

Magtens Spind



En digitalhistorisk analyse af relationen mellem Danmark og
Frankrig
1131-1265



AALBORG UNIVERSITET

Kandidatspeciale: 2025

Laura Aagaard &

Frederik Søndergaard

Vejleder: Torben K. Nielsen

Historie: Middelalderspecialisering

Afleveringsdato: 1. juni 2025

Resume

I vores speciale undersøger vi relationen mellem Danmark og Frankrig i højmiddelalderen, på baggrund af en digitalhistorisk tilgang. Vi kaster lys over relationen gennem eksplorative analyser af vores selvskabte database, som indeholder 2002 digitaliserede breve fra *Danmarks Riges Breve* fra perioden 1131-1265. Vores undersøgelse viser en relation, der i flere tilfælde fremstår fjern mellem de to nationer, men sammenflettet gennem pavedømmet og den katolske kirke. Dette indikerer en svagt funderet relation mellem de to stormagter, som primært består i ægteskabet mellem dronning Ingeborg og Filip. 2. Undersøgelsen belyser også vigtige refleksioner til arbejdet med digitale metoder.

Abstract

This project explores the relationship between Denmark and France in the High Middle Ages, through an explorative digital analysis. The background for this study stands on a desire to open up the field in a digital way and focus on structural elements of communication. The purpose is to provide an understanding of a relationship that is often understood among scholars, to culminate in the marriage between the danish princess Ingeborg and the french king Filip 2. Our data is based on a self-made database of 2002 medieval letters, from the period 1131 to 1265. The analysis includes network visualizations, geographical mapping and text analysis, created through RStudio and Python. Our results showed the relationship was not close as we expected, it did primarily rely on a structural element in the form of the catholic church. The papacy, especially Innocent III and IV, stands out as the main distributors in the letter communication. The analysis also showed the positioning of Denmark, as mainly recipient of letters.

This study demonstrates the potential of using digital methods in historical research, while emphasizing methodological reflection, transparency and reproducibility. Simultaneously showing the complexity of combining a medieval text corpus with digital history.

Forord

Vi ønsker at sende en varm tak til vores vejleder, Torben K. Nielsen, for højt humør og kyndig vejledning gennem de sidste to år. Vi siger tak for, at du har taget imod os med åbne arme hver gang vi mødte uanmeldt op på dit kontor, forvirrede og uforstående for det hele, tak for du altid tog tiden til at snakke med os. Dertil også en stor tak for at lytte med på de mange ideer vi har smidt op i luften, nogle bedre end andre.

Tak til os! Dette er vores sidste projekt på Aalborg Universitet og samtidig det ottende projekt i rækken vi udarbejder sammen. En rejse ingen af os havde forventet da vi tilmeldte os historiestudiet. Det har været en rejse med op- og nedture, men sammen står vi stadig og det bliver vi ved med. Der er meget at sige, men taknemmelighed omfavner det perfekt. En projektmakker blev til en ven for livet. Fred og Laus out **micdrop**.

Indholdsfortegnelse

<i>En felt i udvikling.....</i>	<i>1</i>
<i>Problemformulering.....</i>	<i>3</i>
<i>Kapitel 1</i>	<i>3</i>
Samtidens virkelighed.....	3
Det metodiske spektrum	9
Vores introduktion til de digitale metoder	16
<i>Kapitel 2 - På vej derhen.....</i>	<i>18</i>
Bagsiden af medaljen	18
Den digitale bibel - Gennemsigtighed og tilgængelighed.....	19
Skabelse af databasen	21
Struktureret fortid.....	21
Hvordan fik vi materialet ind i databasen?	22
Saxo – Seneste format.....	23
Konstruering af metadata	25
Kortlagt kommunikation	26
Materialitet i opløsning?	33
Devon R. Mordell og det femte paradigme	34
Er skridt tættere på	35
Afgrænsning - skabelse af database til specialet	37
<i>Kapitel 3 - Analyse</i>	<i>37</i>
Grundstenene.....	39
Pavens brevspind.....	42
Breve, byer og bånd.....	46
Topic Modeling.....	51
Sentimentanalyse.....	57
Delkonklusion.....	66
<i>De 6%.....</i>	<i>67</i>
En åben verden	70
Danmark og Frankrig – Computeren bestemmer	71
Geotracking.....	73
Når magten taler – Brevenes sprog mellem nord og syd.....	81
NER - Named Entity Recognition	88
Det hellige netværk	99

Innocens IV – Manden i midten.....	106
<i>Ctrl+ Alt+ Historie – En refleksion over digitale metoder</i>	107
Når koden taler sandt.....	110
Manipulering med omtanke	111
Tekstens modstand	112
<i>Diskussion - En relation mellem Danmark og Frankrig?</i>	113
Vicarius Christi	118
<i>Konklusion.....</i>	119
<i>Bibliografi.....</i>	121
Litteratur	121
Internetressourcer	123
Kilder	125

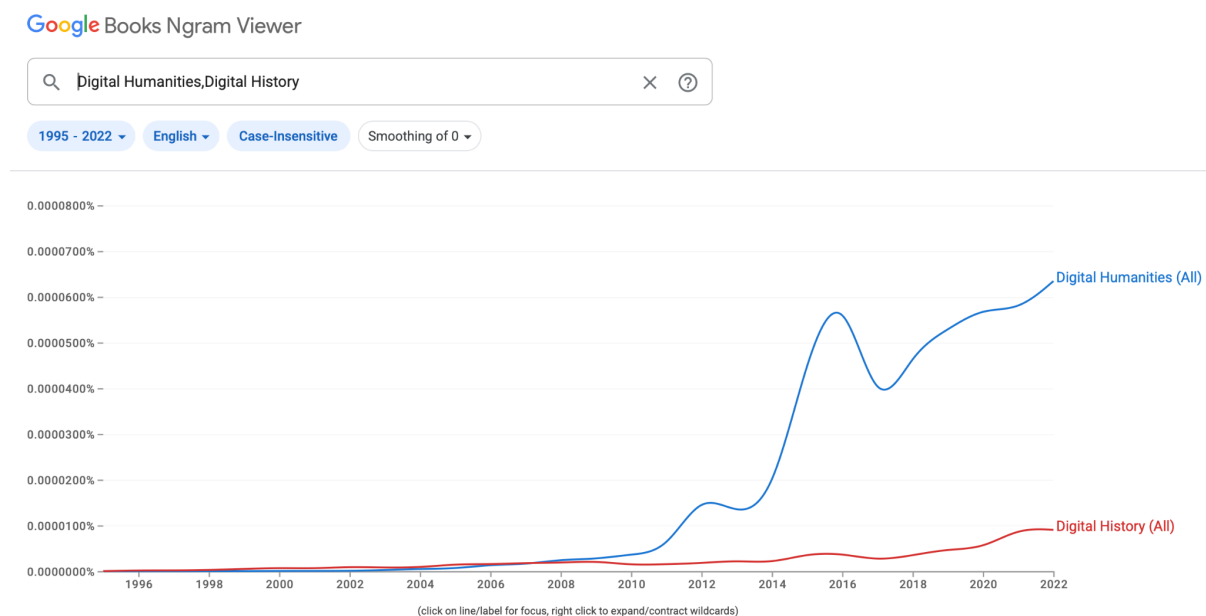
Grafliste

Graf 1) Google Books Ngram Viewer
Billede 1) Kodeudsnit med tekstbokse
Billede 2) Stedregister med 3886 rækker
Billede 3) Renset stedregister
Billede 4) Stedregister med koordinater
Søjlediagram 1) Brevmængde pr. årti
Graf 2) Fordeling af breve i Danmarks Riges Breve
Billede 5) Kodeudsnit udtræk af pave-data
Søjlediagram 2) Pavebreve pr. årti
Søjlediagram 3) Mest nævnte pave pr. årti
Diagram 1) Ordantal pr. brev over tid
Kort 1) Alle afsendere og modtagere (2002)
Heatmap 1) Samlet geografisk berøringsflade
Graf 3) Topord pr. emne (topic)
Diagram 2) Sammenhæng mellem alle brevets længde og sentiment-score
Diagram 3) Sentiment pr. ord i forhold til brevets længde
Kort 2) Sentimentkort (neg <15, pos >50)
Kort 3) Zoom på sentimentkort (neg <15, pos >50)
Kort 4) Kommunikation mellem Frankrig og Danmark
Kort 5) Zoom på kommunikation mellem Frankrig og Danmark
Diagram 4) Sankey-diagram
Graf 4) Omtale med og omtale om Frankrig
Wordcloud 1)
Diagram 5) Danmark/Frankrig sentiment pr. ord i forhold til brevets længde
Netværk 1) Netværk omkring pave Innocens III (1198-1216)
Netværk 2) Netværk omkring pave Innocens IV (1243-1254)
Netværk 3) Netværk mellem danske og franske aktører
Netværk 4) Danmark/Frankrig netværk med sentiment-kategorier
Netværk 5) Netværk i pavebreve (745 breve)
Netværk 6) Netværk i pavebreve med centrale aktører (degree centrality)
Netværk 7) Netværk i pavebreve med sentiment-kategorier
Netværk 8) Zoom på Netværk i pavebreve med sentiment-kategorier
Graf 5) Netværkstæthed i pavebreve over tid.
Netværk 9) Netværk for Innocens IV med sentiment-kategorier
Netværk 10) Netværk i pavebreve uden manipulation
Kort 6) Breve mellem Danmark og Frankrig uden Innocens IV
Kort 7) Afsendere i breve mellem Danmark og Frankrig uden Innocens IV
Kort 8) Modtagere i breve mellem Danmark og Frankrig uden Innocens IV
Netværk 11) Netværk mellem danske og franske aktører uden Innocens IV

En felt i udvikling

Siden 2010 har det digitalhistoriske felt gennemgået en markant udvikling. Denne fremgang afspejles tydeligt i det stigende antal publikationer, der beskæftiger sig med *Digital Humanities* og *Digital History*.

Graf 1) Google Books Ngram Viewer



Det er ikke en overdrivelse at påstå, at vi befinder os midt i en digital revolution inden for historiefaget. Nye digitale metoder og værktøjer fornyer feltet med en utrolig fart. Et tydeligt eksempel er den franske historiker Claire Lemerrier, som i 2021 afholdt en workshop om, hvordan man omsætter historiske kilder til netværksdata. Hendes oplæg blev efterfølgende delt på *The Historical Network Research Community's* YouTube-kanal.¹ Vi kender i forvejen til Claire Lemerriers principper om databasekonstruktion, gennem bogen *Quantitative Methods in the Humanities: An Introduction*, som hun udgav i 2019 sammen med Claire Zalc. Værket er blevet vores metodiske bibel, og er fast inventar for de digitalhistoriske kurser på universitetet.

På blot fire år er Lemerriers idéer om vejen fra kilde til data gået fra at være et oplæg på en forskningsplatform til nu at være fast forankret i undervisningen på AAU. I sin workshop beskriver Lemerrier, at der ofte bliver benyttet bestemte former for kilder i digitalhistorien, da

¹ *From historical source to network data*, Claire Lemerrier, YouTube-video, udgivet af Historical Network Research Community, 8. juli 2021. <https://www.youtube.com/watch?v=ZLgCSKNrTBY>

det er lettest at skabe netværksanalyser ud fra kvantificerbare data.² Vi ser ligeledes, at det især er det homogene kildemateriale, som protokoller fra rets- eller skattevæsenet, der er i centrum i de digitale undersøgelser. Disse kildetyper har ofte en fast struktur og gør det derfor relativt enkelt at omdanne indholdet til strukturerede data. Det er modsat mere kompliceret at bearbejde tekstmateriale kvantitativt.

Siden 2008 har projektet *Programming Historian* arbejdet på at udbrede digitale metoder i historieforskningen gennem peer-reviewed vejledninger. Her kan man eksempelvis tage Igor Sosa Mayor og Nabeel Siddiqui's lektion om visualisering af relationer mellem europæiske venskabsbyer under 2. Verdenskrig. I lektionen lærer man (...) *how to use R's ggplot2 package to create sophisticated data visualizations through a 'grammar of graphics' framework*.³ Til lektionen deler Programming Historian det datasæt, som undersøgelsen skal udføres på baggrund af. Læringsplatformen er tilgængelige, transparente og deler arbejdet fra start til slut. Her ser vi dog stadig, at de bagvedliggende datasæt er karakteristiske for digitalhistoriske undersøgelser, da de er homogene og velstrukturerede. I tilfældet med Mayor og Siddiquis lektion er datasættet bestående af oplysninger om oprindelseslande, geografiske koordinater, befolkningstal og klassificering som enten EU- eller non-EU-land.

Det er svært at finde vejledninger eller undersøgelser, som beskæftiger sig med teksttunge datasæt. Dette skyldes naturligvis materialets struktur, som ikke umiddelbart er let at gøre kvantificerbart.

Måske det er mangel på tilgængelighed eller udfordringen, som gør, at vi finder teksttunge datasæt særligt spændende. Vi ønsker at skubbe til de eksisterende grænser indenfor digitalhistorien, ved at undersøge dansk højmiddelalder gennem digitalhistoriske metoder.

Vi arbejder eksplorativt, da der netop ikke er fastsatte rammer eller vejledninger til, hvordan en sådan undersøgelse bør udføres. Vi følger den vej, som undersøgelsen af materialet leder os. En visualisering af et netværk kan skabe en undren, som vi følger og undersøger nærmere. Den eksplorative tilgang skal være uden forudindtagelser for materialets indhold. Vi har ingen fast opskrift for vores digitale analyse. Den eksplorative tilgang muliggør, at vi kan undersøge, hvor langt vi kan trække en kvantitativ på materialet.

² *From historical source to network data*, Claire Lemerrier, YouTube-video, udgivet af Historical Network Research Community, 8. juli 2021. <https://www.youtube.com/watch?v=ZLgCSKNrTBY> 12:58.

³ "Visualizing Urban and Demographic Data in R with ggplot2", konsulteret d. 1. juni 2025, <https://programminghistorian.org/en/lessons/urban-demographic-data-r-ggplot2#our-data-sister-cities-in-the-european-union>

Da vi ikke kan læne os op ad eksisterende undersøgelser, vil vi prøve at spænde elastikken helt ud. Vi er overbevist om, at en ny metodisk tilgang til et gennemarbejdet materiale vil skabe et positivt afkast og ikke mindst en ny forståelse af kildernes kompleksitet.

Problemformulering

Hvordan kan digitale metoder anvendes til at analysere relationen mellem Danmark og Frankrig i højmiddelalderen, og hvad kan en eksplorativ digital analyse af Danmarks Riges Breve bidrage med?

Kapitel 1

Samtidens virkelighed

Danmark og Frankrig er i højmiddelalderen to kongemagter i udvikling. I Danmark er den statslige konsolidering højt prioritet efter mange år, hvor borgerkrige har skabt splittelse og hærget landet.⁴ Frankrig har derimod fokus rettet mod international positionering. De har mistet store dele af det sydvestlige territorium, da kong Ludvig VII i 1152 lod sig skille fra dronning Eleonora af Aquitanien. Hun giftede sig derefter med den engelske kong Henrik d. 2, og han overtog arveretten til hertugdømmet Aquitanien i det sydvestlige Frankrig.⁵ Her lægges grundstenene til en årelang konflikt mellem Frankrig og England.

I Danmark er borgerkrigen slut, og Valdemar den Store er blevet enekonge i 1157, tre år efter at Henrik d. 2 blev konge af England.

For at konsolidere og fremtidssikre magten så de begge på arveretten til deres magtfulde positioner. Tidligere havde sønnerne delt arven mellem sig. Nu ønskede man ifølge McGuire og Esmark at stramme arvefølgen for at sikre retten til *kontrol med jord, borge og territorial myndighed* (...).⁶ Dette gjorde man ved udelukkende at lade arven gå til *primogenitus* - den

⁴ Kjær, Lars. *Valdemar Den Store - Borgerkrigens barn*. Gads Forlag, 2022. s. 100-102

⁵ "Eleonora af Aquitaine", konsulteret d. 21. maj 2025, [https://lex.dk/Eleonora af Aquitaine](https://lex.dk/Eleonora_af_Aquitaine)

⁶ Esmark, Kim & McGuire, Patrick Brian, *Europa 1000-1300*. Roskilde Universitetsforlag, 1. udgave, 1999. s. 53

førstefødte. Valdemar den Store indførte primogenitur i kongeslægten under kirkefesten i Sankt Bendts kirke i sommeren 1170.⁷

Opgøret med det traditionelle valgkongedømme kunne udløse eksplosive spændinger, som kunne føre til opgør mellem sønner, fædre og brødre. I sådanne situationer argumenterer Esmark og McGuire for, at kvinderne i familien blev en vigtig brik, idet de kunne motivere de frustrerede sønner.⁸ Dette er præcis, hvad Ralph V. Turner ser udfolde sig i samtidens England. I 1995 skriver han artiklen *The problem of Survival for the Angevin "Empire": Henry II's and his Sons' Vision versus late Twelfth-Century Realities*, hvori han giver et skarpt indblik i den komplekse politiske scene i slutningen af 1100-tallet og starten af 1200-tallet mellem det engelske Plantagenet dynasti og det franske kongehus Capet.⁹ Ifølge Turner lider det angevinske imperium af en strukturel skrøbelighed, som han mener er årsagen til Henry II's arvepolitik. For at sikre fremtidig dominans, ønskede Henry II at inddrage alle sine sønner i sin arvepolitik. Han indførte derfor ikke primogenitur, som i Danmark, men skabte i stedet *parage*.¹⁰ Turner inddrager Eleanor Searle, der beskriver *parage* som (...) *a kind of partible inheritance under the chairmanship of one son*.¹¹ Den ældste søn aflægger ed mod *caput*, slægtens hovedsæde, og fungerer som overhovedet samtidig med, at det er (...) *required that younger sons be given some holdings out of their father's possession*.¹² Systemet krævede og var særdeles afhængigt af samarbejde og loyalitet mellem Henrys sønner. Turner belyser, hvordan dette træk fra Henry II var med forhåbningen om stabilitet og sikringen af et dynasti. Det blev i stedet grundlag for konflikt og begyndelsen på enden for det angevinske imperium.

Selvom Henrik II ikke indførte samme arvepolitik som Valdemar den Store, er Esmark og McGuire tese også gyldig for situationen i England. Planen om fordeling af territorium endte som intern splittelse og familiekrig, da Eleonora af Aquitanien motiverede sine sønner til at gøre oprør mod deres far, kong Henrik II.¹³

⁷ Kjær, *Valdemar Den Store*, s. 174-183

⁸ Esmark & McGuire., *Europa 1000-1300*, s. 53

⁹ Turner, Ralph V. "The Problem of Survival for the Angevin "Empire": Henry II's and His Sons' Vision versus Late Twelfth-Century Realities", *The American Historical Review*, Vol 100, No 1, Oxford University Press on behalf of the American Historical Association(Feb 1995). s. 78-96.

¹⁰ Turner, "The Problem of Survival for Angevin "Empire"", s. 83

¹¹ Turner, "The Problem of Survival for Angevin "Empire"", s. 83

¹² Turner, "The Problem of Survival for Angevin "Empire"", s. 83

¹³ Turner, "The Problem of Survival for Angevin "Empire"", s. 86

Krig er i højmiddelalderen ikke motiveret af ideologi eller patriotisme, da dette forudsætter en grad af national identitetsforståelse, som ikke var til stede i samtiden.¹⁴ Måske det i virkeligheden er forkert at tale om et decideret Frankrig, England eller Danmark, da dette giver konnotationer til vores nutidige forståelse af nationer. Disse lande skal ikke forstås som nationer med et bestemt folkefærd, territorial afgrænsning eller fælles identitetsforståelse.¹⁵ Frankrig skal eksempelvis nærmere tænkes som Filip 2. Augustus berøringsflade. Krig er i samtiden motiveret af *dynastiske hensyn og ambitioner (og af religion)*.¹⁶ Slægten er derfor essentiel i højmiddelalderen. Der tages hensyn til slægt og dynastier i alle tænkelige beslutninger, da slægtens sociale struktur dominerede middelalderen. Det enkelte individ forstår sig selv i forhold til sin egen placering i slægtshierarkiet. Familie er ligeledes ikke fastsat til biologisk slægtskab. Dette ser vi eksempel i de kirkelige strukturer, hvor der er såkaldte munkebrødre og en fader i form af abbeden. *I virkeligheden blev alle sociale forhold begrebet i familiære termer som en slags symbolske blodsfælleskaber*.¹⁷ Dette resulterede i, at familie var centralt i højmiddelalderen. Man dyrkede eksempelvis sine slægtninge som helgener, som vi ser gennem Knudsgilderne eller etablerede store monumenter for at sikre slægtens fremtidige ære, som Hvideslægten gjorde med Sorø Kloster.¹⁸

I 1193 forsøgte det danske og det franske kongehus at sikre deres dynastier gennem et ægteskab mellem kong Filip 2. Augustus og kongedatter Ingeborg. I kilderne opstår ægteskabet ud af ingenting. Der er ingen forudgående omtale eller planlægning.¹⁹

Filip 2. Augustus blev konge af Frankrig i 1180, og han udviklede det franske kongedømme til en direkte og effektiv kongemagt. Capetingerne anvendte det feudale system som et strategisk

¹⁴ Ifølge Inge Adriansen opstår national identitet, eller en følelsesmæssig tilknytning til nationen, i 1700-tallet. Hun mener, at kærligheden til fædrelandet opstår som en reaktion på de religiøse opbrud, der sker i samtiden. Den strenge ortodokse kristendom erstattes af mere omfavnende pietisme, som giver rum og plads til, at den tidligere identifikation med religion kan erstattes af en følelsesmæssig tilknytning til fædrelandet. Adriansen, Inge. *Nationale symboler i Det Danske Rige 1830-2000 - Bind II: Fra undersætter til nation*. (Museum Tusculanums Forlag Københavns Universitet, 2003), s. 540

¹⁵ Michael Böss, *Nationalitet*, (Aarhus Universitetsforlag, 2019), s. 4-5

¹⁶ Esmark & McGuire, *Europa 1000-1300*, s. 48

¹⁷ Esmark & McGuire, *Europa 1000-1300*, s. 47

¹⁸ Absalon tildeler hele sin fædrene arv i sit testamente til Sorø Kloster. Hvideslægten etablerer et monument for deres fremtidige ære. De sikrer sig prestige i livet og i efterlivet. Kim Esmark, "Fromhed og ære - Hvideklanens gaver til Sorø Kloster i 11- og 1200-tallet", Årbog for Historisk Samfund for Holbæk Amt. (2003): s. 114

¹⁹ Thomas Heebøll Holm og Lars Kjær, Mægtig Middelalder, Ingeborg, Dronning af Orléans I: Brylluppet i Amiens, [20:51–21:28], udgivet af Spotify, 13. september 2024, <https://open.spotify.com/episode/4UFHyrVojXxpX1ZKrTeRqC?si=995f472822784d44>

redskab, der kunne tage et fragmenteret politisk landskab til et territorium under kontrol af den franske konge.²⁰

I artiklen *Philip Augustus and the Norman Church* fremstiller Baldwin Filip Augustus som en konge, der forstod det spil, som han var en del af. Filip Augustus var ikke blot en, der anvendte sin magt, fordi han kunne, men i stedet en konge som udnyttede sin autoritet til at skabe et netværk af aktører, hvor alle kunne bidrage og samtidigt sikre kongens position. Ideen om den stærke konge er ifølge Baldwin bundet op omkring et individ frem for en institution, da Filips efterfølger ikke formåede at videreføre de relationer, som han havde sikret.²¹

Kong Filip 2. Augustus bliver essentiel for relationen mellem Danmark og Frankrig i højmiddelalderen, da han giftede sig med Valdemar den Stores datter Ingeborg i 1193. Deres ægteskab var ifølge Esmark og McGuire

*(...) et rent politisk anliggende, og kvindens underordning under manden en given ting, som det eksempelvis kom konkret til udtryk i vielsesriten, hvor bruden knælede for gommen som en vasal for sin herre.*²²

Selvom Ingeborg gennem sit ægteskab til Filip Augustus var underlagt sin mand, spillede hun alligevel en vigtig rolle i samtiden. Ved at forene slægter gennem ægteskab sikrede man freden mellem de to implicerede dynastier. Kvinder var ligeledes essentielle for tilblivelsen af arvinger og dermed for videreførelsen af slægten.²³

De mandlige familieoverhoveders magt over kvinderne blev truet, idet ægteskab gennem højmiddelalderen begyndte at blive anset som et helligt sakramente. Det var nu ikke længere muligt at gifte sig inden for familiens egne rækker for at fastholde krav på jord internt, da det, som James Brundage beskriver det, ikke var lovligt at ægte familie. *(...) no one could marry a person to whos s/he was related within seven degrees of kinship by blood or by a fictive relationship.*²⁴

²⁰ Turner, "The Problem of Survival for Angevin "Empire"", s. 88

²¹ Baldwin, John W, "Philip Augustus and the Norman Church", *French Historical Studies* 6, no. 1, Duke University Press (1969): s.1-30. <https://doi.org/10.2307/286242> s. 10

²² Esmark & McGuire, *Europa 1000-1300*, s. 50-51

²³ Esmark & McGuire, *Europa 1000-1300*, s. 52

²⁴ Brundage, A. James, *Law, Sex, and Christian Society in Medieval Europe*, The University of Chicago Press, 1987. s. 238

En del af den interne magt i slægterne blev ligeledes placeret hos kirken, idet der blev skabt lovgivning om ophævelse af ægteskab. Det var nu ikke længere muligt at ophæve et ægteskab, fordi dette kunne gavne slægten politisk, da en ophævelse nu krævede en legitim grund godkendt af kirken.²⁵

Den nye ægteskabslovgivning spillede en stor rolle for ægteskabet mellem kong Filip 2. Augustus og dronning Ingeborg, da Filip frasagde sig Ingeborg efter bryllupsnatten. Nanna Damsholt har i sit værk *Kvindebilleder i dansk højmiddelalder* opsummeret situationen således.

*Philippe-Auguste havde friet til den danske prinsesse for at få arvekrav på England og dansk flådehjælp til den erobring, som han havde planer om. Brylluppet kom i stand, medgiften blev aftalt, bruden fulgte til Frankrig med standsmaessig følge, men efter bryllupsnatten ønskede Philippe-Auguste af ikke kendte årsager ægteskabet opløst.*²⁶

Vi læner os op ad Damsholt beskrivelse, da vi deler hendes tolkning af situationen. I højmiddelalderen kunne kvinder overføre arv, og kvinder som Ingeborg blev derfor (...) *en særlig attraktiv brik i det dynastiske spil.*²⁷ Dronning Ingeborg havde i princippet et arvekrav på England, som kan spores tilbage til Knud den Store, der var konge af England indtil sin død i 1035.²⁸ Ifølge Damsholt var dette krav nyttigt for Filip, da han netop havde planer om at erobre landet. Flere kilder argumenterer ligeledes for, at ønsket om at erobre England var årsagen til, at Filip ægtede den danske prinsesse.²⁹ Da brylluppet fandt sted i 1193, var Valdemar den Stores søn Knud d. 6 konge af det danske rige. Danmark var den stærkeste del af Nordeuropa, og landet havde en mægtig storpolitisk alliance gennem Knud d.6's ægteskab med Henrik Løves datter Gertrud.³⁰ Danmark var derfor en nyttig allieret for Filip.

Selvom Valdemarerne var en forholdsvis ny kongeslægt i Europa, og de samtidig kom fra det kolde nord, var der intet ukultiveret eller uciviliseret over Ingeborg. Hun talte latin og

²⁵ Esmark & McGuire, *Europa 1000-1300*, s. 52

²⁶ Damsholt, Nanna, *Kvindebilledet i dansk højmiddelalder*, Borgen, 1984. s. 237

²⁷ Esmark & McGuire, *Europa 1000-1300*, s. 52

²⁸ "Knud den Store", konsulteret d. 1. juni 2025,

https://lex.dk/Knud_den_Store?gad_source=1&gad_campaignid=20505639983&gbraid=0AAAApCkS9LICB0hlLtOZokZM1_dFpr0w&gclid=Cj0KCQjwlvrvBBhDnARIsAHEQgOTnQknRPZtt02xZZ-eO3D5NMfRKXf0ZzYnk-XV3XbsjvgQel5jLwwIaAjFkEALw_wcB

²⁹ Roesdahl, Else, *Fra Vikingetid til Valdemarstid*, Aarhus Universitetsforlag, 2023. S. 222

³⁰ "Knud d. 6", konsulteret 23. maj 2025, https://lex.dk/Knud_6

beskrives i Rigord's krønike *Deeds of Philip Augustus* som, (...) *a saintly girl adorned with a good character*.³¹

Lars Kjær opstiller en politisk og en mere personlig årsagsforklaring i forhold til det forliste ægteskab. Den personlige forklaring lyder på, at der gik noget helt galt mellem Ingeborg og Filip på bryllupsnatten. Den politiske forklaring lyder, at Filip ville have fat i den medgift, som fulgte med Ingeborg, og at han derefter afviste hende for at straffe danskerne for deres manglende hjælp mod englænderne.³²

På trods af en manglende årsag til den forliste relation, kender vi dog til det retslige efterspil, som Philips ønske om skilsmisse udløste. Ifølge James Brundage, var der inden for ægteskabslovgivningen opstillet to diskurser, der defineres som den franske og den italienske. I den franske skole er ægteskab uopløseligt, hvis alliancen er indgået med enighed og ønske om ægteskab fra begge parter. I den italienske skole er ægteskabet først gyldigt, når aftalen er fuldbyrdet gennem samleje.³³ Det er derfor muligt at forlade ægteskabet og finde en anden at gifte sig med, hvis ægteskabet ikke er endeligt fuldbyrdet. I første omgang argumenterede Phillip for, at han og Ingeborg var for nært beslægtede, og at de derfor kunne lade sig skille. Dette modbeviste Vilhelm af Æbelholt.³⁴ Philip forsøgte i stedet med et argument om, at ægteskabet var ugyldigt, da det, ifølge ham, aldrig var blevet fuldbyrdet seksuelt på grund af trolddom. Dette berører Rigord i sin krønike, når han skriver;

*But a strange thing! On that very day by the work of the devil! The same king, snared, as it is said, by the spells of sorceresses, began to loath the wife whom he had for so long desired.*³⁵

Anklagen om trolddom benægtede Ingeborg hårdnakket. Ingeborg forstod det politiske spil, som hun var blevet en del af, og hun blev efter 20 års kamp anerkendt som retmæssig dronning af Frankrig i 1213.

³¹ Field, Larry F. *The Deeds of Philip Augustus: An English Translation of Rigord's "Gesta Philippi Augusti."* Edited by M. Cecilia Gaposchkin and Sean L. Field. Cornell University Press (2022). kap. 99

³² Kjær, Lars, *Valdemar Sejrs. Imperiebyggeren*, Gads Forlag 1, udgave, 1. oplag, 2024. s. 64

³³ Brundage, Law, *Sex and Christian Society in Medieval Europe*, s. 235-238

³⁴ "Ingeborg: Prinsesse af Danmark og dronning af Frankrig, 1175-1237", konsulteret 25. maj 2025, https://danmarkshistorien.lex.dk/Ingeborg_Prinsesse_af_Danmark_og_dronning_af_Frankrig_1175-1237

³⁵ Field, *The Deeds of Philip Augustus*, kap. 99

Denne sag belyser i høj grad, hvor stærkt pavedømmet var interageret på flere samfundsniveauer. I sagen om Ingeborgs ægteskab er en hovedtematik *magt*. Efter Filips forstødelse af Ingeborg, står den franske kirke klar til at annullere ægteskabet. Her sætter pave Innocens III sig imod.³⁶ Denne sag blev stor, fordi Innocens III helt personligt besluttede, at dette ægteskab skulle gennemføres. I værket *The Capetians* skriver Jim Bradbury; *A problem for Philip was Innocent's intervention in secular affairs, often in business that Philip thought not of papal concern*.³⁷ Relationen mellem Filip og Innocens III danner et stærkt billede af, hvordan den katolske kirke var involveret i alle områder af samfundet. Denne grænse, der virker naturligt at skelne mellem, er svag og tæt på ikke eksisterende. Middelalderens samfund var bygget op omkring tro, både spirituelt og administrativt. Det kan blandt andet ses ved, at store dele af Filips kongelige administration bestod af gejstlige individer, da *They were the educated and literate section of the populace*.³⁸ Middelalderens katolske kirke og særligt paven selv var derfor ikke blot involveret i gejstlige sager, men fungerede aktivt i alle dele af samtiden.

For at undersøge relationen mellem Danmark og Frankrig, er det derfor væsentligt at forholde sig til pavedømmet. Konflikten om Ingeborg og Filips ægteskab er ligeledes en central brik i forståelsen af den dansk-franske relation i samtiden.

Det metodiske spektrum

Dr. Claire Brennan fra James Cook University i Australien skrev i 2018 artiklen *Digital Humanities, digital methods, digital history, and digital outputs: history writing and the digital revolution*. I denne artikel behandler hun flere interessante områder inden for det digitale univers. Den første og særdeles interessante pointe hun kommer med, er hendes diskussion af digital humaniora overfor digital historie. Brennan belyser, hvordan begrebet *digital humanities* kommer fra litteratur- og medievidenskab. Det har medført en skævhed i formidlingen og brugen af digitale værktøjer. Hun pointerer, at selvom *digital humanities* lyder bredt og inkluderende, har det i stedet allerede skabt en splittelse.³⁹ På det grundlag er der skabt

³⁶ Bradbury, Jim. *The Capetians, Kings of France 987-1328*. Hambledon Continuum, 2007. s.183.

³⁷ Bradbury, *The Capetians*, s. 183

³⁸ Bradbury, *The Capetians*, s.185

³⁹ Brennan, Claire, "Digital humanities, digital methods, digital history, and digital outputs: history writing and the digital revolution", *History Compass*, 16 (10) (2018). <https://doi.org/10.1111/hic3.12492>, s. 1

underkategorier af feltet, *digital methods* og *DHSS*.⁴⁰ Førstnævnte refererer til arbejdet med de teknikker og værktøjer der anvendes, såsom textmining og netværksanalyser. DHSS er en underkategori, der strækker sig ud til andre forskningsfelter som eksempelvis historie. Brennan argumenterer for at historikere er at finde i udkanten af digital humaniora, når hun skriver;

*The place of historians within digital humanities is at best peripheral, as demonstrated by an Australian guide to the digital humanities which includes some contributions from historians but which is dominated by the concerns of literary scholars.*⁴¹

Brennan inkluderer hertil det, hun omtaler som “manifeste” til digital humaniora blandt andet *Digital Humanities* af Anna Burdick, Johanna Drucker, Peter Lurenfeld, Todd Presner og Jeffrey Schnapp. Hertil beskriver Brennan, hvordan historikere gerne anvender digitale værktøjer, men ikke nødvendigvis placerer sig inden for manifesternes retningslinjer. En af grundene til dette skyldes de historiefaglige traditioner og historikerens (...) *relationship with sources, ethics, its publics, and the past*.⁴²

Brennan stiller sig mellem to felter i sin artikel, hvor hun bevæger sig frem og tilbage. Hun ser, at der ikke er fastlagte retningslinjer for historikerens arbejde med det digitale, som de tidligere omtalte manifeste. Samtidig belyser hun en pointe om, at digitalt arbejde ikke er nyt for historiefaget. Ganske vidst har det taget gradvise udviklinger, i takt med at computerens ydeevne forbedres. Hun inddrager i sin artikel Gerben Zaagsma, der har lavet undersøgelser af historikerens digitale ageren. Han fastslår, at digitalt arbejde ikke er nyt for historiefaget. Brennan inddrager et citat fra Zaagsma, hvor han skriver;

*‘digital history’ has been part of the practice doing history for a substantial period of time and is certainly less new than the current buzz surrounding digital humanities might suggest.*⁴³

⁴⁰ DHSS – digital humanities and social sciences

⁴¹ Brennan, “Digital humanities, digital methods, digital history, and digital outputs”, s. 1

⁴² Brennan, “Digital humanities, digital methods, digital history, and digital outputs”, s. 2

⁴³ Brennan, “Digital humanities, digital methods, digital history, and digital outputs”, s. 3

Ideen om at det hele er nyt og storslået og et direkte afkast af litteratur- og medievidenskabens forståelse af digital humaniora, er derfor måske ikke helt korrekt. Dette er et felt, som ikke kan dateres eller tages ære for at være, der hvor det startede. Digital historie er ikke en perifer disciplin af en specifik forståelse af digital humaniora, men i stedet en naturlig forlængelse af historiefaget i en digital tidsalder. Det Brennan i virkeligheden søger mod med denne artikel, er ideen om en åbenhed, hvor historiefaget kan få større del af den anerkendelse, som hun mener er på sin plads.⁴⁴

En vigtig del af historiefaget og i særdeleshed i relation til digitale projekter, er de historiske arkiver. Brennan har i sin artikel et særligt fokus på historikerens relation til arkiverne, både fysiske og digitale. Digitaliseringen af arkivmateriale er helt essentiel for historiefagets videre udvikling i det digitale univers. Denne udvikling har derfor også medført, at (...) *archival practice is at the heart of digital history*.⁴⁵ Det gælder både vores metodiske tilgange til bevaring af materiale og formidling af det, samt at historikeren bliver arkivskaber. En pointe der er særligt relevant i vores projekt. Vi er i realiteten arkivskabere, da vi har indsamlet store mængder af middelalder relateret kildemateriale og opsat det i en digital og anvendelig form. Det kan medføre en idé om, at den digitaliserede mængde er den samlede sum af materiale. Det, der mangler, kan derfor nemt blive skjult i de digitale arkiver.

Dertil mener Brennan ydermere, at digitale arkiver er mindre stabile end de fysisk bevarede kopier. Det skyldes vedligeholdelse af materialet, software og lagring af data.⁴⁶ Digitale arkiver er ifølge Brennan med til at ændre den historiske metode. Til det digitale materiale er der ofte indlejret søgefunktioner, der gør det nemt og hurtigt at finde det materiale, som mangler i forhold til besvarelse af en specifik problemstilling. Det medfører samtidig, at forståelsen for kontekst hurtigt kan gå tabt.⁴⁷

Brennan viser i sin artikel, hvordan digitalisering har ændret forståelsen af arkiver og måden, hvorpå historikere anvender og arbejder med materialet. Hovedpointen vi får ud af denne artikel er vigtigheden af at fortsætte med at være historikere, lige så vel som digitale historikere. Der er kerneprincipper i historiefaget som kildekritik og kontekstuel forståelse, der er absolut nødvendige i arbejdet med digitalt materiale.

⁴⁴ Brennan, "Digital humanities, digital methods, digital history, and digital outputs", s. 10

⁴⁵ Brennan, "Digital humanities, digital methods, digital history, and digital outputs", s. 4

⁴⁶ Brennan, "Digital humanities, digital methods, digital history, and digital outputs", s. 5

⁴⁷ Brennan, "Digital humanities, digital methods, digital history, and digital outputs", s. 6

Brennan rejser nogle interessante spørgsmål og metodiske formål i sin artikel, uden at give et reelt bud på arbejdsmetoderne. I 2022 skrev Jesse W. Torgerson derimod artiklen *Historical practice in the era of Digital history*. I artiklen tager Torgerson stilling til problematikkerne og giver et bud på, hvordan vi rent metodiske bør agere indenfor digitalhistorien. Hans formål med sin artikel er, at *provide a theory of the practice of history that encompasses contemporary, traditional, and even ancient historical methods: capturing sources, producing data, and creating facts*.⁴⁸ Torgerson forsøger derfor at tage stilling til de konkrete problemer og manglende manifeste, som Brennan beskrev det. Torgerson anskuer til forståelsen af digital historie, at der er problematikker i feltet. Først og fremmest er det termen digital historie i sig selv, der indikerer en problematik og sammenstød mellem historikerens traditionelle tilgange og de moderne tendenser. Han skriver; *For many, digital history means Digital History: not just something new but something other*.⁴⁹ Her ser vi en opdeling af forståelsen af digital historie, da der både er historikere, som ser det digitale, som et fundamentalt anderledes arbejde, hvortil der ligeledes er historikere, som ser det digitale, som en naturlig forlængelse af feltet.

