Från chiffer till klartext? Temamodeller i statliga offentliga utredningar 1945–1989

Pelle Snickars

Det var en gång för länge sedan som den första datamaskinen uppfanns, ”därmed inleddes en ny epok”. Inledningsraderna i Olof Johannessons *Sagan om den stora datamaskinen: en vision* (1966) har visat sig vara närmast profetiska. Dataålderns ankomst innebar en lika snabb som ”fantastisk utveckling som omvandlat de primitiva fördatiska samhällena och sammansmält dem till vår tids fulländade organisation”, där ”teletotal” länkade samman alla ”dator” i ett nätverk. Johannessons vision var visserligen en överdrift, en sorts datadystopi där datorna tagit makten efter en katastrof. En ”vacker dag strejkade teletotalen”, den superdator som styrde allt. Därefter utbröt panik: all ström försvann. Först genom assistans från ”rymdkolonier” kunde datorna helt överta herraväldet.

Utan skönlitteraturens spegling av kultur och samhälle är det svårt att få korn på den tid som varit (eller kommer). Men hos Johannesson skvallrade rubriker som ”Datan automatiserar” och ”Datastyrd dataservice” om att de formulerats på 1960-talet; automatisk databehandling (adb) var ett av dåtidens modeord. Ändå var Johannessons berättelse ovanligt framsynt. Förmodligen var berättaren själv ”en data”, som funderade över ”problemet människan”. Men säker kan man inte vara. ”När en historiker nått fram till sin egen tid, bör han rimligen lägga ned pennan”: att fortsätta är ”att spekulera om framtiden”. Men det var förstås det som Olof Johannesson – ett *nom de plume* för fysikern (och senare Nobelpristagaren) Hannes Alfvén – ägnade sig åt.[1]

1987 gav det lilla förlaget Pilgrim press ut en nyutgåva av *Sagan om den stora datamaskinen* som då sedan länge varit slutsåld. Den innehöll ett nytt efterord av Alfvén, skriven vid en tidpunkt då persondatorn blivit alltmer populär. Alfvén såg tillbaka på den framtidsvision han en gång formulerat. Sommaren 1966 hade han suttit ”i en liten stuga på Västkusten” och börjat skriva på en saga för ett av barnbarnen. Det märkliga var att ett slags datalogisk tänkande så sakteliga hade smugit sig på och tagit över berättelsen. För Alfvén var det därför ”nog inte helt klart” om Johannesson i slutet av boken var ”en människa, en data eller en hybrid” – en sorts robotkrönikör. Men Alfvén påpekade också att hans vision inte bara var *science fiction*; den hämtade en hel del näring från det nationella forskningsarbete han ägnat sig åt. ”Vid den tiden hade jag suttit i Erlanders forskningsberedning några år.” Den första forskningsberedningen hade bildats 1962 på initiativ av statsminister Tage Erlander för samråd, som det hette, ”om den långsiktiga inriktningen av den svenska forskningspolitiken”.[2] Beredningen var tänkt som ett forum med både formella och informella kontakter mellan forskare och regeringsansvariga; Alfvén var en av femton professorer som medverkade. En av beredningens huvuduppgifter ”var att i Sverige införa datatekniken, som snabbt utvecklats under kriget och som nu svepte över

[1] Olof Johannesson / Hannes Alfvén, *Sagan om den stora datamaskinen: en vision*, Stockholm 1966, s. 9, 137.

[2] ”Berättelse till 1963 års riksdag om vad i rikets styrelse sig tilldragit”, 3/1 1963. Digitaliserat riksdagstryck: https://www.riksdagen.se/sv/dokument-lagar/dokument/framstallning--redogorelse/till-1963-ars-riksdag-om-vad-i-rikets-styrelse_EO40Rberb1 (2021-03-01). Arkivet från 1962 års forskningsberedning (fram till 1986) finns på Riksarkivet, SE/RA/1200/07. Det omfattar 15 hyllmeter och 183 volymer; vissa inkommande handlingar är sekretessbelagda.

civila användningsområden". Det var sådana frågor som bildade upptakten till sagan om den stora datamaskinen. Det var rentav med utgångspunkt i beredningens promemorior som Alféns alter ego gick lös "på den närmaste framtiden ... Det gällde endast att översätta deras okänsliga kanslisvenska till ett klarare och enklare språk". Olof Johannesson har fått många komplimanger för att han förutsåg framtiden och de sätt som datorer blivit oumbärliga, påpekade Alfvén 1987. Men det var egentligen "en enkel sak: översättningen av forskningsberedningens chiffer till klartext".[3]

Utredningar – som data

I den här artikeln är det inte forskningsberedningars promemorior som står i fokus, men väl textuell transformation av ett annat slag: omkodandet av statens offentliga utredningar mellan 1945 och 1989 till maskinläsbar data. Artikelns uppslag låter sig sammanfattas ungefär så här: om alla SOU:er betraktas som en enda enorm text som staten skriver, vilka teman i denna text kan programvara läsa och uppfatta? Huruvida det som i Alfvéns fall handlar om att gå från chiffer till klartext låter jag vara osagt; frågetecknet i artikelrubriken är avsiktligt. Men för de algoritmiska temamodeller jag använder, baserade på den statistiska modellen *latent dirichlet allocation* (lda), är alla resultat numeriska, ofta med mer än femton decimaler. Hur dessa resultat sedan ska tolkas är en annan sak.

Betraktat som ett gigantiskt dataset är SOU:erna ett slösande rikt forskningsmaterial; ett slags omfattande textuellt mediearkiv som kan läsas på många olika sätt. Visserligen ägnade sig kommittéväsendet ganska sällan åt specifika forskningsfrågor. Av de cirka femton hundra betänkanden som exempelvis publicerades under 1960- och 1970-talen rörde bara tio stycken särskilda forskningsfält: näringslära, lant- och skogsbruk, pedagogik, energi- och försvarsforskning, farmaceutisk forskning, beteendevetenskaplig arbetslivsforskning, kriminologisk forskning, samt – kanske lätt förvånande – massmedieforskning.[4] Jag är mediehistoriker och den här artikeln behandlar framför allt medierelaterade teman som (lda-modellerna menar) återfinns i min forskningsdata. Fokus ligger på hur medier och medierelaterade termer förekommer i SOU-datan, men jag kommer inledningsvis också att göra några mer generella analyser av andra temamodeller kring miljö, religion och invandring.

Det var under den tid då jag själv arbetade som forskningschef på Kungliga biblioteket som alla statliga offentliga utredningar mellan 1922 och 1999 digitaliserades (resterande utredningar finns tillgängliga på regeringen.se). Det är en av nationalbibliotekets allra mest lyckade digitaliseringsinsatser, inte minst eftersom utredningarna till sin natur var offentliga och därför inte inskränktes av upphovsrättsliga hinder. Hur de skulle digitaliseras var ett återkommande tema i KB:s ledningsgrupp, men jag kan tyvärr inte ta åt mig äran av detta arbete; idé och initiativ kom från min dåvarande kollega Agneta Holmenmark. 2011 upphandlade KB en digitaliseringsrobot som framgent användes för att skanna alla

[3] Hannes Alfvén, "Efterord", *Sagan om den stora datamaskinen: en vision / av Olof Johannesson*, Stockholm 1987, s. 147, 148, 151. Man kan konstatera att Alfvéns databerättare återkommer i genren; epigonerna är många. Teknofilosofen Manuel de Landa såg exempelvis i början av 1990-talet i boken *War in the age of intelligent machines* en framtid med robothistoriker, "specialized robot historians", med uppgift att förstå och förklara sin egen uppkomst, "dedicated to understanding their historical origins" – man får väl föreställa sig Schwarzeneggers *Terminator* som historiker. Manuel de Landa, *War in the age of intelligent machines*, New York 1991.

[4] För en diskussion, se Mats Hyvönen, Pelle Snickars & Per Vesterlund, "Mediestudiets formering: en introduktion", *Massmedieproblem. Mediestudiets formering*, Mats Hyvönen, Pelle Snickars & Per Vesterlund (red.), Lund 2015.

utredningar, ett arbete som avslutades 2015.[5] Totalt hade då 6129 SOU:er digitaliserats, ett material som gjordes öppet tillgängligt via KB:s webb. Forskare på Linköpings universitet har därefter utvecklat ett gränssnitt för fritextsökning i dokumenten.[6] Materialet är populärt, och sedan 2017 finns det till och med ett hyllningskonto på Twitter ägnat ”den svenska statens utredningskommittéväsende och dess publikationer”, som med viss regelbundenhet muntert twittrar om specifika utredningar, teman eller illustrationer ur SOU-materialet.[7] Det har hävdats att SOU:erna på grund av sin lätttillgänglighet håller på att bli ett ”överanvänt källmaterial som förmodligen kommer att spegla bilden av svensk politik i historiska undersökningar lång tid framåt.”[8] Det må vara hänt – men det får framtiden utvisa. Att materialet är välanvänt råder det inget tvivel på; nedladdningsstatistik från KB anger att över 30 000 SOU-pdf:er laddats ned årligen sedan 2016.[9] I den här artikeln är det dock alla utredningar som ett enda dataset vilket ligger till grund för de iakttagelser och resultat som redovisas.[10]

Statliga offentliga utredningar har, menar Kjell Östberg och Jenny Andersson i deras översiktsverk av svensk efterkrigstid, spelat en ”utomordentligt stor roll för utförandet av 1900-talets politiska reformer”.[11] De utredningskommittér som tillsattes bestod i regel av politiker och experter inom respektive område; de var ofta en sorts arena för samverkan – och ibland konfrontation – mellan olika samhällsintressen. Samtidigt fungerade kommittéväsendet som en kunskapsgenererande verksamhet. Information inhämtades, bearbetades och nytt vetande producerades, har Kent Zetterberg framhållit. Redan tidigt var grundlighet i arbetssätt ett kännetecken för det svenska utredningsväsendet. Den så kallade emigrationsutredningen 1907–13 levererade inte bara ett slutbetänkande på 900 sidor, ett

---

[5] Hur KB:s arbete med att digitalisera SOU:erna framskred var under flera år ett återkommande tema i nationalbibliotekets årsberättelser. När arbetet var färdigt 2015 påpekade riksbibliotekarie Gunilla Herdenberg: ”Forskningen efterfrågar stora mängder data, inte minst inom det starkt växande fältet digital humaniora. Det handlar om att genom digitalisering skapa förbättrad tillgång till specifika källor men också om att utveckla verktyg för kvantitativa analyser för att spåra trender och tendenser i stora datamängder. För att nationalbiblioteket ska kunna producera stora volymer av digital råvara, vilket krävs för att skapa ett brett forskningsmaterial, krävs ordentliga satsningar och särskilda medel för digitalisering. KB har dock i och med digitaliseringen av Statens offentliga utredningar (SOU) redan skapat en samling väl lämpad att testa olika analysverktyg på.” KB:s årsredovisningar finns online, https://kb.se/samverkan-och-utveckling/kbs-uppdrag/arsredovisningar.html (2021-03-01).

[6] Samtliga SOU:er mellan 1922 till 1999 finns tillgänliga i pdf-format på http://regina.kb.se/sou/. SOU-sök via Linköpings universitet är tillgängligt på https://ep.liu.se/databases/sou/default.aspx (2021-03-01).

[7] Twitterkontot @SOU_min_SOU är populärt, det har i början av 2021 mer än 1 700 följare. Det återfinns på, https://twitter.com/sou_min_sou (2021-03-01).

[8] Brita Planck, ”Digital history – A field, a method or just a phase?”, *Historisk tidskrift* 2020:2. Det ska nämnas att det inte är Planck som argumenterar för en sådan överanvändning, utan att detta argument var ett återkommande tema vid konferensen Digital History vid Göteborgs universitet hösten 2019.

[9] KB:s it-tekniker Anders Blomgren har i mail till författaren (14/12 2020) uppgivit att nedladdade SOU:er från filservern weburn.kb.se 2016 uppgick till 30 492, en siffra som 2018 steg till 35 456 för att 2020 landade någonstans kring 27 500 stycken.

[10] Jag ska villigt tillstå att andra forskare tidigare använt utredningsmaterialet på snarlika sätt. KB anordnade rentav ett #SOUhack hösten 2015. Gustav Holmberg har exempelvis bloggat om begreppet ”amatör” i SOU-materialet, http://www.gustavholmberg.com/tomrum/2015/06/02/statens-blickar-mot-amatorforskningen (2021-03-01). Och inte minst skrev Christopher Kullenberg under 2015 ett flertal inlägg på sin blogg Intensifier om SOU:ernas geografi, om nätverk och ”relationer i SOUniversum”. Kullenberg använde dock inte alla utredningar som dataset, även om han påpekade att det skulle vara intressant att göra en ”analys av hela SOU-materialet”. Se, https://christopherkullenberg.se/amnen-i-souniversum-3-topic-modelling.html (2021-03-01).

[11] Kjell Östberg & Jenny Andersson, *Sveriges historia 1965–2012*, Stockholm 2013, s. 118.

tjugotal expertutredningar och forskningsrapporter hann man också med.[12] Mängden utredningar som initierades från regeringen var många: 1922 publicerades till exempel 59 stycken SOU:er (från ett något färre antal kommittéer; flera åstadkom mer än ett betänkande.) Sjuttio år senare var siffran uppe i 150 utredningar per år. Ser man till de SOU:er som finns på nationalbibliotekets webb publicerades det i genomsnitt cirka 80 betänkanden per år; för den period som den här artikeln baserar sig på rör det sig om 72 utredningar varje år.

Bland historiker råder det konsensus om att det svenska utredningsväsendet primära funktion varit att underbygga och göra politiska reformer – ibland av betydande omfattning – praktiskt genomförbara. Även andra nordiska länder använde sig av offentliga utredningar för att förbereda reformer, exempelvis Norge. Men som Klas Åmark påpekat har de svenska utredningarna genomgående varit både utförligare, sakligare och mer övertygande, och därigenom ”betydligt bättre källor för historisk forskning”.[13] Även den statsvetenskapliga forskningen har flitigt ägnat sig åt att studera det nationella kommittéväsendet eftersom utredningar varit ett sätt att få korn på den politiska beslutsprocessens drivkrafter. Utredningar låg ju till grund för politiskt beslutsfattande. Redan 1927 disputerade Gunnar Hesslén på avhandlingen, *Det svenska kommittéväsendet intill år 1905: dess uppkomst, ställning och betydelse*. Arbete i kommittér var, enligt Hesslén, att likna vid en ”fristående utredningsapparat”, en träffande metaforik i detta sammanhang.[14] Om Hesslén avslutade sin studie i och med unionsupplösningen mellan Sverige och Norge, tog Hans Meijer vid i sin avhandling från 1956, *Kommittépolitik och kommittéarbete: det statliga kommittéväsendets utvecklingslinjer 1905–1954 samt nuvarande funktion och arbetsformer*. Idag ägnar sig statsvetare som Carl Dahlström åt snarlika frågeställningar. Grundfrågan numera handlar inte sällan om huruvida kommittéväsendets roll i svensk politik försvagats – eller inte.[15]

Men det är inte kommittéväsendet som intresserar mig i det följande, utan textmassan i de offentliga utredningarna, ”statens röst” som den kallats.[16] Den här artikeln närläser emellanåt vad som skrevs i vissa specifika utredningar, men framför allt betraktar och bearbetar jag dem som en enda text, som ett (av mig) sammansatt dataset. Forskningsperspektiv och tillvägagångssätt blir då annorlunda jämfört med traditionell historievetenskaplig praktik och metod. Jag är disputerad filmvetare och har tidigare arbetat med statliga utredningar som kvalitativ empiri, framför allt i mediehistoriska sammanhang. I likhet med andra kvalitativt skolade humanister har utredningsmaterial i sådana sammanhang i regel handlat om läsning och tolkande av ett fåtal SOU:er.[17] För historiker

---

[12] Kent Zetterberg, ”Det statliga kommittéväsendet”, *Att styra riket: regeringskansliet 1840–1990*, Stockholm 1990, s. 287.

[13] Klas Åmark, *Hundra år av välfärdspolitik*, Umeå 2005, s. 32.

