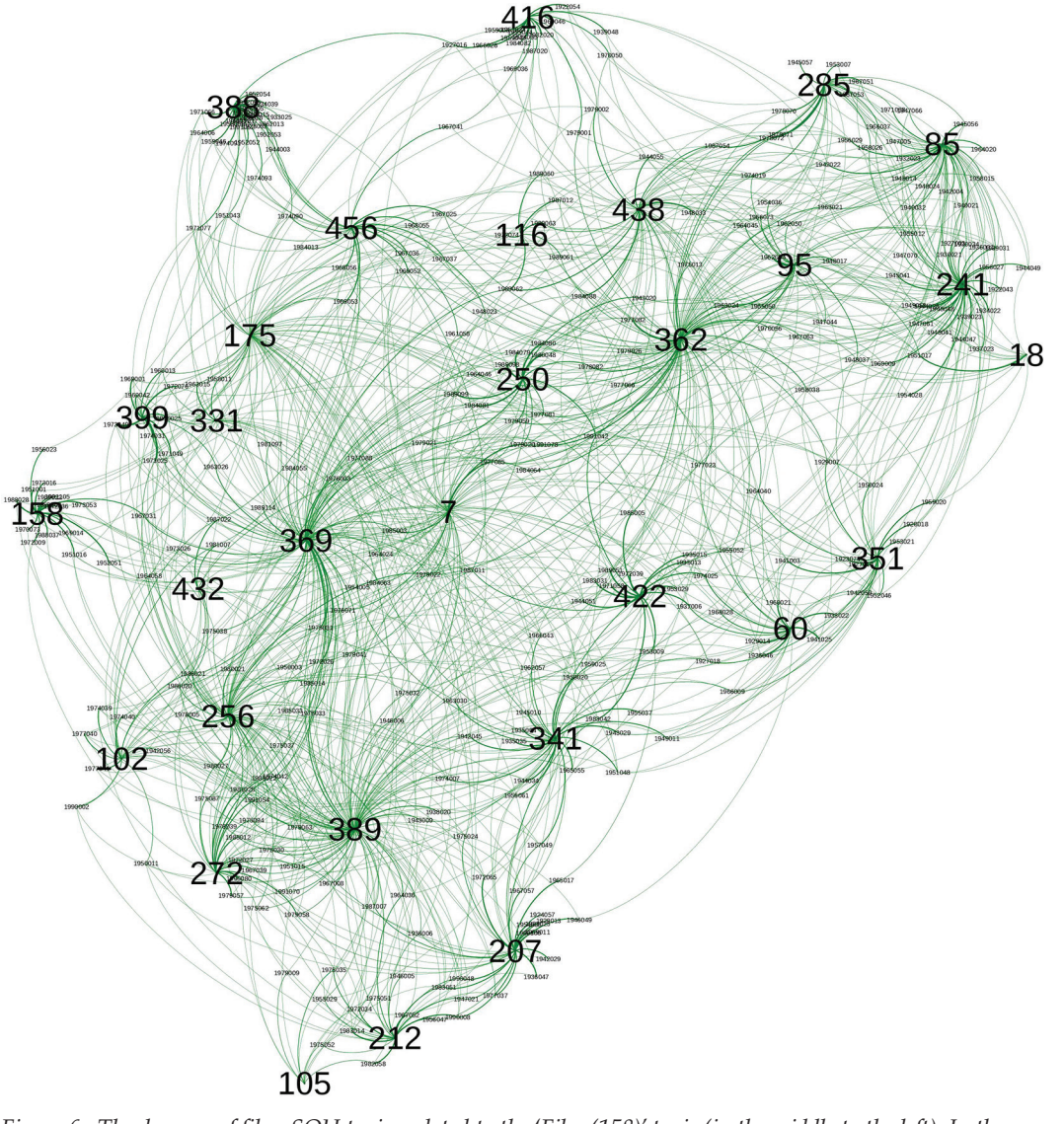

Figure 6: The dangers of film. SOU topics related to the 'Film (158)' topic (in the middle to the left). In the Gephi graph, the film topic is perceived and interlinked as a 'negative medium', and related to topics such as 'Social Welfare (102)' and 'Upbringing (389)'. As is apparent, the film topic was even situated next to the 'Crime (399)' topic as well as distantly related to the topic 'Mental Hygiene (351)' (in the middle to the right).