Det er ikke blot historie med computerens anvendelse, men et andet felt som indeholder andre praksisser og metodiske forståelser. Det har ifølge Torgerson flere konsekvenser. Først og fremmest er integration og sammensætning af et metodisk grundlag besværligt. Det gør det ligeledes sværere for nogle historikere at anskue digitale analyser og de outputs der medfølger, som legitime. Dertil har Torgerson også en speciel vinkel, som sjældent er skitseret i relateret forskningsmateriale. Her står han fra sin position, som underviser på Wesleyan University, og ser et klart problem i forvirringen af undervisningsmateriale. Inkorporeringen af det digitale har medført, at det bliver besværligt at gennemskue, hvad historisk metode egentlig er. Torgerson skriver til dette; *But to change any of this, we first must stop telling our students lies*.⁵⁰

Torgerson giver i forlængelse af ovenstående citat en kritik af historikerens brug af begreberne *primær-* og *sekundærkilde*. Han ser i takt med historiefagets udvikling en generel fejl i vores forståelse af kilder, hvortil han skriver;

⁴⁸ Torgerson, W. Jesse. "Historical practice in the era of digital history", i *History and Theory* 61, no. 4 (December 2022). S. 37-63. DOI: [10.1111/hith.12276](https://doi.org/10.1111/hith.12276) s. 37

⁴⁹ Torgerson, "Historical Practice in the era of digital history", s. 38

⁵⁰ Torgerson, "Historical Practice in the era of digital history" s. 39

*Historians also do not, in actual practice, do anything of the sort: we do not categorize objects of study into primary and secondary sources(...) og Nevertheless, this terminology is as ubiquitous as it is useless.*⁵¹

Torgersons pointe er tydelig, og han mener oprigtigt, at begreberne, som vi benytter til at kategorisere kilder, er forvirrende og metodisk upræcise. Torgersons pointerer, at uanset hvordan vi kategoriserer kilder, så vil det i sidste ende være måden, hvorpå vi anvender dem, som betyder noget. Det er derfor ikke kilden, der definerer, hvad som er en primær eller sekundær kilde. Det er i stedet, hvad vi som historikere vælger at anvende kilden til.

Torgerson læner sig derfor op ad en praksis, der indebærer (1) *capturing sources*, (2) *producing data*, and (3) *creating facts*.⁵²

Capturing sources er Torgersons ide om, at vi ikke starter med færdige kilder i en digital kontekst, men selv er med til at definere kilderne. Det bliver først til en kilde, når det er udvalgt og afgrænset rettet mod bestemte spørgsmål. Hertil inddrager han Johanna Druckers *The Digital Humanities Coursebook*. Drucker udformer i sit værk en (...) *distinction between capta (what is taken or seized) and data (what is given)*.⁵³ Torgerson fokuserer særligt på *capta*, da det er det første metodiske punkt, hvor der vurderes, hvad den kommende analyse skal fokusere på. Vi som historikere vælger aktivt, hvad vi beslutter os for at anvende som data.

Del to (*producing data*) i Torgersons tredelte metodiske fremgang er struktureringen af det udvalgte materiale. Torgerson skriver; *The creation of data out of sources is a distinct, necessary, and inherently valuable step in the historical method*.⁵⁴ Det er her, hvor historikeren skal bearbejde det udvalgte materiale og strukturere det, for at kunne danne et grundlag for de analytiske argumenter. I vores tilfælde vil det eksempelvis være at opsætte kolonner med årstal, geografiske markører samt afsender og modtager. Torgerson konkluderer i sin artikel *Facts are made from data*.⁵⁵ Det er ikke blot en produktion af visuelle elementer, men ligeledes produktionen af datagrundlaget, som er vigtig.

I sidste ende bygges konklusionen på baggrund af det selvskabte datagrundlag. Dette leder mod Torgersons sidste metodiske punkt, *creating facts*. Torgerson beskriver i dette led, at fakta ikke

⁵¹ Torgerson, "Historical Practice in the era of digital history", s. 39

⁵² Torgerson, "Historical Practice in the era of digital history", s. 37

⁵³ Torgerson, "Historical Practice in the era of digital history", s. 46

⁵⁴ Torgerson, "Historical Practice in the era of digital history", s. 60

⁵⁵ Torgerson "Historical Practice in the era of digital history", s. 56

er fundet i kilderne, men er noget som skabes gennem struktureringen af data. Han skriver; *facts are faciunt; they are facta-made, created, constructed-after the process of turning sources into data. Facts are made from data.*⁵⁶ Dette betyder at digitale historiske problemstillinger ikke blot findes i kildematerialet, men er skabt gennem en datadreven analyse. Fakta skabes dermed, når vi som historikere udtrækker viden fra netværksplot, visuelle mønstre og geografisk spredning over tid.

Torgerson bevæger sig i sin artikel væk titlen egentlige indikation; *Historical practice in the era of digital history*. Det ender med ikke blot at være en række pointer, der kan styrke den metodiske forståelse af digital historie, men også en gentænkning af hele historiefagets metode. Han skaber en ramme til det praktiske arbejde i form af *capturing sources, producing data, creating facts*. Ved at have et stort fokus på databehandling, laver han et opråb mod noget, som ofte ikke skildres eller anerkendes. Torgerson mener tydeligt, at databehandlingsprocessen er en del af den nye historiefaglighed og ikke blot et forarbejde til en undersøgelse. Dertil skriver Torgerson sig også ind på en helt generel nødvendighed for digitalt historisk arbejde, nemlig gennemsigtighed. Vi skal fortsat vise, hvordan vi udvælger, strukturerer og bearbejder det aktuelle kildegrundlag.

Brennan tydeliggør de problematikker og faldgruber, der er at finde i integrationen af digitalt arbejde i historiefaget. Torgerson forsøger at give et bud på, hvordan vi tilnærmelsesvis bør omtænke vores måde at opfatte en historisk metode på. Gustavo Fernández Riva giver et bud på, hvordan digitale metoder kan fungere og anvendes i praksis i sin artikel *Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods* fra 2019.⁵⁷

Riva skriver; *The goal of this article is to set the theoretical foundations for the network analysis of shared manuscript transmission.*⁵⁸ Denne problemstilling har en interesse for os, da det ikke er digitale middelalderundersøgelser, der udgives flest af. Rivas ønsker ikke at belyse netværk, som blot værende et visuelt redskab i en historisk analyse. I stedet søger han mod at etablere det som et værktøj, der kan anvendes til at afsløre mønstre, som normalt ikke ville blive spottet gennem traditionel historisk metode.

⁵⁶ Torgerson, "Historical Practice in the era of digital history", s. 56

⁵⁷ Fernández, R. Gustavo, "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods", i *Journal of Historical Network Research* 3 (2019). S. 30-49. DOI: [10.25517/jhnr.v3i1.61](https://doi.org/10.25517/jhnr.v3i1.61)

⁵⁸ Riva, "Network Analysis of Medieval Manuscript Transmission." s. 32

I sin analyse belyser Riva et vigtigt metodisk element i form af de valg, som er taget i databehandlingen.

Da denne artikel fokuserer på det metodiske og teoretiske, anvender han materiale fra en allerede eksisterende online database, *Handschriftencensus*, hvortil der er taget forudgående beslutninger til indholdet. Han skriver; *The resulting network will be profoundly affected by these editorial decisions*.⁵⁹ Udgiveren tager en klar beslutning om, hvordan materialet skal sammensat og dertil, hvad som frafalder.

Her vurderer vi, at Riva bevæger sig i samme retning som Johanna Druckers begreb *capta*.⁶⁰ Data er ikke givet, men i stedet taget og struktureret. Det er et vigtigt element for Riva, da hans foretagne valg vil blive afspejlet i resultaterne. Det bliver derfor ikke en objektiv fremstilling, men i stedet en analyse, som bygger på fortolkning. Til sin fremstilling er det, som før nævnt, vigtigt for ham at påpege netværksvisualiseringers formål.

*The model behind network theory permits not only a useful visualization for exploratory analysis, but also the implementation of different statistical tools and algorithms that shed light on the features of the underlying phenomena.*⁶¹

Det visuelle er en del af en større datamængde og kan derfor bruges på forskellig vis. Det handler ikke blot om at finde mønstre i en visuel formidling, men det er ligeledes vigtigt at bearbejde den data, i form af statistik, som danner hele grundlaget. Det er en proces, hvor de to elementer bør sammenføres til analytisk fortolkning.

Netværksanalysen har muligheden for at kunne bidrage til større strukturelle undersøgelser, da man kan samle hele og forskellige tekstkorpus i en analyse. Denne type undersøgelse muliggør, at man kan bevæge sig ud over tid og i stedet fokusere på de strukturelle elementer i kildematerialet. Denne mulighed forekommer til vores materiale, da vi har opsat vores database på en måde, hvor hvert brev er skilt fra hinanden. Hver række i vores database vil derfor potentielt kunne fungere som en node i en netværksanalyse.

⁵⁹ Riva, "Network Analysis of Medieval Manuscript Transmission." s. 36

⁶⁰ Capta (what is taken or seized)

⁶¹ Riva, "Network Analysis of Medieval Manuscript Transmission.", s. 48

Riva skriver sig ligeledes ind i en gennemgående tematik i den digitalhistoriske forskning, når han belyser sin indsamlings- og behandlingsproces. Før sin analyse har han et afsnit, der er dedikeret til transparens og reproducerbarhed. Her belyser han processen med webscraping af HSC og hans anvendelse af Python og Gephi.⁶² Noget der kan anskues som et metodisk ideal indenfor digital historie, er muligheden for at andre kan genskabe dit arbejde og validere resultaterne. Han har derfor tilgængeliggjort hele sit arbejde i en fodnote:

Files and Python scripts used can be found on GitHub at the following URL:
https://github.com/GusRiva/projects_data/tree/master/HSC.⁶³

Dette understøtter i høj grad hans forståelse af feltet og anerkendelse af validitet, transparens og reproducerbarhed. Ved at han indsætter de koder, som han har anvendt, insinuerer han ligeledes at koden ikke blot bør anses som et værktøj, men som en del af hans analytiske fremstilling.

I vores arbejde stræber vi ligeledes efter transparens og reproducerbarhed. Af denne grund har vi valgt at dele vores kodning i opgaven, så processen bag tilblivelsen af data ligeledes tilgængeliggøres.

Vores introduktion til de digitale metoder

I løbet af vores bachelor blev vi introduceret til spændingsfeltet mellem den klassiske historieforskning og de digitale metoder. På 3. semester forsøgte vi os med disse metoder, da vi benyttede digitale værktøjer til at undersøge og fremstille det aalborgensiske prostitutionsmiljø i slutningen af 1800-tallet. Det afskrækkede os, da vi hverken havde indsigt i det digitalhistoriske felt eller færdighederne til at producere resultater, som vi var tilfredse med. På 6. semester stiftede vi bekendtskab med bogen *Quantitative methods in the humanities: an introduction*, som er skrevet af de to franske forskere Claire Lemerrier og Claire Zalc.⁶⁴ Bogen er overskuelig og konkret, og den gjorde det mere håndgribeligt for os at arbejde med konstruering af databaser. Med udgangspunkt i denne guide fik vi igen mod på at kaste os over

⁶² Riva, "Network Analysis of Medieval Manuscript Transmission.", s. 38

⁶³ Rivas, "Network Analysis of Medieval Manuscript Transmission.", fodnote 18, s. 38

⁶⁴ Lemerrier, Claire og Claire Zalc. *Quantitative Methods in the Humanities - an introduction*. University of Virginia Press, 2019.

det digitale, og vi skabte derfor vores første database med 1527 sager fra Christianshavns Straffeanstalts protokoller. Her fik vi erfaringer med, hvordan der skabes homogenitet i et datasæt, for at indholdet bliver kvantificerbart.

På 7. semester begyndte vi vores kombination af den middelalderlige periode og de digitale metoder, idet vi startede på vores specialiseringskandidat.⁶⁵ Vi digitaliserede de 230 breve, som er i Danmarks Riges Breve Række 1. Bind 3, og fik mange positive erfaringer med fra denne undersøgelse. Herefter begyndte vi at se muligheder for at udvikle på databasen såvel som vores digitale færdigheder.

Da vi nåede sommeren 2024, tog vi beslutningen om at sætte ambitionsniveauet højt i forhold til vores videre arbejde med det digitale.

Den ide, som vi havde brugt det første år af vores kandidat på at tænke igennem, var nu klar til at blive sat i værk. Vi skrev derfor i vinteren 2024 projektet *Fra pergament til Python*. Dette projekt skulle ikke blot være en fortsættelse af vores tidligere arbejde men et metodeudviklende projekt, som skulle optimere vores evner til at bearbejde middelaldermateriale i en digitalhistorisk kontekst. Vi skabte en database med alle breve fra *Danmarks Riges Breve*, *Saxos Danernes Bedrifter*, *Adam af Bremen om Menigheden i Norden*, *Roskildekrøniken*, *Deed of Philip Augustus*, *Deeds af Innocens III samt brevsamlinger fra Cathrine af Siena*, *Gregory VII* og *Saint Bernard*.⁶⁶ I vinteren 2024 havde vi derfor en database med 4.891.232 ord fordelt på 20.193 rækker.⁶⁷

I projektet arbejdede vi særligt med det eksplorative og metodeudviklende, da vi vidste, at vi ville støde på fejkilder i vores konstruering af databasen. I denne sammenhæng er fejl essentielle, idet den metodeudviklende proces er cirkulær og dermed konstant kan optimeres.⁶⁸ Da forskningen spænder bredt på baggrund af en konstant udvikling, var det en længere proces for os at etablere vores ståsted i det digitalhistoriske felt. Derfor skabte vi det særskilte tekniske projekt *Fra Pergament til Python* for at udfolde de principper, som vi arbejder på baggrund af. Det gør vi særligt, fordi vi ønsker en skarp kontinuitet i vores digitale arbejde, transskribering

⁶⁵ Vi har på kandidaten særligt fokus på dansk og international højmiddelalder.

⁶⁶ Se bibliografi.

⁶⁷ Hver række indeholder en kilde, i form af et brev eller et kapitel i en krønike.

⁶⁸ Parland-von Essen, J. (2020). "Building historical knowledge byte by byte: Infrastructures and data management in modern scholarship." in M. Fridlund, M. Oiva, & P. Paju (Eds.), *Digital histories: Emergent approaches within the new digital history*. (s. 89-102). Helsinki: Helsinki University Press.
<http://doi.org/10.33134/HUP-5-5>, s. 93

såvel som kodning, for at arbejdet bliver gennemsigtigt og reproducerbart.⁶⁹ Projektet *Fra Pergament til Python* var derfor et nødvendigt forarbejde, for at vores undersøgelse af relationerne mellem Frankrig og Danmark bliver muligt og tilfredsstillende.

I digitalhistoriske undersøgelser kan den tekniske del være undladt og man kan støde på argumenter som, (...) *we thought it might act as a barrier to our less technical readers*.⁷⁰

Denne præmis køber vi ikke. Vi anser det for essentielt at formidle det tekniske arbejde, da det bliver svært at gennemskue og ikke mindst validere forskningsresultater, hvis grundlæggende data og teknik ikke er redegjort for. Derfor ønsker vi at efterkomme, hvad vi ser som et behov for om mere gennemsigtighed i digitalhistorien, ved at redegøre for etableringen af den database, som vi benytter som kildegrundlag til dette projekts undersøgelser.

I dette projekt benytter vi databasen, som vi skabte i forbindelse med *Fra Pergament til Python*, hvortil vi filtrerer en mindre relationel database fra, som relaterer sig til vores undersøgelse af relationen mellem Danmark og Frankrig.⁷¹

Kapitel 2 - På vej derhen

Bagsiden af medaljen

Arbejdet med digitalt kildemateriale tillader os at træde ind i et digitalt univers, hvor kilder kan analyseres og behandles med en større effektivitet grundet computerens evner. Johan Jarlbrink skriver i artiklen *All the Work that Makes it Work: Digital Methods and Manual Labour*, at det er vigtigt at anerkende, hvordan det digitale ikke blot skaber muligheder, men også hvordan det ændrer den metodiske tilgang. Databehandlingen kræver fokus, idet digital databehandling ikke er fuldkommen. Det er væsentlig ikke at forglemme det arbejde, der ligger i at rense, kvalitetssikre og kontrollere for fejl i dataene.⁷² Selvom disse processer sjældent får meget

⁶⁹ Projektet er vedhæftet som bilag.

⁷⁰ Burrows, Toby, "Reproducibility, verifiability, and computational historical research," I *International Journal of Digital Humanities*, 283-298. (Springer, 2023), s. 290

⁷¹ Vi filtrerer i den primære database af flere årsager. Først og fremmest er det unødvendigt at bruge alt det digitaliserede kildemateriale, da vi har afgrænset os til perioden fra 1131 til 1265. Det ville ydermere være besværligt at bruge den samlede database, da vores computere vil have svært ved at behandle så store datamængder.

⁷² Johan Jarlbrink, "All the Work that Makes it Work: Digital Methods and Manual Labour," I *Digital histories: Emergent Approaches within the New Digital History*, s. 113-126. (Helsinki: Helsinki University Press, 2020), s. 113

fokus, er de vigtige, da (...) *they have a significant impact on results and need to be taken seriously*.⁷³ Jarlbrink beskriver dette arbejde som *det usynlige arbejde*.⁷⁴ For at skabe en gennemsigtig fortælling om datasættets tilblivelse, er det derfor vigtigt at være åben omkring hele tilblivelsesprocessen.

Den digitale bibel - Gennemsigthed og tilgængelighed

Gennemsigthed er generelt et af de vigtigste omdrejningspunkter i diskussionen om digitalhistorie og skabelsen af databaser af kildemateriale. For at opnå det bedst mulige udbytte, er det yderst brugbart at kunne kombinere datasæt fra forskellige databaser. For at dette bliver muligt, er tilgængelighed og gennemsigthed essentielt. Det er vigtigt, at de forskere, som skaber datasæt, beskriver deres proces, så denne kan genskabes af andre. Det er ligeledes relevant at være gennemsigtig i sit arbejde med det digitale, så dem, som ønsker at benytte materialet, har en forståelse for det. Hvis man eksempelvis har filtreret noget fra de oprindelige data, er det væsentligt at dette er beskrevet grundigt.

Den finske historiker Jessica Parland-Von Essen har defineret det, som hun kalder *romantic lonely genius syndrome*.⁷⁵ Når historikeren lider af dette syndrom, forglemmes gennemsigtheden, og dette skaber i sidste ende et benspænd for hele feltet. Hendes tese går ud på, at en historiker, som bliver forblændet af det, der kan ses som hans eller hendes arbejdes potentiale og genialitet, glemmer sine kollegaer. Ved at være så dybt begravet i sin egen gode ide, glemmer forskeren at beskrive processen bag dataens tilblivelse. For den ramte bliver det unødvendigt "da der alligevel ikke er nogle, som forstår den geniale ide bag mit store arbejde...". Når der ikke bliver skabt en forståelse for datasættets tilblivelse, bliver det svært for andre forskere at benytte og arbejde videre med materialet. I sidste ende skaber det et kompliceret workflow, som måske afskrækker.

*Better collaborating and coordinating than working in silence.*⁷⁶

⁷³ Jarlbrink, "All the Work that Makes it Work", s. 113

⁷⁴ Jarlbrink, "All the Work that Makes it Work", s. 114

⁷⁵ Parland-von Essen, "Building historical knowledge byte by byte", s. 94

⁷⁶ Parland-von Essen, "Building historical knowledge byte by byte", s. 94

Tilgængelighed er af samme årsag en af de primære omdrejningspunkter. Ved at gøre data tilgængeligt kan flere forskere kombinere og udvikle deres undersøgelser. Ifølge Kimmo Elo er det metadataens primære funktion at skabe tilgængelighed.⁷⁷

På et møde i 2016 diskuterede de førende eksperter indenfor området. De skabte sammen konceptet FAIR-data. Det står for data som er *Findable, Accessible, Interoperable* og *Reusable*.⁷⁸ Her er tilgængelighed ligeledes en af kerneværdierne. Tilgængelighed har i denne sammenhæng særlig fokus på internettet. Databaser skal publiceres på en måde, hvor det er muligt at downloade dem. Rå-data og metadata skal ligeledes være tilgængeligt i en sådan form, at de fungerer på flere brugerflader. Dataene skal kunne hentes ned fra nettet og fungere i flere programmer som eksempelvis Sheets, Excel og i CSV-format.

Det er ligeledes på internettet, at keyword spiller en central rolle. Kimmo Elo mener, at keyword er essentiel metadata, da det er dem, som gør materialet *findable*.⁷⁹ Selvom Elo er fortaler for at skabe metadata i form af keyword, er der ikke nogen fokus på de problemer, som keywords ligeledes kan være med til at skabe. Disse problemer ser Tim Hitchcock til gengæld. Han påpeger i sin artikel *Confronting the Digital, Cultural and Social History - Or How Academic History Writing Lost the Plot*, at Google i 2009 har opsat parametre for google-søgninger, som er linket til brugernes personlige søgehistorik.

*In other words, Google's innovations mean both that researchers are increasingly being directed to bodies of online material they have used before at the expense of a wider knowledge landscape and, more significantly, that two researchers are unlikely to find the same results returned from the same query.*⁸⁰

⁷⁷ Elo, Kimmo. (2020). "Big Data, Bad Metadata: A Methodological Note on the Importance of Good Metadata in the Age of Digital History." i Mats Fridlund m.fl. (red.), *Digital histories: Emergent Approaches within the New Digital History*. (pp. 103-111). Helsinki: Helsinki University Press. <https://doi.org/10.33134>

⁷⁸ Parland-von Essen, "Building historical knowledge byte by byte", s. 90

⁷⁹ Elo, "Big Data, Bad Metadata", s. 108

⁸⁰ Tim Hitchcock, "Confronting the Digital, Cultural and Social History - Or How Academic History Writing Lost the Plot," *The Journal of the Social History Society* 10 (2013): s. 9, downloaded 10. november 2024, <https://doi.org/10.2752/147800413X13515292098070>. s. 15

Det er dermed ikke sikkert, at de benyttede keyword har den ønskede effekt. Den personlige historie spiller en stor rolle i den enkelte brugeres resultater på en søgning. Det er naturligt, at man bliver ledt mod det materiale, som passer bedst til brugerens tidligere behov eller historik. Hitchcock benytter selv søgeordene *digital humanities*, *digital history*, *standards*, *scholarship*, *referencing*, *OCR* og *search* i sin artikel. Ved en simpel test, er det tydeligt, hvor stor forskel personlig søgehistorik gør. Ved at søge på disse søgeord i en browser som Laura normalt ikke benytter, er det første resultat et link til artiklen *Reading in the mist: high-quality optical character recognition based on freely available early modern digitized books* i tidsskriftet *Digital Scholarship in the Humanities*. Artiklen berører ligesom Hitchcocks arbejde *Digital Humanities*, men er dog udgivet syv år efter Hitchcocks artikel.

Vi mener ikke at søgeord eller keyword forhindre materialet i at være *findable*. De har naturligvis en positiv effekt. Det er dog væsentligt at overveje graden af denne effekt.

Skabelse af databasen

Struktureret fortid

Første skridt til projektet *Fra Pergament til Python* var indsamlingen af kildemateriale. For at dette kunne lykkes, skulle vi være særdeles innovative og ikke blot anvende vores værktøjskasse, men også udvide den. Ambitionen var at skabe et projekt, som kunne omfavne den middelalderlige periode, og som ikke var orienteret imod specifikke spørgsmål eller undersøgelser. Grundet vores specialisering i middelalderhistorie stod en samling frem for os, som den vigtigste: *Danmarks Riges Breve*. I vores tidligere arbejde med brevene, havde vi transskriberet dem i hånden. En sådan fremgangsmåde ville ikke være bæredygtig til et projekt af denne størrelse. Vi begyndte derfor at undersøge, hvad computeren egentlig kunne klare. Central for vores tilgang blev udviklingen af en metode, som kunne anvende tekstgenkendelsesteknologi. Denne proces begyndte i RStudio men blev i sidste ende afsluttet efter flere uger med adskillige fejl. Det blev nødvendigt at udvide vores værktøjskasse i form af nye programmeringsværktøjer. Efter mange timers onlinetilgængelig undervisning på YouTube, stod det klart at Python ville være et værktøj, som kunne tilpasses til vores ambitioner, da dette program fremstod som værende bedst til tekstudtrækning og dokumentbehandling.

To programpakker blev særligt vigtige i udfoldelsen af dette projekt - *PyMyPDF* og *pandas*. *PyMyPDF* er en grundig og stabil kodepakke til udtrækning af tekst fra pdf-filer, som i vores tilfælde var hele *Danmarks Riges Breve*.⁸¹ *Pandas* benyttes til at samle den udtrukne tekst i et samlet dataframe.

Det blev arbejdet med de mange breve i *Danmarks Riges Breve* samt transskriberingen af *Saxos Danernes Bedrifter*, der blev grundstenene for principperne i vores videre digitaliseringsfase. Da vi først havde skabt grundkoderne, kunne disse eksempelvis overføres til transskribering af *Rigord* og *Innocens III's krøniker*.⁸²

Hvordan fik vi materialet ind i databasen?

Efter vi havde udvalgt vores materiale - en proces som var styret af, hvad der var online tilgængeligt - skulle vi have det digitaliseret. Vi downloadede Claus Rønlevs pdf'er af *Danmarks Riges Breve* og havde derefter behov for effektiv tekstudtrækning, OCR samt behandling af store mængder data. Her viste programmet Python sig som det mest effektive og kraftfulde programmeringsværktøj. Ved at benytte Python effektiviserede og automatiserede vi vores proces markant. En manuel transskribering af de 14.973 breve havde været ekstremt tidskrævende og urealistisk som kildegrundlag til et projekt på universitetet. Derfor skabte vi en metodisk tilgang der, gennem *Optical Character Recognition (OCR)*, kunne omsætte de omkring 20.000 scannede sider fra *Danmarks Riges Breve* til et søgbart dataformat.⁸³ Det indebar en kompliceret proces, da kodeopbygningen skulle tilpasses til materialet. Metoden kunne derfor ikke blot bestå af OCR, da det var nødvendigt for os, som historikere, at undersøge materialet for at opsætte de korrekte parametre.

Danmarks Riges Breve har en kontinuerlig struktur, som er gældende i stort set alle bind fra Det Danske Sprog- og Litteraturselskab.⁸⁴ Hvert brevs overskrift er skrevet i fed skrift,

⁸¹ Alle samlingens bind lå tilgængelig som PDF på Claus Rønlevs hjemmeside <https://www.ronlev.dk/>. Vi ønskede at skabe en database med brevenes indhold, så det var muligt at søge i dem og kategorisere efter afsender, modtager, emne, aktører og geografisk indhold.

⁸² Vi inddrager kodning og transskribering af *Saxos Danmarkshistorie*, på trods af at denne ikke indgår i vores relationelle database speciale_projektet. Dette skyldes at kodningen omkring Saxo er essentiel for vores videre arbejde.

⁸³ Claus Rønlevs samling var OCR scannet, men af en kvalitet som havde alt for mange fejl, til at vi fandt det anvendeligt.

⁸⁴ Det Danske Sprog- og Litteraturselskab begyndte samlingen af *Diplomatarium Danicum* i 1932. "Danmarks Riges Breve", konsulteret 7. april 2025, <https://dsl.dk/publication?id=122>

regesten står altid i kursiv og selve brevet er skrevet i standard skrift. Denne viden omsatte vi til det, der i Python kaldes regulære udtryk.⁸⁵ Ved hjælp af vores kode lærte vi vores computer at filtrere mellem alle disse ting. Denne fremgangsmåde resulterede i en struktureret database af samtlige 35 bind af Danmarks Riges Breve. OCR sikrede læsningen af brevene, hvor de regulære udtryk automatisk fordelte overskriften, regesten og brevet i hver deres kolonne på samme række.

Herefter gennemgik vi databasen for at kvalitetssikre. Vi noterede fejl og implementerede rettelser manuelt og ved brug af computeren. Vi havde eksempelvis problemer med, at henvisning til fodnoter blev lavet til et *x*, og at lange ord som *kardinalpresbyter* eller *jurisdiktionssteder* blev splittet i to. Vi havde også problemer med 1. række bind 1, da dette bind var opsat anderledes end de resterende. Værket indeholder ikke regest og har samtidig flere citater i brevtæksten, som er skrevet i kursiv. Derfor virkede de opsatte regulære udtryk ikke på denne del af samlingen. Vi vurderede her, at det var hurtigst og overskueligt at transskribere de 500 breve manuelt. Dette er et godt eksempel på den svenske professor Johan Jarlbrinks pointe om kvalitetssikring - *Still, such procedures need to be monitored in order to ensure the quality of the outcome*.⁸⁶ Hvis vi ikke havde gennemgået databasen manuelt, havde der været væsentligt flere fejlkilder.⁸⁷

Saxo – Seneste format

Et af vores helt store ønsker for databasen var at inddrage Saxos krønike *Danernes Bedrifter*. For at kunne automatisere denne proces, skulle vi tage et lille skift i forhold til den kodeopbygning, som vi havde brugt til transskriberingen af Danmarks Riges Breve. Vi valgte at anvende Peter Zeebergs oversættelse, som ligger tilgængelig på *Litteraturbanken.se*.⁸⁸ Hertil opsatte vi en kode, som var i stand til at følge URL-kodens kronologi og automatisk downloade

⁸⁵ Regulære udtryk forkortes til RegEx, og består af et bestemt mønster i en bestemt rækkefølge. Et regulært udtryk i brevene kunne eksempelvis være datering i form af år, som altid består af en rækkefølge af tre til fire tal.

Vi brugte årstal som et regulært udtryk i arbejdet med brevene, idet vi benyttede følgende kode: `year_pattern = re.compile(r"(789|79[0-9]|8[0-9]{2}|9[0-9]{2}|10[0-4][0-9]|105[0-1]|11[0-9]{2}|12[0-9]{2}|13[0-9]{2}|140[0-9])"`)

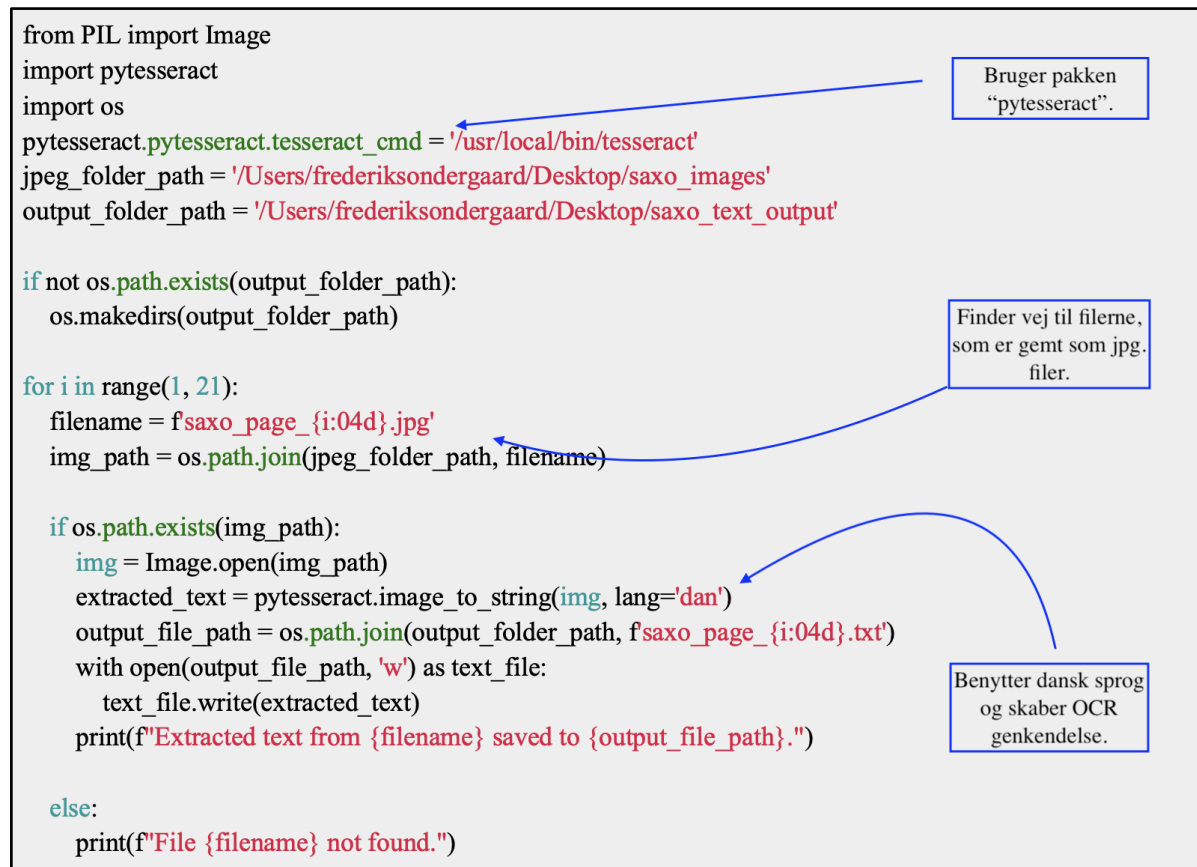
⁸⁶ Jarlbrink, "All the Work that Makes it Work", s. 115

⁸⁷ Da vi arbejder med en stor database, ved vi, at der er fejlkilder i databasen. Det er ikke muligt for os, på den tid som vi har til rådighed til specialet at gennemgå hvert eneste element.

⁸⁸ "Saxo Grammaticus & Peter Zeeberg övers", konsulteret 1. juni 2025, [//litteraturbanken.se/f%C3%B6rfattare/Saxo/titlar/SaxosDanmarksHistorie/sida/6/faksimil](http://litteraturbanken.se/f%C3%B6rfattare/Saxo/titlar/SaxosDanmarksHistorie/sida/6/faksimil)

en side ad gangen og gemme dem på vores computer som en samlet pdf. I nedenstående ses en af koderne, som vi anvendte.

Billede 1) Kodeudsnit med tekstbokse



Ved at køre ovenstående kode på materialet processeres teksten, idet der skabes OCR genkendelse på hver enkelt side. Filerne omdannes derefter til dubletter, som gemmes som en txt.fil. Herefter var vi nødsaget til at tilføje de 903 sider manuelt i Excel, idet uncialerne forhindrede direkte computerstyret transskribering.⁸⁹ Uncialerne var en hindring i en automatisk tekstoverførsel, idet de optræder som et forstyrrende element, da de løber over fire linjer og dermed splitter teksten uhensigtsmæssigt.

Med udgangspunkt i kodningen fra Danmarks Riges Breve og Saxo transskriberede vi yderligere: *Adam af Bremen om Menigheden i Norden*, *Roskildekrøniken*, *Deed of Philip*

⁸⁹ Se bilag. *Fra Pergermanent til Python*, s. 22-23

*Augustus, Deeds af Innocens III samt brevsamlinger fra Cathrine af Siena, Gregory VII og Saint Bernard.*⁹⁰

Vores udvælgelsesproces var styret af tilgængelighed. Vi havde opsat klare parametre for, hvilket materiale vi fandt nødvendigt (DRB og Saxo) og udvalgte derefter, hvad vi kunne få adgang til. Vi har ikke taget stilling til, hvorvidt det indsamlede er repræsentativt for perioden, befolkningsgrupper eller kildetyper. Vi har arbejdet ud fra et håb om, at der senere kunne blive tilføjet yderligere materiale til samlingen. Her er opgavens tidsbegrænsning en afgørende faktor.

Konstruering af metadata

Med databasen samlet kom vi til den måske vigtigste fase, idet vi skulle skabe metadata. Det er ikke tilstrækkeligt kun at have kilderne i deres egen form. For at løfte mulighederne for det analytiske niveau, er det nødvendigt at skabe og tilføje metadata. Metadata kan forsimples til *data about data*.⁹¹ Alle detaljer, som tilføjes til det originale materiale, kan kategoriseres som metadata, idet metadata er alt den nye data om den originale data. Det er derfor både metadata, når vi tilføjer afsender, modtager og ID-nummer til brevene i Danmarks Riges Breve. Det er også metadata, når vi inddeler tekst fra Adam af Bremens *om Menigheden i Norden* i kategorier som fortale og introduktion.

Det er i fasen, hvor skabelsen af metadata finder sted, at historikeren skal tænke sig godt om. Metadataene er vigtige og interessante, da de kan bruges til at beskrive og skabe kontekst i den digitale data, så den kan anvendes på tværs af forskellige analyser. Der er dog forskel på god og dårlig metadata.

I 2020 skrev Kimmo Elo artiklen *Big Data, Bad Data*, hvori han beskriver forståelsen af metadata som et princip. Der findes metadata, som kan være forældet, inkonsekvent og mangelfuld. Det er dårlig metadata, idet det skaber en mulighed for mistolkning og samtidig besværliggør muligheden for at kombinere dataene med andre datasæt. Modsat er god metadata noget, som skaber øget tilgængelighed i materialet og samtidig gør det mere anvendeligt. Det kunne eksempelvis være søgeord eller keywords, som gør dataene lettere at finde på nettet og skaber overblik over materialets indhold. Da vores data er OCR genkendt, er det ikke nødvendigt at tilføje keywords, idet alt tekst er søgbart.

⁹⁰ Se bibliografiske henvisninger til kilder i bibliografi.

⁹¹ Kimmo Elo, "Big Data, Bad Metadata", s. 104 + 108

Vi tilføjede ligeledes “årstal” på hvert brev. Det har vi gjort, fordi vi ser, at en fordeling af brevene både tidsmæssigt og geografisk, vil være relevant i en undersøgelse af højmiddelalderens relation mellem Danmark og Frankrig. Efter vores beslutning om relevante metadata, startede vi processen. Dele af denne tilføjelse skulle gøres manuelt, hvortil andre dele kunne automatiseres. Den korrekte opsætning af årstal, har vi valgt at gøre manuelt, da dette krævede subjektive beslutninger.⁹²

I dette projekt om de danske og franske relationer, ønskede vi at tilføje geografisk metadata til databasen. Vi ville gerne have et defineret sted på *afsender* og *modtager*, samt en liste over de byer, som nævnes i brevene. I nogle af brevene er afsendelsesstedet markeret i overskriften. Vi kunne derfor benytte Python til at udtrække denne metadata. Her skaber vi dog dårlig metadata, idet afsendelsesstedet er markeret på under halvdelen af materialet. Vores metadata bliver derfor ikke konsistent og brugbar. Vi har afgrænset os til perioden 1131 til 1265, for at rammesætte en generation før og efter Ingeborg og Filips ægteskab. Da vi arbejder med 134 år, besluttede vi, at vi ville gøre den dårlige metadata god gennem manuelt arbejde. Vi har derfor læst brevene igennem og markeret sted for afsender og modtager på hvert enkelt brev manuelt. I 292 tilfælde har det ikke været muligt for os at konkludere afsendelsessted eller modtagelsessted, og vi har derfor tildelt brevene NA-værdier.⁹³

Kortlagt kommunikation

Til den geografiske metadata, har vi anvendt vores viden fra tidligere. Vi har gennem de sidste år arbejdet på behandling af geografisk data. Derfor har vi forbedret vores proces gevaldigt, og vores viden indenfor dette område er derfor nu på sit højeste niveau. Vi har haft særlig fokus på at udvikle vores evne til at kode geografiske kort med interaktive brugerflader. Vi har hjulpet

⁹² I flere tilfælde var årstal ikke præciseret. Et brev kan eksempelvis være dateret inden for en bestemt pavens regeringstid. I sådanne tilfælde vil årstal være markeret med firkantede parenteser som eksempelvis: [1213—1233] eller med en beskrivelse som *omkr. år 1170* i brev 1 i 1. række. 3.bind 1170-1199. Vi har konsekvent taget det år, som er benævnt, selvom der er usikkerhed eller med det første år, hvis der har været tvivl om en periode. Dette skaber naturligvis en vis fejlmargen. Vi mener ikke at den mindre usikkerhed om præcis datering vejer højere end det resultat som et “cirka-årstal” kan medvirke til.