[14] Gunnar Hesslén, *Det svenska kommittéväsendet intill år 1905: dess uppkomst, ställning och betydelse*, Uppsala 1927, s. 6.

[15] Hans Meijer, *Kommittépolitik och kommittéarbete: det statliga kommittéväsendets utvecklingslinjer 1905-1954 samt nuvarande funktion och arbetsformer*, Lund 1956. För en vidare diskussion, se Jan Johansson, *Det statliga kommittéväsendet: kunskap, kontroll, konsensu*, Stockholm 1992. För en mer samtida diskussion om kommittéväsendet, se exempelvis Carl Dahlström, Erik Lundberg & Kira Pronin, ”Det statliga kommittéväsendets förändring 1990–2016”, *SNS Analys* 2019:59.

[16] Denna audiella metaforik återkommer exempelvis hos Kullenberg för att beskriva utredningarna som textmassa. Se, https://christopherkullenberg.se/souminarium-pa-kungliga-biblioteket.html (2021-03-01).

[17] Ett exempel är My Klockar Linders avhandling, *Kulturpolitik: formeringen av en modern kategori*, Uppsala 2014, där SOU:er återkommande används som empiri – även om bara tre av dessa listas separat i referenslistan.

som explicit ägnat sig åt exempelvis kulturpolitiska frågeställningar har antalet bearbetade och analyserade SOU:er varit fler – men sällan överstigit trettiotalet utredningar.[18] Roger Blomgren baserade till exempel sin avhandling, *Staten och filmen: svensk filmpolitik 1909–1993* på ett knappt femtontal filmrelaterade SOU:er. Hans ambition var att använda ”alla offentliga utredningar om filmpolitik under undersökningens tidsperioder för fånga upp och analysera regeringars och riksdagsledamöters ståndpunkter och argument”.[19] Blomgren nyttjade också andra källor, och min poäng är här inte att vara polemisk. Men jag vill fästa läsarens uppmärksamhet på att storskalig textanalys av *alla* utredningar som ett sammansatt dataset är väsenskilt som forskningspraktik från närläsning av ett tiotal utredningar. Skalan är radikalt annorlunda, distansläsningens kunskapsformer likaså: ”Distant reading ... *is a condition of knowledge*”.[20]

För en människa är det inte möjligt att läsa flera tusen utredningar och dra slutsatser om teman i dem. Men med datorers hjälp går det. Inte minst är det möjligt att genom nätverksillustrationer visualisera och spåra förbindelser mellan utredningar och teman på ett slags aggregerad nivå. I det följande handlar det därför lika mycket om att visa som att beskriva, och artikeln kan ses som ett bidrag till det framväxande forskningsfältet kring digital historia. I boken *What is digital history?* (2021) ger Hannu Salmi följande definition: ”Digital history is an approach to examining and representing the past; it uses new communication technologies and media applications and experiments with computational methods for the analyses, production and dissemination of historical knowledge.”[21]

En närmast historiefilosofisk frågeställning inom kunskapsfältet digital historia är vad tekniken egentligen uppfattar och urskiljer när (som i mitt fall) algoritmisk textanalys appliceras på ett omfattande historiskt material. Så kallad temamodellering kan framstå som närmast sagolik, men vad som en gång var *science fiction* är idag alltmer en realitet – även om själva forskningspraktiken blivit beroende av omfattande programmeringsinsatser. I Sverige kan man för närvarande ana konturerna av hur kunskapsfältet kring digital historia håller på att ta form, där inte minst tidskriften *Scandia* varit ett forum. Anna Nilsson Hammar publicerade redan 2015 en introduktion till fältet, och en serie bloggposter i tidskriften under 2019 har också tematiserat frågan, ”DigHist – Perspectives on Digital History”. Vid Lunds universitet har därtill nyligen plattformen DigitalHistory@Lund lanserats, och en serie av svenska konferenser om digital historia – häromsistens DHIS2020 i Uppsala – har också anordnats.[22]

Att skriva om digital historia och publicera sig i en analog tidskrift kan lätt uppfattas som en anomali. Den här artikeln trycks i *Scandia* men blir också tillgänglig digitalt (efter en tid). En viktig poäng med texten är att jag i den digitala versionen länkar till all data som mina resonemang bygger på. På plattformen Github har jag samlat den data som texten

---

[18] Se, Pelle Snickars, ”Vad är kvalitet?”, *Citizen Schein*, Lars Ilshammar, Pelle Snickars & Per Vesterlund (red.), Stockholm 2010.

[19] Roger Blomgren, *Staten och filmen: svensk filmpolitik 1909–1993*, Stockholm 1998, s. 23.

[20] Franco Moretti, *Distant reading*, London 2013, s. 48.

[21] Hannu Salmi, *What is digital history?*, Cambridge 2021, s. 7.

[22] För en diskussion, se Anna Nilsson Hammar, ”Digital history”, *Scandia* 2015:2. Tidskriftens bloggposter på temat återfinns på https://journals.lub.lu.se/scandia/announcement. Bland svenskspråkiga pionjärer för digitala perspektiv på det förflutna bör nämnas Kenneth Nyberg och Jessica Parland-von Essen, vilka under flera år drivit bloggen, Historia i en digital värld, https://digihist.se/hdv2/ (2021-03-01). För en vidare diskussion, se även Kenneth Nyberg, ”Forskarrollen i en digital värld”, *Historikerns roller*, David Ludvigsson & Martin Åberg (red.), Möklinta 2021.

använder sig av: samtliga SOU:er i txt-format, resultaten från alla temamodeller (åtta olika lda-modeller i excelformat), där de fyra mallet-modellerna (som jag främst använder) bland annat innehåller procentuella sannolikhetsvärden för alla teman, samt alla mina illustrationer, som är mycket detaljerade och bör betraktas på skärm (snarare än i tryck) i både pdf och csv-format (ett format för tabelldata). Det är inte bara ett sätt att förtydliga mina slutsatser, det handlar också om att ge andra forskare access till den data jag använder. Till yttermera visso utgör denna data-metodologiska öppenhet också ett steg i att göra mina resultat reproducerbara – ett förfarande som fortsatt är ovanligt inom humaniora.[23]

Det dataset jag arbetar med innehåller 3154 statliga offentliga utredningar publicerade mellan 1945 och 1989, vilket utgör tidsramen för forskningsprojektet Välfärdsstaten analyserad.[24] Det rör sig om cirka 87 miljoner tecken. Som ofta är fallet inom digital humaniora är iordningställandet av data en mödosam verksamhet.[25] Nu är SOU-datan (som jag kallar den) av god kvalitet; svart tryck på vita standardiserade boksidor förenklar all ocr-tolkning. Tidigare textfiler av många utredningar, framför allt från 1970-talet, hade visserligen felgenererad text med omkastade ord. De berodde dock inte på dålig ocr utan på att texten extraherats felaktigt från resultatet av ocr-tolkningen. KB rättade till dessa fel under sommaren 2020 genom att ocr-tolka om samtliga SOU:er. På Humlab på Umeå universitet har chefsutvecklare Roger Mähler först kört alla SOU:er som textfiler genom Språkbankens korpuspipeline Sparv för att annotera och märka varje ord i texten med dess ordklass och grundform. Resultatet av både annotering och lemmatisering (sammanföring av olika böjningsformer och varianter av ett ord) har lagrats som csv-filer. Sparv är enkelt att använda, det är bara att klistra in text i ett webbgränssnitt för annotering. Men att köra större dataset tar tid; enbart för ett års utredningar behövs ett par timmar (om körningen fungerar utan problem). Språkbanken har därtill under 2020 utvecklat en ny version av Sparv som ett paket i programspråket Python, vilket gör det avsevärt enklare att installera och köra systemet lokalt, även på sin egen dator.

Temamodellering – som metod

Den metod jag använder för min analys kallas på engelska för *topic modeling*, ett samlingsnamn på en rad algoritmer som kan klassificera stora textkorpusar baserat på tematiska strukturer i textmassan. I likhet med Nina Tahmasebi väljer jag att kalla topic modeling för *temamodellering* på svenska, för det rör sig om ämnen i form av olika (mer

[23] Uppslaget är som sagt ovanligt – men knappast nytt. Jag sitter i redaktionsrådet för tidskriften *Journal of Digital History* som lanseras hösten 2021 och drivs av Luxembourg Centre for Contemporary and Digital History i samarbete med förlaget De Gruyter. Grundtanken är att publicera artiklar samt alla relevanta dataset som forskningen bygger på. Även de digitala metoder som nyttjas kommer att redovisas genom öppen access till så kallade Jupyter notebooks eller snarlik programvara som använts för analys. Utan att vara alltför vetenskapsteoretiskt polemiskt innebär öppen tillgång till både metod, förfarande och data att även humanistisk forskning blir mindre beroende av vem som utför analyserna. I en internationell kontext förblir emellertid språk ett problem; att skriva på engelska är enkelt, men att översätta all data är omöjligt. Bara de temamodeller jag arbetar med i denna artikel innehåller sammanlagt mer än hundratusen svenska ord. Se, https://journalofdigitalhistory.org (2021-03-01).

[24] Det forskningsprojekt jag leder, Välfärdsstaten analyserad. Textanalys och modellering av svensk politik, media och kultur 1945–1989, använder tre stora textsamlingar för att analysera tre centrala sfärer i det svenska välfärdssamhällets framväxt: politikens, mediernas och kulturens. Projektet är finansierat av Vetenskapsrådet (2019–2024) inom ramen för forskningsprogrammet Digarv. För mer information, se https://www.westac.se/.

[25] För en diskussion, se Johan Jarlbrink, ”All the Work that Makes It Work: Digital Methods and Manual Labour”, *Digital histories. Emergent approaches within the new digital history*, Mats Fridlund, Mila Oiva & Petri Paju (red.), Helsingfors 2020.

eller mindre tydliga) teman som algoritmerna urskiljer.[26] Dessa teman baserar sig på ords statistiska sammanhang. Grundfrågan inom all temamodellering handlar om vilka diskurser, motiv eller tematiker som kan ha tänkts skapat de dokument som forskaren intresserar sig för. Vilka teman återfinns exempelvis i en utredning som *Bättre miljöskydd* (SOU 1983:20)? Teman kring miljöarbete förstås, liksom miljöavgifter (tema 132), men tematiker som buller och flygplatser (tema 289) liksom fiskeri och utsläpp (tema 132) visar sig förekomma i betänkandet. Med temamodeller är det därför möjligt att upptäcka både förväntade diskurser – liksom mer oväntade motiv.

Inom all temamodellering finns flera modeller att arbeta med och de kan ställas in på väldigt många sätt. När väl SOU:erna annoterats till csv-format har det i mitt fall skapats åtta olika lda-modeller, fyra så kallade gensim-lda-modeller och fyra mallet-lda-modeller. Utan att gå in på tekniska detaljer så består skillnaden i hur man skattar parametrar i en statistisk modell. Gensim använder en snabbare Variational Bayes samplingsmetod medan Mallet förlitar sig på vad som kallas Gibbsampling som har högre precision (men är något långsammare). För temamodeller hävdas det ofta att Gibbssampling har statistiska egenskaper som genererar tydligare resultat. På Humlab har vi valt att arbeta med modeller om 50, 100, 200 och 500 teman (härefter kallat 100-modellen etcetera). De har inga inherenta beteckningar utan listas enbart i stigande nummerordning (från 0 till 49 eller 0 till 499). All namngivning av teman baserar sig på subjektiva tolkningar (varför jag bara använder siffror i mina illustrationer). Modellerna har gjorts åtkomliga för mig i utvecklingsmiljön Jupyter Lab; via ett användarvänligt gränssnitt kan man där köra (och ändra på) Pythonkod direkt i webbläsaren. I Jupyter Lab har jag kunnat välja mellan gensim- och mallet-modeller (med olika många teman). Generellt har mallet-lda-modellerna genererat bättre resultat, det vill säga teman i SOU-datan vars ord och betydelser tydligt hänger samman. I den här artikeln är det därför främst mallet-lda-modellerna som jag använder.

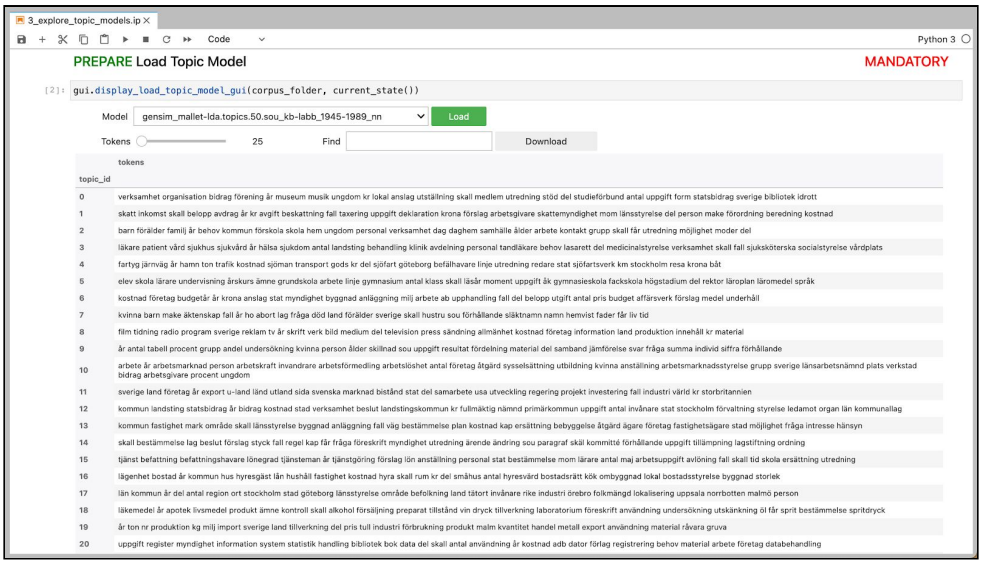


Illustration 1. Skärmdump av analysmiljön för temamodellering i Jupyter Lab. 50-mallet-lda-modellen listar teman – i detta fall med topic-id 0 till 20 – med 25 ord (tokens) för varje tema. Vad teman handlar om är upp till användaren att bestämma, men som synes är det ofta tämligen uppenbart: tema 1 (om skatter), tema 3 (om sjukvård), tema 5 (om utbildning).

[26] För en diskussion, se Nina Tahmasebis serie bloggposter från våren 2020 på Språkbanksbloggen, https://spraakbanken.gu.se/blogg/index.php/author/nina-tahmasebi/ (2021-03-01).

Temamodeller arbetar med statistisk analys av distribution av ord i form av sannolikhetsfördelning över alla ord i en textmassa. Modellerna kan urskilja vilka ord som ingår i ett tema (baserat på ords statistiska fördelning), vilka statliga utredningar som innehåller detta tema, hur det fördelar sig över tid liksom hur temat – eller ord i ett tema – anknyter till andra teman och utredningar. Grundprincipen är att ord samförekommer; det hela är inte konstigare än att det är mer sannolikt att ord som biograf och spelfilm förekommer i en SOU om filmpolitik än termer som fartyg eller jordbruk. Men som André Baltz och Fredrik Norén påpekar i en pedagogisk introduktion av temamodellering så är det ”ords fördelning, inte deras ordning i dokumenten (eller det språk orden tillhör) som har en betydelse för hur teman genereras”.[27] Teman figurerar därför med olika procentuella sannolikhetsvärden. I modeller som listar 200 eller 500 teman är dessa sannolikhetsvärden mycket låga (eftersom det handlar om många teman), medan de i modeller om femtio teman är något högre.