situated next to the 'Crime (399)' topic (interlinked by SOU 1951: 16) as well as distantly related to the topic 'Mental Hygiene (351)'. For the state, however, socio-political worries about the care of children dominated governmental commission work on the matter. Film was seen as a potentially socially damaging medium that needed to be restricted, especially regarding children. Within the topics 'Social Welfare (102)' and 'Upbringing' (389)', for example, 'children' was among the top keyword list of nouns. The film and upbringing topics were connected through the publication *Barn och film* ('Children and film') (SOU 1952: 51). Similarly, the later publication *Några barn- och ungdomsfrågor 1982–1985* ('Some questions concerning children and youth 1982–1985') (SOU 1985: 33) connected both topics and the topic 'Parental education (256)'. Suffice to say, topic modeling can shed light on issues that at first might not be visible. Of course 'film and children' has been a recurring topic within film historiography. Yet topic modeling the SOU material with such an emphasis reveals quite astonishing connections between governmental commission work that situate the discourse on film policy within a much broader legislative framework.

Conclusion

In 1987, the statistician George E. P. Box stated that 'Essentially, all models are wrong, but some are useful' (Box and Draper 1987: 424). At least the latter part of the quote is an apt description of the topic models connected to Swedish film politics that this article has foregrounded and displayed graphically. Although we have emphasized the need to study an entire archival collection (in this case, SOUs from 1922 to 1991), it is not necessary for film scholars to pay attention to all the details in our results. The envisioned methods are what matter, and the scripts we have developed to analyse the SOU genre do have ramifications for more general (and textual) digital humanities research. While the methods are not novel within the digital humanities field, the SOU material as a whole has not until recently been studied through digital and quantitative methods (Norén 2016). Within digital humanities, the latter part of our analysis is also less commonly practiced. Naturally we are aware that our article does not fundamentally change the perception of the history of Swedish film politics. Still, it should be apparent that digital methods can potentially reveal hitherto neglected parts of Swedish film history, especially regarding relations between different SOU publications. We have shown that as a digital method, topic modeling is far more sophisticated and supersedes traditional 'digital searches' of potentially relevant SOUs in the database located online at the National Library of Sweden. Unsupervised text mining methods, in our case LDA topic modeling, can connect seemingly separated topics and discourses that together create a more comprehensive (and literal) depiction of the broader political context of Swedish film politics during the twentieth century, leading to new perspectives and new questions.

Topic modeling should be recognized as a valuable method and research tool for gathering an overview of a major corpus; as a way to pose new and unforeseen research questions; and as a kind of computational support that makes it possible to apprehend major patterns virtually impossible to detect through a traditional archival investigation. In many ways, topic modeling is a method that leads to both anticipated and unanticipated results. It may confirm assumptions, but may also support findings that have not been noted before.

It should be noted that all our computational results needed to be interpreted. It is thus a misapprehension that the digital humanities are only

devoted to clear-cut computational answers. Another general observation is that due to the transformation of media into data, digital methods can today be widely utilized for academic research. At online services or digital platforms, ranging from YouTube to Netflix and Spotify, computational media can be analysed by novel means, albeit with some necessary coding skills. The same is true for archival material, since the rapid digitization of the cultural heritage has turned the latter into data as well. There is thus a need for media-specific readings of the computational base, that is, the mathematical structures underlying various interfaces, surfaces and media modalities that resonate with scholarly interests in technically rigorous ways of understanding the operations of contemporary media technologies. What exactly happens when data is turned into media – and vice versa? Research on computational media, whether in the form of software studies, digital humanities, platform studies or media archaeology, has repeatedly stressed the need for in-depth investigations of how computing technologies work. Our analyses of the SOU genre is specific, but the methods we propose are of general relevance.

However, at present many media scholars, largely unfamiliar with and lacking hands-on experience working with such digital methods, take a critical stance towards computational perspectives on media and cultural research. At the 2016 media and communication conference (ECREA in Prague), for example, senior media scholar Peter Dahlgren in his keynote lecture disapproved of such data-driven research: 'digital positivism' should be abandoned and 'more theory' was instead needed (Dahlgren 2016). We disagree. Over the last ten years, within the extensive field of media and film studies, different types of digital methods have been taken up as key instruments for developing pioneering ways to analyse and understand digital media, both in contemporary and in historical forms (regarding digitized material). If contemporary media platforms and the digitized heritage increasingly serve as key delivery mechanisms for cultural materials, we need more digital methods for relevant scholarly analyses – not fewer.

REFERENCES

Acland, Charles and Hoyt, Eric (eds) (2016), *The Archlight Guidebook to Media History and the Digital Humanities*, Sussex: Reframe Books.