⁹³ Vi bruger kontinuerligt NA for manglende værdier. Dette er standard for både RStudio og Python. (...) NA, der betyder “not available” eller “not applicable”, en manglende observation i et datasæt, fx en manglende aflæsning fra et forsøg.

Larsen, Morten & Peter Sestoft “Noter om R”, <https://cran.r-project.org/doc/contrib/Larsen+Sestof-noter-om-R.pdf>

professor ved Aalborg Universitet Johan Heinsen med at skabe geografisk overblik over avismateriale fra Aalborg Stifts Adresseavis fra årene 1818 til 1827, som indeholder 6400 *tætpakkede sider*.⁹⁴⁹⁵

Den bedste og mest præcise tilgang til dette projekt, var at finde stedregistre fra de forskellige bind af Danmarks Riges Breve og manuelt skabe en ny database over disse. På nedenstående *Billede 2) Stedregister med 3886 rækker* ses vores digitaliserede stedregister, som indeholder 3886 rækker med steder og bynavne.

Billede 2) Stedregister med 3886 rækker

372	3001	Patriacum (forsv. lokalitet, Limousin, Frankrig): 267.	
373	3001	Patrinton (Northumberland, England) :440.	
374	3001	Paunat (Périgord, Frankrig) : 348.	
375	3001	Pavia (Italien): 251.	
376	3001	Peene (flod, Tyskland) : 280, 294, 346, 492.	
377	3001	Pepperness (Kent, England): 414.	
378	3001	Périgord (Frankrig) : 307.	
379	3001	Périgueux (Périgord, Frankrig): 111, 348.	
380	3001	Perrecy (Burgund, Frankrig): 201.	
381	3001	Perret (Bretagne, Frankrig): 157.	
382	3001	Pershore (Worcestersh., England): 431, 497,498	
383	3001	Peschiera (Italien) : 223.	
384	3001	Petegem (nu Petegem-aan-de-Leie, Belgien) :31, 124.	
385	3001	Peterborough Abbey (England) : 334, 335,403, 463 note i.	
386	3001	Pfävers (Tyskland) : 460.	
387	3001	Pied-de-Font (Poitou, Frankrig): 317.	
388	3001	Pierrepont (Ile-de-France, Frankrig): 215note i, 341.	
389	3001	Pisa (Italien): 86 note 1	
390	3001	Pitres (Normandiet, Frankrig): 89, 100, 102,121, 122, 123 note, 152.	
391	3001	Plélan (Bretagne, Frankrig): 150.	
392	3001	Poitiers (Poitou, Frankrig): 188, 317.	
393	3001	Poitou (landskab, Frankrig): 147, 174, 296,327, 347-	
394	3001	Pont-Sainte-Maxence (Ile-de-France, Frankrig):91.	
395	3001	Pont-sur-Seine (Champagne, Frankrig): 128.	
396	3001	Poolhampton (Hampsh., England): 438.	
397	3001	Portisham (Dorset, England): 416.	
398	3001	Poulligny (Ile-de-France, Frankrig): 199.	
399	3001	Prinçay (Poitou, Frankrig) : 52.	
400	3001	Provence (Frankrig): 225 note 1, 235.	
401	3001	Prüm kloster (Tyskland) : 66, 98, 202, 293.	
402	3001	Quercy (landskab, Frankrig) : 307.	

Næste skridt til sikringen af korrekt metadata, var at omdanne stedregisteret til brugbart materiale. Klargøringen af vores stedregister skete med det formål at splitte informationen, så det blev mere anvendeligt samt reproducerbart. Vi ønskede derfor at splitte hver reference i fire kolonner i stedet for de to kolonner som på ovenstående billede. Vi anvendte følgende kode til at automatisere denne proces.

⁹⁴ “Fra avissider til segmenterede artikler”, konsulteret d. 1. juni 2025, https://hislab.quarto.pub/step_4.html

⁹⁵ Link til interaktivt kort fra tidligere forløb, opsat og skabt af Frederik Søndergaard: https://hislabaaui.shinyapps.io/Manuelt_kort/. Her følger link til første forsøg skabt i plenum på kurset. <https://hislabaaui.shinyapps.io/Kort/>

Kodeblok 1

```
library(tidyverse)
library(readxl)
library(stringr)

data <- read_excel("Stedregister.xlsx", sheet = 1)
cat("Column names in the Excel file:\n")
print(colnames(data))

# Rensning og splittelse af dataen
cleaned_data <- data %>%
  mutate(
    # Extract Sted (before "(" or ":")
    Sted_New = str_extract(Sted, "^[^\\(:]+") %>% str_trim(),
    # Extract Ekstra (inside "(" if present)
    Ekstra_New = str_extract(Sted, "\\([^\\)]+\\)") %>%
      str_replace_all("[\\(\\)]", "") %>%
      na_if(""),
    # Extract Brev_ref (after ":" if it contains numbers)
    Brev_ref_New = str_extract(Sted, "(?<=:)\\s*[0-9].*$") %>% str_trim(),
    # Use original Brev_ref if no ":" with numbers exists
    Brev_ref_Final = case_when(
      is.na(Brev_ref_New) ~ Brev_ref,
      TRUE ~ Brev_ref_New
    ),
    # Use original Ekstra if no "(" exists, otherwise use extracted
    Ekstra_Final = case_when(
      is.na(Ekstra_New) ~ Ekstra,
      TRUE ~ Ekstra_New
    )
  ) %>%
  # Opsætning af excel fil med eksisterende og nye kolonner
  select(
    Bog,
    Sted = Sted_New,
    Brev_ref = Brev_ref_Final,
    Ekstra = Ekstra_Final,
    # Include Unnamed columns only if they exist
    any_of(c("Unnamed: 4", "Unnamed: 5", "Unnamed: 6"))
  )

# Kort information for gennemtjek af korrekthed

head(cleaned_data)

# Gem den nye fil
library(writexl)
write_xlsx(cleaned_data, "Renset_Stedregister.xlsx")
```

Ovenstående kode ensrettede hele vores stedregister således, at det var mere anvendeligt. Dette skete først og fremmest, da vi splittede den præcise lokation i kolonnen *Sted* og dertil placerede den tilhørende geografiske viden i kolonnen *Ekstra*. Vi får dermed filtreret den enkelte by fra det område, som byen befinder sig i. Det skaber et overblik, som gør det lettere at matche vores resterende arbejde med forskellige byer med kolonnen *Sted*. Opdelingen gør det ydermere overskueligt at gennemgå og manuelt rette de fejlkilder, som Googles geografiske søgeparametre skaber. På nedenstående *Billede 3) Renset stedregister* ses resultatet af den ovenstående kodeblok 1.

Billede 3) Renset stedregister

90	3001	Clandé	296	Touraine, Frankrig	
91	3001	Canterbury	17, 301, 349, 353, 386, 387 note, 414, 418, 42		
92	3001	Catin	64	Bretagne, Frankrig	
93	3001	Chablis	190, 204	Ile-de-France, Frankrig	
94	3001	Chalgrove	300	Oxfordshire, England	
95	3001	Chalkhill	486	Middlesex, England	
96	3001	Chalonais	200	landskab, Frankrig	
97	3001	Chalon-sur	174, 295	Burgund, Frankrig	
98	3001	Chalons-sur	247 note 1	Ile-de-France, Frankrig	
99	3001	Champlitte	162	Burgund, Frankrig	
100	3001	Chappes	96	Ile-de-France, Frankrig	
101	3001	Charente	457	landskab, Frankrig	
102	3001	Charmois	250	landskab, Frankrig	
103	3001	Charroux	19, 317	Burgund, Frankrig	
104	3001	Chartres	94, 278 no	Frankrig	
105	3001	Chateaudu	278 note 1	Ile-de-France, Frankrig	
106	3001	Chaunois	172, 174	Ile-de-France, Frankrig	
107	3001	Chaussay	274	Touraine, Frankrig	
108	3001	Chertsey	435	Surrey, England	
109	3001	Cheselbour	393	Dorset, England	
110	3001	Cheveley	411	Cambridgeshire, England	
111	3001	Chilham	435	Kent, England	
112	3001	Chisledon	295	Wiltshire, England	
113	3001	Choisille	272	flod, Frankrig	
114	3001	Clain	147	flod, Frankrig	
115	3001	Clovesho	2	forsv. lokalitet, England	
116	3001	Cluny	310	Burgund, Frankrig	
117	3001	Combring	498.	Worcestersh., England	
118	3001	Compidgne	81	Ile-de-France, Frankrig	
119	3001		90, 129, 183, 243, 264, 267, 277, 288, 289.		

Billede 3) Renset stedregister ovenfor viser et udsnit af den komplette digitalisering af de bind fra *Danmarks Riges Breve*, som vi har udvalgt til dette projekt. Da hvert bind indeholder et register, er der naturligvis lokationer, som går igen i de forskellige registre. For at fjerne tekniske vanskeligheder og generelt forbedre outputtet i søgningen af geografiske benævnelser, er en oprydning og skabelse af et relationelt stedregister nødvendigt. Kodeblok 2 gennemgår hver linje i stedregistret, og holder dette op imod hver række i databasen. Når vi fjerner de byer, som optræder flere gange i stedregistret, mister vi ikke information. Vi fjerner derimod unødvendig støj, så vi begrænser en unødigt proces i computerens søgning.

Vi opbyggede en kode, som først kunne rydde op i registreret, således at lokationerne fordelt på de 3886 rækker blot nævnes én gang i vores opdaterede register.⁹⁶ Denne oprydning resulterede i en ny fil med 2703 unikke lokationer, der alle optræder i vores kildesamling. Processen bliver dermed effektiviseret, da computeren slipper for en unødigt gennemgang af de 1183 byer, som går igen.

Næste skridt var at gøre disse lokationer brugbare til vores undersøgelse. Vi konstruerede derfor kodeblok 2, der kunne samarbejde med Google Maps søgeflade og derfra udtrække de korrekte koordinater tilhørende hver lokation.

For at skabe et samarbejde mellem Python og Google Maps, benytter vi en API-nøgle. API-nøgle står for *Application Programming Interface Key* og er en form for adgangskode, som giver adgang til brug af bestemt online software eller et online program. API-nøglen bruges til at kontrollere brugers anvendelse af en bestemt tjeneste samt til at beskytte et system mod misbrug.

I vores tilfælde har vi oprettet en API-nøgle gennem Google, som giver begrænset adgang til den software, som Google Maps benytter. Ved at implementere API-nøglen i Python eller R, åbner vi nu op for muligheden for, at de programmer får adgang til Google Maps software.

⁹⁶ Det gør vi med fokus på den tekniske proces, hvor vores database skal matches op mod stedregistret, for at identificere hver gang et sted nævnes.

Kodeblok 2

```
import pandas as pd
import googlemaps
import time

# API adgangskode til brug af Google Maps
API_KEY = "privat nøgle"
gmaps = googlemaps.Client(key=API_KEY)

# Indlæs datasæt med stednavne
file_path = '/Users/Username1/Desktop/Unikke_steder1.xlsx'
data = pd.read_excel(file_path, sheet_name='Sheet1')
unique_cities = data['Sted'].dropna().unique()

# Geokodning funktion
def geocode(city):
    try:
        result = gmaps.geocode(city)
        if result:
            location = result[0]['geometry']['location']
            return location['lat'], location['lng']
    except Exception as e:
        print(f"Fejl ved geokodning af {city}: {e}")
    return None, None

city_coordinates = {}

for city in unique_cities:
    lat, lng = geocode(city)
    if lat and lng:
        city_coordinates[city] = (lat, lng)
    time.sleep(0.1) # Pause mellem forespørgsler

# Lav dataframe med koordinater
coordinates_df = pd.DataFrame.from_dict(city_coordinates, orient='index',
columns=['Latitude', 'Longitude'])
coordinates_df.reset_index(inplace=True)
coordinates_df.rename(columns={'index': 'Sted'}, inplace=True)

# Gem resultatet på skrivebordet
output_path = '/Users/Username1/Desktop/Stedregister.kord.xlsx'
coordinates_df.to_excel(output_path, index=False)

print(f"Data med koordinater er gemt i {output_path}")
```

I kodeblok 2 ovenfor kan man se, hvordan vi har benyttet API-nøglen til at skabe koordinater til vores stedregister.

Kodeblok 2 er opsat til at gennemgå hver række i datasættet, og derefter søge lokationen frem gennem Google Maps. Herefter kan den udtrække de tilhørende koordinater, og tilføje dem til rækken. Når dette er gjort på alle 2703 rækker, samles det nye datasæt og eksporteres. Denne udvikling kan ses på *Billede 4) Stedregister med koordinater*.⁹⁷

Billede 4) Stedregister med koordinater

702	Handest	56,5696	9,8202	
703	Karlby	57,7156	16,506	
704	Skjellerup	56,6276	9,8622	
705	Skortorp	56,4471	12,9615	
706	Gundsømagle	55,7375	12,1512	
707	Halsnæs	55,9875	12,0082	
708	Harløse	55,9191	12,2017	
709	Jyllinge	55,7485	12,1103	
710	Kirkerup	55,3596	11,4908	
711	Sønderby	55,7349	12,0323	
712	Tørslev	55,8105	12,0229	
713	Finderup	56,4131	9,2836	
714	Gerlev	55,3524	11,3375	
715	Gierslev	55,4908	11,3219	
716	Sønderup	56,8124	9,6521	
717	Ølsemagle	55,4995	12,1626	
718	Napoli	40,8518	14,2681	
719	Holmekloster	55,1532	10,351	
720	Søllested	54,8105	11,2705	
721	Ågerup	55,6944	12,1569	
722	Alsønderup	55,9704	12,2309	
723	Annisle	55,9806	12,1728	
724	Helsingør	56,0208	12,1969	
725	Klønne	52,3872	4,8694	
726	Kregme	55,9484	12,0443	
727	Nejeste	55,9578	12,196	
728	Tåstrup	55,646	12,2979	
729	Hals	56,9966	10,3086	

⁹⁷ Trods automatiseringen var det stadig nødvendigt med en manuel gennemgang for fejl. Der var naturligvis fejl i materialet, af forskellige årsager. Vi har tidligere oplevet, at nogle lokationer antages af Google som værende på det Amerikanske kontinent, grundet deres inspiration af forskellige bynavne. Vi gennemgik dem derfor en for en og holdt det op med vores originale stedregister, hvor deres områder også er nævnt, som det ses ved række 384, Petegem (Nu Petegem-aan-de-Leie, Belgien). På den måde kunne vi validere lokationen og rette det. Der var ydermere også lokationer, hvor der blot efterlodt tomme kolonner til længde-og breddegrader, da stederne ikke eksisterer længere. Hertil har vi valgt at tilføje koordinater manuelt, når der var en nutidig lokation nævnt i stedregisteret. Det kan eksempelvis være som i række, 391 med Plélan. I en situation som den ville vi så indhente koordinater på Bretagne, Frankrig.

Årsagen til vi vælger at have så omfattende et fokus på de geografiske elementer i vores projekt, skyldes ambitioner og vores forståelse af mulighederne for dette materiale.

Ligesom Riva argumenterer for at netværksanalyser bør forstås som en metodisk udvidelse, ser vi ligeledes muligheden for at udvide dette til geografiske analyser. Det muliggør, at vi kan løfte de historiske analyser væk fra den klassiske undersøgelse af enkeltstående steder, med særlige begivenheder tilknyttet. Det digitale skaber en mulighed for os til at lave strukturelle undersøgelser, hvor hver lokalitet kan agere som en aktør. Det er nu muligt for os at gøre komplekse geografiske bundne netværk håndgribelige gennem en visuel formidling. Computeren gør det ligeledes muligt at opfange netværk, knudepunkter og relationer, som vi ikke ville være i stand til at formidle og forstå gennem en klassisk tekstnær analyse.

Materialitet i opløsning?

I arbejdet med digitalisering af middelalderkilder har det hele tiden været vigtigt for os at reflektere over den udvikling, som sker med kildens materialitet. Til vores materiale har det en dobbelthed, da kilden allerede har ændret sig. Vi benytter digitaliserede udgaver af trykte kildesamlinger. Det betyder, at den fysiske materialitet allerede har ændret sig og fremstår helt anderledes end samtidens originalmateriale. Det er alligevel relevant at overveje, hvad der sker når der digitaliseres. Vi har valgt at følge kildens naturlige struktur så langt som muligt. Vi gør dette for at overholde Lemerrier og Zalc's principper om transparens og respekt for rådataene. Lemerrier og Zalc opsætter to stadier, som skal følges i begyndelsen af databasekonstruktionen.⁹⁸ Første stadie er den såkaldte *inputfase*, som efterfølges af *kategoriseringsfasen*. Lemerrier og Zalc's principper går i sin helhed ud på, at man i første omgang transskriberer kilderne uden at tage stilling til eller ændres i deres opsætning og struktur. Derefter fordeles de forskellige oplysninger ud i kolonner og kategoriseres. Det er vigtigt at have respekt for kildes oprindelige struktur og indhold, da en for tidlig manipulation modarbejder transparens og skaber et skævt billede af den oprindelige data. Ved at følge disse to trin i den indledende fase, sikrer vi os, at vi altid har dataene i sin oprindelige form, og dermed altid kan vende tilbage til vores udgangspunkt.

I vores tilfælde kan vi heller ikke komme udenom, at det ville have været ulogisk for os at ændre strukturen på dataene. Det var derfor naturligt at skabe en kode, der kunne følge

⁹⁸ Lemerrier & Zalc, *Quantitative Methods in the Humanities*, s. 51-72

og læse den allerede eksisterende opsætning. Det ville ydermere være meget kompliceret at holde overblik over så store mængder data, hvis vi konstant skulle forholde os til en ny visuel opsætning. Det er vigtigt at skabe en sammenhæng mellem opsætning i database, metadata og det oprindelige kildemateriale. For at sikre, at det altid er muligt at finde tilbage til den kilde som vi digitaliserede, benytter vi ID-nummer og ref. Kolonne *ref* refererer til det sted, hvor vi fandt de trykte kildesamlinger, og fungerer som en kvalitetssikrende faktor for eksterne brugere af datasættet.

Devon R. Mordell og det femte paradigme

Grundet størrelsen af vores indsamlede materiale blev der hurtigt sået tvivl om, hvorvidt vi sad med en database eller et egentligt digitalt arkiv. Denne undren opstod på baggrund af Devon R. Mordells artikel *Critical Questions for Archives as (Big) Data*, hvori hun reflekterer over grænsefladen mellem data og arkivalier. Mordell argumenterer for, at det er relevant at anskue arkiver som dynamiske datasæt frem for statiske opbevaringssteder.⁹⁹ Ved at bearbejde arkivernes indhold med digitale værktøjer muliggøres undersøgelser, som kan tiltrække et nyt publikum og samtidig styrke arkivernes relevans.¹⁰⁰

Mordell reflekterer over de digitale værktøjer i samspil med arkiverne, da hun ser en tendens til, at arkiver omtales som Big Data. Det er svært at komme med en præcis definition på, hvad *Big Data* råder over, da der ikke findes en officiel definition af begrebet. Europa-Parlamentet definerer Big Data som (...) *samlede datasæt, der er så store og komplekse, at de kræver nye teknologier, såsom kunstig intelligens, for at kunne blive behandlet*.¹⁰¹ Når der omtales Big Data hentydes der ofte til datasæt, som består af flere varianter af data, og det er derfor naturligt at se sammenhængen mellem Big Data og arkiver. Med afsæt i en sådan forståelse opfordrer Mordell til, at historikere og arkivarer overvejer og tager stilling til, hvad der sker, når arkiverne bliver datafiseret.¹⁰² Ifølge Mordell opstår der det, som hun definerer som et femte paradigme.

⁹⁹ Devon R. Modell, "Critical Questions for Archives as (Big) Data," *Archivaria* 87 (2019): s. 140-161. downloaded 9. November 2024, <https://orcid.org/0000-0003-1020-5176>, s. 142

¹⁰⁰ Mordell, "Critical Questions for Archives as (Big) Data." s. 143

¹⁰¹ "Big data: Definition, fordele og udfordringer (infografik)", konsulteret d. 9. april 2025, <https://www.europarl.europa.eu/topics/da/article/20210211STO97614/big-data-definition-fordele-og-udfordringer-infografik>

¹⁰² Mordell skriver datafied, men der findes ikke et dansk ord.

*I am suggesting, then, that archives as data may be poised to emerge as a fifth archival paradigm (...).*¹⁰³

Med et femte paradigme skaber Mordell et supplement til Cooks arkivteori om de fire paradigmer (*evidence, memory, identity and community*).¹⁰⁴ Mordell definerer det, som hun mener kunne være et femte paradigme, som et *archives-as-data paradigm*.¹⁰⁵

En misforstået forestilling om dette paradigme kan fremme ideen om de upartiske og objektive arkiver. Det ønsker Mordell bestemt ikke. Data, der ofte beskrives som "rådata", kan fremstå som værende uden bias eller påvirkning fra omverdenen. Her kan tankerne ledes mod Cooks første paradigme, hvor arkiverne er passive opbevaringssteder. Dette er dog en misforståelse. At data fremhæves som "rådata" er i sig selv et oxymoron, idet rådata er et produkt af en proces.¹⁰⁶ Data kan i princippet aldrig være rå, da det er blevet skabt og (...) *like archival description and arrangement, they are always representational and the result of interpretive acts*.¹⁰⁷ Algoritmerne er ligeledes påvirket af bias, da de er konstruerede til at lede efter bestemte sammenhænge.

Er skridt tættere på

Den eksplorative tilgang til sidste semester projekt har givet os mulighed for at opbygge en metodisk tilgang til vores materiale, som er tilpasset materialet. Vi har opbygget koder, som passer til Saxo og Danmarks Riges Breves struktur, og som kan tilpasses til lignende materiale.

Digitalhistorien er et felt i konstant udvikling i takt med optimeringen af teknologier og metoder. Dette kunne eksempelvis være en ny og bedre model i Transkribus, som kan skabe OCR med højere succesrate eller kunstig intelligens, som kan forbedre machine learning. Jo mere der arbejdes med de digitale værktøjer, desto bedre bliver de. Det kan derfor være svært at danne et overblik og ikke mindst finde sit ståsted i det digitalhistoriske felt. I 2013 fremlagde Hitchcock eksempelvis en kritik af OCR-teknologien, som vi i dag ser som forældet og

¹⁰³ Mordell, "Critical Questions for Archives as (Big) Data," s. 145

¹⁰⁴ Mordell, "Critical Questions for Archives as (Big) Data," s. 145.

¹⁰⁵ Mordell, "Critical Questions for Archives as (Big) Data," s. 145.

¹⁰⁶ Mordell, "Critical Questions for Archives as (Big) Data," s. 150

¹⁰⁷ Mordell, "Critical Questions for Archives as (Big) Data," s. 150.

uaktuel.¹⁰⁸ Da vi befinder os midt i en digital revolution, flyttes feltet og der skabes konstant forandring.¹⁰⁹

Derfor er det nødvendigt at tage afsæt i en form for fælles regelsæt. Her benytter vi FAIR-principperne fra 2016.¹¹⁰

Interoperabilitet sikres ved at databasen konstrueres, så dataene fungerer i samspil med andre datasæt. Det skal være muligt at kopiere teksten og inkorporere denne i andre datasæt. Ved at tilføje *ID* og *Ref* til de forskellige kolonner, forbedres evnen til samarbejde mellem datasæt, da det er muligt at følge sporet tilbage til originaldataen.

Der er konsensus om, hvordan man producerer et godt stykke digitalt forarbejde. Vi gør vores bedste for at læne os op ad denne forestilling. Det gør vi ved at sikre at databasen er tilgængelig i et format, som er søgbart og giver mulighed for downloads. Det gør ligeledes databasen reusable. At publicere en database er dog ikke en mulighed, før vi har fået godkendelse fra Peter Zeebergs forlag, som har rettighederne til hans oversættelse af Danernes Bedrifter.

Vi ser dog et behov for, at gennemsigtigheden og tilgængeligheden sikres i de videnskabelige udgivelser. Det er primært for at forskningsresultater kan valideres, men lige så meget for at sætte fokus på det usynlige manuelle arbejde. Der ligger hundredvis af timer i kvalitetssikring og rettelser.

Vi placerer os så gennemsigtighed, tilgængelighed, mulighed for reproduktion og samarbejde er i fokus. Derfor beskriver vi vores metodiske tilgang og deler vores koder. Vi gør dette for at være åbne og gennemsigtige i vores proces, og for at gøre det lettere for andre at tilgå og benytte vores database.

Vi finder det ligeledes særligt vigtigt at skabe mulighed for at undersøgelser kan reproducere, for at de kan kvalitetssikres. Fodnoten mister sin autoritet i det digitalhistoriske, da en henvisning til originalkilde eller en kode ikke kan stå alene. Det er nødvendigt at tage læseren med fra start til slut i processen. Forestillingen om at den tekniske del i digitalhistorien kan være en barriere, finder vi ikke relevant. Det må være forskerens ansvar at løfte formidlingen, så teknikken bag undersøgelsen bliver forståelig.

¹⁰⁸ Hitchcock, "Confronting the Digital, Cultural and Social History" s. 13

¹⁰⁹ Hitchcock, "Confronting the Digital, Cultural and Social History", s. 9

¹¹⁰ Jf. Den digitale bibel – Gennemsigtighed og tilgængelighed

Afgrænsning - skabelse af database til specialet

På trods af at vores database dækker utallige kilder fra årene 789 til 1400, er det nødvendigt at afgrænse kildematerialet, så det passer til undersøgelsen. Til dette projekt ønsker vi at arbejde med relationerne mellem Danmark og Frankrig omkring og efter skandalen om Ingeborg fra 1193 til 1213. Derfor afgrænser vi undersøgelsen til tidsperioden fra 1131 til 1265, så vi rammesætter til generationen før og efter Ingeborg. Dette projekt bliver ikke en undersøgelse af Filip Augustus og Ingeborgs ægteskab. Vi anvender i stedet Ingeborg som en katalysator, der dikterer en relation mellem Danmark og Frankrig. På det grundlag finder vi det interessant at undersøge, hvordan relationen er mellem de to nationer.

Idet vi afgrænser os, skaber vi en relationel database, som vi kalder *Speciale_database*. En relationel database er struktureret i kolonner og rækker, og da den indeholder et unikt ID-nummer, sikre det dataintegritet. Dermed vil det altid være muligt for os at tilgå den oprindelige data i den samlede database. Vi vælger i denne sammenhæng at skabe en database, som kun benytter kilder fra Danmarks Riges Breve. Vores relationelle database *Speciale_database* dækker dermed de 134 år fra 1131 til 1265, og består af 2002 individuelle breve.¹¹¹

Kapitel 3 - Analyse

Formålet med vores analyse er at afdække, hvordan relationen mellem Danmark og Frankrig fremstår i vores digitale tekstkorpus, som dækker perioden 1131-1265. Der er som tidligere belyst essentielle metodikker i arbejdet med digitalhistorie i form af transparens og reproducerbarhed, som vi ligeledes ønsker at efterkomme. De løbende analytiske nedslag vil derfor fortsat suppleres af de tekniske fremgangsmetoder, som vi anvender.¹¹²

Vi bevæger os løbende mellem forskellige programmeringsværktøjer, herunder RStudio og Python. Begge universer har elementer, som påvirker vores proces både negativt og positivt. Python fremstår særdeles effektivt til de større databehandlinger og strukturelle elementer af databasen. Vi ønsker at belyse dette, inden vi bevæger os dybere ned i analysen af vores materiale. I tilblivelsen af vores database var Python helt essentiel, da dette univers, som

¹¹¹ Se database i bilag.

¹¹² Alle koder kan ses i bilag i Script.

tidligere beskrevet, hjalp os til at automatisere en komplet gennemgang af *Danmarks Riges Breve*.

Igennem vores arbejde, som er særligt funderet i det eksplorative, kan det være nødvendigt at splitte databasen op. Det kunne eksempelvis være for at forfølge en undren om bestemte relationer til paven, eller en afgrænsning til breve med mindst 1000 ord. I denne sammenhæng er RStudio særdeles anvendeligt, idet programmet er godt til at splitte, bearbejde og krydse forskellige datasæt.

Da vi ønsker at undersøge den dansk-franske relation i højmiddelalderen, ville det være utænkeligt ikke at fokusere på de geografiske elementer i materialet. Det har været en klar prioritering for os at sikre, at denne information var at finde i vores materiale, da det forhåbentligt vil skildre en relation anderledes end de skrevne ord. Vejen hertil er lang, hård og fordelt i adskillige step-by-step-processer. Vi begyndte ved manuelt at tage samtlige 4004 lokationer for afsender og modtager. En start, vi ser som logisk, da de individer, som skrev sammen uomtvisteligt kan antages at have en relation på den ene eller den anden måde. Det er dog interessant i forhold til vores senere undersøgelse at se hvorvidt disse relationer er bundet op omkring geografiske hotspots, eller om der er forskellige afvigelser samt eventuelle udviklinger, da vores materiale strækker sig over 134 år. Det var denne logiske start, som ligeledes fik os til at anerkende andre ambitiøse muligheder, som lå i det materiale, som vi snart havde arbejdet med i et år.

Som vi tidligere har belyst, gennemgik vi manuelt de tilhørende stedregistre til de bind, som vi anvender til dette projekt. Derfra opsatte vi en kode, der byggede på det samlede stedregister og den information, som lå deri. I stedet for blot at fokusere på den viden, som vi kunne se i overskriften af brevene og regesten, forsøgte vi at køre den igennem kolonnen *text*. Dette gjorde vi først og fremmest for at se, om både Python og vores computer var i stand til det.¹¹³ Det udfoldede sig lykkeligt og gav et afkast i form af alle nævnte geografier, som var nævnt i stedregistret, nu var tagget i en ny kolonne *found_city*.¹¹⁴

¹¹³ Koden gennemgik hele stedregistret i forhold til hver eneste række i databasen. Vi var dog i tvivl om materialet var for stort, og om det derfor ville være for "tung" en kode at køre, idet alle 2703 geografier skulle køres op imod alle 512.377 ord i brevene. Der var 1,38 milliarder potentielle stedordskombinationer. Efter koden havde kørt i fire timer, var den nået frem til et endeligt resultat.

¹¹⁴ Fra vores tidligere bearbejdelser af vores stedregister, havde vi allerede indsamlet tilhørende koordinater til alle de geografiske lokationer. De blev derfor også tilføjet som ny metadata i *speciale_database*, i form af en ny kolonne, *found_city*.

Grundstenene

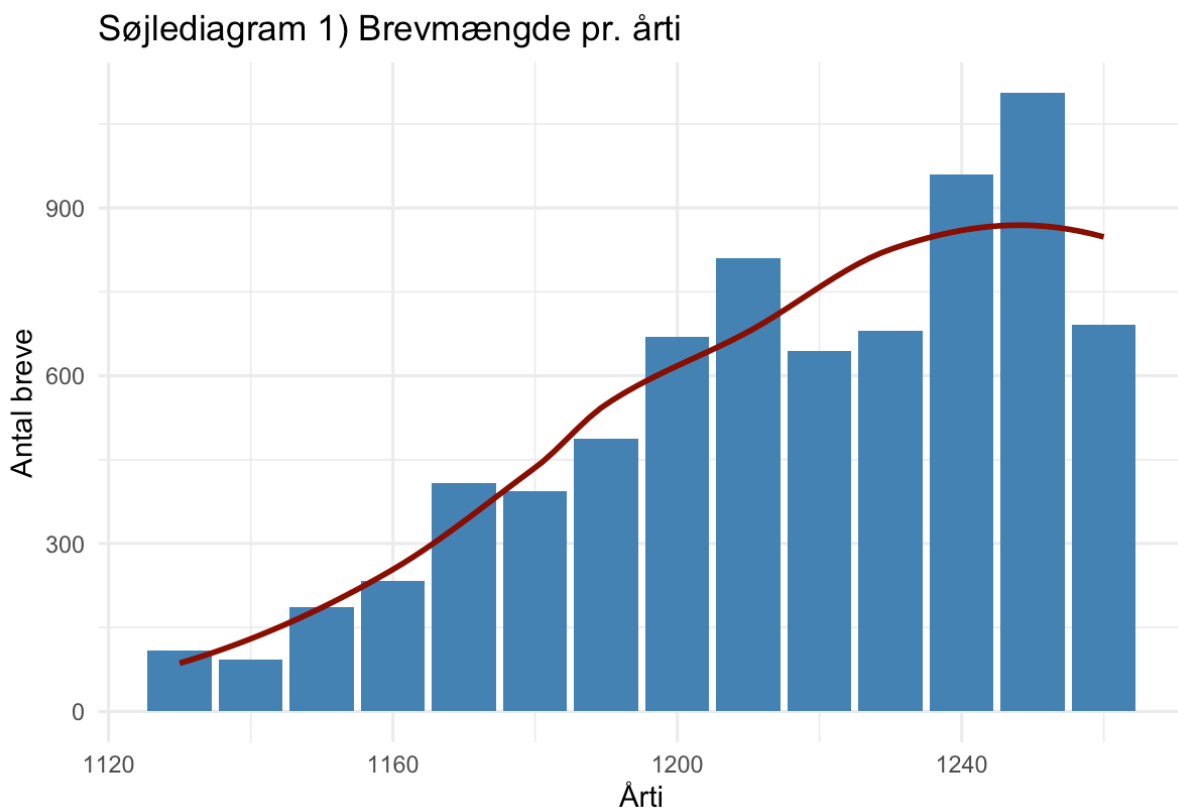
For at skabe et overblik over databasens indhold, indleder vi med at visualisere fordelingen af de 2002 breve. Vi skaber først en dataframe, som ser kolonnen *årstal* numerisk, og skaber derefter et ggplot, som er søjlediagrammet, der ses nedenfor. Vi har gemt vores database under navnet *df*, da denne forkortelse på sigt gør arbejdet mere effektivt.¹¹⁵

Kodeblok 3

```
data <- df %>%
  mutate(årstal = as.numeric(årstal)) %>%
  filter(!is.na(årstal))
brev_per_årti <- data %>%
  mutate(årti = floor(årstal / 10) * 10) %>%
  count(årti, name = "antal_breve")
#HUSK NAVN
ggplot(brev_per_årti, aes(x = årti, y = antal_breve)) +
  geom_col(fill = "steelblue") +
  geom_smooth(method = "loess", se = FALSE, color = "darkred") +
  theme_minimal() +
  labs(
    title = "Søjlediagram 1) Brevmængde pr. årti",
    x = "Årti",
    y = "Antal breve"
  )
```

På *Søjlediagram 1) Brevmængde pr. årti* ses fordelingen af brevene over tid. Hver søjle repræsenterer et årti, og den røde linje fremhæver den generelle tendens på grafen.

¹¹⁵ *df* står for dataframe, og benyttes egentlig kun fordi det er nemmere at skrive *df* i koden end *Speciale_database*. Igennem opgaven vil vi ofte forkorte materialet i R, da dette effektiviserer processen.



På *Søjlediagram 1) Brevmængde pr. årti* kan vi se, at der med tiden sker en generel vækst i antal breve. Vi kan ydermere se, at der sker en stigning i 1200-1219, samt en markant stigning fra 1240-1259. Den generelle stigning kan skyldes mængden af bevaret materiale, da dette kan have påvirket sammensætningen af værket *Danmarks Riges Breve*. Det originale brevmateriale er i de fleste tilfælde gået til, og værket er derfor sammensat af kopier af breve eller kopier af kopier. Det er af denne grund ikke muligt at konkludere præcist, hvor mange breve som blev sendt i perioden. Vi arbejder ud fra brevsamlingen, og er her opmærksomme på at visse redaktionelle valg er taget, da værket blev sammensat. Riva skriver i denne sammenhæng,

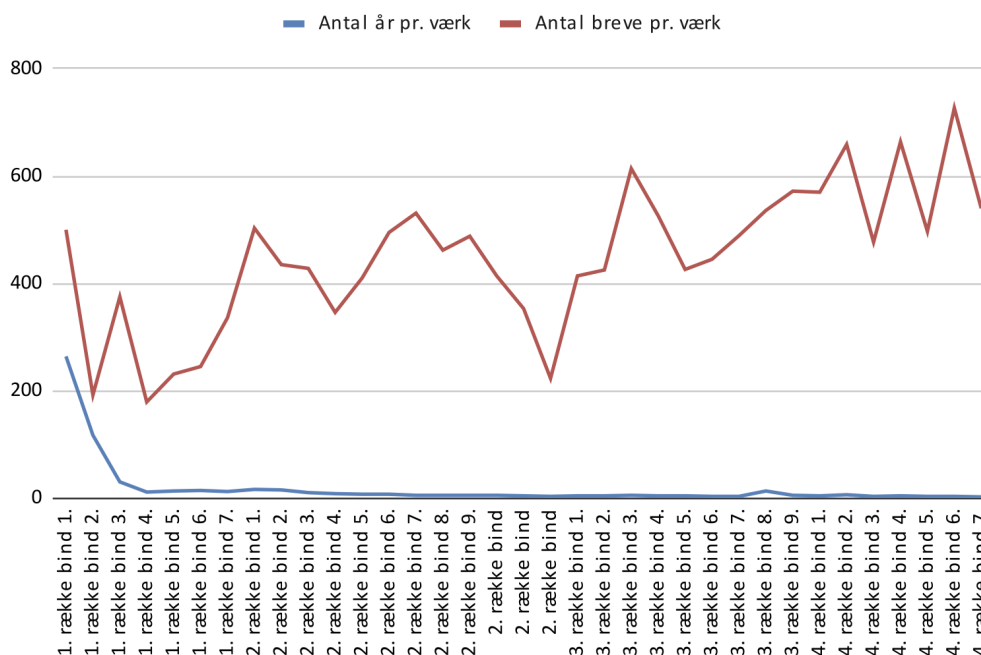
*This editorial decision is perfectly valid for a reference work, but distorts the analysis proposed here. The resulting network will be profoundly affected by these editorial decisions.*¹¹⁶

¹¹⁶ Riva, "Network Analysis of Medieval Manuscript Transmission.", s. 36

Vi er opmærksomme på, at vores svar ligeledes vil være påvirket af de redaktionelle valg, som ligger bag skabelsen af Danmarks Riges Breve.

For at undersøge, om der sker en generel vækst i antal bevaret materiale, undersøger vi fordelingen af alle 15.727 breve i samlingen Danmarks Riges Breve. På *Graf 2) Fordeling af breve i Danmarks Riges Breve* nedenfor ses fordeling af alle breve i de 35 bind af Danmarks Riges Breve markeret med rødt. Med blå er det antal år, som værket dækker, fremvist. Hvert bind dækker i gennemsnit 17,75 år, og indeholder 449,3 breve

Graf 2) Fordeling af breve i Danmarks Riges Breve



På *Graf 2) Fordeling af breve i Danmarks Riges Breve* kan vi se, at der er stor forskel på hvor mange år, de forskellige bøger dækker, da den første bog dækker 263 år, og den sidste kun dækker 2 år. Vi ser ligeledes en tendens til, at der bevares flere breve med tiden. På baggrund af graf 2 må vi udlede, at der løbende sker en vækst i antal gemte breve, og at dette ligeledes gør sig gældende for vores udvalgte materiale. Om dette er en faktisk repræsentation af mængden af kommunikation, kan vi kun gisne om.

Den eksponentielle stigning, som vi ser på *Søjlediagram 1) Brevmængde pr. årti* omkring år 1200, havde vi dog forventet at se. Dette baserer vi på indsættelsen af Pave Innocens III i år 1198, da hans pontifikat traditionelt bliver betragtet som (...) *the climax of papal power in the*

Middle Ages (...).¹¹⁷ Innocens III's eftermæle er særligt bundet op på pavedømmets administrative udvikling. I artiklen *Power, celebration and circuits of legitimation: the local use of papal letters in late twelfth-century Denmark* argumenterer Emil Lauge Christensen, Kim Esmark og Iben Fonnesberg-Schmidt for, at pavebreve i Innocens III's samtid spillede en afgørende rolle i opbygningen af autoritet.¹¹⁸ Brevene blev både anvendt til at understrege pavens nærvær og til at styrke de lokale magthavere. Ved at fremvise et pavebrev offentligt, fungerer dette som bevis på pavens støtte til kongen, hvilket bidrog til legitimering af kongens magt. Pavebrevet skabte samtidig en fysisk forbindelse mellem afsenderen og modtageren, hvorved paven fremstod nærværende og håndgribelig for de troende. Brevenes rolle i opbygning samt styrkelse af autoritet kan være årsagen til, at vi ser en stigning i breve fra Innocens III's første pontifikatår.