Som historiska utsagor finns en temporal nackdel med temamodeller eftersom de baserar sig på all SOU-data oberoende av när utredningar publicerats. Samtidigt går det att studera förändringar över tid eftersom den procentuella förekomsten av ett tema är mer eller mindre stark vid olika tidpunkter. I Jupyter Lab finns kodceller där man som forskare med knappar och manuella reglage kan styra parametersättningen av koden; de kan modifieras och exempelvis plotta valfritt antal ord i varje tema (från 25 till 200), lista teman över tid (eller en specifik tidsperiod), jämföra eller vikta teman med andra snarlika tematiker etcetera. Resultatet förevisas som tabeller, diagram eller enklare nätverk. Arbetsgången handlar i regel om att söka fram teman, se efter vilka SOU:er som ingår, ladda ned tabelldata och därefter visualisera resultaten i Gephi – standardprogramvara för nätverksvisualiseringar – för bättre överblick. Genom att använda KB:s gränssnitt för Digitaliserade SOU:er 1922–1999 är det därtill enkelt att gå från mer övergripande frågeställningar till att läsa de specifika utredningar som behandlas.

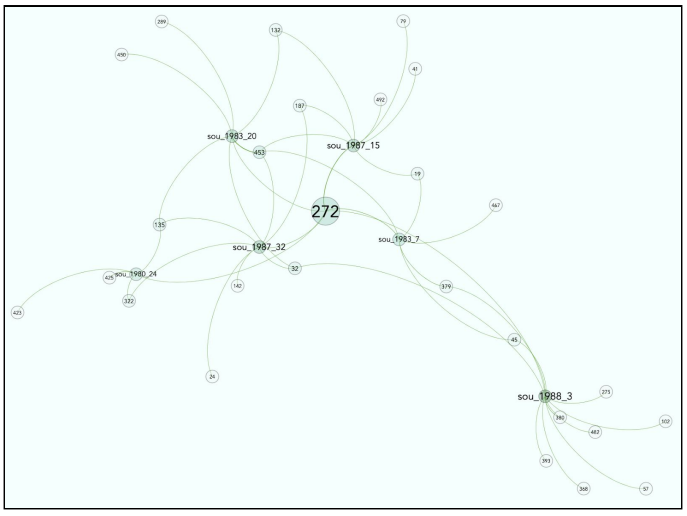


Illustration 2. Nätverk i Gephi (Force Atlas 2 Expansion) för miljötema 272 i 500-gensim-lda-modellen av alla SOU:er mellan 1945–89 (med ett mycket lågt tröskelvärde på en procent). SOU:erna kring noden 272 är alla miljörelaterade – såsom *För en bättre miljö* (SOU 1987:32) eller *Bättre miljöskydd* (SOU 1983:20) – och det gäller även närliggande teman: om naturresurser (135), om miljöavgifter (132), om miljöförvaltning (450). Men noden i höger hörn, SOU 1988:3, handlar snarare om arbetsmiljö och arbetsolyckor, varför intilliggande teman har en sådan karaktär: om arbetstid (275), om arbetsersättning (380), om fängelsestraff (482).

[27] André Baltz & Fredrik Norén, ”Tematisk innehållsanalys med temamodellering”, *Digitala metoder i humaniora och samhällsvetenskap*, Johan Jarlbrink & Fredrik Norén (red.), Lund 2021, s. 150.

Om temamodellering ger indikationer på hur övergripande tematiker fördelar sig över specifika SOU:er, hur de är relaterade och samförekommer – det vill säga, vilka utredningar som delar ett specifikt tema – liksom hur teman och förbindelser förändrar sig över tid, så är det av vikt att notera vilken typ av modell man som forskare använder. Modeller som delar in textmassan av utredningar i 50 teman ger generella resultat, medan modeller med 500 teman är mer specifika. Eftersom det är ord (eller begrepp om man så vill) som modellerna beräknar kan ordsammansättningar ibland förvilla. I 50-mallet-lda-modellen figurerar till exempel termen *miljö* i sju teman. I tema nummer 41 är ordet som starkast – men då handlar det främst om arbetsmiljö. I 100-modellen däremot skiljs arbetsmiljö (tema 65) och miljövård (tema 21) tydligt åt. Och växlar man till 200-modellen figurerar ordet miljö i mer än ett tjugotal teman, där det är främst förekommande i tema 7 (om arbetsmiljö), tema 84 (om miljöpolitik), tema 169 (om miljöfaror) och tema 194 (om arbete och miljö). En nätverksvisualisering i Gephi av de 150 mest förekommande orden i dessa fyra teman med de termer som förbinder dem (där förkortningar och generiska termer rensats bort), bildar en visuellt komplex graf. I centrum befinner sig ett kluster av för alla teman gemensamma ord som hälsa, kontroll och tillståndsmyndighet.

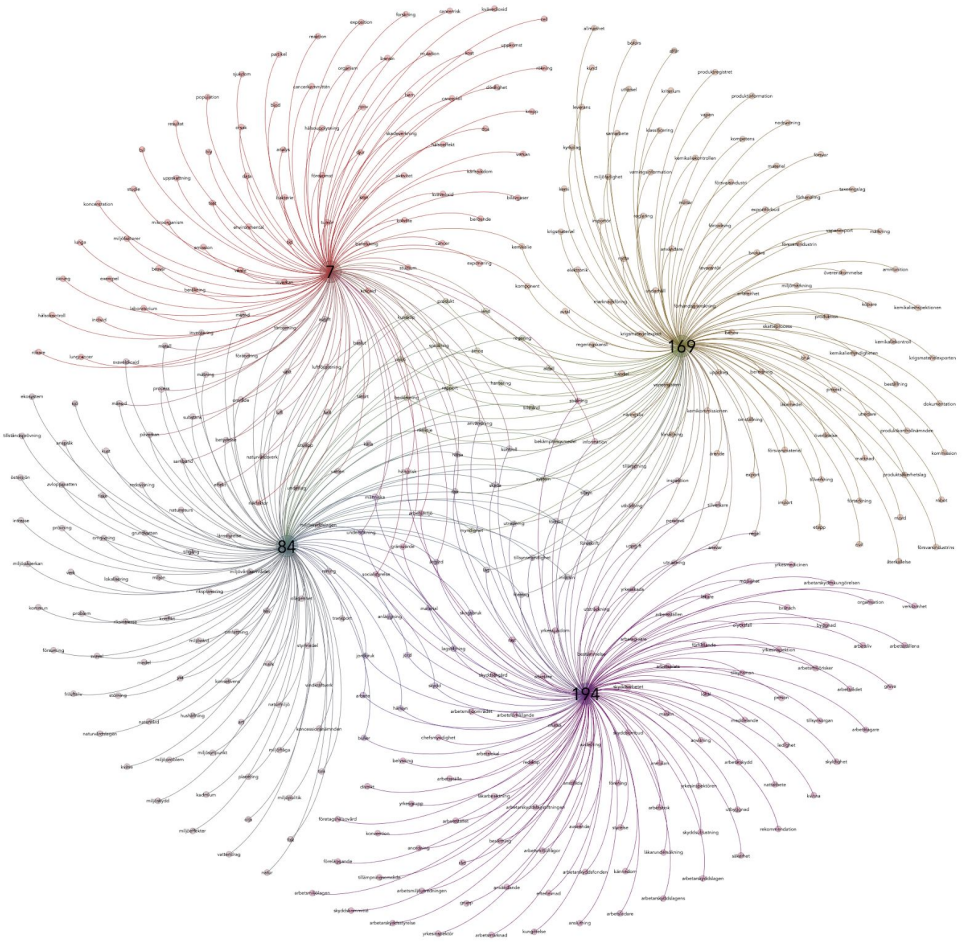


Illustration 3. Nätverk i Gephi (Fruchterman Reingold Expansion) med drygt 400 ordnoder ur fyra olika miljöteman – tema 7 (om arbetsmiljö), tema 84 (om miljöpolitik), tema 169 (om miljöfaror) och tema 194 (om arbete och miljö) – i 200-modellen av alla SOU:er mellan 1945–89. Ett fåtal ord i centrum av nätverket – användning, kontroll, risk, hälsa och skada – är gemensamma för alla teman.

Att använda modeller med olika många teman är ett sätt att zooma in och ut ur SOU-datan – från det övergripande till det mer partikulära, eller från det specifika till det som sammanbinder. Temamodellering kan framstå som abstrakt, men i en Jupyter Lab-miljö är det inte speciellt komplicerat att utföra. Som forskare väljer man tema, tidpunkt eller tidsspann, man kan öka eller sänka viktning och så vidare. Man kan studera relationer mellan SOU:er eller mellan termer – eller båda. I allra högsta grad rör det sig om en explorativ forskningsverksamhet.

Generellt handlar tematiker i modellerna med 50 ord om bredare frågeställningar som återkommer i SOU-datan, som undervisning (elever och skola), sjukvård (patient och läkare), bostadsfrågor (hushåll och hyra), kriminalvård (missbrukare och socialnämnd) eller kyrka (präster och församlingar). De sistnämnda kan få en att höja ögonbrynen, men i 50-mallet-lda-modellen finns ett tydligt kyrkotema och i 50-gensim-lda-modellen innehåller hela 26 procent av alla teman kyrkliga termer: församling, gudstjänst, psalmbok, kyrkoherde, Kristus, himmel, pastorat, begravningsplats – ja, till och med herren själv, Gud, figurerar. SOU-datan är full av kristen terminologi.

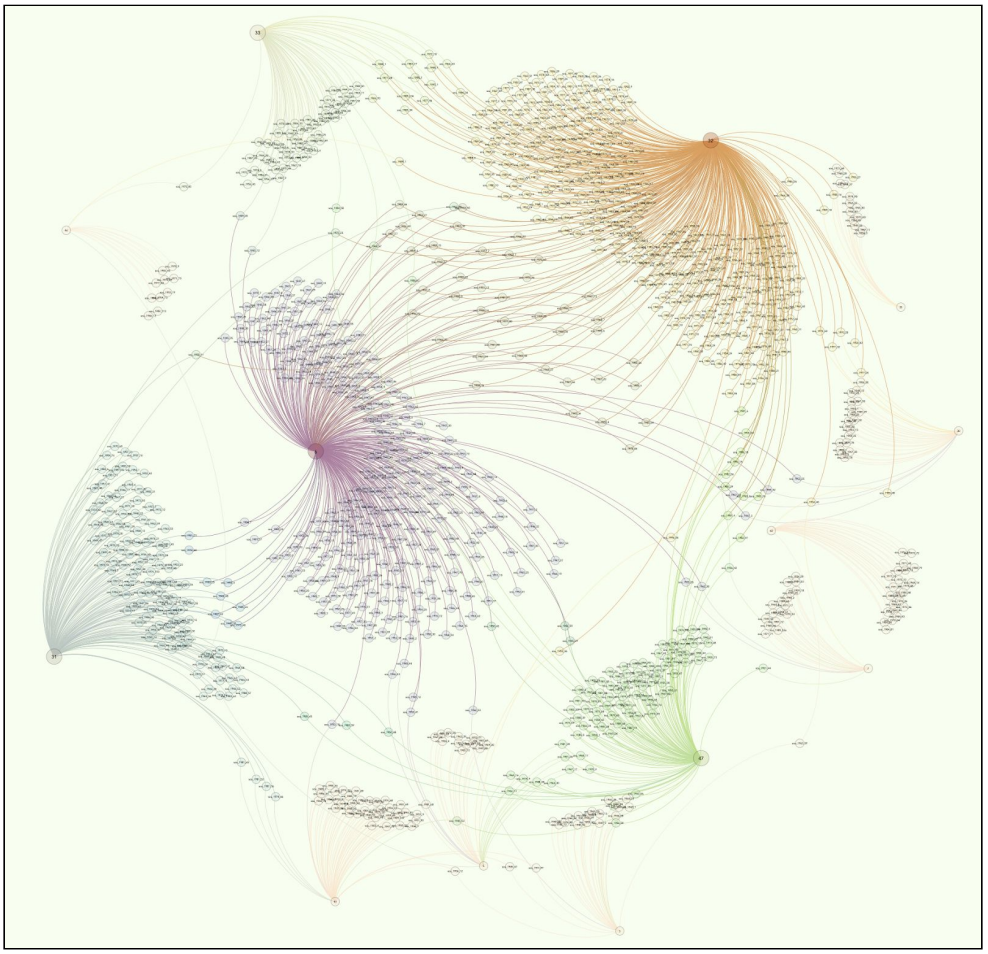


Illustration 4. Sakralt nätverk (Gephi Force Atlas Repulsion Strength 200 000) för 13 stycken kyrkliga teman i 50-gensim-lda-modellen av SOU-datan – samt inte mindre än 477 statliga utredningar (där vissa knyter samman teman) i vilka minst 20 procent av det kristna temat återkommer. Tematiken kan förvåna, men enbart den 1968 tillsatta kyrkohandbokskommittén publicerade fram till 1985 inte mindre än 15 utredningar om *Den svenska kyrkans gudstjänst*, åtta om *Den svenska psalmboken* samt tre utredningar om *Psalmer och visor*.

Eftersom temamodellering handlar om statistisk sannolikhet går det att i en Jupyter Lab-miljö laborera med parametrar och de resultat som modellerna genererar, exempelvis beträffande relationer mellan teman. Sätter man till exempel ett tröskelvärde på 0,3 då plottas alla utredningar som innehåller minst trettio procent av ett specifikt tema. Om tröskelvärdet sänks, så kommer följaktligen fler teman att hänga samman. Omvänt minskar kopplingen mellan teman om samma värden höjs. Men om tröskelvärden sätts alltför lågt (till bara några få procent) så finns risken att teman egentligen handlar om något annat. Exakt vilket värde som bör användas beror på hur många teman en modell har. I en 500-lda-modell är det konkurrens, där behövs låga tröskelvärden för att få korn på relationer. I en 50-lda-modell kan de vara betydligt högre.

Som min kollega Johan Jarlbrink påpekat föreställer sig temamodellering ”textförfattandet något bakvänt: redan från början finns ett antal ämnen, författaren har bara att välja vilka av dem som ska ingå i den enskilda texten, samt blanda dem i olika proportioner. Utifrån detta antagande kan modellen återskapa den ’ursprungliga’ listan på ämnen, och spåra hur de har blandats med varandra i varje enskild text”.[28] Det hela kan förefalla avigt, men påminner historiografiskt om arkeologiska kunskapsanalyser där omedvetna struktursammanhang reglerat och styrt varför det under vissa tidsperioder varit möjligt att skriva (och tänka) om somligt – men inte annat. Jo Guldi har i en analys av politiska diskussioner om brittisk infrastruktur under 1800-talet, baserade på *the Hansard* – transkriberade debatter från Storbritanniens parlament – påpekat att temamodellering gör det möjligt att spåra ”the invisible categories that structure mind, language, priorities, or prejudice in a given corpus”.[29]

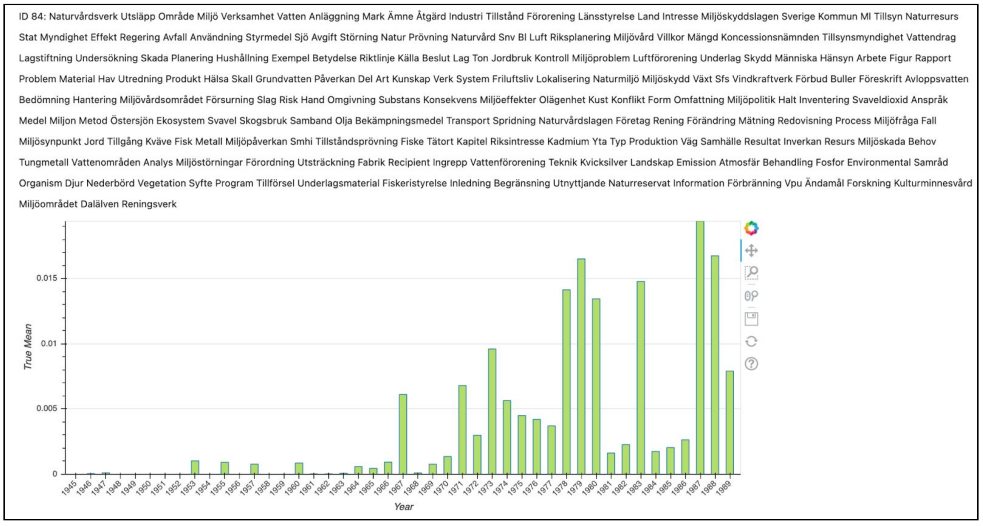


Illustration 5. Skärmdump av stapeldiagram i Jupyter Lab-miljö för temamodell 84 om miljö (med 200 relaterade ord) i 200-mallet-lda-modellen av SOU-datan. Som framgår av (det normaliserade) diagrammet förekommer temat främst från och med 1970-talet och framåt – om än med en viss nedgång några år vid mitten av 1980-talet.