Blei, David, Carin, Lawrence and Dunson, David (2012), 'Probabilistic topic models', *IEEE Signal Processing Magazine*, 27:6, pp. 55–65.

Blei, David, Ng, Andrew and Jordan, Michael (2003), 'Latent dirichlet allocation', *The Journal of Machine Learning Research*, 3:1, pp. 993–1022.

Blomgren, Roger (1998), *Staten och filmen: svensk filmpolitik 1909–1993*, Stockholm: Gidlund.

Box, George E. P. and Draper, Norman R. (1987), *Empirical Model-Building and Response Surfaces*, New York: Wiley.

Clement, Tanya (2016), 'Where is methodology in digital humanities?', in Matthew Gold, *Debates in the Digital Humanities 2016*, Minneapolis: University of Minnesota Press, pp. 153–75.

Dahlgren, Peter (2016), 'Media and civic cultures in Western "late democracy": The new subjectivity and sociality', *ECREA Conference*, Prague, 9-12 November.

Flueckiger, Barbara (2011), 'Analysis of film colors in a digital humanities perspective', *Frames Cinema Journal*, 2012:1, pp. 44–60.

Furhammar, Leif (1991), *Filmen i Sverige: en historia i tio kapitel*, Stockholm: Svenska Filminstitutet.

Garncarz, Joseph and Ross, Michael (2006), 'Die Siegener datenbanken zum frühen kino in Deutschland', *KINtop. Jahrbuch zur Erforschung des frühen Films*, 14&15, pp. 151–63.

Government Offices of Sweden (2016), 'Committees', http://www.government.se/how-sweden-is-governed/committees/. Accessed 19 December 2016.

Hoyt, Eric, Hagenmaier, Wendy and Hagenmaier, Carl (2013), 'Media + history + digital + library: An experiment in synthesis', *E-media studies*, 3:1, http://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/1/xmlpage/4/article/430. Accessed 19 December 2016.

Hyvönen, Mats, Snickars, Pelle and Vesterlund, Per (2015), *Massmedieproblem: Mediestudiets formering*, Stockholm: Mediehistoriskt arkiv.

——— (2017), 'The formation of Swedish media studies, 1960–1980', *Media History*, February, pp. 1–13.

Jockers, Matthew (2013), *Macroanalysis: Digital Methods and Literary History*, Urbana: University of Illinois Press.

Johansson, Jan (1992), *Det statliga kommittéväsendet: Kunskap, kontroll, konsensus*, Stockholm: Stockholm universitet.

Klenotic, Jeffrey, F (1994), 'The place of "new" film historiography: The discourse of corrective revisionism', *Film History*, 6:1, pp. 45–58.

Mimno, David (2012), 'Computational historiography: Data mining in a century of classics journals', *Journal on Computing and Cultural Heritage*, 5:1, pp. 1–19.

Moretti, Franco (2000), 'Conjectures on world literature', *New Left Review*, 2000:1, pp. 54–68.

——— (2013), *Distant Reading*, London: Verso.

Moretti, Franco and Pestre, Dominique (2015), 'Bankspeak: The Language of World Bank Reports, 1946–2012', *Pamphlet*, 9, https://litlab.stanford.edu/LiteraryLabPamphlet9.pdf. Accessed 19 December 2016.

Norén, Fredrik (2016), 'Information som lösning, information som problem: En digital läsning av tusentals statliga utredningar', *Nordicom Information*, 38:3, pp. 9–26.

Piper, Andrew (2016), 'There will be numbers', *Journal of Cultural Analytics*, 23 May, http://culturalanalytics.org/2016/05/there-will-be-numbers/. Accessed 19 December 2016.

Premfors, Rune (1983), 'Governmental commissions in Sweden', *American Behavioral Scientist*, 26:5, pp. 623–42.

Rieder, Bernhard and Röhle, Theo (2012), 'Digital methods: Five challenges', in D. Berry, *Understanding Digital Humanities*, Basingstoke: Palgrave Macmillan, pp. 67–85.

Schein, Harry (1971), 'Lösa stolpar kring tråd-TV', utkast till PM (till Ingvar Carlsson), Harry Scheins personarkiv and Arbetarrörelsens arkiv och bibliotek, volym 4:51.

——— (1973), 'Till Filmutredningens ledamöter och Gösta Werner', Harry Scheins personarkiv and Arbetarrörelsens arkiv, brev 9/3, volym 4:36.

SFI (2016), http://www.filminstitutet.se/en/about-us/swedish-film-institute/mission-and-organisation/history-of-the-film-institute/. Accessed 19 December 2016.