Pavens brevspind

For at undersøge nærmere, om det netop er pavebrevene, som skaber udsving i materialet, filtrerer vi dem fra i databasen, og skaber endnu en relationel database. Ved en simpel manuel gennemgang af databasen, kan vi konkludere, at der optræder 18 paver i vores materiale.¹¹⁹ De 18 paver deler i flere tilfælde navne, og det er derfor kun nødvendigt at definere 12 pavenavne. På *Billede 5) Kodeudsnit udtræk af pave-data* nedenfor, kan det ses hvordan vi skabte den relationelle database med pavebreve, som vi har kaldt *pavebreve*. Vi definerer først de 12 forskellige pavenavne, hvorefter vi skaber et *pave_mønster*. Her søges der efter ordet *Pave*, efterfulgt af de 12 navne, uden forbehold for store eller små bogstaver. Den lille hat ^ før *Pave* definerer, at ordet skal staves med stort P. Da vi filtrerer ud fra regesten, ved vi, at teksten altid indledes med aktører efterfulgt af handling. Eksempelvis *Abbed Bernard af Clairvaux skriver til pave Eugenius 3.* eller *Pave Hadrian 4, tager Tommerup kloster under sin beskyttelse*. Ved at søge efter *Pave* med stort begyndelsesbogstav i regesten, får vi dermed

¹¹⁷ Hoppenbrouwers, Peter & Blockmans, Wim, *Introduction to Medieval Europe 300-1500*, Routledge, third edition, 2014. s. 198.

¹¹⁸ Christensen, Emil Lauge, Kim Esmark og Iben Fonnesberg-Schmidt, "Power, celebration and circuits of legitimation: the local use of papal letters in late twelfth-century Denmark," *Journal of Medieval History* 49:3 (2023): downloaded 4. januar 2024, DOI: 10.1080/03044181.2023.2208906, s. 362

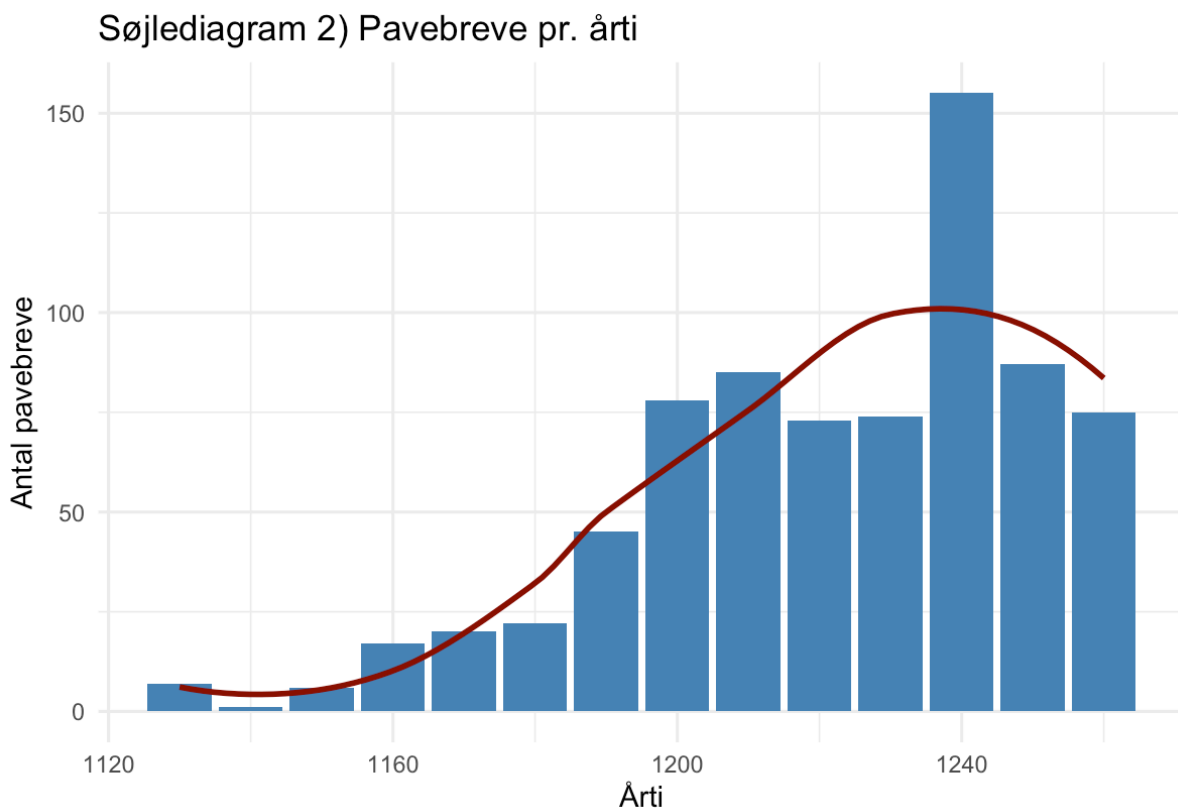
¹¹⁹ I materialet finder vi: Anaclet II (1130-1138), Innocens II (1130-1143), Eugenius III (1145-1153), Anastasius IV (1153-1154), Hadrian IV (1154-1159), Alexander III (1159-1181), Lucius III (1181-1185), Urban III (1185-1187), Gregor VIII (1187), Clemens III (1187-1191), Cølestin III (1191-1198), Innocens III (1198-1216), Honorius III (1216-1227), Gregor IX (1227-1241), Innocens IV (1243-1254), Alexander IV (1254-1261), Urban IV (1261-1264) og Clemens IV (1265-1268).

registreret de breve, hvor det er en pave der optræder som aktør, og ikke omtales eller modtager brevet.

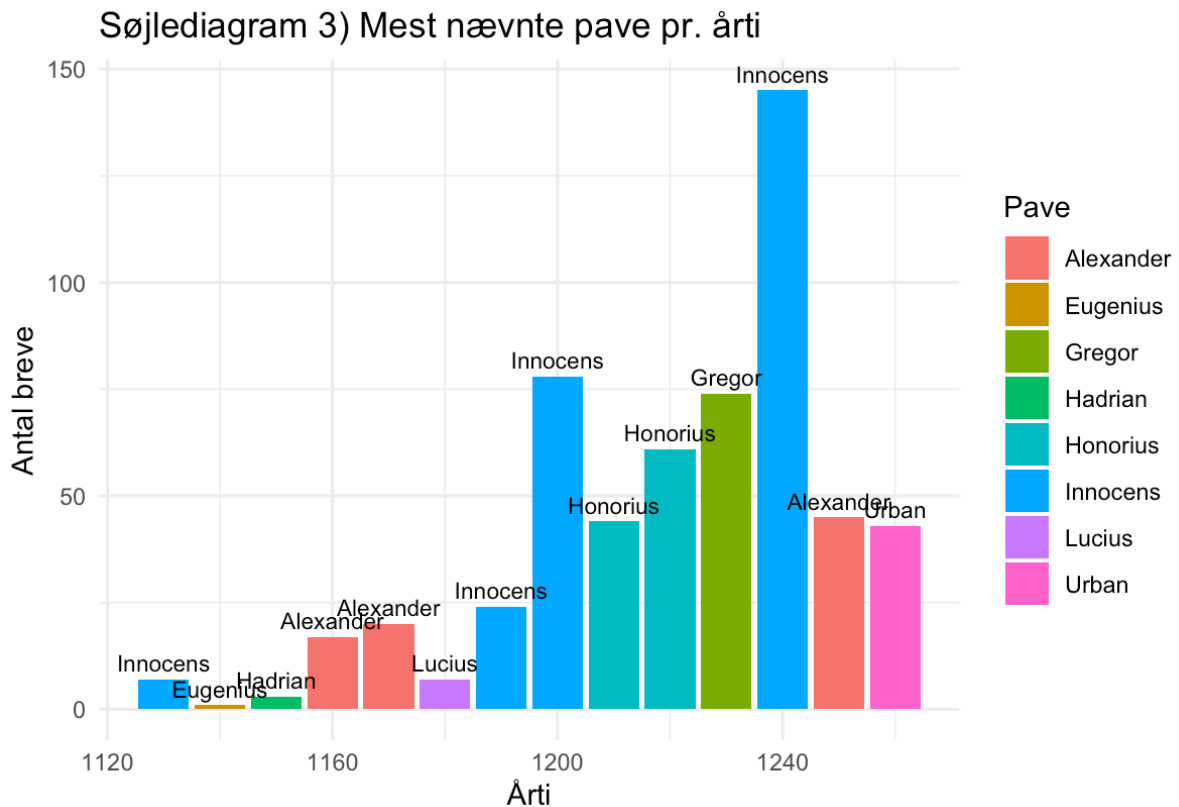
Billede 5) Kodeudsnit udtræk af pave-data

```
31  
32 names(data)[names(data) == "regist"] <- "regist"  
33  
34 paver <- c("Alexander", "Gregor", "Honorius",  
35           "Anaclet", "Innocens", "Eugenius", "Anastasius", "Hadrian", "Lucius",  
36           "Urban", "Clemens", "Cølestin")  
37  
38 pave_mønster <- paste("^Pa\\s+", paste(paver, collapse = "|"), "\\b")  
39  
40 pavebreve <- data %>%  
41   filter(grepl(pave_mønster, regist, ignore.case = TRUE)) %>%  
42   mutate(årstal = as.numeric(årstal)) %>%  
43   filter(!is.na(årstal))  
44
```

På baggrund af ovenstående kode, har vi nu skabt *pavebreve*, som indeholder 745 breve. På *Søjlediagram 2) Pavebreve pr. årti* nedenfor ses fordelingen af de 745 pavebreve over tid. De er ligesom på graf *Søjlediagram 1)* fordelt efter årti. Her kan vi se en tydelig sammenhæng mellem stigningerne på *Søjlediagram 1)* og stigningerne på *Søjlediagram 2)* nedenfor. Vi kan hermed udlede, at der er en sammenhæng mellem stigningen i brevmateriale og pavebrevene.



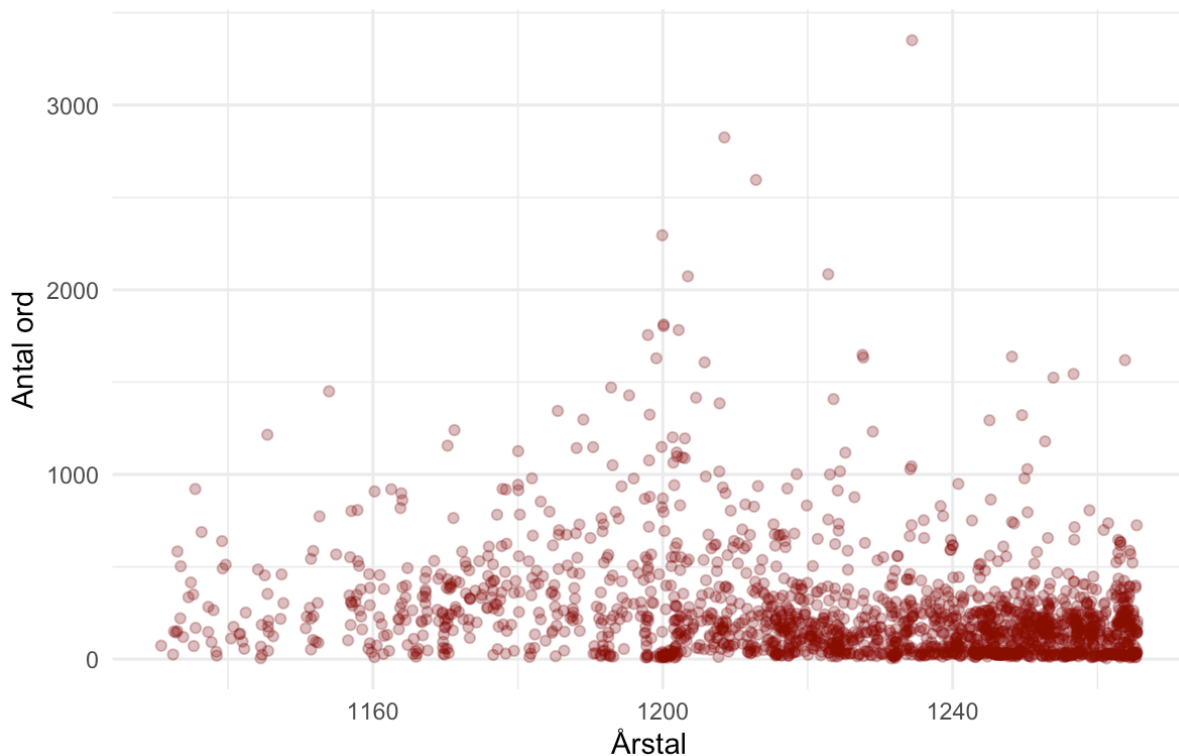
For at undersøge nærmere skaber vi et søjlediagram, som viser fordelingen af de hyppigste afsendere for hvert årti. Her har vi ikke fundet det nødvendigt at tilføje regentnummer til de enkelte paver, da grafen har et tidsmæssigt perspektiv, og vi derfor kan benytte egen viden til at udlede, hvilken pave som er repræsenteret på grafen. En tilføjelse af regentnummer ville ydermere gøre filtreringen af pavebrevene mere kompliceret, da vi ville være nødsaget til at opsætte flere regulære udtryk.



På *Søjlediagram 3) Mest nævnte pave pr. årti* kan vi se, at Innocens IV, som er pave fra 1243 til 1254, er den absolut mest aktive afsender i vores brevsamling. Under hans pontifikat sker der en markant stigning i pavens repræsentation i vores materiale. Vi finder dette interessant og ønsker at se nærmere på paverne og særligt Innocens IV's betydning for de danske relationer i den senere del af analysen.

Først vil vi dog fortsat skabe et grundlæggende overblik over databasens samlede materiale. Databasen indeholder som beskrevet 2002 breve, som er fordelt eksponentielt over 134 år. De 2002 breve har tilsammen et tekstkorpus på 512.377 ord. Fordelingen af ord i de forskellige breve ses på *Diagram 1) Ordantal pr. brev over tid* nedenfor. Her kan vi se, at der er flest breve med 1000 eller færre ord. Vi kan ydermere udlede, at der med tiden sker en koncentration af breve med 500 eller færre ord. Hertil tolker vi, at dette skyldes mængden af materiale, da der med tiden optræder flere breve i samlingen. Vi kan ligeledes se, at de længste breve er fokuseret omkring år 1200.

Diagram 1) Ordantal pr. brev over tid



Vi finder det relevant at undersøge længden på brevene, da denne information er essentiel for den senere undersøgelse. Det skaber en mulighed for, at vi kan undersøge, om der er en sammenhæng mellem længden på brevene og deres indhold. En sådan sammenhæng kan være med til at give udtryk for eksempelvis konflikter eller diplomati.

Breve, byer og bånd

Da det netop er i en geografisk analyse, at vi ser at de digitale metoder har muligheder for at bidrage med nye perspektiver, har vi naturligvis undersøgt geografisk data i brevsamlingen.

Vi har gennem følgende kodeblok 4 skabt et kort, hvor afsender og modtager i de 2002 breve er visualiseret. Vi har skabt kortet i RStudio med pakkerne *dplyr* og *leaflet*. Pakken *dplyr* er essentiel, da den indeholder funktioner, der kan filtrere rækker, vælge kolonner, gruppere data samt flette tabeller sammen. *Leaflet*-pakken benyttes til at opsætte interaktive kort, og skaber derfor visualiseringen af den data, som vi har i de forskellige kolonner.

Kodeblok 4

```
library(dplyr)
library(leaflet)

# Fjern dubletter (ét brev pr. doc_id)
data_unique <- data %>%
  distinct(doc_id, .keep_all = TRUE)

# Opsummer afsendere
afsendere <- data_unique %>%
  group_by(afsender_sted, afsender_breddegrad, afsender_længdegrad) %>%
  summarise(antal_afsendt = n(), .groups = "drop")

# Opsummer modtagere
modtagere <- data_unique %>%
  group_by(modtager_sted, modtager_breddegrad, modtager_længdegrad) %>%
  summarise(antal_modtaget = n(), .groups = "drop")

# Kort
leaflet() %>%
  # Farverigt baggrundskort
  addProviderTiles(providers$Esri.WorldStreetMap) %>%

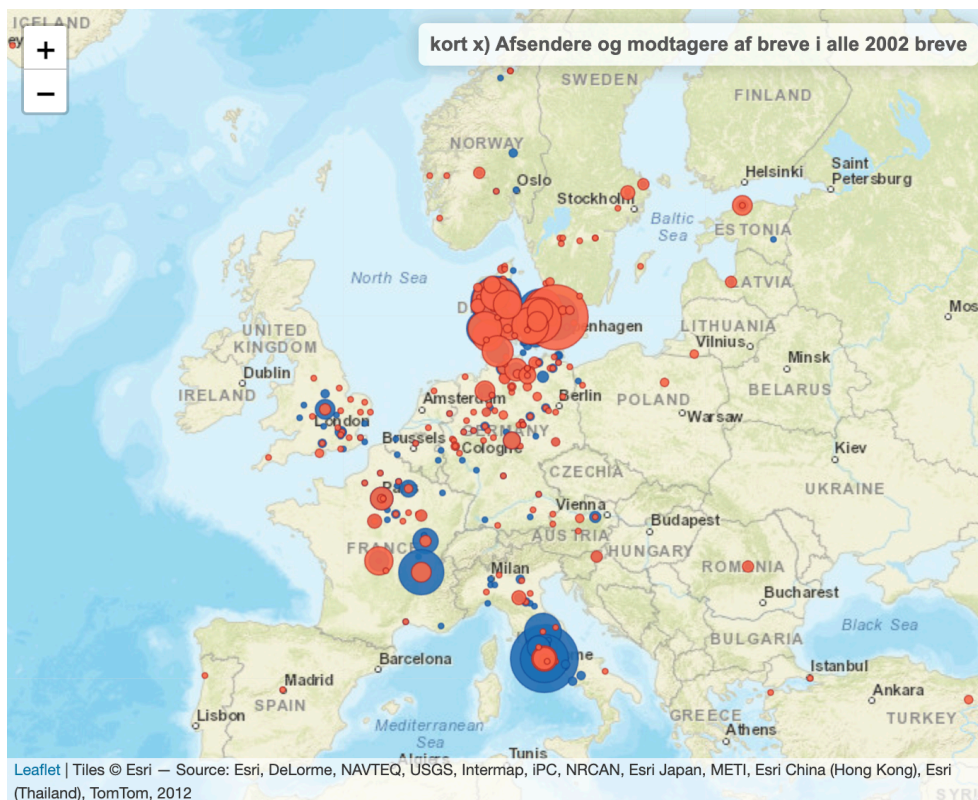
  # Afsender-prikker (mørkeblå)
  addCircleMarkers(
    data = afsendere,
    lng = ~afsender_længdegrad,
    lat = ~afsender_breddegrad,
    radius = ~sqrt(antal_afsendt) * 1.5,
    color = "#084594", # mørkeblå
    fillColor = "#2171b5", # lysere blå fyld
    fillOpacity = 0.9,
    stroke = TRUE,
    weight = 1,
    popup = ~paste0(
      "<strong>Afsendersted:</strong> ", afsender_sted, "<br>",
      "<strong>Antal afsendte breve:</strong> ", antal_afsendt
    )
  ) %>%

  # Modtager-prikker (rød)
  addCircleMarkers(
    data = modtagere,
    lng = ~modtager_længdegrad,
    lat = ~modtager_breddegrad,
    radius = ~sqrt(antal_modtaget) * 1.5,
    color = "#99000d", # mørkerød
```

```
fillColor = "#fb6a4a", # Lysere rød fyld
fillOpacity = 0.9,
stroke = TRUE,
weight = 1,
popup = ~paste0(
  "<strong>Modtagersted:</strong> ", modtager_sted, "<br>",
  "<strong>Antal modtagne breve:</strong> ", antal_modtaget
)
) %>%

# Titel
addControl(
  html = "<b>Kort 1) Alle afsendere og modtagere af breve (2002)</b>",
  position = "topright"
)
```

Dette resulterer i *Kort 1) Alle afsendere og modtagere (2002)*, som kan ses nedenfor. Afsendere er markeret med blå, modtagere er markeret med rød og prikkerne er hertil justeret efter antal repræsentationer.



På *Kort 1) Alle afsendere og modtagere (2002)* fremgår det tydeligt, at det er Rom og Lyon, som er de dominerende positioner for afsendelse, og at Lund er en væsentlig modtagerlokation. Ud fra kortet kan vi udlede, at der er bevaret et stort antal breve, som er adresseret til danske

modtagere. Det fremgår tydeligt, at der er flere røde prikker end blå. I de 2002 breve kan vi ligeledes se, at der er en generel tendens til, at Danmark orienterer sig sydpå.

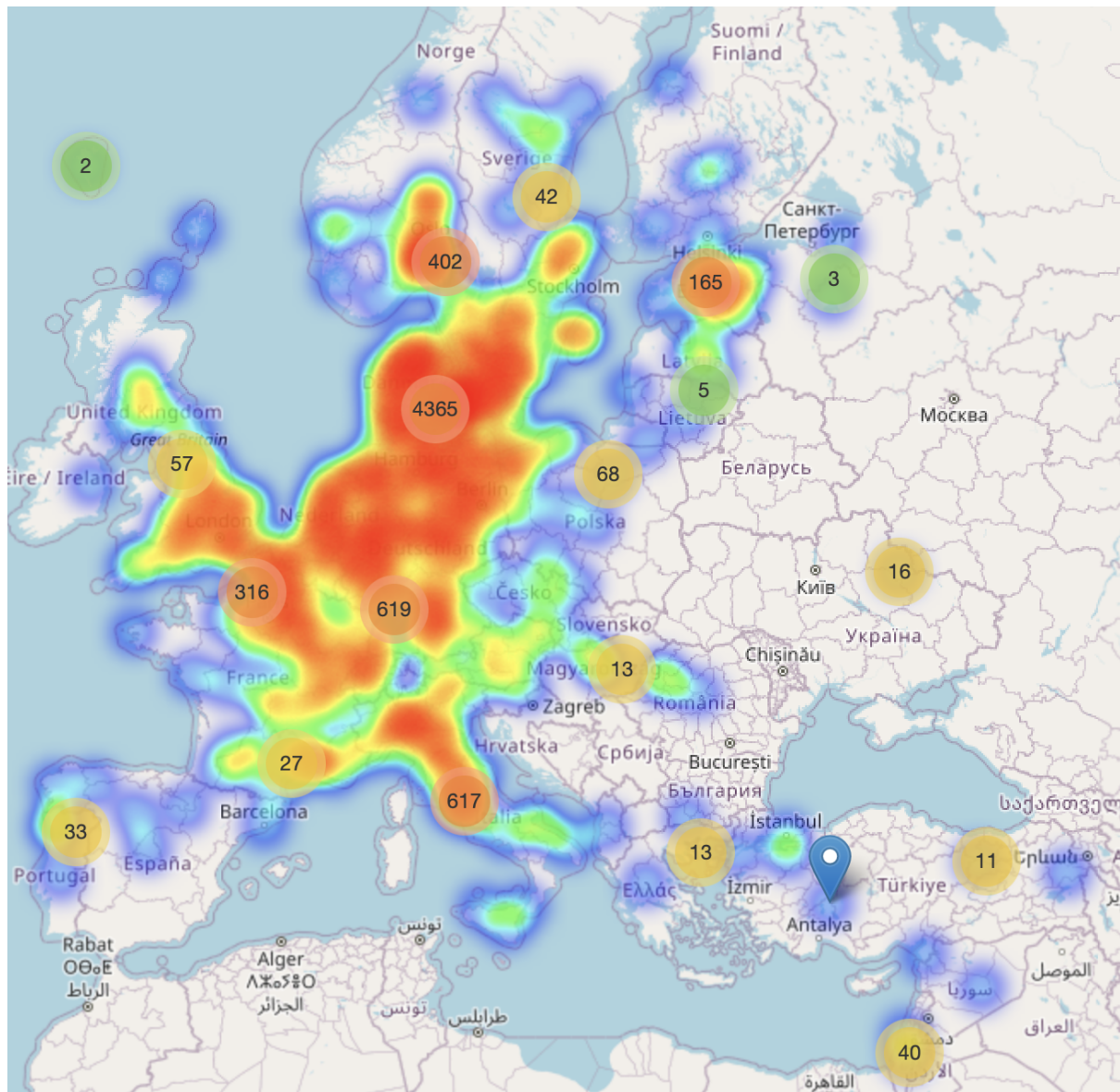
Da vi har arbejdet på at skabe metadata, som relaterer sig til de geografiske elementer i brevmaterialet, kan vi udvide denne geografiske vinkling. Vi har *afsender_sted*, *modtager_sted* og tilknyttede koordinater. Denne data er visualiseret i *Kort 1) Alle afsendere og modtagere (2002)* ovenfor.

Vi ønsker at teste den eksplorative tilgang, og forsøger derfor at skabe et interaktivt kort, som viser alt geografisk information i brevene. Dette gør vi ved først at etablere stedregister og derefter tilknytte koordinater til byerne i registeret.¹²⁰ Herefter matchede vi gennem computeren alle byer i stedregisteret med alle ord i de 2002 brevttekster. Hvis der opstod et match, blev denne by placeret i kolonnen *found_city*. Vi benyttede herefter den førømtalte API-nøgle til at knytte koordinater til byerne i kolonnen *found_city*.¹²¹ Hermed havde vi skabt grundlaget for vores heatmap, som indeholder alle byer der optræder i brevtteksten i de 2002 breve. Heatmappet visualiserer dermed den samlede geografiske berøringsflade, som er tilknyttet materialet.

¹²⁰ JF afsnit om stedregister.

¹²¹ JF om API

Heatmap 1) Samlet geografisk berøringsflade



Ved at se brevenes geografiske indhold visualiseret på *Heatmap 1) Samlet geografisk berøringsflade*, skabes der et håndgribeligt overblik over den geografiske orientering i perioden. Heatmappet er et udtryk for, hvordan man i Danmark har information som orienterer sig fra Grønland til Jerusalem.¹²²

Vores *Heatmap 1) Samlet geografisk berøringsflade* er skabt i en html.fil og fungerer interaktivt. Det er derfor muligt at navigere rundt i mappet online og zoome helt ned på

¹²² Vi har zoomet ind på vores heatmap, så kortet er til at tyde i opgaven.

byniveau, for at undersøge de respektive kilder. Det er muligt at klikke på et punkt og få et pop-up af det brev, som er repræsenteret gennem punktet.

Topic Modeling

For at undersøge indholdet i brevene, kan vi udføre en undersøgelse, som kaldes topic modeling. Topic Modeling kan laves i RStudio, og er en måde hvorpå man kan undersøge potentielle emner i tekstbaserede datasæt. Det er særligt brugbart på store datasæt, da det kan benyttes til at skabe en overordnet forståelse for materialets indhold. Topic Modeling i RStudio fungerer gennem pakken *topicmodels*. Som en del af denne pakke benytter vi LDA-modellen (*Latent Dirichlet Allocation-model*).¹²³ En LDA-model kan finde skjulte emner i tekst ved hjælp af sandsynlighedsregning.

*In machine learning and natural language processing topic models are generative models which provide a probabilistic framework for the term frequency occurrences in documents in a given corpus.*¹²⁴

LDA er en generativ *language processing topic model*, hvilket betyder, at den skaber ny data, som i vores tilfælde er bud på emner i brevtæksten. Grundideen med LDA-modellen er, at den antager, at alle dokumenter kan have flere emner. Modellen er probabilistisk, og den tager derfor forbehold for sandsynlighed. Det kan både være forvirrende og udfordrende at begynde arbejdet med LDA-modeller. For at gøre det mere håndgribeligt, følger der her et eksempel både for vores egen og læserens skyld.

Eksempel: LDA-modellen benyttes på et brev i vores brevsamling, og konkluderer, at der er 75% sandsynlighed for at brevet handler om krig, 10% sandsynlighed for at det handler om religion og 15% sandsynlighed for, at det handler om politiske relationer. Fordi modellen er probabilistisk, tænker den de forskellige emner i grader i stedet for at markerer enten/eller. Brevet bliver derfor noteret til at handle mest om krig, men

¹²³ Grün, Bettina & Hornik, Kurt, "Topicmodels: An R package for Fitting Topic Models", <https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf> s. 2-3

¹²⁴ Grün & Hornik "Topicmodels", s. 1

også som omhandlende religion og politiske relationer. En deterministisk model ville udelukkende markere brevet som omhandlende krig.

LDA-modellen er derfor god til mere komplekse tekster, da disse ofte kan indeholde flere emner.¹²⁵

Modellen benytter en VEM-algoritme (Variational Expectation-Maximization), som er deterministisk. Deterministiske modeller giver det samme resultat hver gang, da de arbejder ud fra en fast formular. Dette er modsat de stokastiske modeller, som beregnes på baggrund af tilfældig sampling, og som derfor får et nyt resultat hver gang. Vi benytter VEM, fordi den er deterministisk og derfor taler ind i gennemsigtighed og mulighed for at genskabe resultater. En væsentlig årsag er ligeledes, at VEM-algoritmen er hurtigere end den stokastiske Gibbs-algoritme, da Gibbs-algoritmen benytter tilfældig sampling og dermed arbejder væsentligt langsommere. Med materialets størrelse har vi ikke tid til at vente flere dage på udregningen gennem Gibbs-algoritmen.

For at kunne udføre topic modeling, fjerner vi først stopord fra brevene. Stopord er ord, som man kan kategorisere som fyldord. Det er eksempelvis ord som *i, dem, de, på, når, vi, at* der ikke bidrager meningsfuldt i en digital bearbejdning og derfor kan fjernes fra brevtteksten.¹²⁶

Der findes stopordslister i RStudio, som vi kunne have benyttet. De er dog skabt til nyere materiale, og vi fandt det derfor bedst selv at skabe en liste til formålet. Vi har rettet listen til under vores undersøgelse, så den til sidst passede perfekt til vores materiale. Listen er skabt ud fra de principper, som Emil Hvitfeldt og Julia Silge beskriver i deres bog *Supervised machine learning for text analysis in R* i kapitel 3.2 *Creating your own stop word list*.¹²⁷ Vi har tokeniseret brevtteksten, og kan dermed få en optælling på de forskellige ord.¹²⁸ Ved at se på hyppighed, har vi skabt listen over de fyldord, som optræder meget i vores materiale, men som

¹²⁵ Vi kunne også have brugt emnemodellen CTM (Correlated Topic Model), som er mere avanceret end LDA. Den tildeler ikke blot emner til tekstmateriale, men kan også undersøge hvordan emner korrelerer med hinanden. Denne model er dog ret kompliceret og ligeledes langsom, fordi modellen udregner på flere parametre. Vi har derfor valgt at holde os til LDA, da denne fungerer godt som grundmodel til topic modeling.

¹²⁶ I en tekstnær analyse ville ord som *os* og *dem* naturligvis bidrage meningsfuldt.

¹²⁷ Hvitfeldt Emil og Julia Silge, "Supervised Machine Learning for Text Analysis in R", konsulteret d. 12 maj 2025, <https://smilar.com>

¹²⁸ Når man tokeniserer i RStudio, deler man teksten op i mindre dele, som i vores tilfælde er ord. Hvert ord bliver dermed til en enhed, en såkaldt token. Når ord er deres egen individuelle enhed er det blandt andet muligt at tælle dem.

ikke er betydningsskabende i sig selv. Vores liste indeholder 253 ord, og kan ses i bilag. Vi benytter funktionen *anti_join* til at fjerne stopord og vi filtrerer derefter også tal og enkeltstående bogstaver fra. Der renses i denne proces 307.352 ord fra, hvilket svarer til 59,98% af tekstkorpus. Selvom det kan fremstå voldsomt af fjerne mere end halvdelen af ordene i brevtteksterne, vurderer vi ikke, at vi har filtreret for meget fra. Det er forventeligt, at mellem 40% og 60% af tekstkorpus skal fjernes.¹²⁹

¹²⁹ “Supervised Machine Learning for Text Analysis in R”, Kapitel 3, konsulteret d. 12 maj 2025.
<https://smltar.com>

Kodeblok 5

```
# Ét brev per ID og behold tekst
data_topics <- data %>%
  distinct(ID, .keep_all = TRUE) %>%
  select(ID, text)

# Tokenisering og fjern stopord
data_tokens <- data_topics %>%
  unnest_tokens(word, text) %>%
  filter(!str_detect(word, "[0-9]+$")) %>%           # fjern rene tal
  filter(nchar(word) > 1) %>%                       # fjern enkeltbogstaver
  anti_join(stopord, by = c("word" = "ord"))         # fjern stopord

# Opret term-frekvens matrix
dtm_data <- data_tokens %>%
  count(ID, word) %>%
  cast_dtm(document = ID, term = word, value = n)

# Kør LDA med 6 topics
lda_model <- LDA(dtm_data, k = 6, control = list(seed = 123))

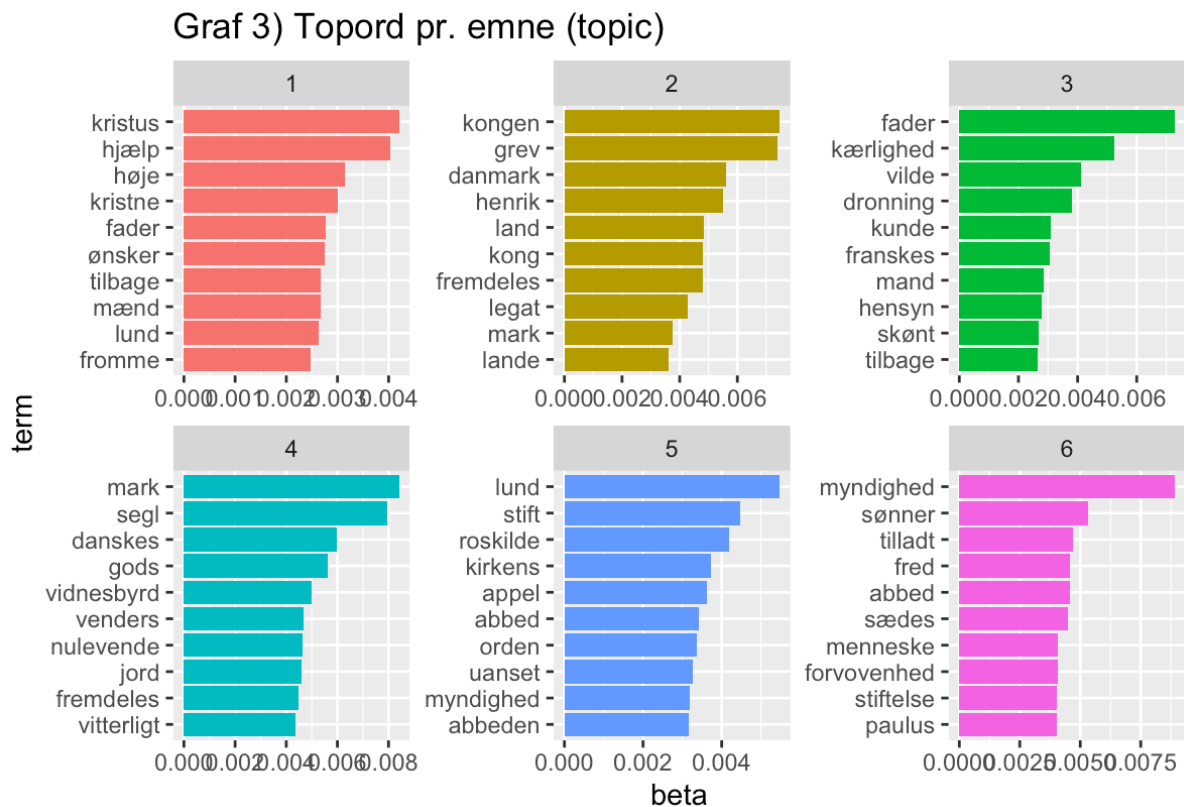
library(tidytext)

topics <- tidy(lda_model, matrix = "beta")

# Vis de 10 mest sandsynlige ord pr. topic
top_terms <- topics %>%
  group_by(topic) %>%
  slice_max(beta, n = 10) %>%
  ungroup()

# Plot
library(ggplot2)

top_terms %>%
  mutate(term = reorder_within(term, beta, topic)) %>%
  ggplot(aes(beta, term, fill = factor(topic))) +
  geom_col(show.legend = FALSE) +
  facet_wrap(~ topic, scales = "free") +
  scale_y_reordered() +
  labs(title = "Graf 3) Topord pr. emne (topic)")
```



Vi får nu *Graf 3) Topord pr. emne (topic)*, som vi har opsat til seks mulige emne kategorier med 10 ord i hver. Det er svært at få et klart indtryk over emnerne i de 2002 breve, ved udelukkende at se på resultaterne af vores topic modeling. Resultaterne giver ikke klart definerede emner, så det er derfor nødvendigt at vi bruger vores historiefaglige viden om perioden for at kunne tolke outputtet. Vi tolker hertil at kategori 1 fremstår som omhandlende tro, kategori 2 virker til at være præget af aktører, kategori 3 tænker vi umiddelbart relaterer sig til Ingeborg sagen, kategori 4 virker til at omhandle materiel værdi og samtidig en form for stadfæstelse gennem ord som “vidnesbyrd” og “vitterligt” (her kunne der være tale om testamenter), kategori 5 tolker vi som kirkens aktører og ligeledes kategori 6 som religiøs magt.

De forskellige emner filtrer sig ind i hinanden, og skaber ikke et klart overblik eller en skarp opdeling af tekstens indhold. Dette skyldes måden man skriver på i perioden, idet der hyppigt gøres brug af symbolik og indirekte henvisninger i brevtteksten. Vi kan dog udlede at religion, kirkelige aktører, politiske aktører, slægt og jord fylder meget i brevene, og at disse emner hænger sammen.

Et eksempel på en skrivemåde, som dækker over emner, der derfor ikke bliver synlige i vores topic modeling undersøgelse, finder vi brev 255 fra 1199.¹³⁰ På grund af de konflikter, som

¹³⁰ Danmarks Riges Breve 1. række bind 3 brev 255

sagen om Ingeborg og Phillips ægteskab havde skabt, blev dele af Frankrig pålagt interdikt. I brev 255 skriver Pave Innocens III til den franske gejstlighed, for at pålægge dem, at denne interdikt skal overholdes. I brevet står der blandt andet; (...) *skal I ikke imod alles håb kunne sammenlignes med dem, om hvem man siger: Stumme hunde, der ikke kan gø.*¹³¹ Her refereres der til det gamle testamente, hvor der i Esajas' Bog kapitel 56 står;

*Alle vogterne er blinde, de forstår intet;
de er alle sammen stumme hunde, der ikke kan gø.
De ligger og drømmer, de elsker at sove.*¹³²

Verset omhandler kritik af religiøs ledelse, da kirkens vogtere ikke agerer som de vagthunde, de burde være. De drømmer og gør dermed ingenting. Når Innocens III skriver således til den franske gejstlighed, er det en skarp opfordring til, at de skal sige fra over for Filip 2. Augustus. Den franske gejstlighed må ikke være tavse i mødet med den synd, som Filip udfører. Innocens III påpeger konsekvenserne ved derefter at skrive, at kongen (...) *hellere skal vælge at adlyde vore, eller rettere Guds formaninger end at mærke kirkens strengthed.*¹³³ Her truer Innocens III med at pålægge hele Frankrig interdikt.