På så vis kan temamodeller spåra hur en diskurs växer fram – och ibland ändrar karaktär. Miljö och invandring är illustrativa exempel. I 50-modellen figurerar miljötermen ungefär

[28] Johan Jarlbrink, ”Från masugn till mikrovågsugn. Historiografisk topic-modellering av årsboken *Daedalus* (1931–2017)”, *Digitala modeller. Teknikhistoria och digitaliseringens specificitet*, Jenny Attemark-Gillgren & Pelle Snickars (red.), Lund 2019, s. 160.

[29] Jo Guldi, ”Parliament’s debates about infrastructure: An exercise in using dynamic topic models to synthesize historical change”, *Technology and Culture* 2019:1.

lika ofta i teman om familjemiljö och arbetsmiljö (som i miljövård), och det är först under 1970-talet som gröna miljöfrågor blir vanliga i SOU-datan. I 100-modellen är samma mönster tydligt; tema 21 om miljövård återfinns exempelvis enbart i två utredningar före 1950, SOU 1946: 85 (om köttkontroll) och SOU 1949:3 (om livsmedel). Temat rör sig sedan över tid från livsmedelskontroll mot miljövård; de fem sista åren av 1980-talet återkommer det i nio utredningar som till exempel *För en bättre miljö* (SOU 1987:32). Ser man till 200- och 500-temamodellerna är bilden densamma: det är först kring 1970 som gröna miljöfrågor blir vanliga. I sin bok *Den gröna vändningen* skriver David Larsson Heidenblad att miljöfrågor fick sitt samhälleliga genomslag under 1970-talet. Hans breda och gröna kunskapshistoria bekräftas i så måtto av flera, olika temamodeller kring miljö i SOU-datan.[30] I den historiska forskning som bedrivits med hjälp av temamodeller brukar det ofta påpekas att modellerna inte sällan bekräftar tidigare forskningsrön. Vän av ordning kan då invända att metoden framstår som historievetenskapligt redundant. Men man kan lika gärna hävda att modellerna verkligen fungerar, och som i exemplet med miljömodellerna och Larsson Heidenblad tydligt indikerar att miljödiskursen i Sverige ändrades vid en specifik tidpunkt.[31]

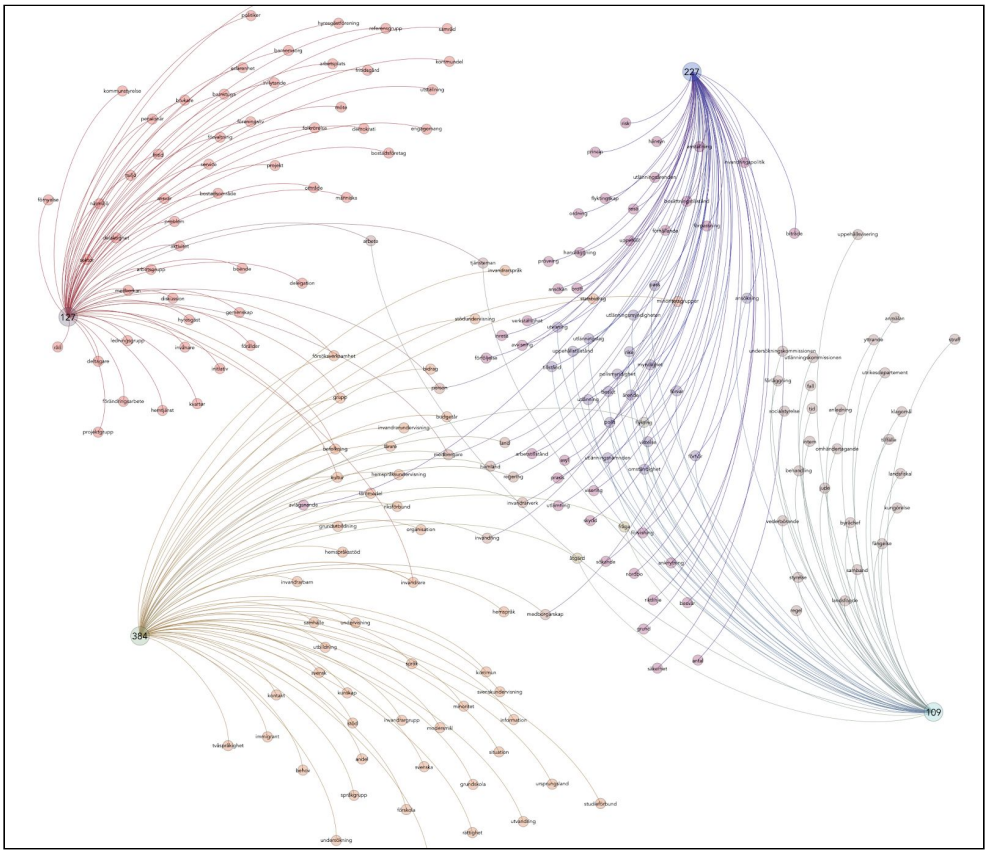


Illustration 6. Nätverk i Gephi (Force Atlas Expansion) med drygt 200 ordnoder ur fyra olika invandringsteman – 109 (flykting och polis), 127 (invandring och bostad), 227 (om invandringspolitik) och 384 (invandrare och språk) – i 500-modellen av alla SOU:er mellan 1945–89.

[30] David Larsson Heidenblad, *Den gröna vändningen. En ny kunskapshistoria om miljöfrågornas genombrott under efterkrigstiden*, Lund 2021.

[31] För en diskussion, se exempelvis Guldi 2019 eller Fredrik Norén & Pelle Snickars, ”Distant reading the history of Swedish film politics in 4500 governmental SOU reports”, *Journal of Scandinavian Cinema* 2017:2.

Ett tema som påminner om det kring miljö är temamodeller kring invandring och flyktingar. SOU-datan innehåller många utredningar som tangerar detta tema – från SOU 1946:36 (om flyktingärenden och säkerhetstjänst) till SOU 1989:111 (om invandrare i storstad). Söker man på besläktade termer i de olika lda-modellerna blir man snart varse att invandringstemat figurerar i en rad konstellationer i SOU-datan. I 50-modellen finns exempelvis ett tydligt tema kring asyl och invandring (nummer 23), men de termer som är vanligast är snarare utlänning, polis, flykting och brott. Byter man till en 200-modell är brottslighet och invandring inte längre samförekommande, snarare förknippas temat (nummer 91) med medborgarskap och undervisning. Och zoomar man ut ytterligare och använder en 500-modell, så är teman kring utlänning och invandrare ofta skilda åt. Tema 109 behandlar flyktingar, utlänningar och polismyndighet, tema 127 är ägnat åt invandring och bostadsområden, medan tema 384 handlar om invandrare, språk och undervisning. Noterbart är att dessa teman förekommer i helt olika tidsspann; tema 109 är strikt relaterat till efterdyningarna av andra världskriget, medan tema 127 och 384 knappt förekommer alls före 1970. Alla dessa tre teman är emellertid sammankopplade med tema 227 som handlar om invandringspolitik, det vill säga om både utlänningar och flyktingar, liksom polismyndighet och utlänningslag – ett tema som därför förekommer under hela den undersökta perioden. I en nätverksvisualisering av samförekommande ord i dessa fyra teman är det uppenbart att tematiker dels skiljer sig; i tema 109 återfinns ord som landsfiskal och jude, medan tema 127 plottar ord som hyresgästförening och barnomsorg. Andra ord som uppehållstillstånd, invandrarverk, medborgarskap och flykting är dock gemensamma.

Tema medier

Alltsedan början av 1900-talet har den svenska staten intresserat sig för medier. Till en början handlade det mest om skatter och regleringar; Statens biografbyrå började exempelvis 1911 sin censurverksamhet. Några år senare infördes (efter protester) en nöjesskatt på biografverksamhet, teatrar och konserter. Staten fick in betydande summor via denna nöjesskatt; bara för biografbranschen handlade det 1919 om cirka 75 miljoner kronor i dagens penningvärde.[32] Under 1920-talet började medier att diskuterades i mer positiva termer. I Folkminneskommitténs *Betänkande med förslag till ett systematiskt utforskande av den svenska allmogekulturen* (SOU 1924:26), lyftes medier exempelvis fram som moderna dokumentationsinstrument, och utredningen hade en närmast begeistrad uppfattning om film. Andra utredningar behandlade framställning av tidningspapper (SOU 1922:36), radioverksamhet inom Telegrafverket (SOU 1925:39), liksom förslag till ändringar i lagstiftningen kring rättigheter till litterära och konstnärliga verk (SOU 1929:37).[33] Efter 1930 blev statliga medieutredningar alltmer standardiserade; vad som i regel publicerades var gedigna utredningar med mediepolitiska förslag. Främst ägnade sig mellankrigstidens medieutredningar åt regleringar – *Förslag till förordning angående biografer och biografföreställningar* (SOU 1930:28) handlade till exempel om ”att förebygga eldsolyckor vid biografföreställningar”[34] – och mer sällan om statlig

[32] ”6,8 miljoner i nöjesskatt”, osignerad, *Svenska Dagbladet* 4/11 1920. För en diskussion om nöjesskatt, biografbyrå och tidig statlig filmpolitik, se Pelle Snickars, *Kulturarvets mediehistoria: Dokumentation och representation 1750–1950*, Lund 2020, s. 498–503.

[33] *Förslag till ändringar i lagstiftningen om rätt till litterära och konstnärliga verk* (SOU 1929:37), s. 7.

[34] *Förslag till förordning angående biografer och biografföreställningar* (SOU 1930:28), s. 3.

medieutveckling; *Utredning och förslag angående rundradion i Sverige* (SOU 1935:10) tillhör ett undantag.

På Nordicoms webbsida återfinns 99 stycken medierelaterade SOU:er publicerade mellan 1945 och 1989 – från *Betänkande angående rundradion i Sverige* (SOU 1946:1) till *Översyn av lagen om pliktexemplar* (SOU 1989:89).[35] Men listan är knappast komplett; ett flertal kommittér och utredningar kopplade till automatisk databehandling är till exempel inte inkluderade. Antalet medierelaterade SOU:er är därför betydligt fler. Det exakta antalet är nu inte det väsentliga; det beror på hur mediebegreppet definieras. Under alla förhållanden kan det konstateras att i takt med den svenska välfärdsstatens formering så utgjorde det statliga utredningsväsendet en kraftfull motor för mediepolitiska initiativ. Utredningarna om film, såsom *Statligt stöd åt svensk filmproduktion* (SOU 1951:1) och *Filmstöd och biografnöjesskatt* (SOU 1959:2) lade grunden för filmreformen 1963. *Televisionen i Sverige* (SOU 1954:1) gjorde detsamma för tv-mediet – en utredning som följdes upp av 1960 års Radioutredning som 1965 publicerade två omfattande utredningar om radion och televisionens framtid (SOU 1965:20/21). Vid ungefär samma tidpunkt sattes även dagspressen under lupp. Det var framför allt rapporteringen om ”tidningsdöden”, som den upprepade företeelsen av nedlagda tidningar då började kallas, som oroade tidens politiker. Termen presstöd är belagd sedan 1965, och det var utredningen *Dagstidningarnas ekonomiska villkor* (SOU 1965:22) som argumenterade för att staten borde ta ett större ansvar för den svenska dagspressens överlevnad. I övrigt var radioutredningen och pressutredningen mycket olika; som jag och kollegor skrivit om i boken *Massmedieproblem* handlade de bägge om massmedier, men refererade bara undantagsvis till andra medier än de som utreddes.[36] Pressutredningen frågade sig bland annat vilken roll ”dagstidningarna har för den politiska åsiktsbildningen. Ibland framföres sålunda uppfattningen, att denna betydelse är ringa och att andra media, främst televisionen, är avgörande för opinionsbildningen”.[37]

I början av 1970-talet var det så dags för publikation av en rad omfattande SOU:er om film – liksom (ånyo) dagspress. Filmutredningen 1968 publicerade fyra betänkanden om *Samhället och filmen* och detsamma gjorde 1972 års pressutredning. *Svensk press* i fyra digra volymer diskuterade både statlig presspolitik, presstöd och dagspressens funktion i samhället. Samtidigt började andra medieformer att uppmärksammas från statligt håll, exempelvis kring automatisk databehandling. Med tanke på tidens socialdemokratiskt färgade kulturpolitik, där 1974 års kulturproposition tänkte sig motverka kommersialismens negativa verkningar inom kulturområdet, var reklam något som också behandlades i flera utredningar. Reklamutredningen resulterade i betänkanden som *Tv-reklamfrågan* (SOU 1973:10), medan Integritetsskyddskommittén publicerade *Reklam och integritet* (SOU 1976:48).

---

[35] Se, https://www.nordicom.gu.se/sites/default/files/Dokument/Mediefakta/Medieregler ing/medierelaterade_souer_1924-_se.pdf (2021-03-01). Nordicom är ett centrum för nordisk medieforskning vid Göteborgs universitet. På hemsidan återfinns medierelaterade utredningar publicerade i de nordiska länderna. Generellt har sådana utredningar använts tämligen frekvent inom medie- och kommunikationsvetenskaplig forskning; inte sällan har betänkanden också inkluderat forskningsrapporter. Inom mediehistorisk forskning har äldre SOU:er också använts flitigt, till exempel i skriftserien Etermedierna i Sverige. Torsten Thurén ägnade bland annat i sin bok, *Medier i blåsväder* (1997), ett helt kapitel åt 1960 års Radioutredning. I mycket begränsad omfattning har emellertid den medie- och kommunikationsvetenskaplig forskningen använt de omfattande kommittéarkiv kring alla SOU:er som finns bevarade på Riksarkivet.

[36] Hyvönen, Snickars & Vesterlund 2015, s. 27.

[37] *Dagstidningarnas ekonomiska villkor* (SOU 1965:22), s. 89.

Även under 1980-talet var reklamfrågan aktuell, men då kopplad till ett annat medieformat: *Videoreklamfrågan* (SOU 1982:8). Men framför allt riktades då intresset mot de så kallade nya medierna. Informationsteknologiutredningens betänkande 1981 fick just titeln, *Nya medier* (SOU 1981:45) med fokus på bland annat tele-data och text-tv. Två år senare publicerade Massmediekommittén *Via satellit och kabel* (SOU 1984:65). Det var också vid denna tidpunkt som det blev uppenbart att samtidens datapolitik nog borde betraktas som en form av utvidgad mediepolitik. I slutbetänkandet för Dataeffektutredningen – som 1978 påbörjat arbetet med att utreda datateknikens inverkan på produktion och arbetsliv – stod det därför en hel del om datoriseringens förändringar av både livsstil och medier liksom kultur. Vid en tidpunkt då persondatorer började bli populära aktualiserades alltså de frågeställningar som Alfvén diktat om i *Sagan om den stora datamaskinen* – eller som Dataeffektutredningen framhöll 1984:

> Den tekniska utvecklingen skapar [numera] informationsteknik som kan användas i olika fritidsaktiviteter. De skiljer sig från TV genom att de ger möjligheter till en tvåvägskommunikation. Datatekniken kommer [därför] sannolikt att inverka på människors sätt att tänka [...] dvs hela vår livsstil kan förändras. Effekten av sådana förändringar kan inte observeras på kort sikt utan först i efterhand i ett historiskt perspektiv […]. Kommer en ökad datorisering av samhällets relationer till medborgarna att försvaga en naturlig värdegemenskap eller kommer det ökade informationsutbudet att möjliggöra ökade sociala kontakter?[38]

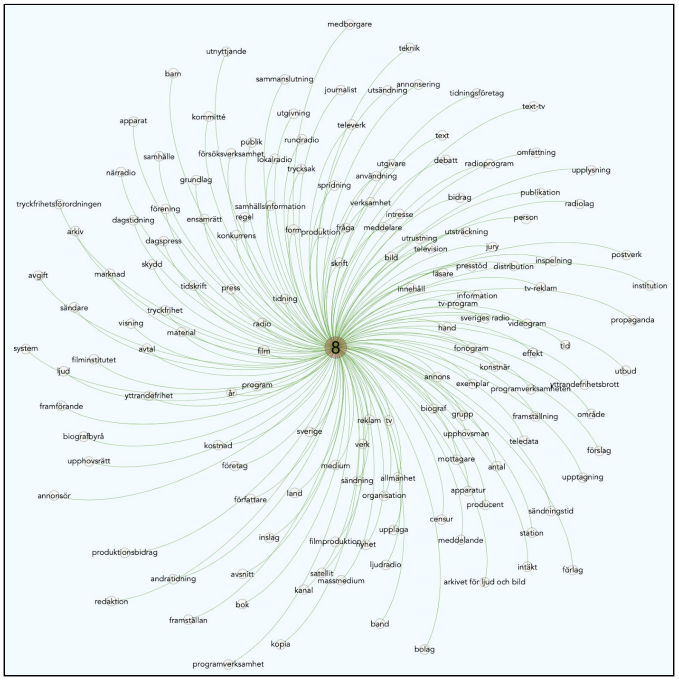


Illustration 7. Illustration i Gephi (Force Atlas 2) för tema 8 (i 50-modellen) som behandlar medier (och relaterade termer) i alla SOU:er mellan 1945–89. Det vill säga, när modellen ställs in att hitta 50 teman i alla SOU:er, så handlar ett av dem om medier. För varje tema kan modellen sortera ut max 200 ord (substantiv och namn). De bör dock kureras och rensas på alltför generiska termer (som exempelvis månader) och förkortningar (kg, kr, nr). Illustrationen innehåller därför 150 ord.