Snickars, Pelle (2010), 'Vad är kvalitet?', in Lars Ilshammar, Pelle Snickars and Per Vesterlund, *Citizen Schein*, Stockholm: Mediehistoriskt arkiv, pp. 162–78.

——— (2015), 'Mediestudiets infrastruktur: Om etableringen av Arkivet för ljud och bild', in Mats Hyvönen, Pelle Snickars and Per Vesterlund (eds), *Massmedieproblem: Mediestudiets formering*, Stockholm: Mediehistoriskt arkiv, pp. 55–105.

SOU (1930), *Förslag till förordning angående film*, Stockholm: Kommunikationsdepartementet.

——— (1942), *Statligt stöd åt svensk filmproduktion*, Stockholm: Nordiska bokhandeln.

——— (1951), *Filmcensuren*, Stockholm: Ecklesiastikdepartementet.

——— (1952), *Barn och film*, Stockholm: Ecklesiastikdepartementet.

——— (1959), *Filmstöd och biografnöjesskatt*, Stockholm: Esselte.

——— (1965), *Radions och televisionens framtid i Sverige*, Stockholm: Hæggström.

——— (1967), *Filmens inflytande på sin public*, Stockholm: Hæggström.

——— (1973), *Samhället och filmen*, Stockholm: Utbildningsdepartementet.

——— (1974), *Svensk press*, Stockholm: Finansdepartementet.

——— (1977), *Forskning om massmedier*, Stockholm: Allmänna förlaget.

——— (1983), *Översyn av upphovsrättslagstiftningen*, Stockholm: Allmänna förlaget.

——— (1984), *Via satellit och kabel*, Stockholm: Allmänna förlaget.

——— (1985), *Några barn – och ungdomsfrågor 1982–1985*, Stockholm: Allmänna förlaget.

——— (1988), *Statens ansvar för visning och spridning av värdefull film*, Stockholm: Allmänna förlaget.

——— (1999), *Bevara dokumentärfilmens kulturarv åtgärdsförslag samt förslag till ett centrum för dokumentärfilm och en filmvårdscentral*, Stockholm: Fakta info direkt.

——— (2009), *Vägval för filmen*, Stockholm: Edita Sverige.

Tingsten, Herbert (1955), 'Stability and vitality in Swedish democracy', *Political Quarterly*, 26:2, pp. 140–51.

Tsivian, Yuri (2009), 'Cinemetrics, part of humanities' Cyberinfrastructure', in Michael Ross, Manfred Grauer and Bernd Freisleben (eds), *Digital Tools in Media Studies: Analysis and Research. An Overview*, Bielefeld: Transcript Verlag, pp. 93–100.

Veit, Sylvia (2009), 'National experiences with general forms of ex ante evaluation of legislation: The cases of Germany and Sweden', in Jonathan Verschuuren (ed.), *The Impact of Legislation: A Critical Analysis of Ex Ante Evaluation*, Boston: Martinus Nijhoff Publishers, pp. 175–98.

Verhoeven, Deb (2012), 'New cinema history and the computational turn', *World Congress of Communication and the Arts Conference*, Guimarães, Portugal, 15–18 April, http://dro.deakin.edu.au/eserv/DU:30044939/verhoeven-newcinema-2012.pdf. Accessed 19 December 2016.

SUGGESTED CITATION

Norén, F. and Snickars, P. (2017), 'Distant reading the history of Swedish film politics in 4500 governmental SOU reports', *Journal of Scandinavian Cinema*, 7:2, pp. 155–176, doi: 10.1386/jsca.7.2.155_1

CONTRIBUTOR DETAILS

Fredrik Norén is a Ph.D. candidate in media and communication with a digital humanities profile at Umeå University. He is affiliated with Humlab (Umeå University) where he is working on several collaborative text mining projects. In 2016, he was a visiting student researcher at the Center for Spatial and Textual Analysis at Stanford University and co-arranged a symposium on topic modeling at the National Library of Sweden in 2017.

Contact: fredrik.noren@umu.se

Pelle Snickars is professor of media and communication studies, a chair directed towards the digital humanities, at Umeå University, where he is also affiliated with the digital humanities centre, Humlab. He is currently in charge of two major research projects: Streaming Heritage: Following files in digital music distribution (focusing on Spotify) and Digital Models: Technohistorical collections and the digital humanities, in cooperation with the Technical Museum in Stockholm.

Contact: pelle.snickars@umu.se

Fredrik Norén and Pelle Snickars have asserted their rights under the Copyright, Designs and Patents Act, 1988, to be identified as the authors of this work in the format that was submitted to Intellect Ltd.