Denne reference vil ikke være synlig i vores topic modelling, da betydninger tolkes i ordene. Da der ikke er stor variation i måde, hvorpå man skriver, kan topic modeling ikke skabe en klar skitsering af emner i vores 2002 breve. Hvis vi derimod benytter vores historiefaglige viden til at tolke resultaterne, kan undersøgelsen give et generelt indblik i aktuelle emner i perioden. Topic modelling kan ligeledes benyttes til at skabe overblik over emner i store mængder tekst. Vores konklusion er hertil, at undersøgelsen ikke kan stå alene og må suppleres af faglighed og periodemæssig kontekst.

¹³¹ Se Speciale_database række 7488 - 7490

¹³² "Esajas' Bog Kapitel 56", konsulteret 25. maj 2025,
<https://www.bibelselskabet.dk/brugbibelen/bibelenonline/es/56>

¹³³ Se Speciale_database række 7488 - 7490

Sentimentanalyse

Da vi vil analysere relationen mellem Danmark og Frankrig i Danmarks Riges Breve, og vi hertil ønsker at undersøge relationens karakter, er det væsentligt at se nærmere på følelser og stemning i brevene. Dette kan lade sig gøre, ved at vi udfører en såkaldt sentimentanalyse.

At analysere sentiment er en metode inden for tekstanalyse, som man benytter til at definere følelsesmæssige udtryk. Dette gøres ved at vurdere om teksten er præget af en positiv, neutral eller negativ tone. Metoden er ordbaseret, da analysen går ind og sammenligner hvert enkelt ord med en score eller værdi i en sentimentordbog. Der findes flere sentimentordbøger i RStudio, hvor vi udfører analysen. Julia Silge og David Robinson skriver om de tre mest brugte sentimentordbøger (*AFINN*, *bing* og *nrc*) på deres hjemmeside *Text Mining with R: A Tidy Approach*.¹³⁴ *AFINN*-ordbogen er udviklet af den danske forsker Finn Årup Nielsen fra Danmarks Tekniske Universitet og er sammensat således:

*The AFINN lexicon assigns words with a score that runs between -5 and 5, with negative scores indicating negative sentiment and positive scores indicating positive sentiment.*¹³⁵

Sentimentordbøgerne *AFINN*, *bing* og *nrc* er dog skabt til engelsksproget materiale og ligeledes til nyere tekst som produktanmeldelser og opslag fra sociale medier. Sproget og de tildelte sentiment er derfor ikke kompatibelt med vores materiale. Vi læner os dog op af Finn's metode, idet vi ligeledes inddeler sentiment i score mellem -5 og +5.

For at få en sentimentordbog, som er velegnet til dansk og samtidig til sprogbrugen i vores materiale, var vi nødsaget til at sammensætte den selv. Sentimentordbogen *AFINN* indeholder 2477 ord. For at få et grundlag for vores egen sentimentordbog, der matcher kapaciteten af dem, som vi ser i RStudio, søgte vi efter lignende datasæt på dansk.

Vi fandt *Danish Sentiment Lexicon* fra 2022, som er skabt af Det Danske Sprog- og Litteraturselskab og Center for Sprogteknologi ved Københavns Universitet. Datasættet ligger offentligt tilgængeligt og indeholder 14.008 ord, samt alle bøjninger af de 14.008 ord.¹³⁶ Datasættet er struktureret med en score fra -5 til +5, og var et særligt godt udgangspunkt for

¹³⁴ "Kapitel 2: Sentiment analysis with tidy data", konsulteret d. 13. maj 2025, <https://www.tidytextmining.com/sentiment>

¹³⁵ "Kapitel 2: Sentiment analysis with tidy data", konsulteret d. 13. maj 2025, <https://www.tidytextmining.com/sentiment>

¹³⁶ "Danish Sentiment Lexicon", konsulteret d. 1. juni 2025, <https://github.com/dsldk/danish-sentiment-lexicon>

vores konstruering af en sentimentordbog til breve fra middelalderen. Vi vurderede, at det var nødvendigt at tilpasse forskellige score i datasættet, da det er skabt til nyere dansk sprogmateriale. Ordet *hellig* havde eksempelvis en positiv score på 2 og ordet *interdikt* var slet ikke inkluderet på listen. Vi fjernede stopord fra vores brevmateriale, og skabte i RStudio en liste over de 100 mest hyppige ord. 10 af disse var ikke brugbare til sentimentanalyse, og vi tildelte derfor de resterende 90 ord værdi fra -5 til +5.¹³⁷ Ordet *interdikt* optræder eksempelvis 117 gange i vores brevmateriale, og det har ligeledes en markant følelsesmæssig værdi. Vi tildelte derfor ordet *interdikt* en negativ score på -5. Vi opdelte de forskellige bøjninger, så de optrådte som individuelle ord, og vi endte til sidst med en sentimentordbog, som indeholder 87.418 ord.¹³⁸

For at kunne omdanne denne proces til et nyt niveau, skal sentimentordbogen inkorporeres i RStudio sammen med de relevante pakker *ggplot2*, *tidytext*, *dplyr*, *stringr* og *tidyr*.¹³⁹ Nedenstående kodeblok 6 sammensætter viden fra ordbogen med vores database og giver et output i form af *Diagram 2) Sammenhæng mellem alle brevets længde og sentiment-score*.

¹³⁷ Tildelingen af værdi til de 90 ord er skabt manuelt og besluttet ud fra en faglig vurdering fra vores side. Vi har læst ordet i kontekst i materialet, og skabt en selektiv vurdering. Det er derfor vores vurdering som dikterer værdiansættelsen, da vi ikke har kompetencer til at skabe værdiansættelsen på baggrund af leksikografiske metoder.

¹³⁸ Vi har ikke fjernet nyere danske ord, som vi ikke forventer at se i materialet, da en manuel gennemgang af de 87.418 linjer ville tage for lang tid. Da disse ord ligeledes ikke optræder i materialet, spiller det ingen rolle, om de optræder på sentimenordbogslisten. Dette gælder ord som eksempelvis *smatso*, *jubelscene*, *kanonpopulær*, *megasucces* og *smadderskæg*.

¹³⁹ *dplyr*: Datahåndtering. *tidyr*: tjekker for manglende værdier og sikrer datastruktur. *tidytext*: tekstanalyse. *stringr*: tæller ordene i hvert brev. *ggplot2*: samler produktet fra hele koden til en visualisering.

Kodeblok 6

```
# 1. Tokeniser teksten til enkelte ord
brev_sentiment <- brevdatab %>%
  mutate(doc_id = row_number()) %>% # hvis du ikke allerede har en unik ID-
  kolonne
  unnest_tokens(word, text) %>%
  inner_join(sentiment, by = "word") %>%
  group_by(doc_id) %>%
  summarise(sentiment_score = sum(value)) # summer score for hvert brev

# 2. Join sentiment-score tilbage til hoveddata
brevdata <- brevdatab %>%
  mutate(doc_id = row_number()) %>% # samme doc_id som før
  left_join(brev_sentiment, by = "doc_id") %>%
  mutate(sentiment_score = replace_na(sentiment_score, 0)) # hvis intet match,
  så 0

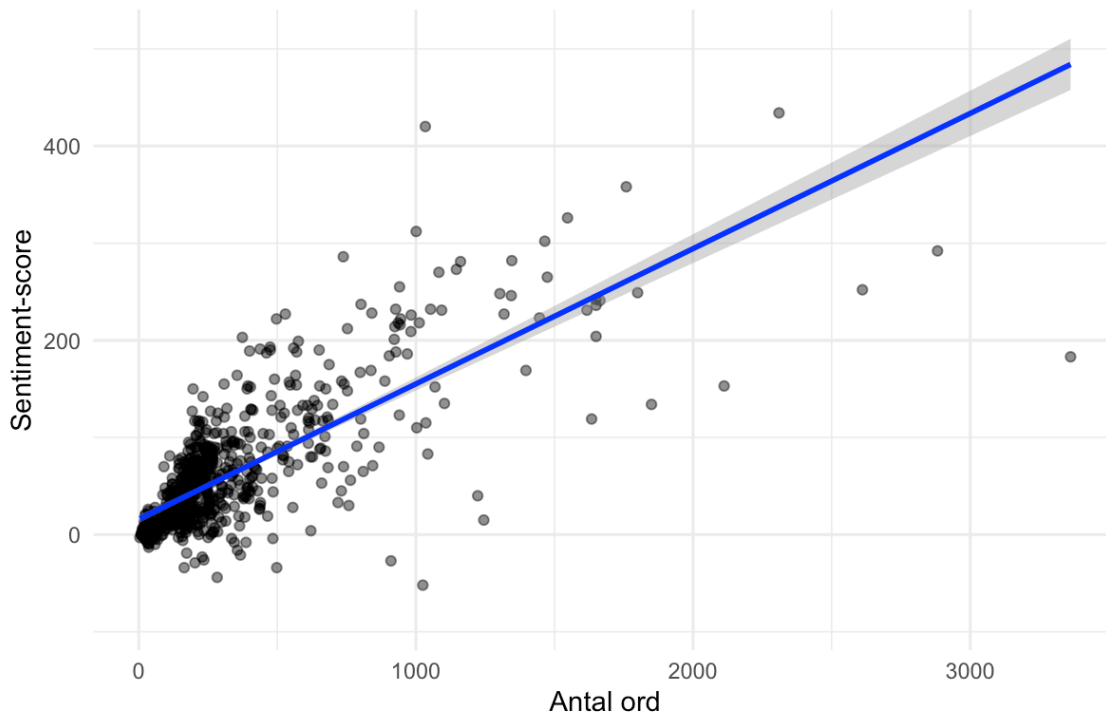
library(ggplot2)

ggplot(brevdata, aes(x = ord_antal, y = sentiment_score)) +
  geom_point(alpha = 0.5) +
  geom_smooth(method = "lm", se = TRUE, color = "blue") +
  labs(title = "Diagram 2) Sammenhæng mellem brevets længde og sentiment-score",
       x = "Antal ord",
       y = "Sentiment-score") +
  theme_minimal()
```

Det er relevant at undersøge sammenhængen mellem længden på brevene og sentiment-score, da dette giver en indikation på den overordnede tendens i materialet. Vi finder det relevant at undersøge, hvordan man kommunikerede i samtiden for at få en grundlæggende forståelse for materialet. Dette skaber ligeledes et grundlag for den senere undersøgelse af breve mellem Danmark og Frankrig, da vi hermed får etableret et referencepunkt. Det er væsentligt at vide, om materialet er domineret af neutral, positiv eller negativ kommunikation.

Diagram 2) Sammenhæng mellem alle brevets længde og sentiment-score nedenfor viser sammenhængen mellem sentiment-score, og længderne på de 2002 breve. Den blå linje symboliserer den generelle tendens på visualiseringen.

Diagram 2) Sammenhæng mellem brevets længde og sentiment-score



Her ser vi, at der er en generel tendens til, at brevene bliver mere positive jo længere de er. Ligesom på *Diagram 1) Ordantal pr. brev over tid*, kan vi se at størstedelen af brevene er fordelt i intervallet 0-800 ord. Disse breve har generelt en sentimentscore mellem 0 og 200, hvilket indikerer neutrale eller lette positive breve. Vi undersøger hertil korrelationen mellem sentimentscore og antal breve, og får et resultat på 0.8.

```
> cor(df$ord_antal, df$sentiment_score, use = "complete.obs")  
[1] 0.8031474
```

Korrelation måler, hvor stærk sammenhæng er mellem to variable. Hvis værdien er +1 er der en perfekt positiv sammenhæng. En korrelation på 0 indikerer ingen sammenhæng, og hvis korrelationen er -1, er der en perfekt negativ sammenhæng.

Vores samling har en høj korrelation og det indikerer derfor en stærk positiv sammenhæng mellem positive breve og længde på brevene. Her er det dog væsentligt at huske på, at korrelation ikke altid betyder kausalitet. Det vil sige, at vi ikke kan udlede, at længden på brevene forårsager brevenes sentimentscore, blot fordi de to ting følges ad. For at undersøge om det hænger sammen således, at brevene bliver mere positive, jo længere de er, undersøger vi

sentiment per ord i forhold til brevets længde. Vi kan risikere, at de lange breve simpelthen indeholder flere af de ord, som er tildelt sentiment, og at de derfor får en høj score.

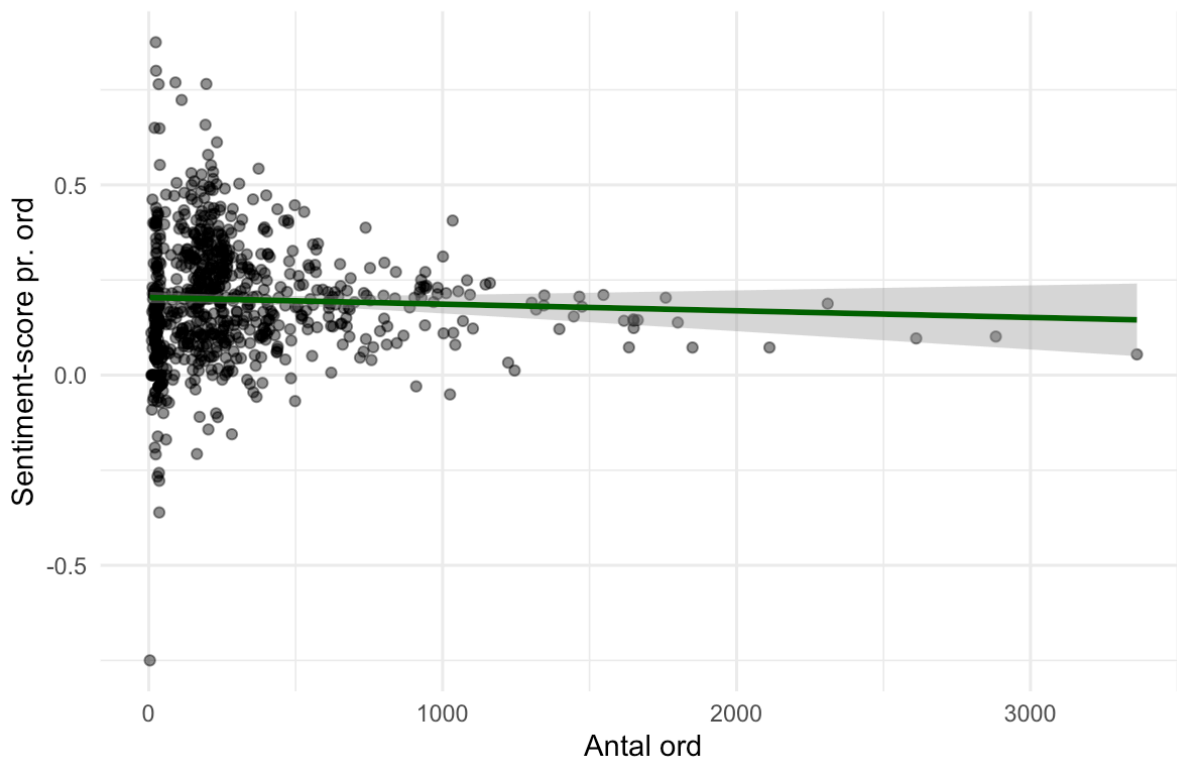
Vi kører følgende kodeblok 7, der ikke ser brevet som helhed, men analyserer det enkelte ord.

Kodeblok 7

```
brevdata <- brevddata %>%  
  mutate(sentiment_pr_ord = sentiment_score / ord_antal)  
  
ggplot(brevdata, aes(x = ord_antal, y = sentiment_pr_ord)) +  
  geom_point(alpha = 0.5) +  
  geom_smooth(method = "lm", color = "darkgreen") +  
  labs(title = "Diagram 3) Sentiment pr. ord i forhold til brevets længde",  
        x = "Antal ord",  
        y = "Sentiment-score pr. ord") +  
  theme_minimal()
```

Dette giver os *Diagram 3) Sentiment pr. ord i forhold til brevets længde* nedenfor, som er opsat efter samme visuelle parametre som *diagram 2* ovenfor. Her er den generelle tendens fremhævet med grøn, for at de to grafer lettere kan skilles fra hinanden.

Diagram 3) Sentiment pr. ord i forhold til brevets længde



På *Diagram 3) Sentiment pr. ord i forhold til brevets længde* kan vi nu se, at der ikke er en positiv sammenhæng mellem længde på brevet og sentimentscore. Korrelationen er ændret markant og faldet til $-0,1$, så sammenhængen i stedet er lettere negativ. Ud fra *diagram 3* kan vi derfor udlede, at jo længere breve bliver, desto mere neutrale bliver de. Vi tolker de neutrale breve som officielle diplomatiske korrespondancer, da vi har erfaret, at disse former for korrespondance er længere end de resterende breve.

Vi vurderer her, at den tidligere positive korrelation (med blå) skyldes længden på brevene. Da de lange breve naturligvis indeholder flere ord, er der større chance for, at de ord som tildeles sentiment optræder. Derfor kommer breve med eksempelvis 1500 ord, til at fremstå meget positive. Dette skyldes ligeledes måden, hvorpå teksten er udformet. De lange diplomatiske korrespondancer indeholder ofte formuleringer som *Pave X*, *Guds Tjeneres Tjener*, *hilser sine ærværdige Brødre*, *Ærkebiskop X* og *sin elskede Søn Magister X*, med den *apostoliske Velsignelse*. Her bliver ord som elskede, ærværdige og velsignelse tildelt positiv sentimentscore. Jo længere brevene er, jo flere standardiserede hilsner indeholder de.

For at undersøge fordelingen af brevene i relation til deres tildelte sentimentscore, har vi skabt *Kort 2) Sentimentkort (neg <15, pos >50)*. Her har vi valgt, at score under 15 skal kategoriseres som negativt og markeres med røde prikker. Breve med en sentimentscore mellem 15 og 50 bliver kategoriseret som neutrale breve og markeres med sort. Til sidst markeres breve med en sentimentscore over 50 som positive og med grønne prikker. Vi har opsat skellet mellem de neutrale, negative og positive ud fra ovenstående undersøgelse. Da der benyttes en venlig tone i materialet, og da brevene ofte er forholdsvis lange, er det nødvendigt at opsætte dette skel. Hvis vi markerede negative breve med sentimentscore under 0, ville vi ikke kunne finde de negative breve, da disse stort set altid indeholder ord med positiv sentimentscore grundet sproget i materiale. Vi testede forskellige opsætninger, og fandt frem til, at det nuværende skel på <15, 50, 50> repræsenterede materialets indhold på bedst mulig vis.

Kort 2) Sentimentkort (neg <15, pos >50) er opsat på baggrund af nedenstående kodeblok 8. Koden kører på en *leaflet*-opsætning og dikteres af tilhørende afsender og modtager koordinater.¹⁴⁰ Vi har hertil set muligheden for at kunne opskalere niveauet af denne fremstilling. Det gør vi ved at drage nytte af vores geografiske metadata i kombination med de nye metadata relateret til sentimentanalysen. Ved at kombinere geografi med fordelingen af sentiment er skaber vi en visuel opdeling af de negative, positive og neutrale breve. En visuel fremstilling muliggør nye perspektiver og skaber et andet overblik.

Kodeblok 8

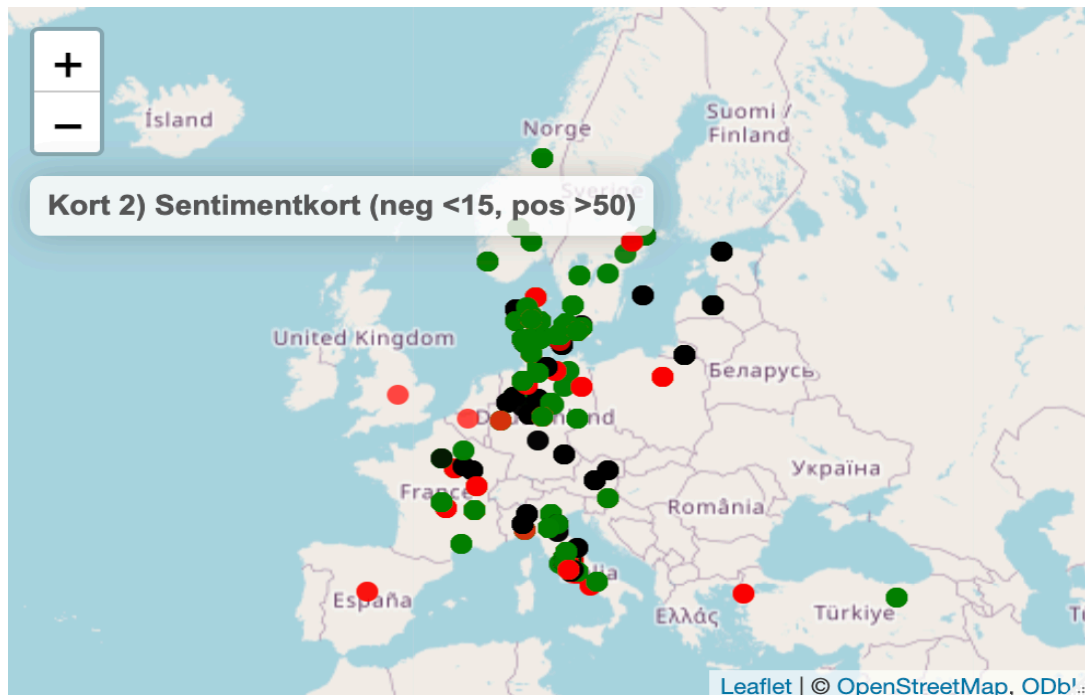
```
# Klassificér sentiment i kategorier
data <- data %>%
  mutate(sentiment_kategori = case_when(
    sentiment_score > 50 ~ "positiv",
    sentiment_score < 15 ~ "negativ",
    TRUE ~ "neutral"
  ),
  farve = case_when(
    sentiment_kategori == "positiv" ~ "green",
    sentiment_kategori == "negativ" ~ "red",
    TRUE ~ "black"
  ))
```

¹⁴⁰ “Kapitel 5: Visualisering 2: Kort”, konsulteret d. 16. maj 2025,
<https://johanheinsen.quarto.pub/programming-for-historikere/kort.html>

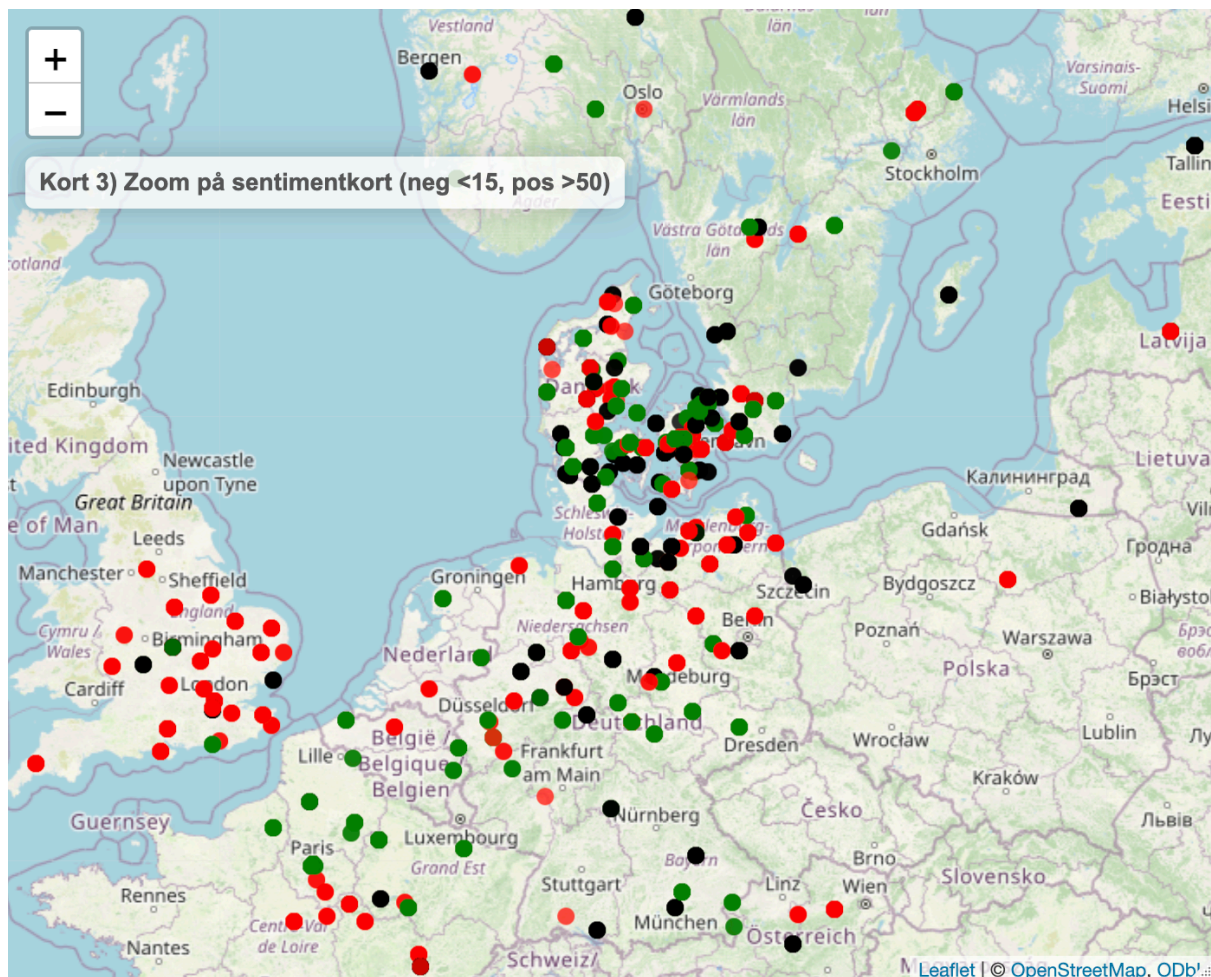

```
# Lav en ny data frame med punkter (én for afsender, én for modtager)
punkter <- data %>%
  select(doc_id, sentiment_score, sentiment_kategori, farve,
         afsender_breddegrad, afsender_længdegrad,
         modtager_breddegrad, modtager_længdegrad) %>%
  # Afsender-punkter
  mutate(type = "Afsender",
         lat = afsender_breddegrad,
         lon = afsender_længdegrad) %>%
  select(doc_id, lat, lon, sentiment_score, sentiment_kategori, farve, type) %>%
  bind_rows(
    # Modtager-punkter
    data %>%
      mutate(type = "Modtager",
             lat = modtager_breddegrad,
             lon = modtager_længdegrad) %>%
      select(doc_id, lat, lon, sentiment_score, sentiment_kategori, farve, type)
  )
#kort med farvede punkter for negativ, neutral og positiv.(laver bare zoom på
det her og gemmer 2)
leaflet(punkter) %>%
  addTiles() %>%
  addCircleMarkers(
    lng = ~lon,
    lat = ~lat,
    color = ~farve,
    radius = 5,
    stroke = FALSE,
    fillOpacity = 0.8,
    popup = ~paste0(
      "<strong>", type, "</strong><br>",
      "Brev ID: ", doc_id, "<br>",
      "Sentiment score: ", sentiment_score, "<br>",
      "Kategori: ", sentiment_kategori
    )
  ) %>%
  addControl(
    html = "<strong> Kort 2) Sentimentkort (neg <15, pos >50)</strong>",
    position = "topleft"
  )
```

Ud fra *Kort 2) Sentimentkort (neg <15, pos >50)* nedenfor kan vi udlede, at der er en stor fordeling i de tre kategorier. Der er flest neutrale breve omkring Danmark, og vi kan samtidig se en koncentration af negative breve omkring England. De positive breve er fordelt rundt i Europa. Det er her væsentligt at have for øje, at punkterne på kortet repræsenterer

kommunikation over en periode på 134 år. På trods af den lange tidshorisont, skaber kortet en fremstilling af, hvilke områder som kommunikerer i perioden. Vi får ligeledes en forståelse for de danske relationer til forskellige dele af Europa.



For at kunne få et bedre overblik over, hvad som gemmer sig i den cluster, der er omkring Danmark på *Kort 2*, har vi zoomet ind. Dette kan ses på *Kort 3) Zoom på sentimentkort (neg <15, pos >50)* nedenfor. Hermed bliver fordelingen i Europa tydeligere. Vi kan se, at der er en god fordeling af både neutrale, positive og negative breve inden for det danske rige. Der er en koncentration af negative såvel som positive breve i Frankrig, henholdsvis syd og nord for Paris. Kommunikation nord fra Danmark er fordelt inden for de tre kategorier. På *Kort 3) Zoom på sentimentkort (neg <15, pos >50)* er det ydermere meget tydeligt, at der er en overvægt af negative breve i England.



Kort 2 og Kort 3 indikerer ikke tendenser, som for os fremstår iøjenfaldende for den dansk-franske relation. Der er en generel fordeling rundt i Europa, hvilket indikerer et generelt niveau af kommunikation. Dertil afspejler kortet også en periode på 134 år, hvortil der sker flere ting i Europa. Det er derfor forventeligt at se forskellige sentimentscore fordelt rundt i Europa.

Delkonklusion

På baggrund af vores indledende digitale analyse af materialet kan vi fastslå følgende: Antallet af bevarede breve stiger over tid, hvilket indikerer en øget dokumentationspraksis eller bevaringsgrad. I de 2002 breve finder vi 745 breve sendt af paver, som ligeledes følger den eksponentielle stigning. I pavebrevene er Innocens IV den mest aktive afsender, og da disse fylder en væsentlig del af materialet, må vi udlede at paverne var centrale aktører i det relationelle netværk i Danmarks Riges Breve.

Ud fra sentimentanalysen kan vi fastslå, at vi ser en tendens til, at længere breve ofte udviser et mere neutralt sprogligt udtryk. Hertil vurderer vi, at de længste breve består af diplomatiske, officielle dokumenter.

Den geografiske analyse viser, at brevmaterialet rummer klare indikationer på en sydlig orientering mod især Frankrig og Italien, hvor det især er byerne Lyon og Rom, der er repræsenteret. Dette indikerer en direkte forbindelse til pavestolen. Relationen til England viser sig ligeledes interessant, da den signalerer en involvering i et bredere europæisk netværk.

Afslutningsvis kan vi udlede at de digitale analyser i flere tilfælde skal suppleres af en historiefaglig eller kildekritisk vurdering. I topic modeling analysen er computerens evner ikke til at tolke på symbolik. Derfor er en supplerende nødvendig i tekstanalytisk arbejde med Danmarks Riges Breve.

De 6%

Da vi ønsker at undersøge relationen mellem Danmark og Frankrig med udgangspunkt i brevmaterialet, vælger vi nu at skabe endnu en relationel database. Vi vil i denne del af analysen udelukkende fokusere på de breve, som relaterer sig til relationen mellem Danmark og Frankrig.

For at kunne skabe en database, som indeholder breve, der er væsentlige i den dansk-franske relation, har vi arbejdet med *nation_tags*. *Nation_tags* er metadata, som vi tilføjer til hvert enkelt brev. Det bliver skabt på baggrund af kodeblok 9 nedenfor.

Kodeblok 9

```
import pandas as pd
import requests
import time

filnavn = "/Users/Username1/Desktop/df_ner_analyseret.tekst.xlsx"
df = pd.read_excel(filnavn)

# Oprydning i kolonne Sted for eventuelle fejl
print("Rydder op i kolonnen 'Sted'...")
df['Sted'] = df['Sted'].fillna("").apply(lambda x: x.strip())

# Google Maps API
def find_land_google(stednavn, api_key):
    base_url = "https://maps.googleapis.com/maps/api/geocode/json"
    params = {"address": stednavn, "key": api_key}
    response = requests.get(base_url, params=params)
    result = response.json()

    if result['status'] == 'OK':
        for comp in result['results'][0]['address_components']:
            if 'country' in comp['types']:
                return comp['long_name']
    return None

# Aktivisering af API-nøgle
api_key = "privat nøgle"

# For hvert sted søges der efter nation
sted_til_land = {}
print("Slår stednavne op mod Google Maps API...")
for sted in df['Sted'].unique():
    if sted:
        try:
            land = find_land_google(sted, api_key)
            sted_til_land[sted] = land
            print(f"{sted} -> {land}")
            time.sleep(0.2) # Undgå at overskride API-begrænsning
        except Exception as e:
            print(f"Fejl ved {sted}: {e}")
    else:
        sted_til_land[sted] = None

# Skab ny kolonne nation_tag
print("Tilføjer land-tags for hver række...")
df['nation_tag'] = df['Sted'].map(sted_til_land)
```

```
# Gem den nye fil
output_fil = "breve_med_nationaltags.xlsx"
df.to_excel(output_fil, index=False)
print(f"Fil gemt som: {output_fil}")
```

Denne kode bygger på samme princip som koden om *Found_city*, som vi anvendte i tilføjelsen af koordinater i relation til vores kolonne *Sted*.¹⁴¹ Koden går ind og finder stednavne, hvorefter den tildeler det respektive land, som stednavnet hører til ifølge Google Maps. Google Maps arbejder naturligvis med de landegrænser, som eksisterer i dag. Derfor har vi været nødt til at lave en manuel kontrol, og samtidig en redigering af særligt Slesvig og Lund. De er gennem Google Maps blevet anerkendt som tilhørende henholdsvis Tyskland og Sverige, hvortil vi har rettet alle disse tags til *nation_tag denmark*.

Her kan arbejdet fremstå som det samme arbejde, der blev udført, da vi fandt de forskellige byer i materialet. Vi benytter samme tilgang i kodningen, men undersøger her lande frem for byer. I denne sammenhæng ser vi derfor geografien som berøringsflade frem for kommunikationsmønstre, da vi arbejder med den geografi, der er omtalt i brevene og ikke den geografiske data, som er tilknyttet afsender og modtager.

Efter at vi har lavet *nation_tags* på alle brevene, har vi en fordeling af top ti, som er således:

Top 10 Nation_tag

	Nation_tag	Antal
1	Denmark	3746
2	Germany	801
3	France	694
4	Italy	634
5	Sweden	385
6	United Kingdom	142
7	Estonia	101
8	Norway	90
9	Austria	38
10	Latvia	35

¹⁴¹ Se i script.

Da vi er særligt interesserede i relationen mellem Danmark og Frankrig, ønsker vi for nu at arbejde med disse 694 + 3746 linjer. I en gennemgang af de linjer, som har fået tildelt franske *nation_tags*, kan vi se, at Vilhelm af Æbelholt spiller en afgørende rolle. Ud af de 694 franske *nation_tags*, kommer 375 tags fra Vilhelm af Æbelholts brevsamling. Da han står for mere end 50% af brevene, vurderer vi, at han er en primær aktør i relationen mellem Danmark og Frankrig. Det er hertil vigtigt at have for øje, at selvom samlingen kommer fra det danske kloster i Æbelholt, er Abbed Vilhelm af fransk afstamning. Det er derfor forventeligt, at store dele af hans relationer og netværk er stærkt bundet op omkring Frankrig.

På fordelingen af *nation_tags* kan vi ydermere se, at Tyskland og Italien spiller en stor rolle i brevsamlingen. Vi tolker den store repræsentation af udenlandske tags som et udtryk for en geografisk bred berøringsflade. Vi finder det relevant at undersøge den geografiske berøringsflade, som kommer til udtryk i kildernes brevttekst, da vi gennem denne vinkling kan få et indtryk af den danske orientering i perioden. Selvom denne orientering både kan være kulturelt, politisk eller religiøs, skaber det en idé om det danske riges væren i Europa.

En åben verden

Nation_tags kan ydermere give et overblik over hvor stor en del af databasen, som er funderet nationalt og internationalt. Dette gør vi ved at lade RStudio skabe et overblik over, hvad den definerer som “internationale” og “nationale” breve. I nedenstående kodeblok 10 tager vi udgangspunkt i hvert enkelt ID-nummer og ser på alle de steder, der nævnes i hvert enkelt brev. Koden er opsat til at vurdere, hvorvidt alle *nation_tags* er ens eller forskellige.

Kodeblok 10

```
library(tidyverse)
library(readxl)

nationer_per_brev <- data %>%
  filter(!is.na(nation_tag)) %>%
  group_by(ID) %>%
  summarise(
    antal_nationer = n_distinct(nation_tag)
  )

international_breve <- nationer_per_brev %>%
  mutate(type = if_else(antal_nationer > 1, "International", "National")) %>%
  count(type)

print(international_breve)
```

Når vi kører kodeblok 10 får vi en optælling på 905 breve, som indeholder geografi fra mere end en enkelt nation, og disse breve defineres derfor som internationale. Vi får hertil 840 breve, som kun indeholder geografisk omtale, der relaterer sig til en enkelt nation. Disse breve bliver kategoriseret som nationale.¹⁴² Vi kan hermed udlede, at næsten 52% af de breve, som indeholder information om geografi, er internationale.¹⁴³ Vi tolker dette som et udtryk for, at de forskellige dele af Europa er opmærksomme på hinanden i samtiden, og at der er en vis kommunikativ forbundenhed.

Danmark og Frankrig – Computeren bestemmer

Af de 2002 breve er 1320 tagget tilhørende nationen *Denmark* og 230 tagget til nationen *France*. Der sker et overlap mellem 121 unikke ID-numre. Disse 121 har computeren dikteret som direkte forbindelser mellem Danmark og Frankrig, og vi finder det derfor naturligt at indlede med disse 121 breve, som vores kildegrundlag til vores videre undersøgelse af relationen mellem Danmark og Frankrig.

¹⁴² Vi har 257 breve, som ikke har geografiske værdier tilknyttet.

¹⁴³ Det svarer til at 45,2% af alle brevene er internationale, 41,96% er nationale og 13,74% er uden geografisk information.

Der er en tendens til, at historiebøgerne bruger ægteskabet mellem Filip og Ingeborg, som et argument for Danmarks position i højmiddelalderens Europa. *Brylluppet i 1193 mellem kong Knuds søster Ingeborg og Frankrigs kong Filip Augustus markerede rigets nye stormagtsstatus*.¹⁴⁴ Der er ingen tvivl om, at det danske rige var en indflydelsesrig nation og en central magtfaktor under kong Knud d. 6. Her undrer vi os dog. Hvis et samarbejde mellem to af Europas største magtfaktorer markeres gennem dette ægteskab, hvorfor er det så kun 6% af vores kildemateriale, som indeholder identificerbare markører mellem de to nationer? Dette stemmer ikke helt overens med det narrativ, som er fremstillet i litteraturen.

Med afsæt i denne undren, vil vi undersøge hvilket netværk, der ligger til grund for denne relation. Vi vil derfor undersøge de 121 breve, og se om en eksplorativ digitalhistorisk tilgang kan bidrage med nye vinkler og perspektiver. Vi ønsker hertil at benytte en kvantitativ netværksbaseret analyse til at undersøge relationens karakter.

Med de digitalhistoriske metoder, er det muligt at skabe visuelle outputs, som kan bidrage til analysen. De visuelle elementer bliver en form for kulmination af måneders arbejde med at sammensætte og skabe metadata. Processen er langvarig, da vi har udført store dele af metadata-skabelsen manuelt. Metadata er data omkring dataene, og de kan være yderst brugbare til at fremstille nye tendenser i kilderne.¹⁴⁵ At skabe god metadata er dog omfattende og i vores tilfælde ekstremt tidskrævende.¹⁴⁶

Vores forhåbning er, at et visuelt blik på kilderne måske kan belyse nye tendenser eller perspektiver, som kildernes brevttekst ikke anskueliggøre på samme vis. Vores udgangspunkt i det eksplorative og digitalhistoriske er stærkt forbundet med det visuelle, da vi netop ser at digitalhistoriens muligheder for at bidrage med nye perspektiver, er at finde i det visuelle.

Præmissen for den eksplorative tilgang er at teste nye ideer og skubbe til eksisterende grænser og forestillinger om materialets begrænsninger. Derfor følger vi forskellige ideer og lader de eksplorative output diktere vores vej gennem analysen.