Med tanke på den mängd medieutredningar som var verksamma mellan 1945 till 1989, och den veritabla flod av rapporter, del- och slutbetänkanden som publicerades – i sig en tydlig

[38] *Datorer och arbetslivets förändring* (SOU 1984:20), s. 37.

exponent för hur medier beskrevs, studerades, behandlades och reglerades eller understöddes inom den politiska sfären – så är det knappast förvånande att det existerar ett medietema i SOU-datan. I 50-modellen, som alltså delar in textmassan i alla 3154 SOU:er i femtio teman, så handlar tema 8 tydligt om medier. Temats sannolikhetsvärde är 1,37027961279058 vilket innebär att medierelaterade spörsmål diskuteras i 1,37 procent av alla utredningar. I nätverksvisualiseringen nedan har de ord som är närmast centrum högst sannolikhet att förekomma i medietemat, medan de i periferin är mindre sannolika. Ord som tidning, radio, film, reklam och tv befinner sig följaktligen i centrum, medan termer som tryckfrihetsförordningen, programverksamhet och propaganda är mer perifera.

Som framgår av grafen nedan var följande 15 ord procentuellt mest framträdande i medietema 8: film, tidning, radio, program, sverige, reklam, tv, år, skrift, verk, bild, medium, del, television och press. Film var alltså det medium som flitigast omtalades i medietemat, det förekom i 1,81 procents utsträckning, att jämföra med 0,49 procent för dagspress. Samtidigt är det uppenbart att det inte är helt enkelt att dra en sådan slutsats; tidning är ju det ord som följer efter film. Även om temamodellens beräkningar är mycket noggranna och SOU-datan städats och lemmatiserats, så begriper inte modellen att tidning och press *kan* referera till samma medium. I än högre grad gäller det orden tv och television, som för en mänsklig läsare i princip har samma betydelse. Och dessutom är några av de procentuellt mest förekommande termerna i grafen diverse småord av mindre betydelse – år och del – vilka inte har någon större kontextuell relevans.

Som forskare bör man därför vara uppmärksam på de resultat som temamodellering genererar. Det hindrar inte att flera iakttagelser kan göras; noterbart är till exempel den framträdande roll som reklam har i medietemat. Reklam och medier är tätt lierade under hela perioden. Men om bioreklam och annonser i dagspressen var centrala inkomstkällor för respektive mediebransch var etermedier och framför allt public service något som skulle fredas från reklam, därom var de flesta utredningar ense – åtminstone före 1989; den reklamfinansierade TV3 påbörjade sin sändningar via satellit 1987. Genom reklam finns också en förbindelse till tema 35 (i 50-modellen) som handlar om företagande och näringsliv, ett tema med betydligt högre sannolikhetsvärde (än medier) på 1,80 procent, och därigenom långt mer frekvent förekommande i SOU-datan.

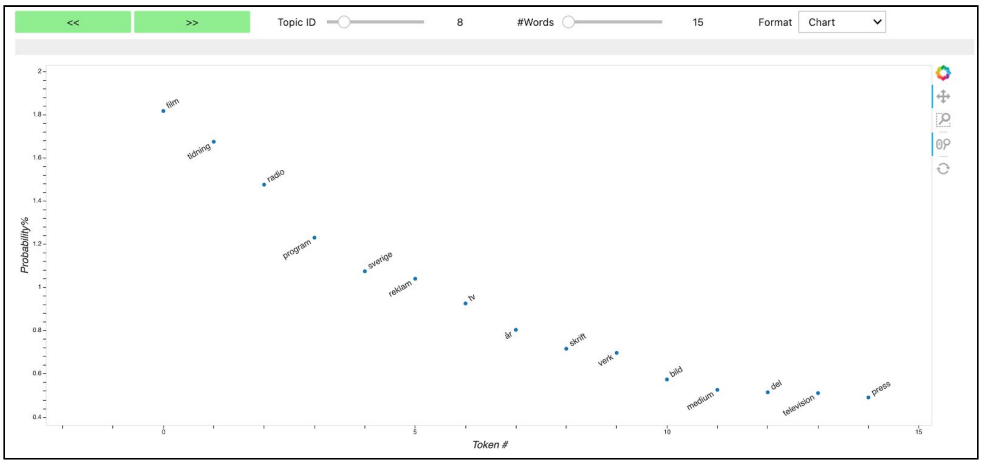


Illustration 8. Skärmdump av temamodellering i Jupyter Lab-miljö för tema 8 om medier i 50-modellen av SOU-datan – med procentuell distribution av de 15 mest förekommande orden.

Som jag tidigare påpekat om miljöteman började gröna tematiker att få sitt genomslag först kring 1970 i SOU-datan. Givet 1960-talets omfattande utredningarna om press, radio och tv

kan man tro att detsamma skulle gälla för medier. Men av stapeldiagrammet nedan framgår att medietemat återfinns under hela perioden 1945–89. Ett flertal utredningar om film kring 1950 gör att temat redan då var vanligt förekommande, därefter är det som procentuellt allra starkast mellan 1965 och 1975, samt mot slutet av 1980-talet. Man kan också notera att medietemat knappt förekommer alls under fyra specifika år (1948, 1957, 1958 och 1986).

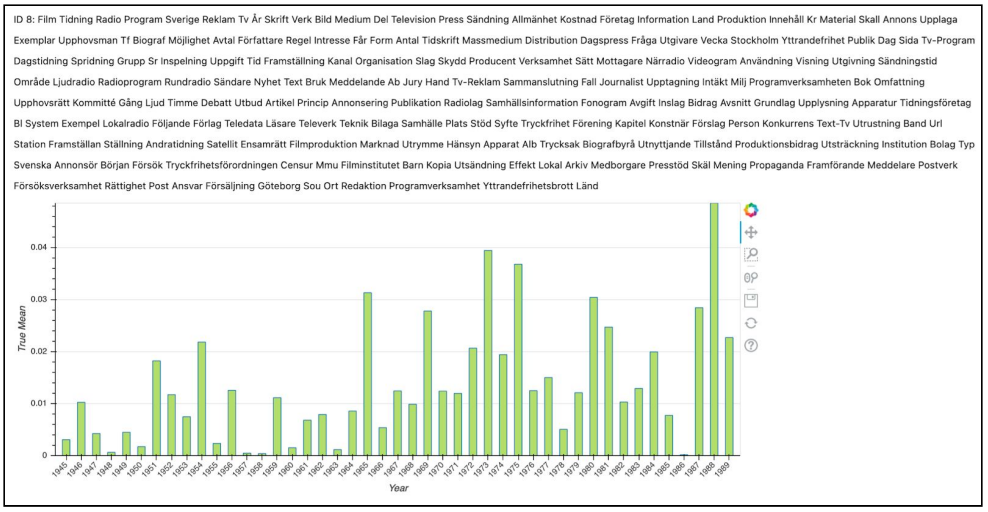


Illustration 9. Skärmdump av stapeldiagram (med 200 relaterade ord) för hur medietema 8 förekommer mellan 1945–89 – med tydliga toppar i början av 1950-talet, vid mitten av 1960- och 1970-talet samt i slutet av 1980-talet.

SOU-datan är som sagt mångfacetterad och studerar man vilka utredningar (och teman) som förbinder medietema 8 i en nätverksvisualisering så dominerar de från perioden 1965–75. De tidigare nämnda radio- och tv-utredningarna liksom dagspressutredningarna SOU 1965:22 och SOU 1974:102 befinner sig alla nära nätverkets centrum. Med ett tröskelvärde på 0,1 – det vill säga, en förbindelse upprättas mellan SOU:er om medietemat återfinns i åtminstone 10 procent av dem – så ingår hela 115 utredningar i detta kluster. Noterbart är att ganska få andra teman är sammanlänkade. Flest förbindelser finns till tema 39 (om utredningsväsende), dessutom figurerar förbindelser till tema 28 (om administration), tema 20 (om arkiv och adb), tema 35 (näringsliv) liksom till tema 14 (om lagstiftning); en publikation från Våldsskildringsutredningen, *Videovåld 2* (SOU 1988:28) utgör exempelvis en sådan länk.

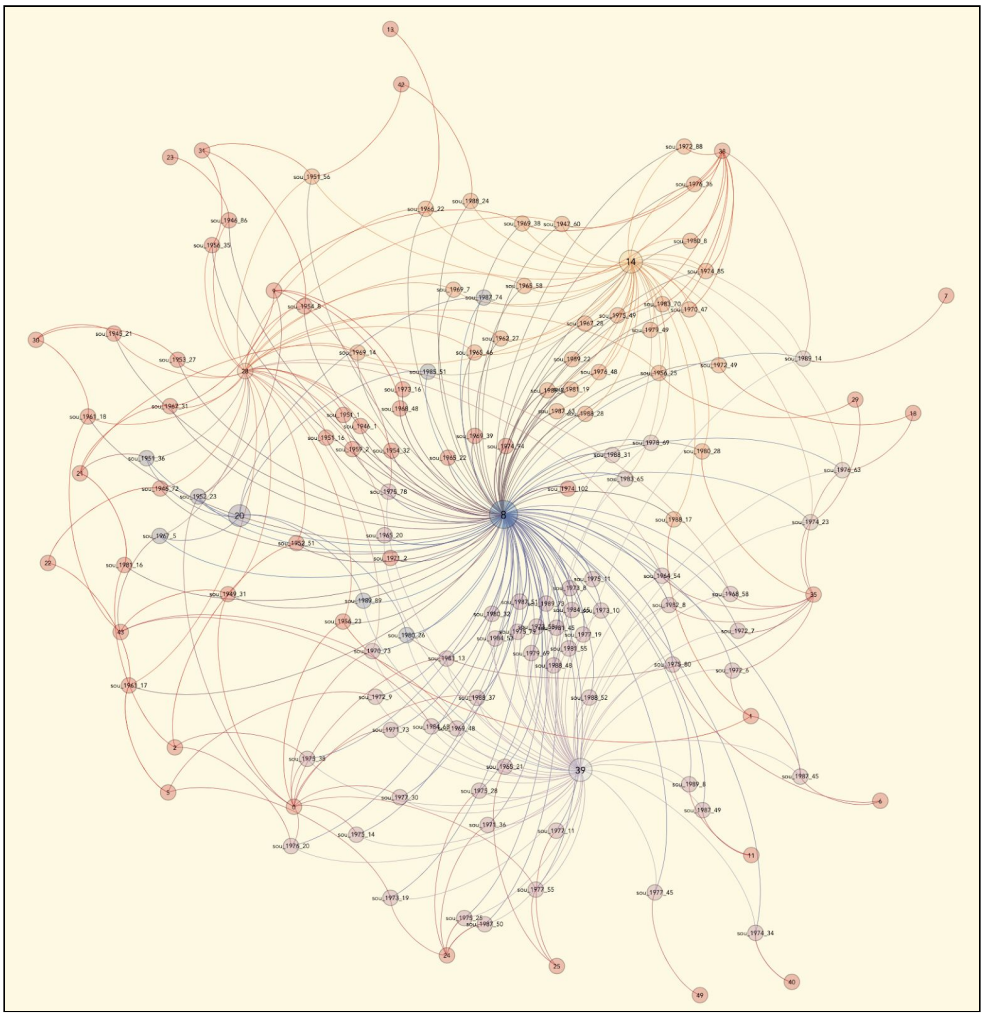
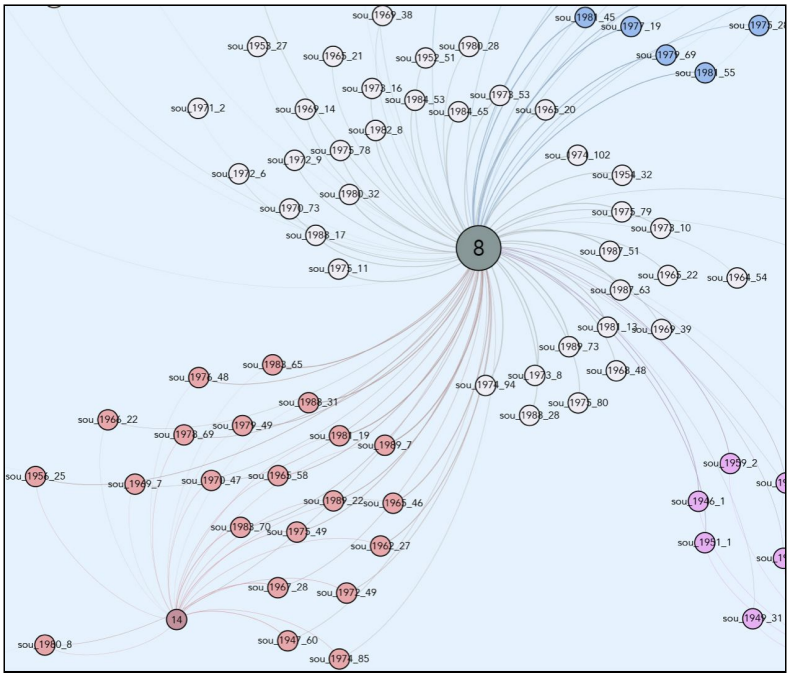


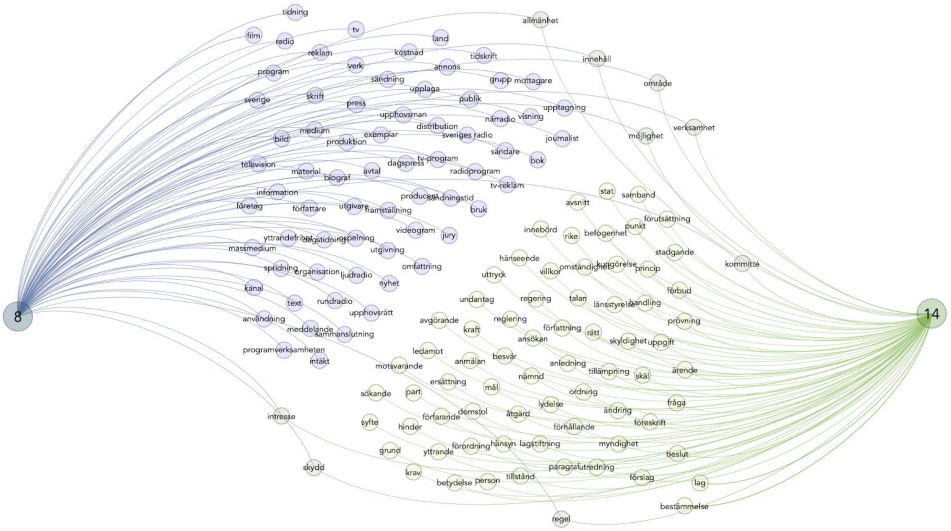
Illustration 10. Medietema 8 i SOU-datan med kopplingar till 115 utredningar. Illustration i Gephi (Force Atlas) för tema 8 (i 50-modellen) med ett tröskelvärde på 0,1.

Nätverksvisualisering av medietema 8 utgör ett sätt för att få korn på hur en temamodell hänger samman med en annan. Ofta är dessa förbindelser självklara, ibland inte. Det går då att upptäcka detaljer som SOU-datan uppenbarar. Att det finns förbindelser mellan hur medier och lagstiftning hänger samman är knappast förvånande. Men en nätverksillustration kan ge insikter i omfattning, tidsdjup samt vilka och hur många utredningar det har handlat om. Ökar man tröskelvärdet till 0,2 – för att säkerställa att medier verkligen utgör tema i SOU-datan – så är 90 utredningar sammanlänkade med medietema 8. Zoomar man in i en visualisering på de specifika relationerna mellan tema 8 och tema 14 (om lagstiftning) som framgår att hela 21 SOU:er innehåller bägge teman – och det över lång tid, från *Förslag till tryckfrihetsförordning* (SOU 1947:60) över *Skydd mot avlyssning* (SOU 1970:47) till *Översyn av radiolagen* (SOU 1981:19).

På sajten lagen.nu – som drivs av den akademiska datorföreningen Lysator vid Linköpings universitet – finns lagar, förordningar, myndighetsföreskrifter, vägledande rättsfall samt äldre lagstiftning i digital form. Lagen.nu innehåller även statliga utredningar, varför det är möjligt att spåra relationer mellan betänkanden, lagförslag och lagstiftning. Utredningen om en översyn av radiolagen, SOU 1981:19, länkar där naturligtvis till Radiolag (1966:755) – men även framåt i tiden till Lag (1981:508) om radiotidningar:

”Radiotidningen skall vara ett radioprogram avsett för synskadade.”[39] Om grafiska analyser av SOU-datan ger besked om att medier och lagstiftning varit sammanlänkade teman inom svenskt utredningsväsende så kan distansläsning ge hänvisningar till var närläsning därefter bör ta vid – exempelvis på lagen.nu. Till yttermera visso vore det förstås mediehistoriskt intressant att utföra temamodellering av *Sveriges rikes lag* för att ytterligare få korn på relationer mellan lagstiftning och medier.

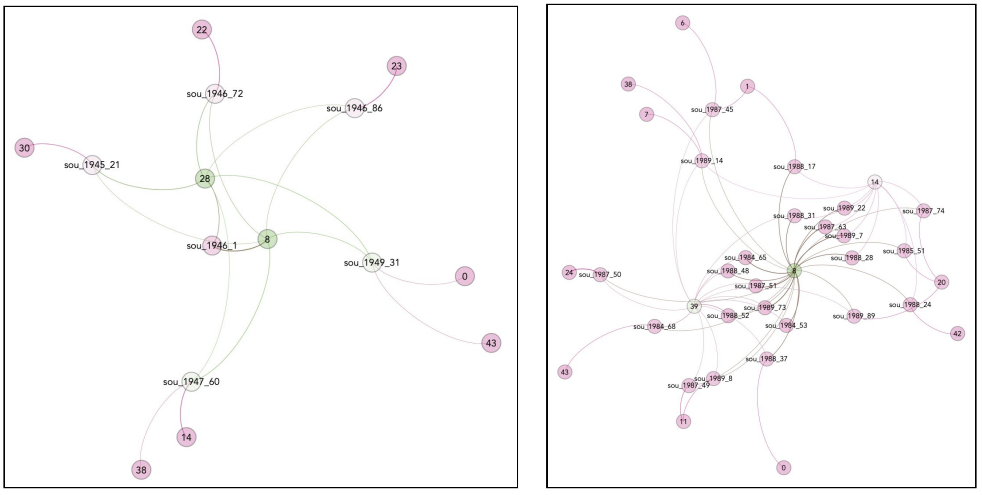




Illustrationer 11 & 12 Förbindelser mellan medietema 8 och tema 14 om lagstiftning (med 21 förbindelselänkar i form av utredningar), liksom nätverk med drygt 100 ordnoder från tema 8 och tema 14 – bägge hämtade från 50-modellen av SOU-datan. Illustrationer i Gephi: Force Atlas Expansion och Force Atlas 2, med ett tröskelvärde på 0,2.

[39] §1., Lag (1981:508) om radiotidningar, https://lagen.nu/1981:508 (2021-03-01).

Med utgångspunkt i stapeldiagrammet kring hur frekvent medietema 8 förekommer är det exempelvis en smal sak att i Gephi visualisera hur dess förbindelser förändrar sig över tid – liksom vilka teman som relaterar till varandra. Delar man in undersökningsperioden i tre delar med 15 års intervall, så innehåller den första perioden, 1945–60, mest förbindelser till (det något intetsägande) tema 28 (om administration). Under perioden 1960 till 1975 förekommer samma admin-tema, liksom främst tema 39 (om utredningsväsende) och tema 14 (om lagstiftning). I princip gäller detsamma för perioden 1975–89. Generellt förändrar sig förbindelser till andra teman i 50-modellen mycket litet. Men eftersom dessa uppdelade nätverksvisualiseringar innehåller färre utredningar är de trots allt av intresse eftersom färre noder ger bättre överblick. Jämför man exempelvis medietema 8 mellan två korta, femåriga tidsintervall i början och slutet av undersökningsperioden, så framträder tydligt de utredningar som förbinder teman med varandra. Att nätverket är större under 1980-talet är inte förvånande; det publicerades trots allt betydligt fler SOU:er då. Som framgår av illustrationerna nedan, utgjorde under perioden 1945–50 den tidigare nämnda SOU 1947:60 den enda länken till lagstiftningstema 14, men också till tema 38 som handlar om brottmål, straff och fängelse. Under den senare perioden 1984–89 är det fler utredningar som förbinder dessa teman, närmare bestämt nio stycken som exempelvis *Vidaresändning av satellitprogram i kabelnät* (SOU 1987:63), *Upphovsrätt* (SOU 1988:31) och *Vidgad etableringsfrihet för nya medier* (SOU 1989:7).



Illustrationer 13 & 14. Förbindelser till medietema 8 – vid två olika tidpunkter, 1945–50 och 1984–89. Illustrationer i Gephi (Force Atlas Expansion) med ett tröskelvärde på 0,1.

Eftersom det finns ett tydligt medietema i 50-modellen kan man sluta sig till att även 100-modellen kommer att innehålla mediala termer – i mer preciserad bemärkelse. 100-modellen delar alltså upp textinnehållet i alla 3154 SOU:er i 100 teman; söker man på ordet medium i Jupyter Lab-miljön så resulterar det i fem temamodeller: tema 13 (om reklam och marknadsföring), tema 39 (om bibliotek och museum), tema 62 (om radio och tv), tema 69 (om information, arkiv och adb), och tema 93 (om tidningar och böcker). Dessutom ingår tema 59 om film i samma konstellation, även om inte det specifika ordet medium figurerar just där. Till skillnad från 50-modellen delar alltså 100-modellen in SOU-datan i mer specifika kategorier. Filmtemat är det tydligaste exemplet, men även tema 62 om etermedier och public service har tydligt urskiljbara medietermer som program, kanal och publik liksom kabel-tv, tryckfrihetsmål och underhållning – den sistnämnda figurerar långt ute i temats periferi. Tema 62 återfinns redan under 1950-talet, men växer

fram i tilltagande styrka efter 1960. Man kan lägga märke till att det inte är television som är temats vanligaste ord – utan radio.

Men vad man främst bör lägga märke till är att dessa sex medieteman inte är kopplade till varandra. Utredningar om dagspress är exempelvis inte alls länkade till filmutredningar, och bara i ett enda fall – genom utredningen *Massmediekoncentration lagförslag och motiv* (SOU 1980:28) – delar de tematik med temat kring radio och tv. Inom kommittéväsendet utreddes medier var för sig, bara i enstaka fall fanns beröringspunkter sinsemellan mediebetänkanden. En nätverksillustration visualiserar tydligt denna brist på konvergens; till och med vid ett tämligen lågt tröskelvärde på 0,1 – det vill säga, en tio-procentig förbindelse mellan teman – så är det bara fyra SOU:er som knyter samman radio och tv med film. SOU-datan bekräftar här den bild som tidigare mediehistorisk forskning framhållit angående det stuprörstänkande som präglade svenskt kommittéväsende, vilket även dåtida utredare kritiserade. När Harry Schein 1971 arbetade deltid som sakkunnig på utbildningsdepartementet – han skulle året därpå publicera debattboken, *Inför en ny mediapolitik: kabeltelevision – bildskval eller kommunikation?* – skickade han ett PM till Ingvar Carlsson (som då var landets utbildningsminister): ”Det finns en filmutredning, en litteraturutredning, en massmediautredning, presstödsutredning, etc. Dessutom håller jag på med kabel-TV. Är det riktigt att på detta sätt isolera frågorna från varandra? Krävs det inte en sammanhängande informations- och kommunikationspolitik?”[40]

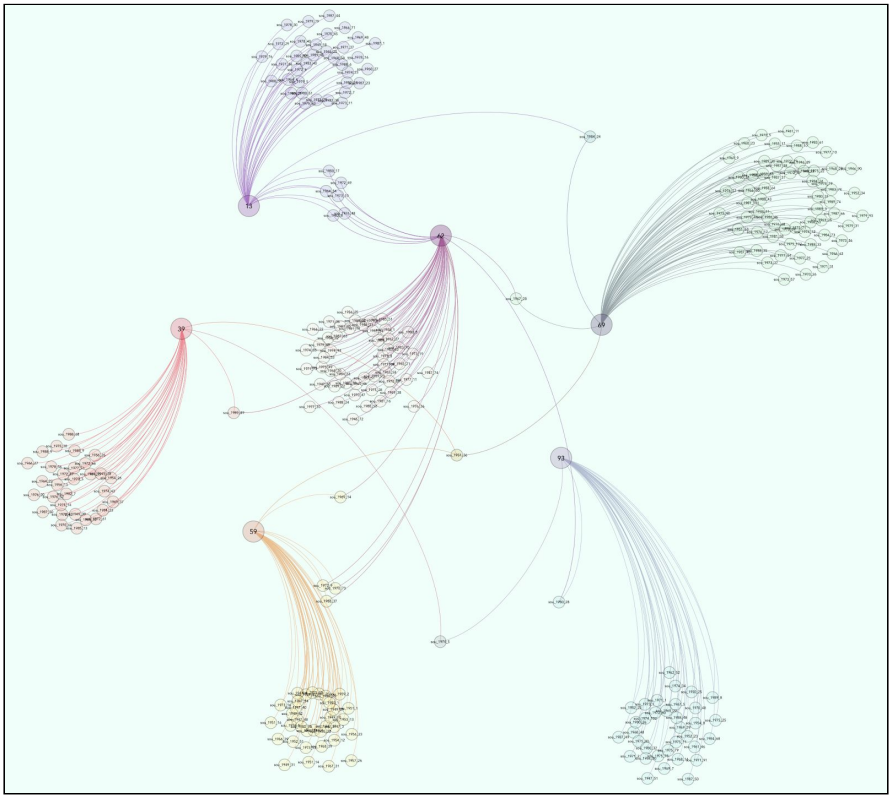


Illustration 15. Förbindelser mellan sex olika medieteman – tema 13 (om reklam och marknadsföring), tema 39 (om bibliotek och museum), tema 62 (om radio och tv), tema 69 (om information, arkiv och adb), tema 93 (om tidningar och böcker) och tema 59 om film. Temanoderna har manuellt dragits närmare varandra; nätverk i Gephi (Force Atlas 2) baserat på 100-mallet-lda-modellen (tröskelvärde 0,1) av SOU-datan 1945–89.

[40] Harry Schein, ”Lösa stolpar kring tråd-TV”, utkast, volym 4:51, Harry Scheins personarkiv, Arbetarrörelsen arkiv och bibliotek. För en vidare diskussion, se Hyvönen, Snickars & Vesterlund 2015, s. 36–37, samt Norén & Snickars 2017.

Om tidigare forskning uppmärksammat att medier behandlades i sinsemellan avskilda statliga utredningar, så har den likväl inte kunnat fastlägga vilket medium som egentligen ägnades mest intresse. Mellan 1945 och 1989 publicerades det ett femtontal utredningar om radio och tv, ungefär lika många om film och något färre om dagspress. Det ligger nära till hands att tro att det var för perioden nya kommunikationsformer, som television eller data-medier, som främst uppmärksammades inom hela utredningsväsendet. Men SOU-datan ger tydligt besked om att film var periodens mest omskrivna medium – och detta trots att mediet 1945 var allt annat än nytt. I 50-modellen var film som påpekats procentuellt allra mest framträdande i medietema nummer 8, och i 100-modellen var filmen det enda medium som textalgoritmerna urskiljde som en egen, separat medieform. I 200-modellen blir denna specificering än tydligare; här framträder med precis vokabulär hur film blir till periodens ledmedium – åtminstone för det svenska utredningsväsendet.