¹⁴⁴ Roesdahl, *Fra Vikingetid til Valdemarstid*, s. 221

¹⁴⁵ Elo, "Big Data, Bad Metadata", s. 103-111

¹⁴⁶ Jf. Konstruering af metadata

Geotracking

For at belyse forbinder i materialet, benytter vi kolonnerne *afsender_sted* og *modtager_sted* og deres tilhørende koordinater. Det er særligt i denne del af arbejdet med det digitalhistoriske, at det forudgående arbejde med konstruering og tilføjelse af metadata kommer til sin ret. Her begynder vi at kunne bearbejde alle de lokationer, som vi har fundet på baggrund af vores manuelle gennemgang af afsender- og modtagersteder, samt de geografier, som vores kode kunne udtrække på baggrund af det digitaliserede stedregister. Hertil bliver vores opsætning af samarbejdet mellem Python, Google Maps og aktiveringen af API-koden helt essentiel. Denne formidling af materiale ville ikke være mulig, hvis vi ikke havde gjort grundarbejdet, der indebar adskillige steps, for at kunne skabe præcise geografisk visuelle referencer. De kort vi vælger at inddrage i det kommende afsnit, er bygget op omkring kodeblok 11.

Kodeblok 11

```
import pandas as pd
import folium

# 1. Indlæs dataframe med 121 breve med både DK og FR tags
df =
pd.read_csv("/Users/Username1/Desktop/Speciale/Breve_Danmark_og_Frankrig_121.csv", sep=";")

# 2. Fjern dubletter, så der kun fokuseres på enkeltstående ID
df_unique = df.drop_duplicates(subset="ID")

# 3. Tilpas ensartede koordinater, gælder kommaer og decimaler
for col in ["afsender_breddegrad", "afsender_længdegrad", "modtager_breddegrad", "modtager_længdegrad"]:
    df_unique.loc[:, col] = df_unique[col].astype(str).str.replace(",", ".", regex=False).astype(float)

# 4. Filtrér kun rækker med fulde koordinater for både modtager og afsender
df_filtered = df_unique.dropna(subset=[
    "afsender_breddegrad", "afsender_længdegrad",
    "modtager_breddegrad", "modtager_længdegrad"
])

# 5. Opret kort centreret efter gennemsnit af koordinaterne
center_lat = df_filtered["afsender_breddegrad"].mean()
center_lon = df_filtered["afsender_længdegrad"].mean()
```

```
kort = folium.Map(location=[center_lat, center_lon], zoom_start=5)

# 6. Skab Linjer mellem punkterne
for _, row in df_filtered.iterrows():
    punkt_a = [row["afsender_breddegrad"], row["afsender_længdegrad"]]
    punkt_b = [row["modtager_breddegrad"], row["modtager_længdegrad"]]

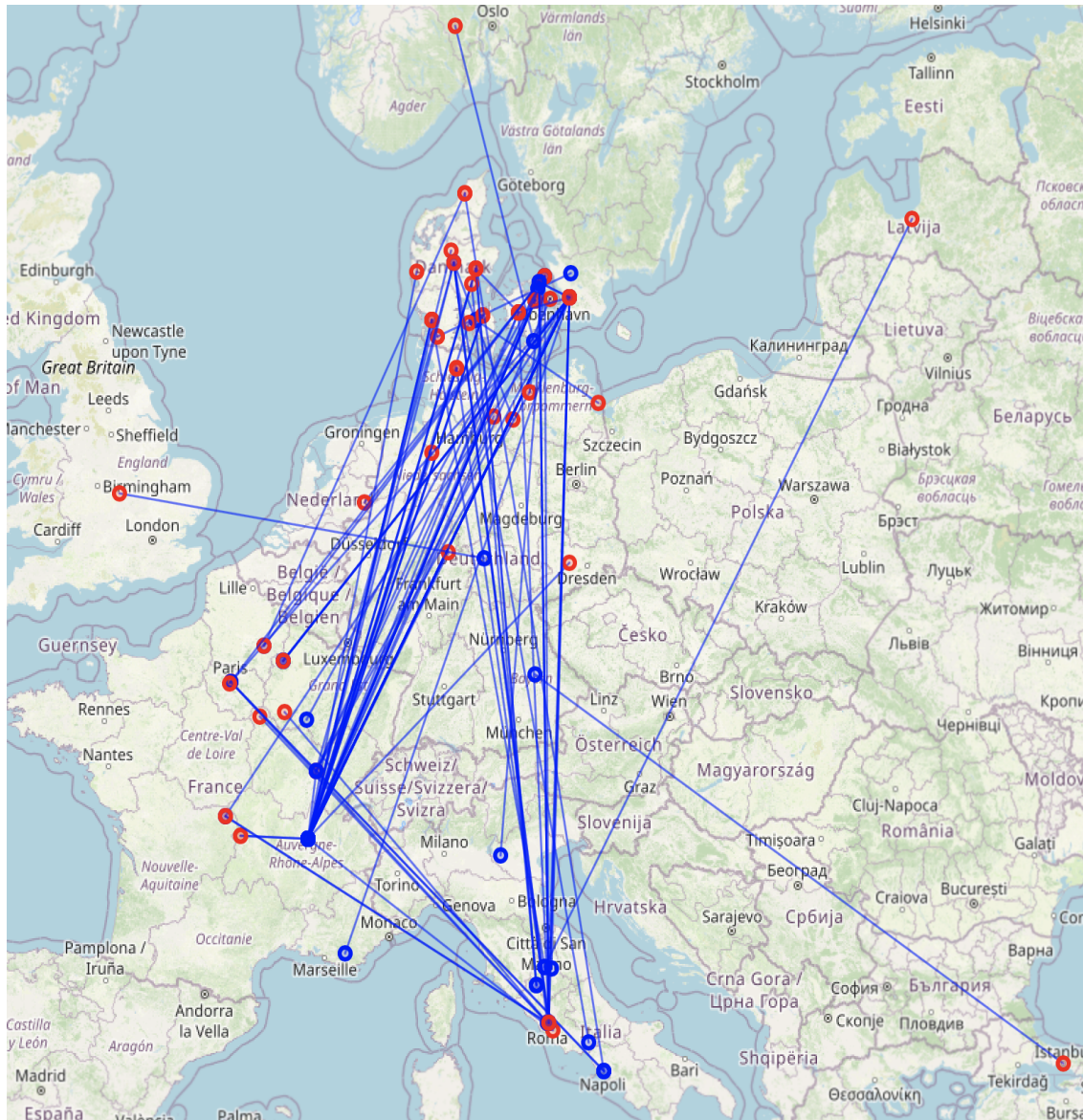
    folium.PolyLine([punkt_a, punkt_b], color='blue', weight=1.5,
opacity=0.6).add_to(kort)
    folium.CircleMarker(punkt_a, radius=4, color='blue', fill=True).add_to(kort)
    folium.CircleMarker(punkt_b, radius=4, color='red', fill=True).add_to(kort)

# 7. Gem kortet som html fil på vores lokale server
kort.save("/Users/Username1/Desktop/Speciale/Brevrelationer_DK_FR_121.html")
```

Koden fungerer ved blot at fokusere på enkeltstående ID-numre og de afsender- og modtagerkoordinater, som er tilgængelige i den relevante række.¹⁴⁷ Disse informationer bliver visuelt formidlet via forbindelseslinjer på et interaktivt kort. Det gør, at hvert brev bliver til en visuel aktør i det historiske netværk, som vi undersøger.

¹⁴⁷ Punkt 2 i koden fjerner dubletter fra brevmængden. Grundet det omfattende geografiske materiale i brevene, er de 121 breve dubleret til 1410 rækker i den relationelle database. For hvert stednavn i *found_city* blev materialet dubleret en række, modtager- og afsendersted og er derfor ens i hver række i de specifikke brev ID. Dette punkt nedskalerer derfor mængden igen, således det blot er 121 rækker, der skal behandles.

Kort 4) Kommunikation mellem Frankrig og Danmark



Ovenstående Kort 4) Kommunikation mellem Frankrig og Danmark visualiserer de 121 breve mellem Danmark og Frankrig. De blå prikker indikerer på dette kort afsendere og de røde prikker repræsenterer modtagere. Når kortet skabes i Python, er det interaktivt i en browser. Det er her muligt at zoome ind og ud og filtrere i de forskellige år samt undersøge de enkelte punkter. Da det ikke er muligt at fremhæve i opgaven, er vi nødsaget til at benytte statiske screenshots, har vi skabt en oversigt over fordelingen af de forskellige afsendere og modtagere.

Afsender (blå)	Antal	Modtagere (rød)	Antal
Lyon	39	Roskilde	7
Lateranet	6	NA	7
Æbelholt	5	Lund	6
Citeaux	3	Esrom	5
Danmark	3	Slesvig	5
Orvieto	3	Sorø	5
Esrom	2	Århus	5
Perugia	2	Ribe	4
Reims	2	Danmark	3
Ribe	2	Lateranet	3

På Kort 4) *Kommunikation mellem Frankrig og Danmark* kan vi udlede, at der er stor geografisk spredning i brevene mellem Danmark og Frankrig. Der er modtagere og afsendere placeret andre steder end i de to lande. Ud fra kort 4 kan vi ydermere konkludere, at der er markant større spredning i geografi for modtagere end der er for afsendere.

Det er hertil interessant, at det i vores kildesamling er Lyon, der er den betydeligt største afsender. Det tyder på, at en stor del af kommunikationen, hvor både Danmark og Frankrig er tagget udspringer fra paven, da Innocens IV er placeret i Lyon. Havde det været breve med kongeligt afsæt, ville det med stor sandsynlighed været noteret, som sendt fra Paris. Vi bygger denne antagelse på Rigords krønike, hvori han anerkender Paris som et administrativt centrum for Phillips regeringstid.¹⁴⁸ Ydermere optræder Lateranet i de ti mest hyppige afsender og modtagere, hvilket indikerer at store dele af de fællestagede breve har et paveligt aspekt. At Lateranet også optræder som modtager, har dog en fortsat interesse i undersøgelsen. Det indikerer, at flere breve med både *Denmark* og *France* tagget går i begge retninger. Det fremstiller også den dansk-franske relation som en, der i flere tilfælde fungerer igennem Rom. Pavestolen bliver dermed en væsentlig aktør i forståelsen af den dansk-franske relation i samtiden.

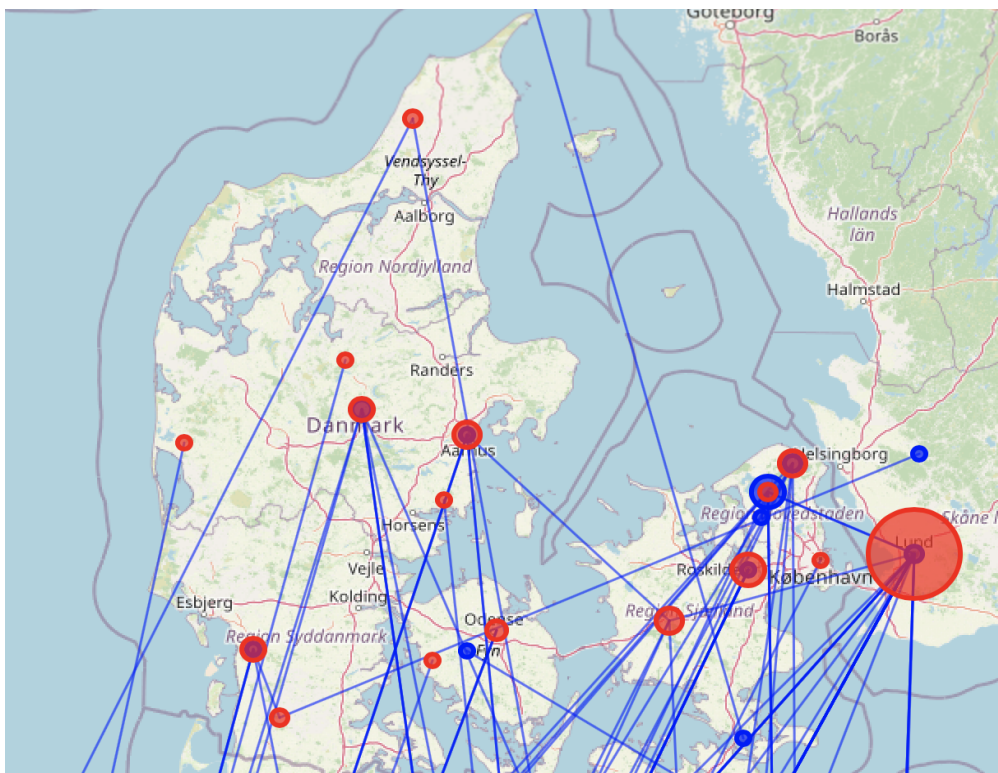
De 121 breve er tydeligt præget af nordgående tendenser. Størstedelen af brevene udgår fra Frankrig og pavemagtens udøvende områder, som Lyon, Lateranet og Orvieto, og sendes til danske kirkelige institutioner, som Roskilde og Lund. Kort 4) *Kommunikation mellem*

¹⁴⁸ *Paris was now the home of a permanent administration (...)* "The Capetians", Bradbary s. 187

Frankrig og Danmark bekræfter et mønster i materialet, som viser et netværk, der geografisk opbygges fra syd mod nord, hvor kommunikationsinitiativet i højere grad udspringer fra franske og italienske områder.

Kortets umiddelbare visuelle formidling, fremstiller Danmark med langt flere breve modtaget end sendt, hvilket er interessant i en overvejelse af initiativ og interne magtpositioner. Det europæiske billede gør det interessant for os at zoome ind på Danmark.

Kort 5) Zoom på kommunikation mellem Frankrig og Danmark



Ved at zoome ind på Danmark i kortet, fremstår der ligeledes tydelige mønstre. Mønstre, der i en grad gengiver de europæiske tendenser, blot indenfor det danske rige.¹⁴⁹ Vi ser, hvad vi tolker som et kommunikativt hierarki. Æbelholt, Århus, Esrom og Lund optræder alle som både modtager og afsender. Dette indikerer et netværk i netværket, da de ikke blot deltager i den grænsekrydsende kommunikation, men ligeledes står for at formidle og kontrollere den indbyrdes videre kommunikation. Disse lokationer fungerer dermed som hotspots i et dansk

¹⁴⁹ Tommerup på Fyn er ligeledes markeret blå. Dette punkt dækker blot over et brev, og vi vælger derfor ikke at anse dette som et centralt element i det kommunikative netværk.

kommunikationsnetværk, relateret til indhold, hvor Frankrig er repræsenteret. Det er vigtigt for os at pointere, at de aktører, som indgår i det danske netværk, er individer ved kirkelige institutioner. Vi ser derfor, at det indbyrdes kommunikationsnetværk i de 121 breve er religiøst funderet.

De strukturelle elementer fra *kort 4* og *kort 5* viser ikke blot geografiske sammenhænge, men antyder også noget, som vi kan betegne som et informationshierarki. De franske informationer bliver fordelt rundt i landet, men det er få lokationer, hvor der optræder en gensidig aktivitet. I Æbelholt finder vi et relativt stort blå punkt, som repræsenterer de afsendte breve fra Vilhelm af Æbelholt. Det er her tydeligt, at han er den mest markante afsender placeret i Danmark.

I takt med at vi vælger at betegne vores projekt og særligt analysen som værende eksplorativ, er det også naturligt, at vi afsøger andre områder, der måske ikke ville være vores første indskydelse. En anden måde at visualisere ovenstående relationer på baggrund af afsender og modtager, kan gøres i en software kaldet *Sankey*.¹⁵⁰ Den data vi har brugt til ovenstående kort og optællinger, har vi ligeledes forsøgt at kombinere med *Sankey* pakken i Python, i kode kodeblok 12.

¹⁵⁰ Sankey er en diagramsoftware, som kan formidle strømninger interaktivt.

Kodeblok 12

```
import pandas as pd
import plotly.graph_objects as go

# 1. Indlæs filen
df =
pd.read_csv('/Users/Username1/Desktop/Speciale/Breve_Danmark_og_Frankrig_121.csv', sep=';')

# 2. Vælg nødvendige kolonner og fjern dubletter
df_unique = df[['ID', 'afsender_sted',
'modtager_sted']].drop_duplicates(subset='ID')

# 3. Fjern rækker med manglende værdier
df_unique = df_unique.dropna(subset=['afsender_sted', 'modtager_sted'])

# 4. Optæl forbindelser

link_counts = df_unique.groupby(['afsender_sted',
'modtager_sted']).size().reset_index(name='count')

# 5. Opret nodemapping
nodes = pd.Series(pd.concat([link_counts['afsender_sted'],
link_counts['modtager_sted']])).dropna().unique()
node_map = {name: i for i, name in enumerate(nodes)}

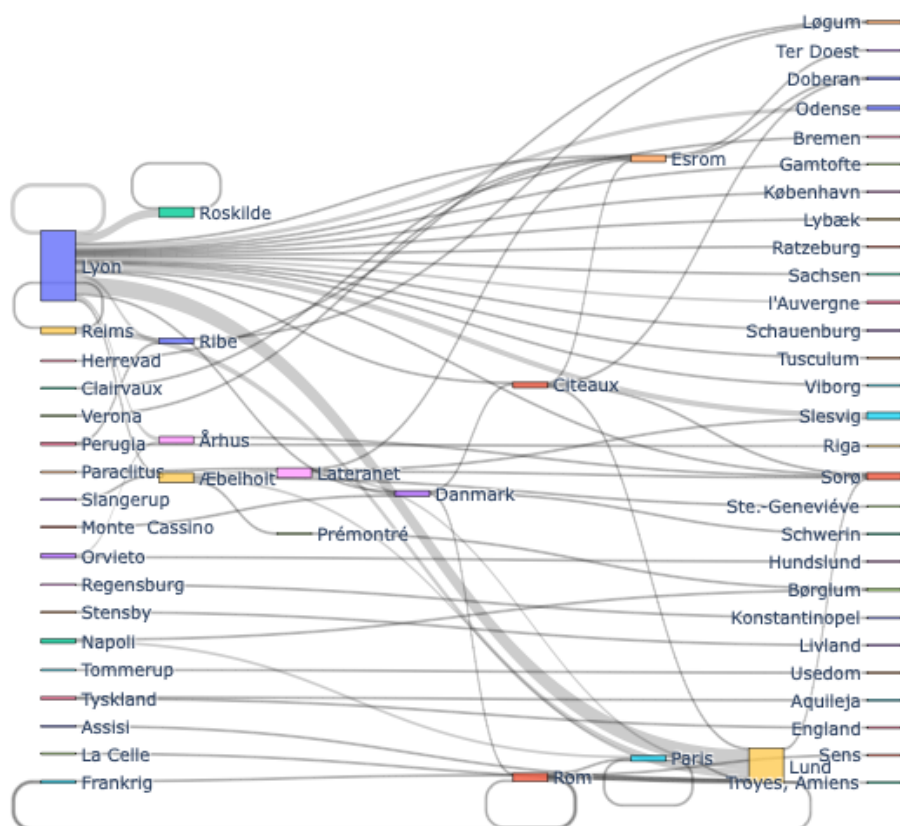
# 6. Plot Sankey-diagrammet
fig = go.Figure(data=[go.Sankey(
    node=dict(
        pad=15,
        thickness=20,
        line=dict(color="black", width=0.5),
        label=list(nodes)
    ),
    link=dict(
        source=link_counts['afsender_sted'].map(node_map),
        target=link_counts['modtager_sted'].map(node_map),
        value=link_counts['count']
    )
)])

fig.update_layout(title_text="Sankey: Brevstrømme mellem afsender og modtager (DK/FR)", font_size=10)
fig.show()
```


Kodeblok 12 gav et output i form af et Sankey-diagram, der kan ses nedenfor på *Diagram 4)* *Sankey-diagram*.¹⁵¹ Et Sankey-diagram er et flowdiagram, som kan fremhæve strømninger mellem kategorier. I vores tilfælde er det kommunikationsstrømninger, det kunne i princippet være penge eller energi.

Diagram 4) Sankey-diagram

Sankey: Brevstrømme mellem afsender og modtager (DK/FR)



De 121 breve med både fransk og dansk tag er alle indlejret i *diagram 4*. I venstre side af diagrammet er alle afsendere noteret og brevenes modtagere til højre. Linjerne, der forbinder de to steder, repræsenterer antallet af breve sendt og modtaget. Tykkelsen af linjen dikteres af antallet af breve, som er sendt fra lokation A til B. De bevæger sig hen af akserne alt efter hvor

¹⁵¹ Efter vores anvendelse af koden oprettes diagrammet i en URL, som kun kan vises på skaberens lokale enhed. Når vi tilgår diagrammet er det interaktivt og hver node kan undersøges. Til hver node er optællingen tagget, hvor til man kan se hvor mange breve der kommer til og fra den nævnte lokation.

de placerer sig mellem afsender og modtager. Jo mere aktiv i både afsendte og modtagne breve, jo længere ind mod midten.

På *Diagram 4) Sankey-diagram* ser vi igen Lyons fremtrædende position i kommunikation mellem Danmark og Frankrig. Fra 1244 er pave Innocens IV placeret i Lyon, da han flygtede fra Rom i frygt for den tysk-romerske kejser Frederik d. 2.¹⁵² Dermed kan vi fra denne visualisering også udlede, at det er gejstlige aktører, dominerer afsendere såvel som modtagere i det dansk-franske kommunikative netværk.

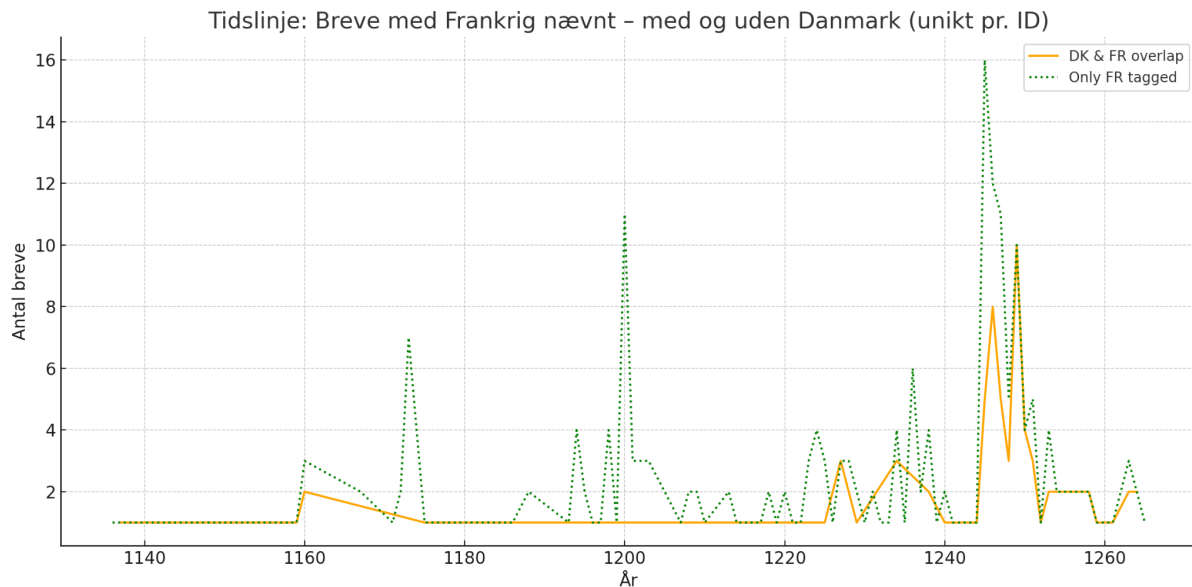
Når magten taler – Brevenes sprog mellem nord og syd

For at udvide vores horisont og teste den eksplorative tilgang, ændrer vi nu vores fokus til at undersøge det tidsmæssige perspektiv. Vi vil ændre datagrundlaget, og forsøge at lade computeren diktere relationen. Vi indleder med at benytte alle breve, hvor Frankrig er markeret i kolonnen *nation_tag*. Vi filtrerer alle disse breve ud, og matche dem med kolonnen *årstal*.¹⁵³ Det giver os en optælling, der visuelt formidles i *Graf 4) Omtale med og omtale om Frankrig* nedenfor. Den gule linje i tidslinjen på diagrammet viser de breve, hvor både Danmark og Frankrig er tagget, som vi behandlede i ovenstående. Den grønne linje viser den samlede sum af breve, hvor Frankrig er tagget, altså uden en direkte dansk geografisk forbindelse. Dermed kan man se kommunikation om Frankrig (grøn) op imod kommunikation med Frankrig (gul).

¹⁵² Lyon var udenfor Frederik d. 2's. autoritet og dermed sikkert for Innocens IV grundet hans relation til Louis IX. "First Council of Lyons – 1245 A.D.", konsulteret d. 1. juni 2025, <https://www.papalencyclicals.net/councils/ecum13.htm>

¹⁵³ Vi har valgt at beholde dataen med de overlappende breve i grafen, da den giver en god indikation på en relationel basislinje.

Graf 4) Omtale med og omtale om Frankrig



På Graf 4) *Omtale med og omtale om Frankrig* kan vi se, at der generelt er mere omtale af Frankrig, end direkte kommunikation. Der er store udsving i den grønne linje omkring år 1200, hvor den gule linje ikke følger med. De udsving, som vi ser i 1240'erne, følger dog hinanden. Grafen giver et overblik over graden af kommunikation mellem de to lande over en periode på 134 år. Det kan fremstå som om, at der sker store ændringer i omtale samt graden af kommunikation mellem de to lande, da udsvingene er markante. Her er det væsentligt at orientere sig mod grafens opsatte parametre. På y-aksen ser vi antal breve, som går fra 1 til 16. På hundrede år svinger den gule linje mellem en til to breve om året og op til maks ti breve. Procentvis er udsvinget markant, men reelt er der ikke så voldsomme forandringer, som det ser ud til på visualiseringen. Diagrammet er baseret på meget få breve, fordelt over meget lang tid. Vi kan derfor udlede, at graden af kommunikation mellem de to lande er lav. Vi kan ydermere udlede, at der i perioden omkring Phillip og Ingeborgs ægteskabskrise sker en markant stigning i omtale af Frankrig. For at undersøge dette nærmere, filtrer vi brevene fra 1200 til 1209 fra og undersøger dem separat.

Kodeblok 13

```
data <- read_csv2("DF_EKSAMEN.csv")

#finder data fra 1200-1209, hvor nation_tag er Frankrig, og fjerner dubletter.
df_1200s <- data %>%
  left_join(df_sentiment, by = "doc_id") %>%
  filter(årstal >= 1200, årstal <= 1209, nation_tag == "France") %>%
  distinct(doc_id, .keep_all = TRUE)

# Tæl antal breve pr. afsender og modtager
afsender_modtager_1200s <- df_1200s %>%
  group_by(afsender_sted, modtager_sted) %>%
  summarise(antal_breve = n(), .groups = "drop") %>%
  arrange(desc(antal_breve))

#splitter brevene op i enkelte ord.
ord_1200s <- df_1200s %>%
  unnest_tokens(word, text)

#indlæser stopord igen. Har justeret dem til 232 stk.
stopord <- read_excel("stopord.xlsx")

# Fjern dine egne stopord, tal og korte ord
ord_1200s_renset <- ord_1200s %>%
  filter(!word %in% stopord$ord,
    !str_detect(word, "^[0-9]+$"),
    str_length(word) > 1)

#tæller ord.
ord_freq <- ord_1200s_renset %>%
  count(word, sort = TRUE)

#Visualiser
wordcloud(words = ord_freq$word,
  freq = ord_freq$n,
  min.freq = 1,
  max.words = 600,
  random.order = FALSE,
  colors = brewer.pal(8, "Dark2"))
```

Der bliver sendt 27 breve i perioden. Her er den hyppigste interaktion mellem Rom og Frankrig, hvor der bliver sendt 8 breve.

Ved at køre koden ovenfor, ser vi på de 27 breve med omtale af Frankrig fra 1200 til 1209. Vi splitter de forskellige breve op i enkelte ord og fjerner stopord.¹⁵⁴ Herefter skaber vi et Wordcloud i R, som er en visualisering af de hyppigste ord i materialet. Her er størrelsen på ordene et udtryk for, hvor ofte de optræder i brevene. Dette resulterer i *Wordcloud 1*.

Her fremstår det tydeligt, at brevene har en forbindelse til sagen om Ingeborg og Philips ægteskab, da ordet *dronning* er markant og placeret i centrum. Vi finder ligeledes ord som *franskes*, *konge*, *hustru*, samt *sag*, *ret*, *befaling* og *dom* der



alle kan tolkes som værende relateret til sagen. Vi kan ud fra *Wordcloud 1* konkludere, at udsvinget i omtale af Frankrig i årene 1200-1209, skyldes kommunikation om sagen vedrørende Phillip og Ingeborgs ægteskab.

I 1240'erne ser vi ligeledes en markant stigning i graden af breve, hvori Frankrig omtales. Stigningen i kommunikation mellem Danmark og Frankrig følger med. Hvis man gennemgår hvad, som berører det danske rige i 1240'erne, ville en umiddelbar konklusion være Kirkekampen, som begynder i 1241. Niels Skyum-Nielsen beskriver pavens involvering i kirkekampen, idet han skriver, at *Roskilde var det ene brændpunkt i konflikten, Lyon i det fjerne Burgund det andet*.¹⁵⁵ Det er forventeligt, at den markante stigning skyldes breve til og fra pave Innocens IV, da både konger og biskopper *brought their cases before Innocent*.¹⁵⁶ Der bliver sendt 137 breve i 1240'erne, som omtaler Frankrig og dertil 43 breve mellem Danmark og

¹⁵⁴ Da der er få ord i brevene fra 1200-1209, fjernede vi nogle stopord, og arbejdede med en reduceret stopordliste på 232 ord.

¹⁵⁵ Skyum-Nielsen, Niels, *Kirkekampen i Danmark 1241-1290*, Munksgaard, 1963, s. 37

¹⁵⁶ Skyum-Nielsen, *Kirkekampen i Danmark*, s. 343.

Frankrig. De 43 breve er naturligvis repræsenteret i de 137, da de netop omtaler Frankrig, idet de er sendt til/fra Frankrig. I de 137 breve kan vi gennem en analyse holdt op imod de tidligere sentimentudregninger, finde 3 negative breve, 78 neutrale og 56 positive. Her ville vi forvente flere negative breve, hvis de relaterede sig til kirkekampen.

Kodeblok 14

```
bigrams <- df_1240s %>%
+   unnest_tokens(bigram, text, token = "ngrams", n = 2)
+
+   # Split bigrams og fjern stopord
+   bigrams_separate <- bigrams %>%
+     separate(bigram, into = c("word1", "word2"), sep = " ") %>%
+     filter(!word1 %in% stopord$ord,
+            !word2 %in% stopord$ord)
+
+   # Saml og tæl
+   bigrams_filtered <- bigrams_separate %>%
+     unite(bigram, word1, word2, sep = " ") %>%
+     count(bigram, sort = TRUE)
+
+   # Vis som tabel
+   head(bigrams_filtered, 15)
# A tibble: 15 x 2
  bigram                n
  <chr>                <int>
1 apostoliske sædes    43
2 apostoliske sæde     40
3 elskede søn          31
4 kære søn             29
5 elskede sønner       28
6 mark sølv            26
7 tjeneres tjener     26
8 romerske kirkes      20
9 sædes legat          20
10 erik danmarks        19
11 jesu kristi          19
12 menige konvent       18
13 kristus erik         17
14 apostolisk velsignelse 16
15 danmarks berømmelige 16
```

Betyder antal

BIGRAM (resultat)

Vi undersøger derfor brevene fra 1240'erne nærmere ved at skabe et bigram, som en måde at se på sammenhæng i tekstkorpus. Bigrams kombinerer to unigrams (enkelte ord) efter hvilke ordpar, som hyppigst optræder sammen. I dette tilfælde optræder resultatet inde i kodeblok 14.

I vores bigram kan vi se, at de hyppigste ordpar fremstår positive. Ud fra vores kendskab til pavebreve og deres standardiserede opsætning, kan vi udlede at par 1 til 5 samt 7 til 9 er typiske fraser fra pavebreve. Dette stemmer overens med det faktum, at Innocens IV er afsender på 63 af de 137 breve. Ordpar 10, 13 og 15 ser ud til at omhandle den danske kong Erik Plovpenning. I kirkekampens indledende år er konflikten mellem kong Erik Plovpenning og bisp Niels Stigsen af Roskilde. Vi tolker derfor hans optræden som en tegn på, at brevene kan relaterer sig til konflikten. Skyum-Nielsen beskriver hvordan *Innocens paabød en undersøgelse og betroede hvervet som undersøgelsesdommer til en franciskaner, Simon fra Auvergne*.¹⁵⁷ Vi laver derfor en simpel søgning på ordet Simon, for at se om han optræder i de 137 breve. I brevenes tekstkorpus optræder de fem bogstaver *s i m o n* ved siden af hinanden 12 gange.

```
> antal_simon <- str_count(df_1240s$text, regex("\\bSimon\\b", ignore_case =  
TRUE))  
> sum(antal_simon)  
[1] 12
```

Innocens IV's undersøgelsesdommer er dermed til stede i brevene fra 1240'erne som omhandler Frankrig. Vi finder det derfor sandsynligt, at brevene til en vis grad relaterer sig til kirkekonflikten. Hertil kan vi dog ikke udelukke, at brevene i 1240'erne ligeledes omhandler andre emner, da de er overvejende neutrale eller positive.

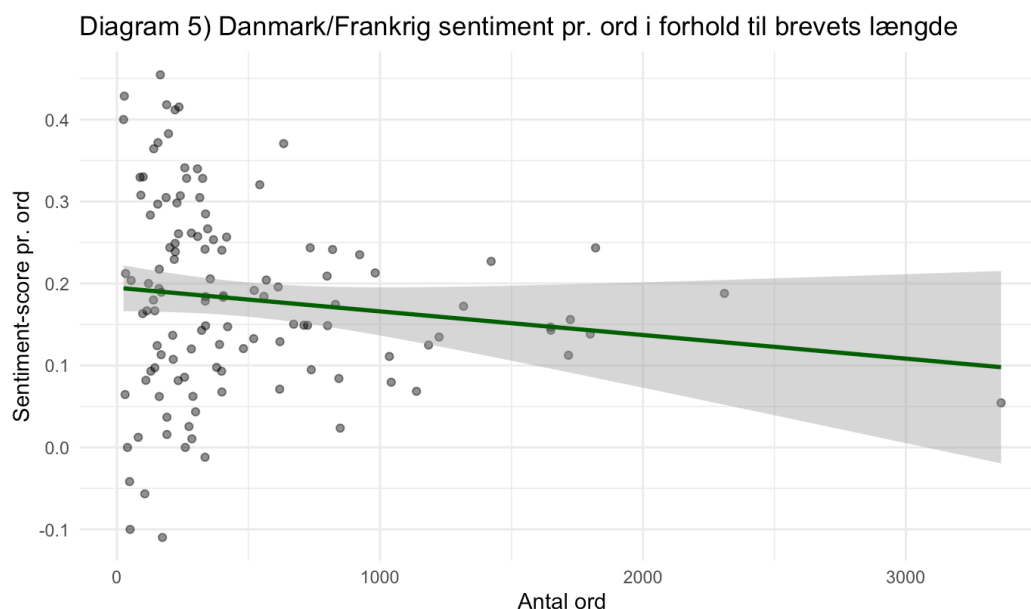
I analysen af det samlede materiale undersøgte vi, om der var en sammenhæng mellem brevenes følelsesmæssige indhold og deres længde.¹⁵⁸ Her fandt vi frem til, at de lange breve indeholder flest positive sentiment, og at de derfor fremstod positive. Dette skyldes dog ophobning af ord. Da vi derefter undersøgte brevene på tættere ordniveau, kunne vi konkludere, at jo længere brevene var, jo mere neutrale blev de. Vi udførte samme undersøgelse op brevene mellem Frankrig og Danmark for at udforske, om brevene mellem de to lande afspejler den

¹⁵⁷ Skyum-Nielsen, *Kirkekampen i Danmark*, s. 32

¹⁵⁸ Jf. Sentimentanalyse

generelle tendens i perioden. Det samlede materiale havde en korrelation på 0.8031474 mellem brevene i deres fulde længde og sentimentscores. Her får brevene mellem Frankrig og Danmark en næsten identisk korrelation på 0.8031293.

Vi kan dermed udlede, at brevene i deres fulde længde afspejler en overordnet tendens, og at der dermed ikke er noget ekstraordinært i den dansk-franske relation i relation til fordeling af sentiment. Vi undersøger naturligvis også ord i brevene i forhold til længde, for ligeledes at undersøge kausalitet i dette tilfælde. Dette resulterer i *Diagram 5) Danmark/Frankrig sentiment pr. ord i forhold til brevets længde* nedenfor, hvor den generelle tendens i de 121 breve mellem Danmark og Frankrig er markeret med grøn. I det samlede materiale var korrelationen -0.1049078. I brevene mellem de to lande er korrelationen -0.1292425 og der fremstår derfor mere negative.



Ved at holde de to analyser op imod hinanden kan vi udlede, at brevene mellem Danmark og Frankrig er mere negative jo længere de bliver. På denne vis adskiller de sig fra den samlede brevsamling. Graden af negativt indhold er dog ikke markant anderledes end tendensen for det samlede materiale, og vi kan derfor vurdere, at relationen mellem de to lande afspejler den generelle tendens i perioden.

I vores analytiske og historiske nedslag omkring Ingeborg-sagen og 1240'erne, dannes der hidtil et billede af en struktur, hvor Frankrigs rolle ikke kan ansæses direkte i

kommunikationen. Det er i stedet en relation, som går igennem forskellige enkeltstående aktører, hertil særligt bundet op omkring paven og den katolske kirke. Frankrig fremstår derfor i høj grad som en indirekte aktør, der alligevel indeholder et element af omdrejningspunkt i perioder med spændinger. Vi er fortsat nysgerrige på forholdet i denne relation og hvad den reelt består i, ud over et ægteskab.

NER - Named Entity Recognition

Vi har hidtil fokuseret på de geografiske og sproglige elementer af vores database, som har givet en tydelig indikation på de institutionelle strukturer, som optræder i brevkommunikationen mellem Danmark og Frankrig.

Vores visualiseringer har tjent det formål at vise, hvordan relationen mellem de to nationer sjældent bygger på en direkte kommunikation. Det er i stedet et større netværk, hvor Frankrig og Danmark agerer i hver deres del, men sammenføres af det helt centrale knudepunkt, kirken. Knudepunkter i vores materiale er først og fremmest repræsenteret ved Rom, Lateranet og Lyon. De anvendte kort og Sankey-diagrammet har ydet den funktion at vise netværkets struktur og hierarki.

Præmissen for en eksplorativ undersøgelse er at teste ideer og forsøge at skubbe grænser. Vores første indskydelse herfra er at lede analysen videre mod, hvordan relationen ser ud, når vi flytter fokus fra det geografiske, og over til entiteter som personer og organisationer. For at muliggøre en analyse, som denne, har vi positioneret os helt ude på grænsen, af hvad der tidligere er forsøgt analyseret på materiale som dette.

Vi har til denne videre undersøgelse gjort brug af værktøjet Named Entity Recognition (NER). Et værktøj der er muligt på grund af vores tidligere digitalisering og sikring er OCR. Miguel Won, Patricia Murrieta-Flores og Bruno Martins skriver i deres artikel *Ensemble Named Entity Recognition(NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora*.

*(...) NER has become of increasing interest in Digital Humanities due to its potential for information extraction and analysis at large scale from historical and literary documents.*¹⁵⁹

Potentialet fremstår særdeles stort og succesfuldt i flere grene af historiefaget. De belyser, hvordan der indenfor flere områder af historisk forskning kommer flere projekter, hvor NER-software er anvendt på historiske datasæt.¹⁶⁰ Softwaren fanger vores interesse, da den fremstår som den nyeste udvikling indenfor digital humaniora.