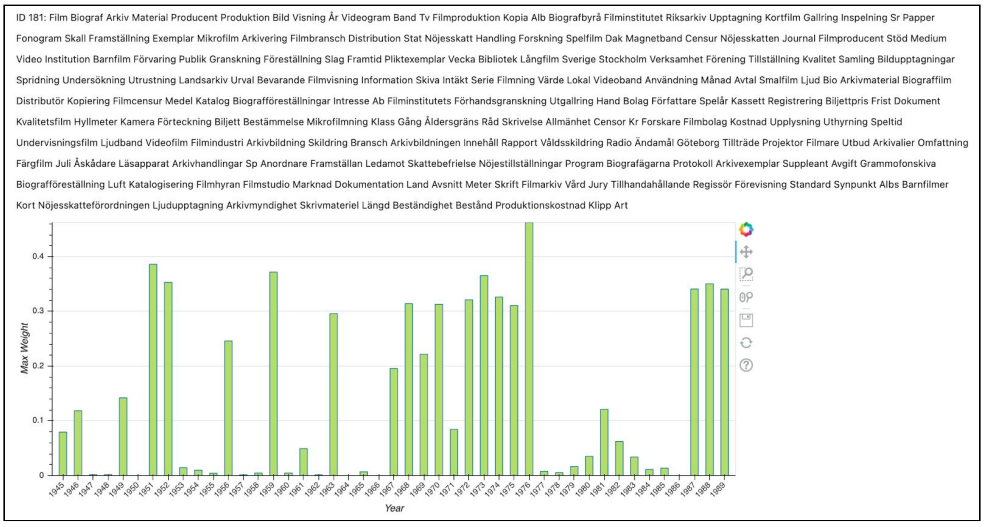
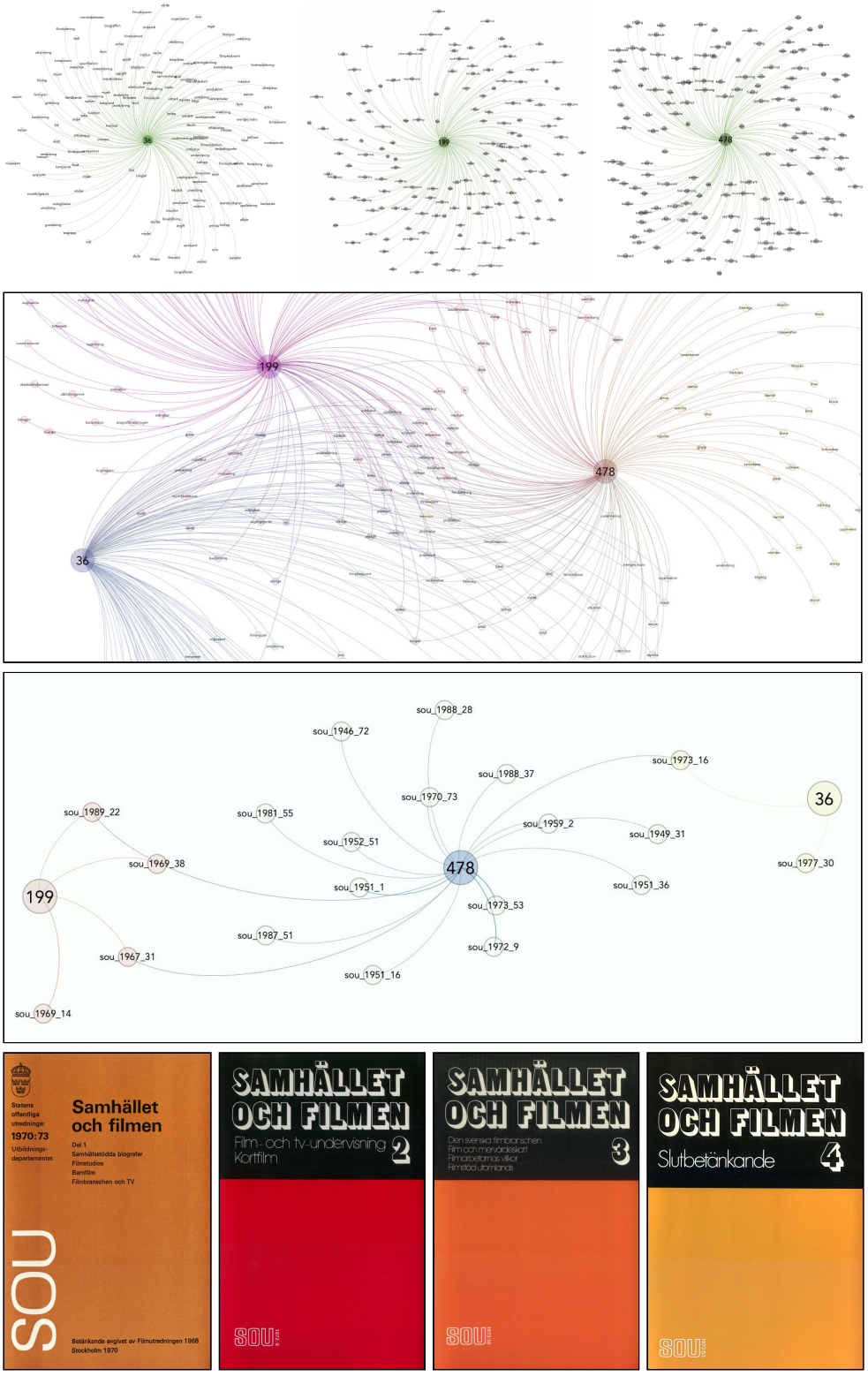


Illustration 16. Skärmdump av stapeldiagram – med de 200 mest vanliga orden – för filmtema 181 samt hur det förekommer mellan 1945–89 i 200-modellen av SOU-datan.

Att filmen var av vikt för statsmakterna är förstås ingen nyhet; akademiskt påpekades det redan av Oliver Krems i sin avhandling, *Schwedische Filmpolitik der sechziger bis frühen achtziger Jahre* (1988).[41] Scheins filmreform 1963 var därtill ett mediepolitiskt snilledrag där den gamla nöjesskatten slopades och ersattes med en avgift om tio procent på varje såld biobiljett som fonderades i den nya stiftelsen Svenska filminstitutet för att stödja kvalitetsfilm; det var alltså Hollywood som finansierade auteuren Ingmar Bergman. Termen kvalitetsfilm förekommer rentav i SOU-datan; termen var den 110:e vanligaste i filmtema 181 (i 200-modellen). Att film var den mest diskuterade medieformen 1945–89 är faktiskt ett av den här artikelns mest tydliga forskningsresultat. Av stapeldiagrammet ovan framgår att film var ett medium som ägnades uppmärksamhet under hela perioden. Tydligast är det för perioden 1967 till 1976, men filmpolitiken formerades redan under 1950-talet. Det fanns både en 1949 års filmkommitté – som mest ägnade sig åt filmcensur, men även åt barnfilm i *Barn och film* (SOU 1952:51). Denna filmkommitté ska inte blandas ihop med 1950 års filmutredning som ungefär samtidigt publicerade, *Statligt stöd åt svensk filmproduktion* (SOU 1951:1), vilken i sin tur följdes upp av 1957 års filmstödsutredning, som bland annat resulterade i *Filmstöd och biografnöjesskatt* (SOU 1959:2).

[41] Oliver Krems, *Schwedische Filmpolitik der sechziger bis frühen achtziger Jahre*, Bochum 1988.



Illustrationer 17–25. A. tre temamodeller om film (med 150 ord vardera) i 500-gensim-lda-modellen i SOU-datan från 1945–89: 36 (om filmbransch), 199 (om filmcensur) och 478 (om filmproduktion) med Gephi Force Atlas 2 Expansion; B. nätverk över hur orden i dessa teman samförekommer (Gephi Force Atlas 2); C. SOU-förbindelser mellan samma teman (Gephi Force Atlas 2 med tröskelvärde 0,11); D. omslag till fyra filmutredningar från 1970-talets början – vilka alla är länkade till filmtema 478.

Även när man använder 500-modellen är film ett tacksamt medium att studera. I den modellen återfinns ordet film i hela 27 olika teman. Tre av dem – nummer 36, 199 och 478 – handlar mer explicit om rörlig bild. Om 200-modellen hade ett enda tema om film, så återfinns det alltså tre tydliga filmteman i 500-modellen, men med sinsemellan olika innebörd, en sorts differentierad filmdiskurs om man så vill. Som påpekats tidigare går det i så måtto att få korn på hur ett tema fördelar och förgrenar sig i de större modellerna (som innehåller flera hundra teman). Som framgår av temamodellerna nedan – där jag istället för mallet-modeller använt gensim-lda-modeller med gott resultat – behandlar tema 36 hela filmbranschen, medan tema 199 mer handlar om filmcensur. Tema 478 är i sin tur ägnat Svenska filminstitutet med fokus på filmproduktion. Om film i alla tre visualiseringarna koncentriskt befinner sig i mitten, följs det av termer som kortfilm, biografbyrå eller filmskapare (för tema 36), mediepåverkan, kontroll och oanständighet (för tema 199) och barnfilm, produktion och filmskapare (för tema 478).

I ett stapeldiagram i Jupyter Lab-miljön framgår att filmtema 478 var det enda som återkom under hela undersökningsperioden, med toppar de år då specifika filmutredningar publicerades, framför allt i början av 1970-talet (då även tema 36 och 199 var som starkast). Det var då som den så kallade Filmutredningen 1968 presenterade ett flertal betänkanden – och att de figurerar i temamodellerna om film är om inte annat ett tecken på att även gensim-algoritmerna fungerar som de ska. Filmutredningen 1968 hade tillsatts för att undersöka hur Scheins filmreform kunde vidareutvecklas; ”statsrådet Palme erinrade ... om de krav på en vidareutbyggnad av reformen som rests i den allmänna debatten”, varför utredningen främst kom att ägna sig åt ”en kartläggning och bedömning av filmbranschens ekonomiska struktur”.[42]

Som i fallet med medieteman i 100-modellen var förbindelser filmteman emellan inte heller speciellt framträdande i 500-modellen. Men man kan notera att fyra SOU:er utgör länk mellan temat om filmproduktion (478) och temat om filmcensur (199). Sätter man fokus på det senare går det att spåra förbindelser mellan filmmediet och diskurser kring yttrandefrihet, osedlighet och pornografi – kort sagt, mot tämligen tidsbundna kontexter där filmen fortsatt ansågs vara ett farligt medium. Filmhistorisk forskning har i detalj utrett turerna kring filmens samhälleligt föränderliga funktion under 1960-talet. Produktion av kvalitetsfilm med estetisk finess eller samhällskritiskt uppsåt, samt en allt mer radikal tidsanda, gjorde företeelser som filmcensur obsolet menade vissa. Andra hävdade att spelfilmer som Vilgot Sjömans ungdomsfilm *491* – vilken 1963 enligt Statens biografbyrån var ”förråande och sakligt upphetsande” och därför (till en början) totalförbjöds – var moraliskt förkastliga och sedeslösa; ibland görs gällande att partiet Kristen Demokratisk Samling delvis bildades i opposition till just den filmen.[43] 1964 tillsattes därför en statlig utredning, Filmcensurutredningen, för att ”utreda den statliga filmcensurens ställning och därmed sammanhängande frågor”. Ur direktiven framgick att den samtida debatten om film spelat roll:

> Den under senare tid pågående intensiva diskussionen rörande filmcensuren har visat, att en betydande oklarhet råder om censur verkligen är motiverad för vuxen publik och hur denna verksamhet i så fall lämpligen bör vara utformad. Bland dem som anser att censuren bör bibehållas godtages i allmänhet den i nuvarande regler för filmgranskningen fastslagna målsättningen, att ingrepp mot film ... inte bör

[42] *Samhället och filmen, del 1* (SOU 1970:73), s. 7.

[43] För en vidare diskussion, se Anders Åberg, *Tabu: filmaren Vilgot Sjöman*, Lund 2001.

> grundas på andra skäl än sådana som syftar till att förebygga brott eller verka som mentalhygieniskt skydd. Däremot är meningarna mycket delade i frågorna om hur olika åldersgrupper påverkas av film och huru påverkan från film skiljer sig från påverkan av liknande slag från andra massmedia.[44]

Chefen för Svenska filminstitutet, Harry Schein – som länge varit kritisk till filmcensurens verksamhet – tillsatte och finansierade därtill ett slags skuggutredning genom den så kallade filmforskningsgruppen. Under andra halvan av 1960-talet var den verksam på filminstitutet och undersökte filmens mediepåverkan utifrån olika vetenskapliga discipliners horisont. Filmcensurutredningen och filmforskningsgruppen kom dock framgent att samarbeta; 1967 publicerades ett första delbetänkande: *Filmens inflytande på sin publik* (SOU 1967:31). Som Per Vesterlund framhållit var de resultat som filminstitutets filmforskningsgrupp torgförde ”otvetydigt. Det gick inte att generellt påvisa några som helst effekter av filmen på dess åskådare.”[45]

Illustrationer 26–28. Film, mediepåverkan och censur var under slutet av 1960-talet omdiskuterade frågeställningar – så exempelvis i dessa tre SOU:er från 1967 och 1969.

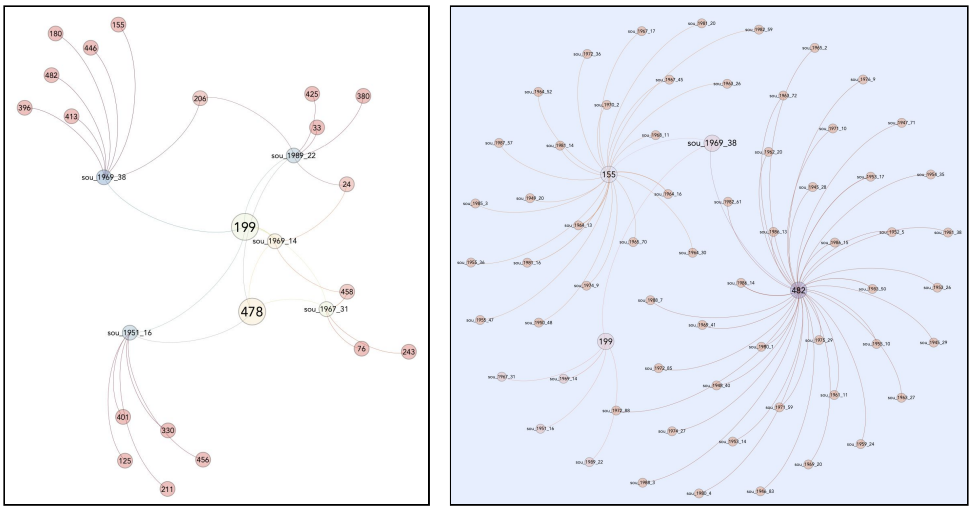
I sitt slutbetänkande 1969 kom även Filmcensurutredningen till snarlika slutsatser, och ansåg följaktligen att åtminstone vuxencensuren för film borde avskaffas (vilket inte blev fallet; det skulle dröja 42 år).[46] Vid sidan av dessa två utredningar arbetade även Kommittén för lagstiftningen om yttrande- och tryckfrihet med koppling till Justitiedepartementet. Deras uppdrag bestod i att se över ”bestämmelserna om sårande av tukt och sedlighet samt brott mot trosfrid”, liksom huruvida sådana framställningar i exempelvis film, radio och tv borde straffbeläggas, alternativt att begränsa rätten att sprida ”sedlighetssårande alster.” 1969 författade kommittén ett delbetänkande kring ”uppvigling”, och samma år publicerades även, *Yttrandefrihetens gränser. Sårande av tukt och sedlighet* (SOU 1969:38). Kommittén föreslog där att staten borde kunna ingripa gentemot ”förråande framställningar ... främst de grövsta våldsskildringarna i olika media [som] anses vara kränkande för den allmänna känslan för anständighet”, vilket även inkluderade ”allvarliga

[44] *Filmen – censur och ansvar* (SOU 1969:14), s. 3, 14

[45] Per Vesterlund, ”Samhället i medierna? Harry Schein och medieforskningen”, *Massmedieproblem. Mediestudiets formering*, Mats Hyvönen, Pelle Snickars & Per Vesterlund (red.), Lund 2015, s. 360.

[46] SOU 1969:14, s. 130.

angrepp mot den kränktes rätt till religionsfrihet”. Två kommittéledamöter var kritiska mot dessa skrivningar och reserverade sig. Det är knappast rimligt menade de, ”att personers åsikter och föreställningar i religiösa frågor skall ha ett speciellt skydd och att deras irritation eller möjligen lidande skall förebyggas med brottsbalksbestämmelser”.[47]



Illustrationer 29 & 30. Temanätverk kring noden 199 om filmcensur i SOU-datan 1945–89 – med ett lågt tröskelvärde på 0,05, visualiserat i Gephi Force Atlas Expansion, liksom temanätverk mellan noder 199 (om filmcensur), 155 (om religion och kyrka) och 482 (om brott och straff) – med ett lågt tröskelvärde på 0,05, visualiserat i Gephi Fruchterman Reingold Contraction.

Tema 199 om filmcensur inbegriper alla dessa utredningar. I en nätverksillustration befinner sig därför tukt och sedlighetskommitténs SOU 1969:38 precis till vänster om huvudnoden 199, medan Filmcensurutredningens SOU 1969:14 ligger strax intill temats centrum, och filmforskningsgruppens SOU 1967:31 något till höger, med koppling till filmtema 478 (om filmproduktion). Om man i Jupyter Lab-miljön plottar alla noder med koppling till filmtema 199 över hela undersökningsperioden, så innefattar det också äldre betänkanden som *Filmcensuren* (SOU 1951:16) – med intilliggande teman som 401 (om barn och föräldrar) och 211 (om äktenskap) – liksom senare utredningar som *Censurlagen – en modernisering av biografförordningen* (SOU 1989:22). Det mest intressanta med nätverksillustrationen är emellertid de sex teman som hänger samman med tukt och sedlighetskommitténs SOU 1969:38. Tre av dem behandlar juridiska frågor: 180 (om bestämmelser och böter), 396 (om lagar och föreskrifter) och 446 (om sociala ersättningar och hushåll). Temamodell 155 handlar i sin tur om religiösa frågor kring termer som kyrka, statskyrka, religionsfrihet, kyrkomöte, präst och skilsmässa – medan det sista sammanlänkade temat 482 behandlar brottslighet, fängelse, domstol och straffskalor.