Da softwaren netop er en helt ny metode, er der begrænsninger for, hvor tilpasset teknologien er til undersøgelsen og dette resulterer i store udsving i succesraten. Særligt med historiske datasæt og kildesamlinger, er der en fejlmargen. I artiklen beskriver de hvordan;

*Historical documents pose multiple challenges. These include language changes over time, spelling variations, transliterations, OCR errors, and sources written in multiple languages.*¹⁶¹

NER er et værktøj, der er indlejret i NLP (*Natural Language Processing*), og det kan bruges til at læse og danne antagelser på materialet. Softwaren anvendes til automatisk at identificere og kategorisere enheder i den undersøgte tekst. Den software, som vi har anvendt i Python, er spaCy-pakken *da_core_news_lg*. Denne pakke er sat op til at kunne forstå dansk tekst og identificere personer (PER), lokationer (LOC) og organisationer (ORG). Formålet med anvendelsen af NER er at kunne gå dybere med materialet og anvende teksten på nye måder, end det hidtil har været muligt. NER-algoritmen skaber ligeledes muligheder for at bearbejde store datasæt, og særligt store datasæt, der primært består af et tekstkorpus.

Til vores brug og implementering af NER-algoritmen, har vi anvendt følgende kodeblok 15.

¹⁵⁹ Won M, Murrieta-Flores P og Martins B (2018). "Ensemble Named Entity Recognition (NER): Evaluating NER Tools in the Identification of Place Names in Historical Corpora". *Frontiers in Digital Humanities*. 5:2. [doi: 10.3389/fdigh.2018.00002](https://doi.org/10.3389/fdigh.2018.00002), s.3

¹⁶⁰ Won et. al. "Ensemble Named Entity Recognition", s. 3

¹⁶¹ Won et. al. "Ensemble Named Entity Recognition", s. 1

Kodeblok 15

```
import pandas as pd
import spacy

# Indlæs NER pakke
nlp = spacy.load("da_core_news_lg")

# Indlæs dataframe
df = pd.read_excel('/Users/Username1/Desktop/df-2.xlsx')

# Hvilken kolonne i filen df skal analyseres
tekst_kolonne = 'text'

# Gem entiteter fra hver række
def extract_entities(text):
    doc = nlp(str(text))
    return [(ent.text, ent.label_) for ent in doc.ents]

df['entiteter'] = df[tekst_kolonne].apply(extract_entities)

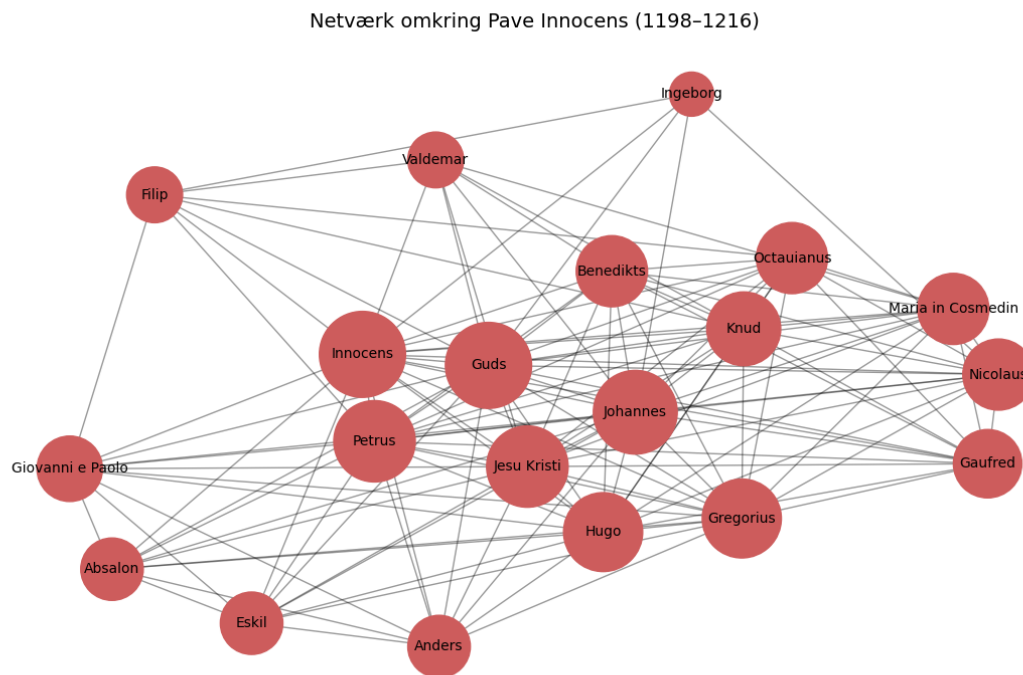
# Gem resultat
df.to_excel('/Users/Username1/Desktop/df_ner_analyseret.tekst.xlsx',
index=False)
```

Denne kodeblok 15 gennemgår hver række i databasen og tagger de entiteter, som den er i stand til at genkende. Det giver et output i form af en ny kolonne kaldet *entiteter*. Vi har her kørt NER på alle 2002 breve og er endt med 256.676 entiteter.

Indholdet i kolonnen er eksemplificeret her med et eksempel fra et enkelt brev: [('Guds', 'ORG'), ('Århus', 'LOC'), ('Sorø Gimlingetorp', 'ORG'), ('Gimlinge', 'LOC'), ('Guds domstol', 'ORG'), ('Guds', 'ORG'), ('Jesu Kristi', 'PER'), ('Kristus', 'MISC')].

Vi benytter NER-algorithmens resultater til at skabe de to netværk *Netværk 1) Netværk omkring pave Innocens III (1198-1216)* og *Netværk 2) Netværk omkring pave Innocens IV (1243-1254)*, som visualiserer netværk omkring Pave Innocens III og Pave Innocens IV. Netværkene viser alle *PER_entiteter*, som brevene indeholder, og hvordan disse relaterer sig til de to paver.

Netværk 1) Netværk omkring pave Innocens III (1198-1216)

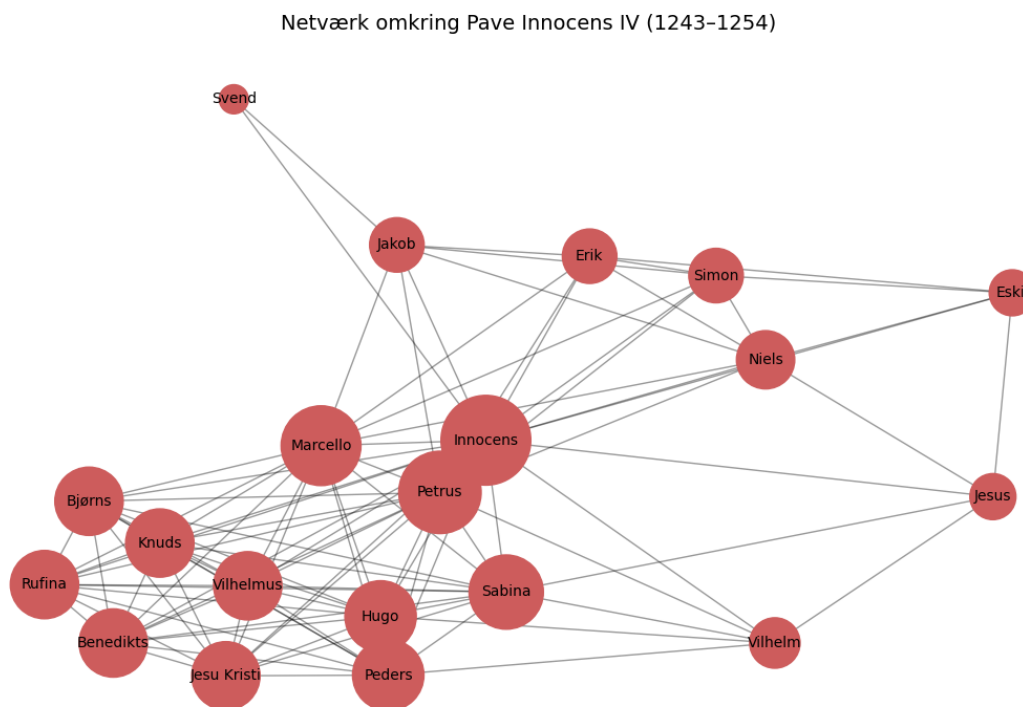


I Innocens III's netværk er der flere centrale aktører, som er tæt forbundet. I dette *netværk 1* er det interessant, at de mest fremtrædende er *Petrus* og *Johannes*. Dette vidner om Innocens III's brug af biblen, da der refereres til apostlen Peter og apostlen Johannes.¹⁶² I toppen af netværket ser vi Filip, Valdemar og Ingeborg, som alle er tilknyttet Innocens III. Hertil ser vi dog også direkte forbindelser mellem Ingeborg og Filip samt Filip og Valdemar. Selvom der er forbindelser mellem dem, behøver dette ikke være et udtryk for direkte kommunikation. NER koden bygger på entiteter pr. brevrække og tager ikke højde for afsender og modtager. Outputtet er derfor en visuel formidling, der viser sammenhæng mellem individer i form af omtale.

På Innocens IV's netværk *Netværk 2) Netværk omkring pave Innocens IV (1243-1254)* støder vi ligeledes på Petrus, som er tagget som en central person. I dette tilfælde er der tale om kardinalpresbyter Petrus.

¹⁶² Apostlen Peter optræder nærmest konsekvent i pavebrevene, da han som kirkens første pave og nøglebærer, benyttes som en central figur til legitimering af pavens autoritet.

Netværk 1) Netværk omkring pave Innocens IV (1243-1254)



Her skaber NER-algoritmen et netværk, hvor Marcello, Petrus og Sabina er placeret centralt. Vi fejlkilder her på samme vis, som vi så fejlkilder i Innocens III's netværk. Vi er generelt forbeholdne over for det output, som NER giver os. LOC fungerer godt, og skaber få fejlkilder. Vi er dog skeptiske overfor resultaterne vedrørende PER og ORG, da vi her ser adskillige fejlkilder, som eksempelvis Marcello og Sabina. Disse er begge fejlkilder, da de ikke bør opfattes som to individer. Marcello bør blot referere til et individ, da der i DRB_1771 skrives *Jeg Petrus, kardinalpresbyter af S. Marcello*, og Marcello derfor er et sted og ikke en person.¹⁶³ Samme fejl opstår i relation til Sabina, som computeren ligeledes tror er en person. Her refereres der i virkeligheden til en kirke i Rom gennem teksten *Til Hugo, Kardinalpræst af Sta. Sabina*.¹⁶⁴ Her optræder en væsentligste fejlkilde, som Won, Murrieta-Flores og Martins belyser i deres artikel, i form af forkerte tags.

I vores arbejde med NER står det klart, at de computationelle muligheder fortsat er begrænset. Det er derfor vigtigt at inddrage historiefagligheden i en undersøgelse som denne, for at fejltags opdages. Det er nødvendigt at tjekke og gennemgå data samt validere outputtet.

¹⁶³ DRB_1771 database

¹⁶⁴ Brev 91 fra d. 20. januar 1253, 2. række 1. bind (1250-1265.)

Efter gennemgang af vores entiteter, bliver kompleksiteten af vores materiale tydelig. Det kan drages tilbage til Won, Murrieta-Flores og Martins pointe om begrænsningerne ved NER.¹⁶⁵ De belyser ligeledes i deres artikel, hvordan de største muligheder for anvendelsen af denne software er at finde i fremtiden. De skriver;

*Although the Spatial Humanities has advanced very quickly in term of the analysis of place names mentioned in corpora, as the field grows and scholars look for expedite techniques to support the exploration of large scale digital datasets, more research is need to solve particular problems related to the correct identification of place names in historical datasets.*¹⁶⁶

Vi anerkender i særdeleshed de fremtidige muligheder, men må vurdere NER til ikke at være udviklet på et højt nok niveau endnu, til vi kan anvende det yderligere i vores analyse. Grundet tidsbegrænsninger for vores opgave, er det ikke en mulighed for, at begynde processen af at skabe vores egen NER model, som er specifikt til dansk middelaldermateriale.¹⁶⁷

Her kunne en løsning være en manuel gennemgang og rettelser af de forskellige NER-tags. Da vi har 256.676 forskellige entiteter, er dette dog urealistisk og fuldstændig uoverskueligt.

Vi vælger at inddrage og fremlægge vores NER-analyse, da arbejdet rummer positiv indsigt såvel som fejlslagne resultater. Vi ønsker at fremstille arbejde på trods af den betydelige fejlmargen, da dette er med til at belyse nødvendigheden af at supplere de digitale metoder med historiefaglig indsigt. At undlade NER-analysen ville efter vores vurdering være en misvisende fremstilling af de digitale metoder, særligt når undersøgelsen fungerer i relation til identifikation af lokationer. Vi ønsker med inddragelsen af NER at belyse, at de digitale metoder ikke skal ansues som en fejlfri løsning, men at de i stedet kræver kritisk refleksion og faglig kontekstualisering.

Da NER ikke kunne benyttes til at identificere aktører, så vi muligheden for at gå i en alternativ retning til vores videre undersøgelse af netværkene i vores materialesamling. Vi har manuelt

¹⁶⁵ Jf. NER (Named Entity Recognition)

¹⁶⁶ Won et. al. "Ensemble Named Entity Recognition", s. 10

¹⁶⁷ Denne proces kan perspektiveres til skabelsen af en sprogmodel i Transkribus, dog mere omfattende.

gennemgået brevene og tagget afsender og modtager.¹⁶⁸ Det muliggør en netværksanalyse, hvor to punkter specifikt bindes sammen.

Magtens mellemlid

For at skabe et overblik og samtidig undersøge strukturer i netværket mellem Danmark og Frankrig, har vi skabt et netværk i RStudio, som kan ses nedenfor. For at kunne forbinde de enkelte individer, var vi nødsaget til at skabe yderligere manuel metadata. Vi har i ovenstående kunne forbinde geografiske lokationer, organisationer og steder for afsendelse og modtagelse af brevene. Vi ønskede i denne sammenhæng at visualisere netværket mellem aktørerne. Derfor har vi gået igennem brevene, læst teksten, og markeret hvem som er afsender og hvem, der er modtager på hvert enkelt brev. Det var nødvendigt med en manuel gennemgang, da der ofte omtales mange personer i brevene. Af denne grund, var det nødvendigt med en selektiv vurdering, for at konkluderer, hvem som var den egentlige modtager af et givent brev.

Efter vi havde markeret afsender og modtager, skabte vi følgende kodeblok 16, som fungerer i RStudio.

¹⁶⁸ Vi har her tagget de 121 afsender og de 121 modtager i breve mellem Danmark og Frankrig. Vi har ligeledes tagget de 745 afsender samt de 745 modtagere i pavebrevene. Vi har afholdt os fra at tage hele materialet, da dette var for tidskrævende.

Kodeblok 16

```
#årstal numerisk og retter to fejl, hvor vi har skrevet Pave
df <- df %>%
  mutate(
    årstal = as.numeric(årstal),
    afsender = gsub("^Pave\\s+", "", afsender),
    modtager = gsub("^Pave\\s+", "", modtager)
  )

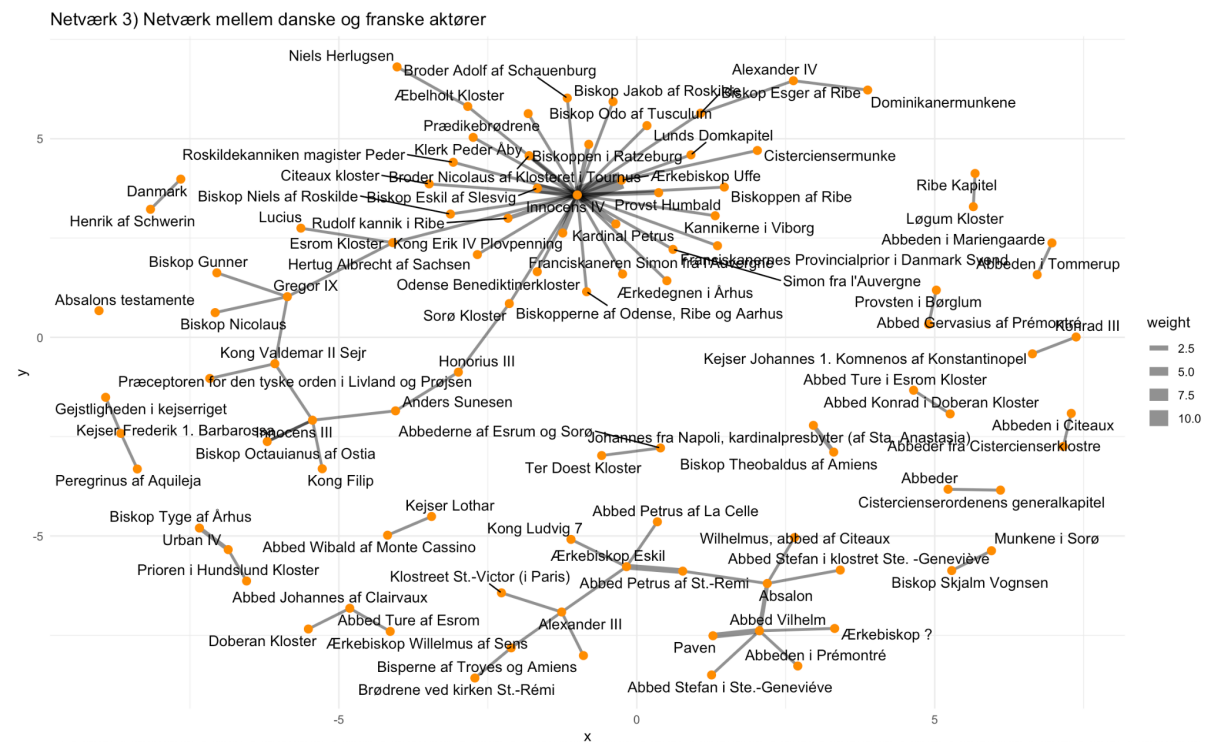
# Fjern dubletter så vi kun har ét brev per doc_id
df_clean <- df %>%
  arrange(doc_id) %>%
  group_by(doc_id) %>%
  slice(1) %>%
  ungroup()

#Hvem skriver med hvem, og hvordan er relationerne?
# Lav netværksdata: noder og kanter
edges <- df_clean %>%
  select(afsender, modtager) %>%
  filter(!is.na(afsender) & !is.na(modtager)) %>%
  group_by(afsender, modtager) %>%
  summarise(weight = n(), .groups = "drop")

# Lav grafen
g <- graph_from_data_frame(edges, directed = TRUE)

# Plot netværket
ggraph(g, layout = "fr") +
  geom_edge_link(aes(width = weight), alpha = 0.5) +
  geom_node_point(color = "darkorange", size = 2.5) +
  geom_node_text(aes(label = name), repel = TRUE) +
  theme_minimal() +
  ggtitle("Netværk 3) Netværk mellem danske og franske aktører")
```

Koden skaber *Netværk 3) Netværk mellem danske og franske aktører* med orange punkter, som repræsenterer enkelte aktører. *Netværk 3* viser, hvordan disse aktører er forbundet gennem kommunikation, samt graden af den indbyrdes kommunikation. Tykkelsen på strengen er sat efter antal breve mellem de forbundne aktører.



Det er tydeligt, at Innocens IV er centrum i en stor del af det kommunikative netværk. Det netværk, som han er forbundet i midten af, forgrener sig ud i et mindre netværk omkring Innocens III og et omkring Gregor IX. Vi ser endnu et netværk, som er isoleret fra det primære netværk i toppen af billedet. Her er netværket forbundet mellem danske og franske gejstlige aktører, og der er en særlig stærk forbindelse mellem *Abbed Vilhelm* og *Paven*, samt *Abbed Petrus* og *Ærkebiskop Eskil*. De mindre relationer mellem to enkelte aktører, giver udtryk for at være kommunikation mellem to enkelte individer, frem for religiøse instanser.

På netværket ser det ud til at det institutionelle er bundet op omkring paverne, og de mindre netværk og korrespondancer er bundet op omkring individer.

Dette netværk 3) Netværk mellem danske og franske aktører er interessant fordi vi kan se tilhørsforhold mellem aktører. Netværket fremstår mere nært end de tidligere visualiseringer, hvor det er svært at få en forståelse for personerne bag de forskellige geografiske lokationer.

For at lade alle vores tidligere undersøgelser gå op i en højere enhed, sammensætter vi vores aktørnetværk med den tidligere udregnede sentimentdata samt kommunikationsvej. Dette sker i kodeblok 17 her:

Kodeblok 17

#kode til sentimentfarve

1. Sammenfat data og beregn gennemsnitlig sentiment pr. relation

```
edges <- df_full %>%
  group_by(afsender, modtager) %>%
  summarise(
    n_breve = n(),
    avg_sentiment = mean(sentiment_score, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(
    sentiment_kategori = case_when(
      avg_sentiment < 15 ~ "negativ",
      avg_sentiment > 50 ~ "positiv",
      TRUE ~ "neutral"
    )
  )
```

2. Lav nodeliste

```
nodes <- unique(c(edges$afsender, edges$modtager)) %>%
  tibble(name = .)
```

3. Lav graf

```
graph <- graph_from_data_frame(d = edges, vertices = nodes, directed = TRUE)
```

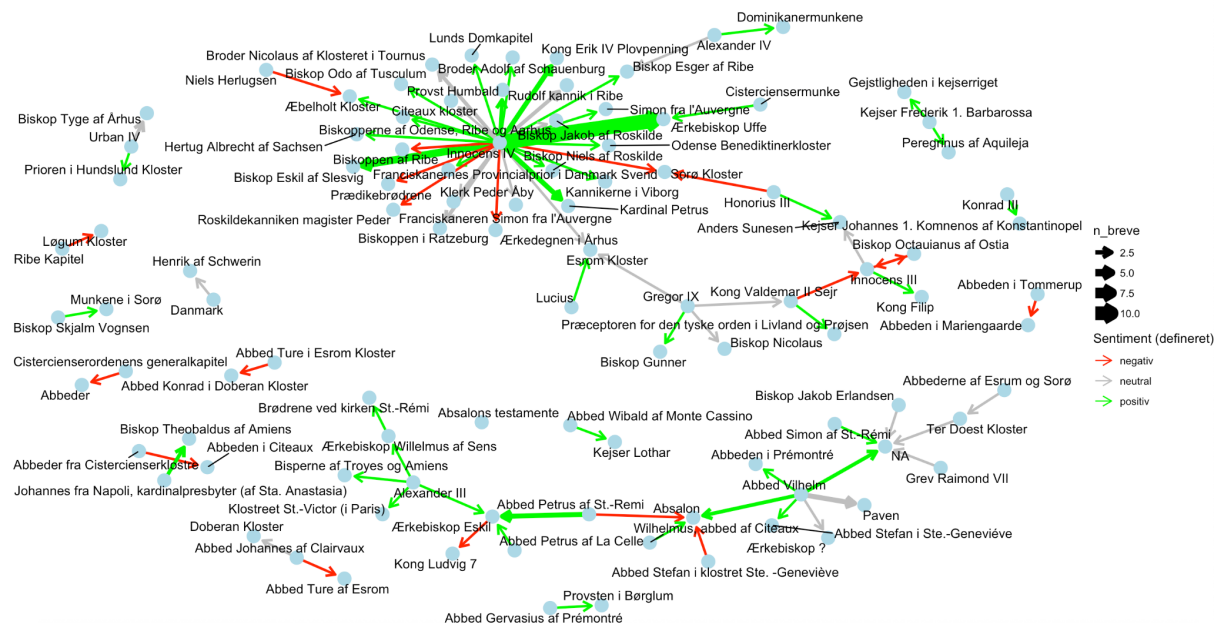
4. Visualisér

```
ggraph(graph, layout = "fr") +
  geom_edge_link(
    aes(width = n_breve, color = sentiment_kategori),
    arrow = arrow(length = unit(3, 'mm')),
    end_cap = circle(3, 'mm'),
    alpha = 1
  ) +
  scale_edge_color_manual(
    values = c("negativ" = "red", "neutral" = "gray", "positiv" = "green"),
    name = "Sentiment (defineret)"
  ) +
  geom_node_point(size = 5, color = "lightblue") +
  geom_node_text(aes(label = name), repel = TRUE) +
  theme_void() +
  ggtitle("Netværk 4) Danmark/Frankrig netværk med sentiment-kategorier")
```

Når alt data er sat sammen, får vi *Netværk 4) Danmark/Frankrig netværk med sentiment-kategorier*, som viser aktørnetværket i de 121 breve mellem Danmark og Frankrig, samt det

følelsesmæssige indhold i brevene. På netværk 4 er der pile mellem afsender og modtager som peger den vej, som brevet rejser.

Netværk 4) Danmark/Frankrig netværk med sentiment-kategorier



Her ser vi sentimentscore indlejret i pilene, så de røde pile repræsenterer negativt indhold i kommunikationen, de grønne repræsenterer positivt og de grå repræsenterer neutralt. Pilenes størrelse viser ydermere graden af kommunikationer i form af antal breve sendt mellem de to aktører. *Netværk 4) Danmark/Frankrig netværk med sentiment-kategorier* kan fremstå stokastisk, fordi det ikke ligner ovenstående *netværk 3) Netværk mellem danske og franske aktører* 1:1. Her er der dog tale om et deterministisk netværk, fordi forbindelserne er koblet op på kommunikation mellem afsender og modtager. Netværk 4 er blot kørt i en algoritme, som roterer det samlede netværk. Det er derfor den samme visualisering som blot har ændret farve og har en tilfældigt roteret opsætning.

Her finder vi det særligt interessant, at den negative kommunikation går ud fra Innocens IV's netværk, og at der samtidig ikke er noget negativt, som går direkte ind i Innocens IV's netværk. De røde pile og dermed negative breve er særligt mellem nationale aktører, som eksempelvis Ribe kapitel og Løgum kloster eller Abbeden i Tommerup og Abbeden i Mariengarde.

Det er ligeledes interessant, hvordan det kun er i Innocens III's netværk, at der skildres en negativ relation, som går begge veje. Han er den eneste, der både modtager og sender negativ kommunikation, i hans korrespondance med biskop Octavianus af Ostia. Der er overvejende

flest positive breve i kommunikationen mellem Danmark og Frankrig, og dette kan være årsagen til, at de negative breve ikke optræder tydeligt i netværket.

På *Netværk 4) Danmark/Frankrig netværk med sentiment-kategorier* er kommunikationen mellem Innocens IV og ærkebiskop Uffe fremhævet, da graden af kommunikation er høj. Pilen er i dette tilfælde grøn. Dette betyder ikke, at der ikke er negative breve mellem de to. Der er blot flest positive breve, så disse dækker de negative breve mellem Innocens IV og ærkebiskop Uffe.

Det er ydermere interessant at dvæle ved Innocens III's kommunikation til Filip Augustus, der fremstår som værende positiv. Som tidligere belyst havde Filip og Innocens III en konfliktfyldt relation, som dog blev hjulpet af, at den franske kirke agerede mellemmand for Filip interesser. Det ville derfor have været forventeligt med en rød markør imellem denne relation. I dette tilfælde fylder de positive breve mest, og de overskygger derfor de negative.

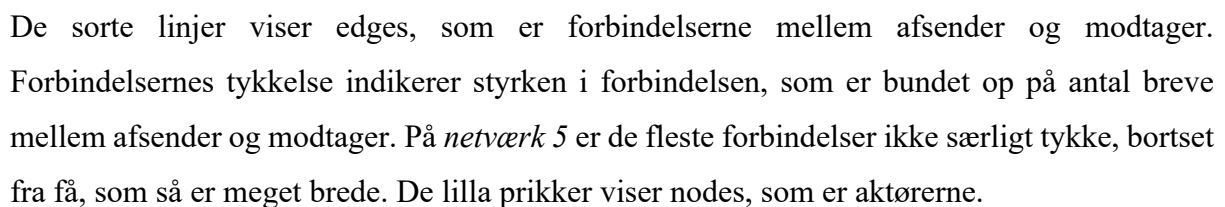
I undersøgelsen af relationen mellem Danmark og Frankrig, finder vi det væsentligt at se nærmere på paverne i samtiden. Paverne i middelalderen er magtfulde overnationale aktører, og optræder ofte som mellemmand mellem magter. Brevene fra paverne kan ligeledes afspejle konflikter, alliancer og diplomati. Det er derfor relevant at undersøge pavestolen og dens relation til og mellem de to riger.

Her laver vi et skifte i datagrundlaget, og bevæger os væk fra de 121 breve mellem Danmark og Frankrig, og tilbage til de 745 pavebreve, som vi filtrerede fra i de indledende faser af analysen. De 745 pavebreve vil udgøre datagrundlaget for denne sidste del af analysen.

Det hellige netværk

Som tidligere belyst finder vi en del pavebreve i vores materialesamling, idet der optræder 745 breve fordelt på 18 paver. I denne sammenhæng så vi, at Innocens IV var den absolut mest aktive afsender. Efter at have taget de 745 afsender og 745 modtagere, skaber vi ligeledes et netværk over pavebrevene, som kan ses nedenfor *Netværk 5) Netværk i pavebreve (745 breve)*. Dette *netværk 5* er væsentligt sværere at afkode, end netværket i de dansk-franske breve, da der er flere nodes og edges.¹⁶⁹

¹⁶⁹ Vi har benyttet kodeblok 16. Vi har filtreret i netværket, for at det er muligt at tyde de forskellige labels.



For at se nærmere på forbindelserne mellem de væsentligste aktører, har vi filtreret i netværket ud fra beregninger om *degree centrality*.¹⁷⁰ Degree centrality er en beregning for, hvor mange forbindelser et givent punkt har i et netværk.

100

Kodeblok 18

```
#prøver at fremhæve centrale aktører ved at udregne betweenness og centrality
edges <- df_paver %>%
  filter(!is.na(afsender) & !is.na(modtager)) %>%
  count(afsender, modtager, name = "weight")

G <- graph_from_data_frame(edges, directed = TRUE)

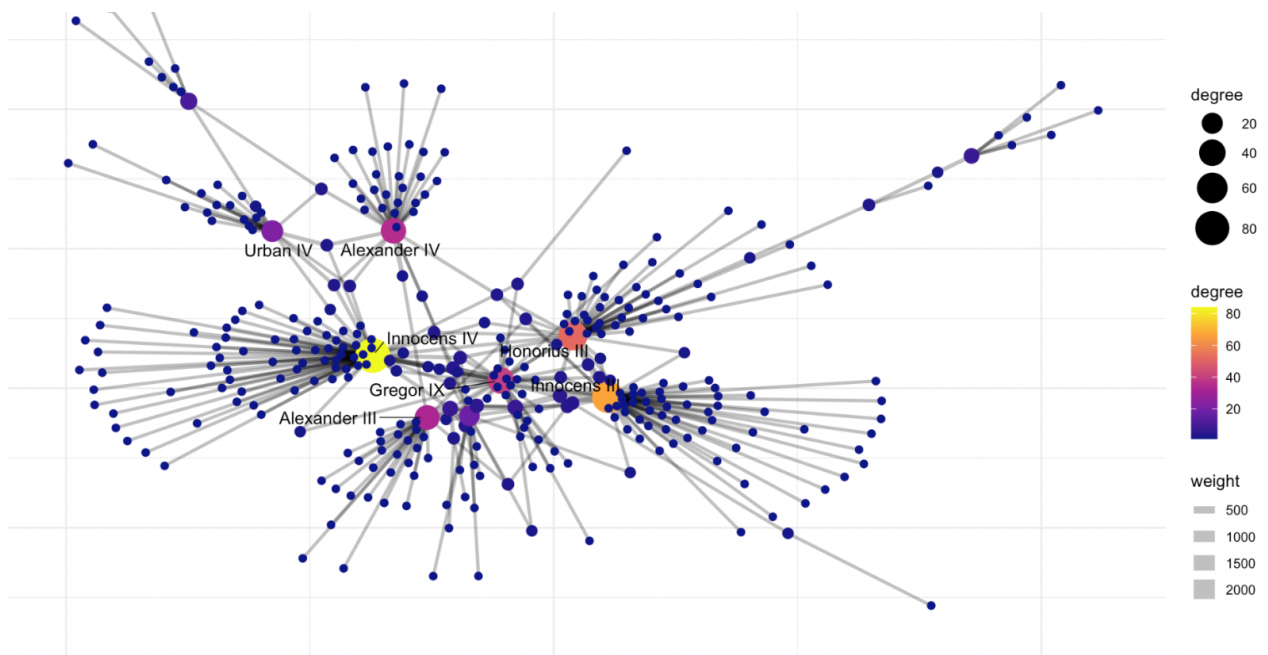
V(G)$degree <- degree(g, mode = "all")

V(G)$betweenness <- betweenness(g, directed = TRUE)
V(G)$eigen <- eigen_centrality(g)$vector

# Plot med centralitet som størrelse og farve
ggraph(G, layout = "fr") +
  geom_edge_link(aes(width = weight), alpha = 0.3) +
  geom_node_point(aes(size = degree, color = degree), show.legend = TRUE) +
  scale_color_viridis_c(option = "plasma") +
  scale_size(range = c(2, 10)) +
  geom_node_text(aes(label = name), repel = TRUE, max.overlaps = 10) +
  theme_minimal() +
  ggtitle("Centrale aktører i netværket (degree centrality)")
```

De vigtigste aktører er fremhævet med en farve, som stiger i degree, jo mere varm (mod det gule) farven bliver. Når et punkt har høj *degree centrality*, har punktet mange forbindelser i netværket. I fordelingen af afsendere af pavebreve så vi, at Innocens IV sendte flest breve. Vi ser nu, at han har et helt gult punkt, og dermed den højeste *degree* og derfor flest forbindelser. Vi tolker det som naturligt, at den pave, som har den højeste grad af kommunikation, også har flest modtagere. På netværket neden, kan vi ligeledes se, at Innocens III er den anden aktør, som har størst kommunikationsvifte fra sit punkt. Han er her markeret med orange, som også har en høj degree.

Netværk 6) Netværk i pavebreve med centrale aktører (degree centrality)



Ud fra *Netværk 6) Netværk i pavebreve med centrale aktører (degree centrality)*, kan vi tolke, at de mest centrale paver i pavebrevene er Innocens IV, Innocens III, Honorius III og Alexander IV. De har hver især et vifteformet netværk, som stråler ud fra deres punkt. Punkterne i disse vifter er umiddelbart ikke forbundet med andet end knudepunktet (Innocens IV, Innocens III, Honorius III og Alexander IV).

Da vi har identificeret den mest centrale aktører, får vi en bredere forståelse for de centrale paver i perioden. Vi sætter lighedstegn mellem graden af kommunikation og stadfæstelse af legitimitet, da vi ser, at paverne sender breve for at udøve, stadfæste og styrke deres legitimitet.

Vi tildeler nu *Netværk 5) Netværk i pavebreve (745 breve)* sentimentdata, for at få en forståelse for kommunikationens følelsesmæssige perspektiv. På *Netværk 7) Netværk i pavebreve med sentiment-kategorier*, kan vi se samme netværk som *Netværk 5) Netværk i pavebreve (745 breve)*. Her er det tydeligt, at Innocens III og Innocens IV sender flest negative breve, at Gregor IV sender flest neutrale breve og at Alexander IV sender klart flest positive breve.

Netværk 8) Zoom på Netværk i pavebreve med sentiment-kategorier



De er placeret tæt på hinanden i midten af netværket, hvor de er forbundet i mange retninger. I filtreringen forsvinder de dog, da de statistisk set ikke er personer af særligt stor vigtighed for paverne i samtiden. Graden af direkte kommunikation er ikke høj nok til at visualisere dem i det filtrerede netværk. Ved at køre nedenstående kodeblok 19 undersøger vi netværkstætheden over tid.

Kodeblok 19

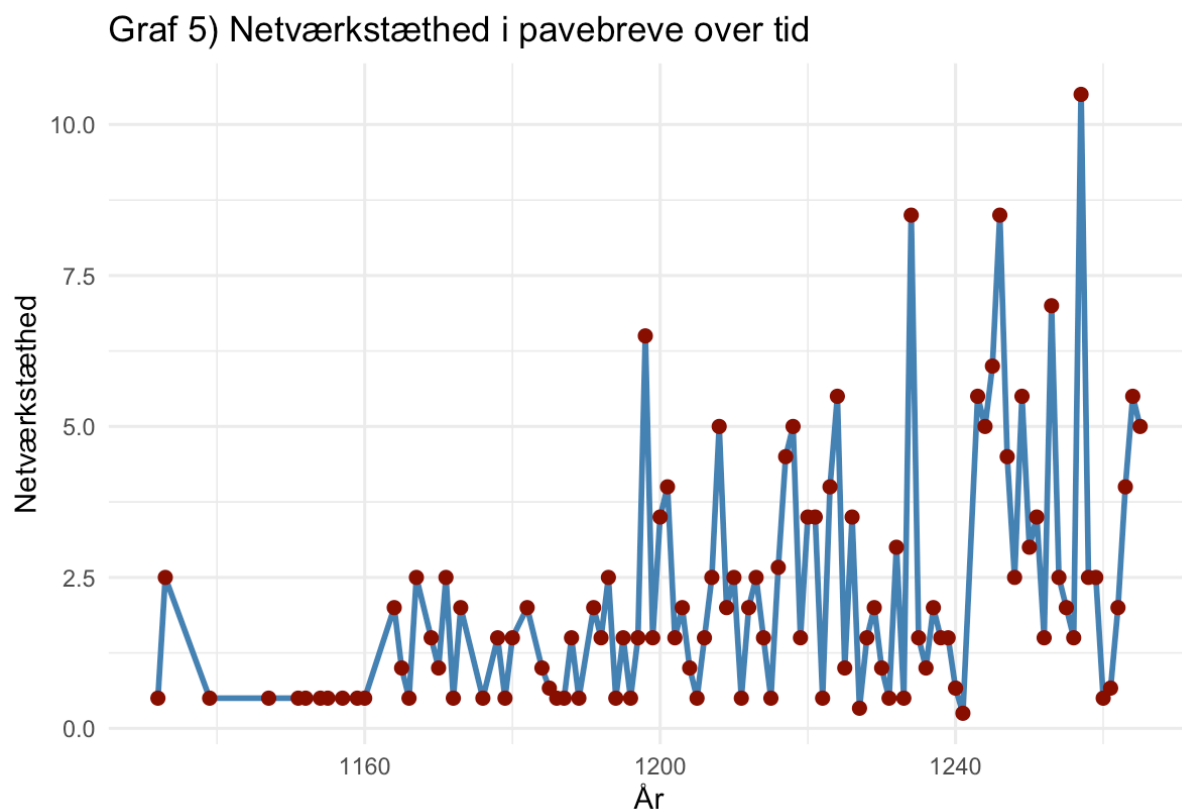
```
# Split per år og beregn netværkstæthed
net_density_by_year <- edges_unique %>%
  group_split(årstal) %>%
  map_dfr(function(df_year) {
    year <- unique(df_year$årstal)

    # Skal være mindst 2 noder for at give mening
    if (nrow(df_year) >= 1) {
      g <- graph_from_data_frame(df_year, directed = TRUE)
      data.frame(
        årstal = year,
        density = edge_density(g, loops = FALSE)
      )
    } else {
      data.frame(
        årstal = year,
        density = NA
      )
    }
  })

ggplot(net_density_by_year, aes(x = årstal, y = density)) +
  geom_line(color = "steelblue", size = 1) +
  geom_point(color = "darkred", size = 2) +
  labs(
```

```
title = "Graf 5) Netværkstæthed i pavebreve over tid",  
x = "År",  
y = "Netværkstæthed"  
) +  
theme_minimal()
```

Dette resulterer i følgende *Graf 5) Netværkstæthed i pavebreve over tid*, som viser hvor tæt og spredt relationerne i netværket er over perioden fra 1131 til 1265.



Graf 5) Netværkstæthed i pavebreve over tid bliver væsentligt tættere med tiden. Der sker en markant fortætning under Innocens III og igen omkring Innocens IV. *Graf 5* bekræfter Innocens III's måde at formidle pavestolens magt, da vi kan se, at pavedømmets relationelle netværk bliver stærkere i denne periode.¹⁷¹ At netværkets tæthed daler drastisk lige inden år 1240, er et udtryk for mindre kommunikation mellem aktørerne. Dette kan skyldes en periode med stilstand såvel som en periode med få bevarede breve.

¹⁷¹ Jf. Grundstenene

Vi tolker netværkstæthedens udvikling som et udtryk for pavedømmets styrkelse af deres relationelle netværk til Danmark. Da netværkene er skabt på baggrund af DRB, har brevene en vis relation til det danske rige. Vi kan hermed se, at netværket mellem Danmark og pavedømmet bliver tættere gennem tiden.