Med temamodellering går det alltså att få syn på hur filmmediet förknippades med såväl religiösa sedlighetsfrågor liksom med rättsliga frågeställningar, samt exakt vilka betänkanden som var mest centrala. De blir än tydligare om man visualiserar förbindelserna mellan tema 199, 155 och 482. Vid ett tröskelvärde på 0,05 är det enbart SOU 1969:32 om yttrandefrihetens gränser som binder samman dessa tre temamodeller. Men det blir också uppenbart vilka andra utredningar – hela 63 stycken – som vidgar diskursen kring filmcensur i SOUniversum (för att använda Kullenbergs term) – som till exempel betänkanden från Fängelsestraffkommittén (SOU 1986:14) eller Strafflagberedningen (SOU

[47] *Yttrandefrihetens gränser. Sårande av tukt och sedlighet* (SOU 1969:38), s. 3, 85, 86, 88.

1953:17), för att inte tala om mängder av utredningar kopplade till ”1958 års utredning kyrka-stat”, bland andra *Historisk översikt Kyrkobegrepp* (SOU 1964:16). Gör man till sist en cirkulär nätverksvisualisering av de ord som förbinder dessa tre teman blir kopplingen mellan film, brott och kyrka än tydligare. Invid respektive temanod befinner sig ord som pornografi, censor och publik (199), dagsbot, frigivning och livstid (482) samt domkapitel, samfund och kristendom. I centrum hittar man de 40 begrepp som dessa teman har gemensamt: våld, samhälle, ordning, regel, problem och bestämmelse – en minst sagt konservativ gemenskap för mediet film.

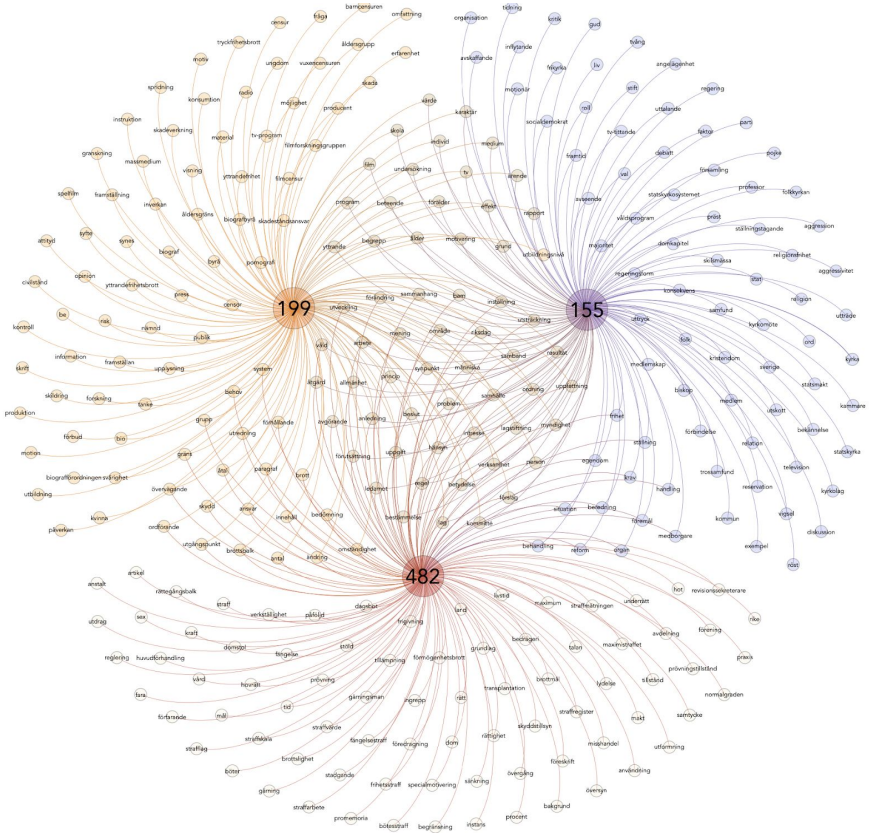


Illustration 31. Nätverk i Gephi (Fruchterman Reingold) med drygt 300 ordnoder ur tre olika teman – 199 (om filmcensur), 155 (om religion och kyrka) och 482 (om brott och straff) – i 500-modellen av alla SOU:er mellan 1945–89.

Avslutning – i mediearkivet

Att temamodellering är ett sätt att få syn på förbindelser mellan diskuser i SOUniversum torde härvidlag vara uppenbart. Genom att modellera SOU-datan kan samförekomster lokaliseras, också mellan diskurser som vid en första anblick inte alls verkar ha med varandra att göra – som mellan filmcensur, brottslighet och kristendom. Genom modellering kan man också få fram i vilka specifika utredningar som ett tema är allra mest framträdande. Gemensamt för alla mina modeller är att de kan lista de betänkanden som procentuellt starkast innehåller ett tema. I 50-mallet-temamodellen är till exempel följande fem SOU:er vanligast förekommande i medietema 8: *Ljud och bild för eftervärlden* (SOU 1987:51), *TV-politiken* (SOU 1989:73), *Svensk press* (SOU 1974:102), *Bevara ljud och bild* (SOU 1974:94) och *Radio i utveckling* (SOU 1973:8). Noterbart är att det rör sig om utredningar kring medier som press, radio och tv – men inte film, som ju annars var det

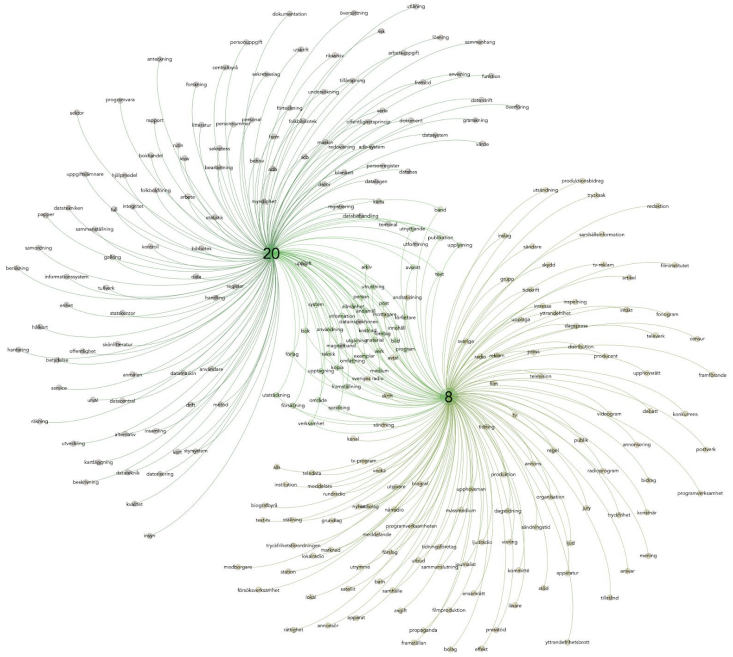
medium som var mest omtalad i SOU-datan. Dessutom är flera av dessa betänkanden publicerade mot slutet av den period som jag intresserat mig för. Men vad som är mest slående är att två av dessa fem betänkanden handlar om hur medier skulle bevaras. Om alla digitaliserade SOU:er betraktas som ett gigantiskt mediearkiv, så handlar detta textarkiv i medierelaterade frågor i stor utsträckning om just bevarande, en metamedial diskurs om man så vill. Flera medieutredningar var förstås inriktade på framtidsfrågor – framför allt hur medier skulle stödjas eller regleras – men arkiveringen av samma medieformer var alltså något som staten också återkommande intresserade sig för.

Illustration 32–36. I dessa fem betänkanden förekommer medietema 8 (i 50-modellen) som starkast i hela SOUniversum; noterbart är att två av dem behandlar hur medier ska sparas och arkiveras.

Jag har i ett annat sammanhang skrivit om etableringen av myndigheten Arkivet för ljud och bild, inrättat 1978/79 som ett statligt mediearkiv (med fokus på audiovisuella medier) och framgent en central forskningsinfrastruktur för det nationella mediestudiets framväxt.[48] Upprinnelsen till denna nya myndighet var arbetet inom Dataarkiveringskommittén (DAK), som påbörjade sitt arbete 1968 med uppgiften att ”utreda frågor om arkivering [...] av fonogram, den automatiska databehandlingens media och övriga lagringsmedia tillhörande informationsbehandlingens område”.[49] Här finns med andra ord tydliga förbindelser mellan medier, arkiv och adb, ett slags digital mediekonvergens *avant la lettre*. I 50-modellen är det därför inte förvånande att medietema 8 har flera förbindelser med tema 20, som handlar om arkiv, information och adb. I nätverksillustrationen nedan mellan dessa teman återfinns hela 38 gemensamma (ljusgröna) ordnoder: medium, arkiv, teknik, spridning, information och program. Samförekomsten mellan medier, arkiv och adb är också framträdande i andra temamodeller. I 500-mallet-modellen går det exempelvis att spåra en mängd förbindelser mellan medieteman, arkiv och automatisk databehandling. Där finns rentav ett mediearkivariskt tema (nummer 343) med termer som arkiv, film, material, handling, bibliotek, mikrofilm och upptagning. I en nätverksillustration av sju relaterade teman i 500-modellen – tema 9 (om etermedia), tema 31 (om dagspress), tema 181 (om film), tema 248 (om reklam), tema 343 (om mediearkiv), tema 394 (om adb) och tema 492 (om massmedieforskning) – framträder en tät väv av fler än 500 ordnoder. Om ett forskningsresultat i denna artikel är att film diskuterades flitigt i SOUniversum, så är ett annat att medier ofta förknippades med hur de skulle bevaras med hjälp av nya lagringsmedier.

[48] *Bevara ljud och bild* (SOU 1974:94), s. 3.

[49] Pelle Snickars, ”Mediestudiets infrastruktur. Om etableringen av Arkivet för ljud och bild”, *Massmedieproblem. Mediestudiets formering*, Mats Hyvönen, Pelle Snickars & Per Vesterlund (red.), Lund 2015.



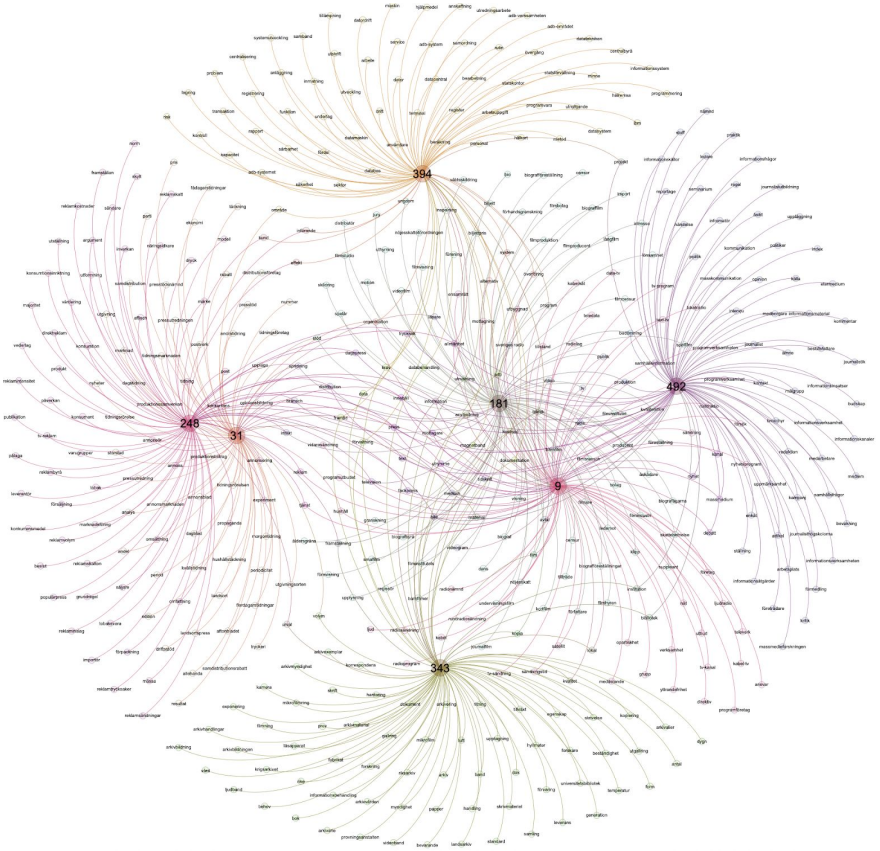


Illustration 37 & 38. Nätverk i Gephi (Fruchterman Reingold) med drygt 260 ordnoder ur två olika teman – 8 (om medier) och 20 (om arkiv och adb) – i 50-modellen av alla SOU:er mellan 1945–89. Nätverk av drygt 500 ordnoder i Gephi (Fruchterman Reingold Contraction) av sju teman i skärningspunkten mellan medier, arkiv och adb i 500-mallet-modellen: tema 9 (om etermedia), tema 31 (om dagspress), tema 181 (om film), tema 248 (om annonsering), tema 343 (om mediearkiv), tema 394 (om adb) och tema 492 (om massmedieforskning).

Den forskningsberedningen som Hannes Alfvén ingick i på 1960-talet hade som påtalats till uppgift att bland annat införa datateknik och automatisk databehandling i Sverige. Om begreppet databehandling ursprungligen handlade om datorn som ett slags bearbetande maskineri så var det (för somliga) redan på 1960-talet uppenbart att dator och medier framgent skulle konvergera. Sådana förbindelser går att spåra i SOU-datan, och i *Sagan om den stora datamaskinen* finns också flera tydliga mediala kopplingar. Ett mycket viktigt framsteg, enligt denna datadystopi, var konstruktionen av teletotal, ”en kombination av automattelefon, radio och tv” som gjorde att människor kunde arbeta hemmavid eftersom teletotalen ”överförde bild och tal till den centrala datan”. När så ”datorna utvecklats vidare och teletotaler massfabricerats i tillräckligt antal, kunde en genomgripande rationalisering av samhället påbörjas”. Genom teletotal kunde man när som helst ”etablera kontakt med en central biblioteksdata och från henne omedelbart få all kunskap man behövde ... Det var alltså onödigt att belasta sin hjärna med en massa minneskunskaper, som ju i realiteten förvarades säkrare och tillförlitligare i en datas minnesenhet”.[50]

Ska man tro Alfvén handlade *Sagan om den stora datamaskinen* om att översätta ”forskningsberedningens chiffer till klartext”. Ordet chiffer kommer från franskans *chiffre*, som betyder siffra, och arabiskans *sifr* som betyder noll. Ett chiffer kan vara en följd av numeriska tecken i hemlig skrift eller ett meddelande vars språkliga form är så förvriden att dess semantiska innebörd förblir obegriplig för en utomstående. Temamodellering är en statistisk metod som behandlar ord och deras samförekomster som siffror; metoden kan därtill framstå som ett obegripligt chiffer, där humanister oförstående lämnas därhän. Men i den här artikeln har jag försökt att förklara vad temamodellering är, hur metoden kan användas i historisk forskning och i relation till närläsningsmetoder, samt hur statistisk analys av distribution av ord i form av sannolikhetsfördelning över alla termer i en mycket omfattande textmassa kan generera intressanta resultat. Mina resonemang har ställvis varit schematiska och kortfattade; resultat har både beskrivits och bokstavligen förevisats visuellt. Min poäng har varit att både visa och lyfta fram att SOU-datan är rik; förbindelser går att spåra i väldigt många olika riktningar. Som forskare måste man emellertid lära sig att arbeta med datan; att ladda olika modeller, att göra körningar med varierande ingångsvärden, att ändra parametrar och ofta kurera utfall på ett sätt som skiljer sig från traditionell historievetenskaplig praktik. Det är därför till stor gagn om man har förkunskaper kring de teman man önskar titta närmare på. Jag har därför koncentrerat mig på temamodeller kopplade till medier. Ambitionen har varit att visa hur temamodellering kan ge nya mediehistoriska insikter, men också mer generellt ge inblickar i hur en metodologiskt färgad analys av ett storskaligt dataset kan te sig. För som mina korta exempel kring kyrka, miljö, invandring, medier, film, adb och mediearkiv antyder så återfinns temamodeller i SOU-datan om de flesta områden som utredningskommittéer ägnat sig åt. Och det är som bekant väldigt, väldigt många.

[50] Johannesson / Alfvén 1966, s. 58,59, 62, 63.