Innocens IV – Manden i midten

Vil gerne undersøge Innocens IV breve individuelt, da vi tidligere har konkluderet, at han er den pave, der optræder som hyppigst afsender i vores materiale. Vi ser ydermere at han er central i *Netværk 3) Netværk mellem danske og franske aktører mellem Danmark og Frankrig* og den mest centrale pave i *Netværk 6) Netværk i pavebreve med centrale aktører (degree centrality)*.

Netværk 9) Netværk for Innocens IV med sentiment-kategorier



Netværk 9) Netværk for Innocens IV med sentiment-kategorier viser alle modtagere af breve fra Innocens IV, og vi kan her se, at der er en markant forbindelse til Sorø Kloster. Netværk 9 er opsat efter samme parametre som de ovenstående i forhold til sentiment og tykkelse på forbindelser.

På *Netværk 9) Netværk for Innocens IV med sentiment-kategorier* kan vi se negative breve til biskop Niels af Roskilde og kong Erik, hvilket vi tolker som et udtryk for Innocens IVs involvering i kirkekampen.¹⁷²

Netværk 9 afspejler hvordan pavemagtens kommunikation med og styrede det kirkelige landskab i Danmark. Dette ser vi gennem forbindelser til danske klostre og biskopper. Netværket bliver dermed et udtryk for den pavelige administrations indflydelse på gejstligheden i Danmark. Det fremstår centraliseret, og vi ser Innocens IV som repræsentant for en administrativ organisation, som gør ham til et kommunikativt knudepunkt. Denne pointe kan perspektiveres til de resterende paver i ovenstående netværk.

I netværkene ovenfor ser det ud til, at en af de stærkeste kommunikative relationer mellem pavedømmet og Danmark, optræder i form af Sorø Kloster og de tilhørende munke. Om dette er et udtryk for, at Sorø Kloster har været særligt dygtige til at bevare materiale, kan vi ikke vurdere. Kim Esmark beskriver i sin artiklen *Fromhed og ære: Hvidernes gaver til Sorø Kloster i 11- og 1200-tallet*, at munkene i Sorø Kloster var opmærksomme på skriftlig dokumentation af aftaler og bevaring af breve, da de oplever problemer med, stridigheder mellem arvinger i forhold til testamenterede donationer til klostret.¹⁷³

Den markante tilknytning til pavedømmet kunne ligeledes skyldes de politiske strømninger i perioden. Hvideslægten er toneangivende i Danmark, og skaber Sorø Kloster som en fysisk manifestation af deres slægts storhed. For særligt Absalon bliver Sorø Kloster et projekt, som skal sikre ham prestige i livet såvel som i døden.¹⁷⁴

Ctrl+ Alt+ Historie – En refleksion over digitale metoder

I vores *Speciale_database* havde vi struktureret lidt over 15.000 breve fra *Danmarks Riges Breve*, som vi til dette projekt filtrerede ud i en relationel database. Den relationelle database indeholder de 2002 breve, som vi har behandlet i dette projekt. Til vores relationelle database har vi gennemført en stor mængde databehandling. Denne behandling er relevant at sætte i kontekst til Torgersons artikel *Historical practice in the era of digital history* og særligt hans inddragelse af Johanna Druckers begreb *capta*. Torgerson opsætter en metodisk fremgangsskabelon, ved at opdele databehandlingen i *capturing sources, producing data*,

¹⁷² Jf. Når magten taler – Brevenes sprog mellem nord og syd

¹⁷³ Kim Esmark, *fromhed og ære* s. 108-126

¹⁷⁴ Esmark, "Fromhed og ære", s. 113.

creating facts.¹⁷⁵ Heri ligger der et argument for, at det er det forudgående arbejde med kilderne, som skaber fakta. Riva argumenterer ligeledes for, at det er opsætningen af data, som skaber hele grundlaget for undersøgelsen.¹⁷⁶

Mange af de valg, som vi har foretaget, er et produkt af akkumuleret viden. Dette gør, at vi har kunnet vurdere, hvilke metodiske valg der for vores arbejde fremstod logiske og naturlige. Drucker har teoretiseret processen og tillagt det begrebet *capta*. Vores etablering af et analytisk grundlag har derfor været et arbejde med *capta*.

I stedet for at forstå materialet, som data der allerede eksisterer, er vores analyse bygget op omkring det, som vi har taget ud og organiseret. Her har vi opsat en eksplorativ præmis for projektet og derfor forsøgt at inddrage forskellige elementer, for at kunne opbygge en bredt funderet analyse.

Vi valgte derfor at involvere elementer som afsender, modtager, årstal, steder og sentiment. Grundlaget for disse valg kommer i forlængelse af vores forståelse som historikere og det særlige kendskab til materialet i *Danmarks Riges Breve*, som vi har fået i takt med vores middelalderspecialisering. På baggrund af tidligere dybdegående kvalitative og kvantitative undersøgelser, havde vi et forspring i forhold til at se hvor mulighederne var, samt hvor vi kunne skubbe til grænserne. Vores forståelse af materialet og *capta*-begrebet, er blandt andet særligt vigtig i vores arbejde med sentimentanalyse. Her gik vi ind og skabte en sentimentordbog og tilpassede den til kildematerialet. Ordet *interdikt* påvirkede vi direkte og tilskrev det værdien -5. Denne proces er en aktiv anvendelse af *capta*, da vi gik ind og skabte en mening i dataen i stedet for blot at læse og forstå dataene direkte. Anvendelsen af *capta* kan overføres til resten af vores databehandling, da vi selv trækker data ud fra kilderne, som ikke er givet på forhånd, idet vi skaber metadata. Vores produktion af metadata over de sidste 10 måneder, skaber grundlaget for analysen og dermed det nye perspektiv på materialet.

Jessica Parland-von Essen beskriver hvordan forskeren gennem sit digitale arbejde, kan rammes af *romantic lonely genius syndrome*.¹⁷⁷ Et syndrom hvor forskeren bliver opslugt af genialiteten i eget arbejde og derfor glemmer at redegøre for dataens tilblivelsesproces. Dermed

¹⁷⁵ Torgerson, "Historical Practice in the era of digital history" s. 37

¹⁷⁶ Riva, "Network Analysis of Medieval Manuscript Transmission" s. 38

¹⁷⁷ Jf. Den digitale bibel

forsvinder gennemsigtheden i arbejdet og det skaber et kompliceret workflow for videre arbejde med datasættet.

For at undgå *romantic lonely genius syndrome* i vores arbejde, har vi fra starten været opmærksomme og fokuserede på at redegøre for hvordan vi skabte dataene. Det er årsagen til, at vi gennem opgaven har redegjort for hvilke software og kodepakker, som vi har benyttet. Vores egne koder, som vi har konstrueret gennem analysen, har vi ligeledes delt, for at være gennemsigtige i vores arbejde og for at gøre reproduktion muligt. Vi har ydermere beskrevet vores metodiske arbejde grundigt, for at sætte fokus på det usynlige arbejde, som ligger i databehandlingen.¹⁷⁸ Det er vigtigt ikke at forglemme det usynlige arbejde, da denne proces har stor påvirkning på resultaterne.

Vi har opsat en metodik for projektet, hvor det digitale arbejde inddrages lige så vel som de visuelle elementer, det har medført. Det er derfor særdeles vigtigt, at vi i denne refleksion også har et fokus på de problematikker, som kan opstå.

I takt med databehandling af en materialesamling i denne størrelse, er objektivitet svært at opnå. Vi har blandt andet opsat *nation_tags* på baggrund af en computers forståelse, da vi har kørt koden gennem Google Maps software. Denne software tager ikke forbehold for andet end nuværende grænser og tilhørsforhold, og kan derfor ikke forholde sig til fortidige nationale opdelinger.

Vi har været opmærksomme på elementer som dette i hele vores proces og opsat en præmis for os selv. For på bedst mulige vis at kunne opnå et eksplorativt kriterium, har vi i store omfang forsøgt at se hvor langt computeren har kunne føre os. Når tidspunktet kom og vi havde nået grænsen for de eksplorative elementer, som computeren kunne bidrage med, har vi som historikere trådt ind og fortsat behandlingen. Det gælder blandt andet ved Google Maps *nation_tags* af Lund og Slesvig, som blev noteret som *Sverige* og *Tyskland*. Her gik vi ind og tilpassede resultaterne, så vores endelige data stemte overens med den historiske kontekst.

Dette kan både anses som en fejkilde og en afvigelse fra den eksplorative præmis for opgaven. Det ville omvendt have været en massiv fejkilde i undersøgelsen, hvis vi ikke havde rettet disse *nation_tags*.

¹⁷⁸ Jf. Bagsiden af medaljen

Vi har flere gange måtte omregistrere i forhold til den geografiske metadata, da *nation_tags* i flere tilfælde ikke var i stand til at identificere ældre stednavne. Denne situation opstod flere gange, da der i materialet omtales byer, som ikke længere eksisterer eller hvor bynavnet har ændret stavemåde. Denne proces tilpassede vi på baggrund af en faglig præmis, da vi sideløbende havde digitaliseret brevsamlingens stedregister. I stedregistret er der redaktionelle valg fra udgiveren af kildesamlingen, til de geografiske referencer. Det sker eksempelvis ved Wawe i Estland, hvor der i stedregistret er tilføjet (*nu Wayga, Estland*). Her valgte vi at finde koordinaterne manuelt og tilføje dem til de steder, hvor Wawe nævnes i materialet. Denne pointe er væsentlig. Pointen omkring de manuelle tilpasninger er helt essentielle for de endelig fakta, som kommer af dataen. I processen står det tydeligt for os, at det fortsat er vigtigt at læne sig op ad sin historiefaglighed. Vi kan ikke lade computeren diktere retningen og alene bygge en analyse på baggrund af den. Forståelse og tolkning bliver helt essentiel for den viden, som i sidste ende uddrages fra datagrundlaget.

Når koden taler sandt

I forlængelse af vores arbejde med capta og hele databehandlingen, har vi haft fokus på et særligt element - *transparens*. Transparens og reproducerbarhed er to absolutter i den forskning, som vi anvendte til vores metodeudviklende projekt *Fra Pergament til Python* samt i Gustavo Rivas behandling af netværksanalyser.

Vi har i høj grad bestræbt os på at efterleve disse kriterier. Vi anser det for nødvendigt at digitalhistoriske undersøgelser kan reproducere, for at resultater kan kvalitetssikres. Fodnoten har ikke samme værdi i det digitalhistoriske felt, som i den traditionelle kildenære historieforskning, da en henvisning til originalkilde eller en kode ikke kan stå alene. Vi finder det helt essentielt at tage læseren med fra start til slut i processen.¹⁷⁹

Vi har indsat vores koder, for at vi kan stilles til ansvar i en metodisk kontekst. Vores inddragelse af koder i projektet, skal i en grad kunne fungere som fodnoter til vores eget arbejde. På denne måde sikres der en transparens, hvor de analytiske nedslag, som vi laver på baggrund af en omfattende databehandling og kodeopsætning, kan valideres.¹⁸⁰ Med disse

¹⁷⁹ Jf. Kapitel 2: På vej derhed

¹⁸⁰ For at møde disse kriterier begyndte det allerede i projektet *Fra Pergament til Python*, hvor vi dokumenterede hvert skridt i konverteringen af materiale fra online-tilgængelige pdf'er til rækker i et regneark.

Torgerson læner sig op af Parland-von Essen, når han skriver (...) *i propose that we finally explain how it is that we turn sources into facts*.¹⁸¹ Ved at være gennemsigtige ønsker vi at fremlægge, hvordan vi har skabt datagrundlaget for vores resultater. Vi skal som digitalhistorikere, turde stå på mål for de beslutninger, som vi har truffet gennem hele processen. Vi mener, at det faglige fundament forsvinder, hvis vi ikke tør fremstille vores arbejde med fuld gennemsigthed.

Der er en specifik årsag til, at vi advokerer for gennemsigtighed. Det er i de fleste tilfælde nødvendigt at manipulere med visualiseringer, for at få et læsbart resultat. Vi har ligeledes foretaget en række valg i fremstillingen af vores visualiseringer. Disse parametre er altid indlejret i koden.

[illegible]

111

Det er her umuligt at aflæse de forskellige labels, da de overlapper i en voldsom grad. Et output som dette får dermed ingen analytisk værdi, da netværket ikke kan aflæses.

I en visuel formidling som dette, er det derfor vigtigt, at vi ikke glemmer at være historikere, blot fordi vi arbejder digitalt. Her bliver det nødvendigt med en grundlæggende indsigt i materialet, så vi kan træffe beslutninger, der ikke giver skæve outputs.

Tekstens modstand

Vi har brugt computerens evner, som omdrejningspunkt i vores projekt. Dette har vi gjort, da vi så, at en digitalhistorisk bearbejdning af brevene ville skabe muligheder for nye vinkler og perspektiver på et ældre gennemarbejdet materiale.

En computerstyret tekstbehandling volder dog store problemer, når der arbejdes med ældre tekstmateriale. Der findes flere programmer, som er yderst effektive og trænet til at finde filologiske mønster i eksempelvis annoncer eller opslag på sociale medier. De software- og datapakker, som kan benyttes til større tekstkorpus, er dog ikke avancerede nok til at skabe tilfredsstillende resultater på materiale af ældre dato.¹⁸²

Vores topicmodeling-undersøgelse fejlede ikke, men den bidrog heller ikke med skarpe resultater. Dette skyldes at softwaren endnu ikke er avanceret nok. Samme problematik kan overføres til vores fejlslagne NER-analyse. Problemet kan opdeles i to kategorier. Først og fremmest er der ikke udført mange digitale undersøgelser på kildemateriale fra middelalderen, og der er derfor ikke trænet machine learning på sådanne materiale. Jo flere undersøgelser som skabes, jo bedre bliver de digitale analyser, da softwaren lærer materialet at kende og dermed skaber færre fejlkilder. Det andet problem består i måden man skriver på, da computeren ikke er i stand til at forstå de intellektuelle, symbolske og indirekte referencer i brevmaterialet. Når softwaren engang kan holde brevteksten i Danmarks Riges Breve op imod indholdet i Biblen, kan den finde de symbolske referencer til eksempelvis *de stumme hunde, der ikke kan gå*. Tekstgenkendelsen er simpelthen ikke god nok endnu. I takt med at AI og machine learning udvikler sig, kan det på sigt være muligt at skabe et teknisk grundlag, der kan spore og anerkende sammenhængen i højere grad end den gør nu.

¹⁸² Med materiale af ældre dato mener vi naturligvis breve fra middelalderen. I dette tilfælde gælder ældre materiale ikke materiale fra 1700- og 1800-tallet, som der findes gode transkribusmodeller til at bearbejde.

For nu må vi nøjes med at skabe undersøgelsen på den software, som er tilgængelig, og dertil benytte vores historiefaglighed til at validere og tolke resultaterne.

Af denne grund kan den digitale metode, ifølge os, ikke stå alene. Vi er fortalere for en tilgang, der er analog → digital → analog.

Ved at have kendskab til kildematerialet, er det muligt at skabe en meningsfuld digitalisering og datasæt, som kan bearbejdes med digitale metoder. Resultaterne skal hertil forstås, tolkes, diskuteres og bearbejdes af historikken igen.

Diskussion - En relation mellem Danmark og Frankrig?

Da vi har benyttet Ingeborg og Philips ægteskab som vores katalysator, havde vi en forventning om, at vi ville kunne spore en relation mellem Danmark og Frankrig i perioden før det mislykkede ægteskab. Vi forventede at brylluppet var blevet sat i stand på baggrund af en eller anden form for relation mellem de to lande, og vi ønskede at undersøge brevene, for at se på, hvordan konflikten omkring ægteskabet påvirkede relationen. Efter at vi har skabt vores digitalhistoriske analyse, sidder vi tilbage med en undren. Er der i virkeligheden en direkte dansk-fransk relation i den undersøgelse, som vi har opsat?

I vores samling af 2002 breve fandt vi 121 breve, som kunne placeres i en computer dikteret relation mellem Danmark og Frankrig. Det undrede os yderligere at kommunikationen mellem de to lande kun fylder 6% af materialet, da sagen om Ingeborg og Filips ægteskab var en international krise under Knud d. 6 regeringstid, hvor Danmark var en væsentlig magtfaktor i Europa.¹⁸³

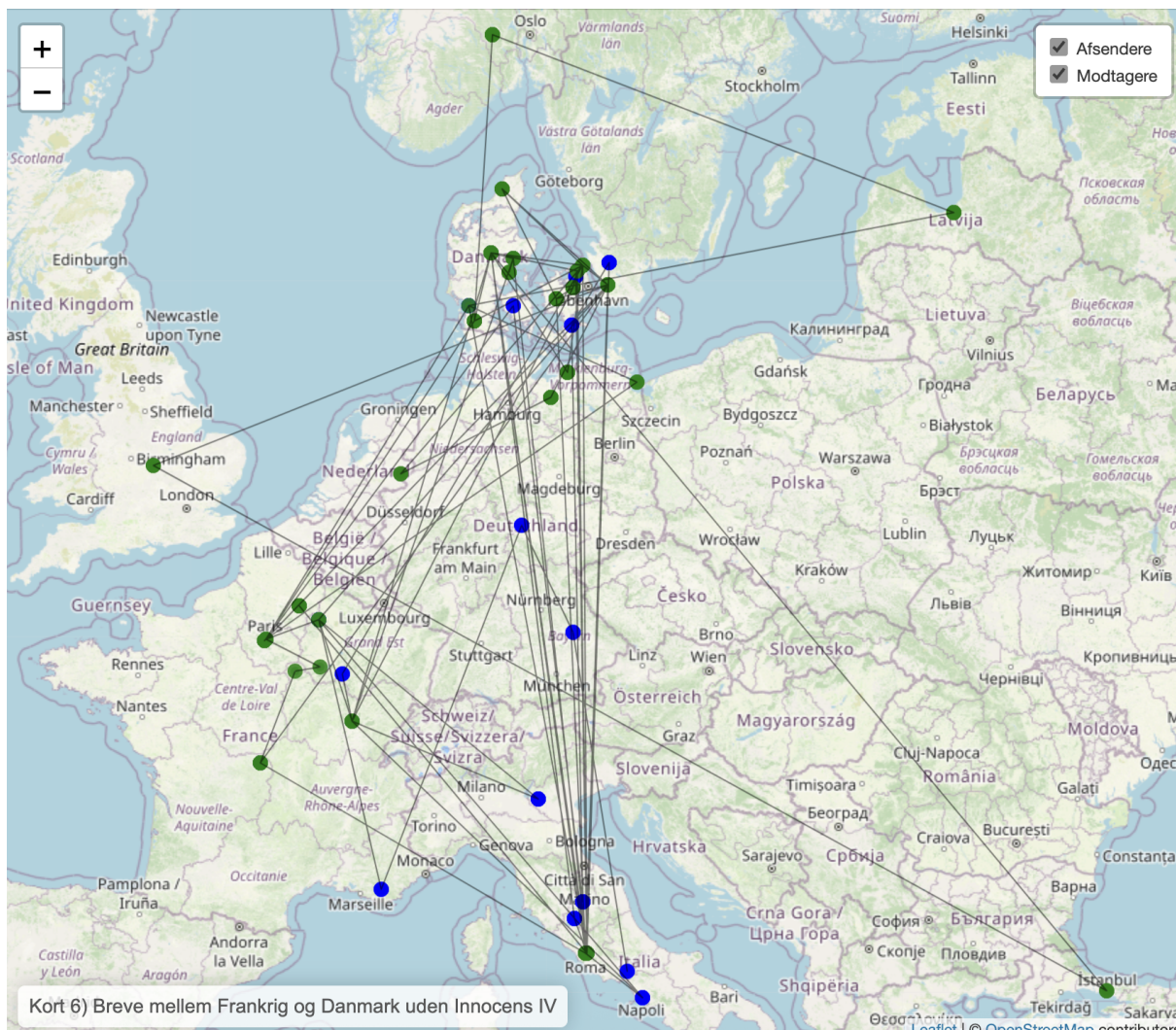
I vores analyse af de 121 breve så vi på kommunikationsmønstre, aktører og følelser i brevtteksterne. Vi visualiserede ligeledes kommunikationsstrømme mellem Danmark og Frankrig i perioden fra 1131 til 1265. Det blev klart for os, at Danmark var modtager af markant flere breve end de sendte. Kommunikationen er dermed dikteret fra syd mod nord. Da vi så på afsendere og modtagere, blev Lyons rolle tydelig, da hvert tredje brev i samlingen er markeret med en afsender i Lyon. I Sankey-diagrammet blev det yderligere tydeligt, at Lyon

¹⁸³ Danmarks Leksikon. "Knud d. 6." Konsuleret 23. maj 2025.
https://lex.dk/Knud_6.?gad_source=1&gad_campaignid=17600838629&gbraid=0AAAAABeGnWrL5mNFll4KE6maOCKqa2PtA&gclid=CjwKCAjwruXBBhArEiwACBRtHTEhE6FY8XwNpvhTbK288Cr99bZJfnMrhvPCgSe_6jy0aNhkOeK3hoCGPsQAvD_BwE

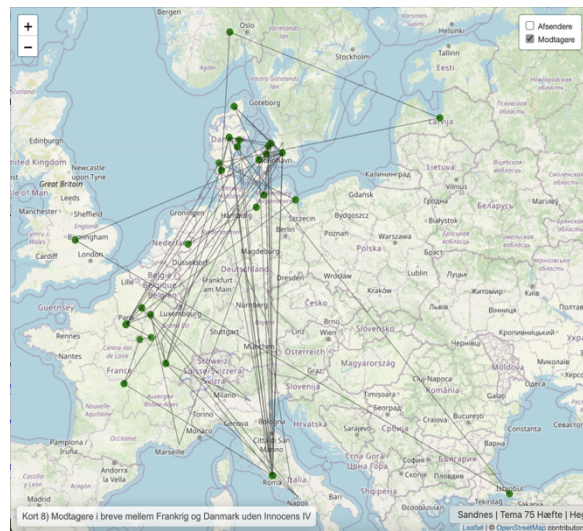
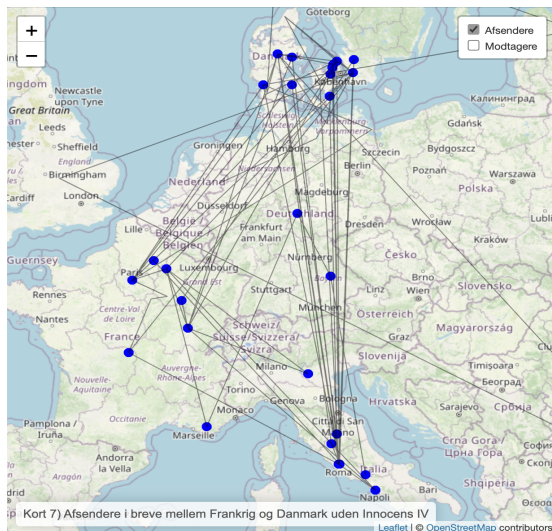
var et knudepunkt i kommunikationsnetværket. Diagrammet visualiserer graden af kommunikation, som kommer fra Lyon, og samtidig de mange steder, hvor den ender. Her optræder Lund, som den hyppigste modtager af breve fra Lyon. Byerne, som Sankey-diagrammet fremhæver, kan defineres som vigtige i en religiøs kontekst, da pave Innocens IV i perioden er placeret i Lyon og det danske ærkebispesæde er i Lund.

Da vi skabte et kort over kommunikationsstrømmene, så vi det, som vi definerer som et kommunikativt hierarki. Brevene i den dansk-franske reaktion bliver sendt fra syd, og modtages i særligt Æbelholt, Århus, Esrom og Lund. Disse byer står for videre distribution til modtagere i Danmark. Da modtagerne i Æbelholt, Århus, Esrom og Lund alle er tilknyttet gejstlige institutioner, styrker dette argumentet om, at den dansk-franske relation er bundet op omkring et religiøst netværk. Det er ikke diplomatiske forbindelser mellem konger som dominerer netværket, men derimod paver, præster, abbeder og bisper. I forlængelse af dette, er det relevant at omfavne sagen omkring dronning Ingeborgs ægteskab med Filip 2. Augustus. De tendenser, der træder frem fra brevmaterialet her, viser ligeledes det samme som resten af analysen. Det er ikke kommunikation mellem det danske og franske kongehus, men i stedet formidlet og styret i høj grad af pave Innocens III. Det viser os, at den dansk-franske relation ofte eksisterer indirekte og er styret af kirkelige aktører.

Det dansk-franske netværk fremstår generelt som et netværk, der består af få forbindelser mellem få individer. Pave Innocens IV er fremtrædende i netværket, som repræsentant for den katolske kirke. Han har her en afgørende rolle, da han optræder som et knudepunkt i det geografiske netværk, og da han ydermere er afsender på stort set hvert tredje brev i samlingen. For at få et overblik over brevene uden at tage stilling til pave Innocens IV, har vi skabt et kort, hvor han er filtreret ud. Her er afsendere markeret med blå og modtagere med grøn. På *Kort 6) Breve mellem Danmark og Frankrig uden Innocens IV* nedenfor ser relationen pludselig anderledes ud, da netværkets knudepunkt nu har flyttet sig til Rom. Lund og Æbelholt er stadig centrale områder i Danmark.



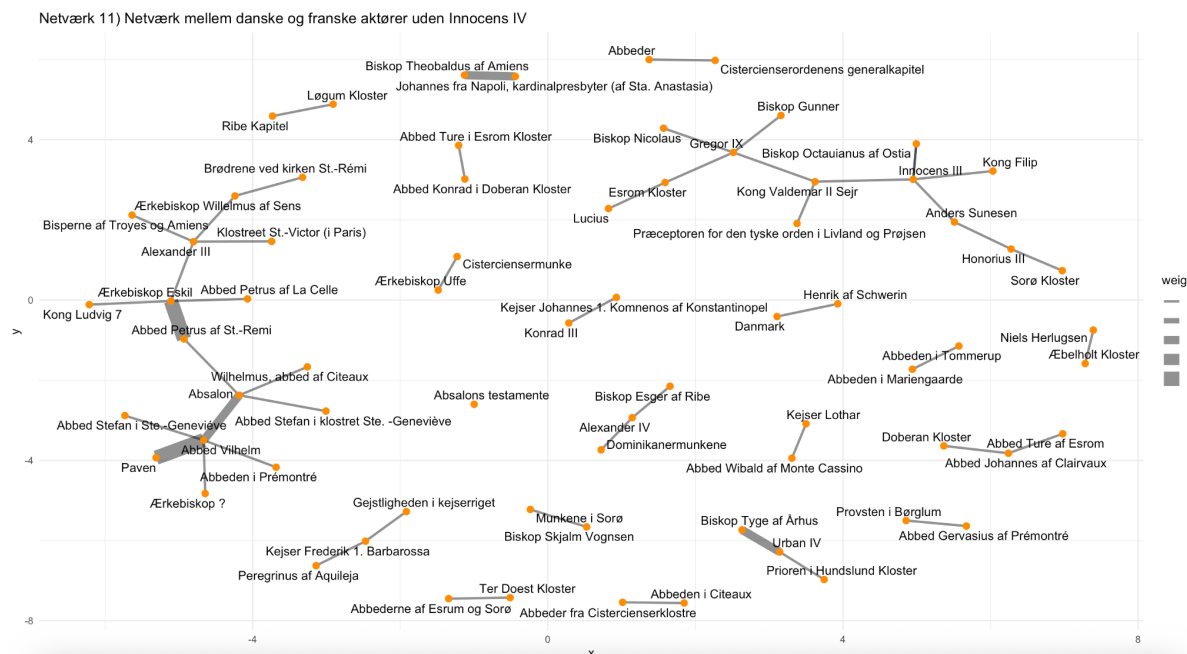
Når vi ser, at det primære knudepunkt flytter position fra Lyon til Rom, tolker vi dette som et udtryk for pavens rolle og betydning i relationen. Ved at tage pave Innocens IV ud af netværket, bliver det primære knudepunkt skiftet til en ny pave. På *Kort 6) Breve mellem Danmark og Frankrig uden Innocens IV* ovenfor, er Rom dog markeret med grøn og dermed som modtager. For at fremstille afsender og modtager, uden at de forskellige farver overlapper, har vi taget screenshots af de forskellige lag. Kortet er interaktivt, og man kan dermed filtrere i lag med afsender og modtager. De to separate lag kan ses nedenfor i *Kort 7) Afsendere i breve mellem Danmark og Frankrig uden Innocens IV* og *Kort 8) Modtagere i breve mellem Danmark og Frankrig uden Innocens IV*. Når vi ser på lagene separat, bliver det tydeligt, at Rom både er en central modtager og afsender af breve i relationen mellem Danmark og Frankrig.



Relationen mellem de to lande fremstår ikke som præget af diplomatiske bånd. Det er derimod abbeder, klostre og biskopper som kommunikerer, og deres forbindelser fremstår samtidigt svage, da de generelt ikke sender mange breve til hinanden.

Ud over den geografiske analyse af relationen mellem de to lande, har vi arbejdet med aktørerne i brevene. Som tidligere beskrevet, har vi tagget afsender og modtager manuelt, for at kunne foretage en selektiv vurdering af, hvem som sender og modtager brevene. Vi finder denne tidskrævende proces nødvendig, da NER-analysen ikke var i stand til at skabe tilfredsstillende resultater af aktørerne i brevene.

Når vi ser på aktørerne i *Netværk 3) Netværk mellem danske og franske aktører*, er Innocens IV igen central. Den stærkeste relation mellem aktørerne i *netværket 3* er ligeledes mellem pave Innocens IV og ærkebiskop Uffe. Derfor har vi genskabt *Netværk 3* og filtreret Innocens IV fra. Dette resulterer i *Netværk 11) Netværk mellem danske og franske aktører uden Innocens IV*. Det er her vigtigt at bemærke, at linjernes tykkelse er sat efter en *weight*, som på det nye netværk går fra 1 til maks 3. På det netværk (3) hvor Innocens IV er inkluderet, går *weight* fra 1 til 10. Selvom streger er lige tykke, repræsenterer de ikke samme mængde information og dermed ikke samme grad af forbindelse.



Når vi fjerner Innocens IV, går netværket fra at være centraliseret til at blive tydeligt decentraliseret. Det består nu af flere relationer mellem enkelte aktører, og kong Valdemar II Sejr og Abbed Vilhelm bliver de mest centrale aktører i *netværk 11*. Abbed Vilhelm har sendt 10 breve og modtaget 2, mens Valdemar II Sejr har sendt 2 og modtaget 1. De centrale aktører i netværket er nu bundet op på to individer, som ikke kommunikerer tilnærmelsesvis lige så meget som Innocens IV gjorde. *Netværk 11*'s tæthed er markant lavere, og det fremstår skrøbeligt uden Innocens IV, som det bindende led i midten. Det fremstår dog mere personligt, da netværksstrukturen ikke er bundet op omkring den pavelige administration, men i stedet på individer fra samtiden. Vi tolker individer til at spille en væsentlig rolle i netværket, da vi vurderer, at klostrene kommunikerer på et individuelt niveau, da graden af kommunikation er så lav.

Ud fra *Netværk 11) Netværk mellem danske og franske aktører uden Innocens IV* fremstår relationen mellem Danmark og Frankrig skrøbelig, da netværket mister store dele af dets forbindelser, når en enkelt aktør udelukkes. Vi tolker netværket så skrøbeligt, at der kun skal fjernes få relationer, for at netværket falder fra hinanden. Da visualiseringen af aktørnetværket mellem Danmark og Frankrig ydermere dækker 134 år, ser vi det som værende ekstremt skrøbeligt.

Hertil opstår der ydermere den ulempe, at vi ikke kan definere hvilken pave, som Abbed Vilhelm skriver til, da der ikke er tidsmæssige markører i hans brevsamling. "Paven" kan derfor dække over flere forskellige paver, og dette kan forårsage at den skrøbelige relation mellem Danmark og Frankrig, bliver forværret.

Vicarius Christi

Det er, ud fra ovenstående, ikke til at komme udenom, at pavedømmet er dominerende i brevmaterialet. Ud af de 2002 breve i vores database, var 745 af dem pavebreve, hvilket giver en klar indikation af kirkens betydning. Ud fra vores analyse af brevene mellem Danmark og Frankrig, er pavedømmet ligeledes et afgørende bindeled. Paven bliver den lim, som holder netværket sammen, da gejstlige institutioner fra begge lande kommunikerer gennem paven. Vores analyse viser derfor ikke direkte diplomatiske kommunikationsmønstre mellem de to stormagter i perioden før, under eller efter ægteskabskonflikten. De to lande bliver i stedet to aktører i et netværk, som er bygget op omkring kirken.

Vi vil argumentere for, at relationen kan og bør forstås som et spindelvæv og ikke som en lige linje mellem to aktører. Den generelle opfattelse af en relation er noget, som går fra den ene ende af linjen til den anden i form af eksempelvis et brev fra en konge til en anden konge. Vores materiale viser tværtimod relationen som et spindelvæv bundet af flere elementer. *Diagram 4) Sankey-diagram* er en god forlængelse af denne metaforiske fremstilling. De 121 breve er en blanding af direkte kommunikation og relationer som går igennem steder som Lateranet, Rom, Lyon og Æbelholt.¹⁸⁴

Danmark hænger i den ene del af det store europæiske spindelvæv og Frankrig i den anden. Vibrationer i Frankrigs del af spindelvævet, bevæger sig gennem de forskellige linjer, og skaber i sidste ende en påvirkning i Danmarks del af spindelvævet. I spindelvævet placerer vi forskellige aktører, som består af verdslige og gejstlige individer og institutioner. Paven, som repræsentant for den katolske kirke, ser vi som en stor edderkop, der med sine otte ben bevæger sig rundt i spindelvævet og trækker i trådene.

¹⁸⁴ se sankey-diagram

Konklusion

Med afsæt i problemformuleringen - *Hvordan kan digitale metoder anvendes til at analysere relationen mellem Danmark og Frankrig i højmiddelalderen, og hvad kan en eksplorativ digital analyse af Danmarks Riges Breve bidrage med?* - har vi undersøgt 2002 breve fra perioden 1131 til 1265. Her har vi haft fokus på geografi og aktørnetværk, for at belyse relationens karakter gennem nye perspektiver.

Computeren dikterede 121 breve som værende en direkte relation mellem Danmark og Frankrig. Disse breve indeholdt ikke store følelsesmæssige udsving og fremstår neutrale, som den resterende del af materialet. Ud fra et følelsesmæssigt perspektiv, ser vi derfor ikke relationen stå i kontrast til det samlede udtryk og den gennemgående tendens.

På baggrund af vores analyse vurderer vi, at der ikke ses tydelige tegn på en egentlig relation mellem Danmark og Frankrig i perioden fra 1131 til 1265. Den franske tilstedeværelse i materialet er tilnærmelsesvis symbolsk, og fremstår som omdrejningspunkt i en strukturel kontekst. Vi ser ikke Frankrig i en direkte kommunikativ diplomatisk relation med Danmark.

Vi benytter Ingeborg som vores indgangsvinkel, men vi ser ingen tegn på en relation mellem danske og franske kongemagter inden Ingeborg og Filips ægteskab, og efterfølgende ser vi ligeledes ikke varige bånd. Det forliste ægteskab, og dermed alliance, gør ikke noget ved den videre relation, da relationen mellem Danmark og Frankrig i perioden ikke er bundet op på diplomatiske bånd. Vi kan ud fra vores digitalhistorisk analyse konkludere, at det er den fælles tilknytning til den katolske kirke, som forbinder de to lande.

En årsag til at vi vælger at vurdere materialet på denne måde, sker i høj grad grundet situationen omkring Lyon i 1240'erne. Det er en atypisk situation, hvor paven flytter sin magtudøvelse fra Rom til Lyon. Det skaber derfor også en betydelig fejlmargen i vores materiale. Grundet vores dataopsætning, kan det hurtigt give et billede af en stærk og tæt relation mellem Danmark og Frankrig, alene på baggrund af geografiske markører. Havde Innocens IV ikke haft problemer med den tysk-romerske kejser Frederik 2., ville han antageligt have udøvet sin magt fra Rom i stedet. Det ville have resulteret i et relationelt netværk, som havde set væsentligt anderledes ud.

At Frankrig nævnes i Danmarks Riges Breve, vurderer vi derfor ikke som et udtryk for en relation mellem de to lande. Vi ser det nærmere som et udtryk for den danske placering i en europæisk og religiøs struktur, hvor Frankrig er centrale. Det gælder både i en relation til

klosternetværk, der udspringer fra Frankrig, men i særdeles høj grad den atypiske situation med pave Innocens IV's position i Lyon. Den geografiske og aktørmæssige analyse bliver derfor et udtryk for den religiøse magt i Europa i samtiden.

Vi har opsat vores undersøgelse som et metodetestende projekt, i forlængelse af vores arbejde *Fra Pergament til Python*. Da vi tidligere havde undersøgt de metodiske udfordringer, der opstår ved digitalt arbejde i en historiefaglig kontekst, satte vi os for at give et solidt funderet bud på, hvordan en digitalhistoriske fremstilling bør opsættes, for at kunne omfavne de generelle metodiske principper som transparens og reproducerbarhed.

Denne proces og konceptuelle opsætning af vores speciale udgør, hvad vi definerer, som et nødvendigt fundament, for at kunne omfavne den digitale teoretiske værktøjskasse, der fortsat er under udvikling. Vi kan ydermere slå fast, at digitale metoder ikke kan stå alene i en historisk undersøgelse. Det er en nødvendighed at have kendskab til traditionelle metoder og vigtigst af alt indgående kendskab til kildemateriale. Det er ikke muligt at opstille en grundig og fagligt funderet undersøgelse på en materialesamling som vores, uden at kunne identificere kerneelementerne i den anvendte kildesamling.

Bibliografi

Litteratur

- Adriansen, Inge. Nationale symboler i Det Danske Rige 1830-2000 - Bind II: Fra undersåtter til nation. Museum Tusculanums Forlag Københavns Universitet, 2003.
- Baldwin, John W, "Philip Augustus and the Norman Church", French Historical Studies 6, no. 1, Duke University Press (1969): s.1-30. <https://doi.org/10.2307/286242>
- Böss, Michael. Nationalitet. Aarhus Universitetsforlag, 2019.
- Brennan, Claire (2018), "Digital humanities, digital methods, digital history, and digital outputs: history writing and the digital revolution", History Compass, 16 (10) . <https://doi.org/10.1111/hic3.12492>
- Brundage, A. James, Law, Sex, and Christian Society in Medieval Europe, The University of Chicago Press, 1987.
- Burrows, Toby. (2023). "Reproducibility, verifiability, and computational historical research." I [International Journal of Digital Humanities \(2023\) vol 5 :283–298. Springer. https://doi.org/10.1007/s42803-023-00068-9](#)
- Claire, Brennan "Digital humanities, digital methods, digital history, and digital outputs: history writing and the digital revolution", History Compass, 16 (10) (2018) . <https://doi.org/10.1111/hic3.12492>
- Elo, Kimmo. (2020). "Big Data, Bad Metadata: A Methodological Note on the Importance of Good Metadata in the Age of Digital History." i Mats Fridlund m.fl. (red.), Digital histo-ries: Emergent Approaches within the New Digital History. (pp. 103-111). Helsinki: Helsinki University Press. <https://doi.org/10.33134>

- Esmark, Kim. "Fromhed og ære - Hvideklanens gaver til Sorø Kloster i 11- og 1200-tallet". Årbog for Historisk Samfund for Holbæk Amt (2003): 108-126.
- Esmark, Kim & McGuire, Patrick Brian, Europa 1000-1300. Roskilde Universitetsforlag, 1. udgave, 1999.
- Fernández, R. Gustavo, "Network Analysis of Medieval Manuscript Transmission. Basic Principles and Methods", i Journal of Historical Network Research 3 (2019). S. 30-49. DOI: [10.25517/jhnr.v3i1.61](https://doi.org/10.25517/jhnr.v3i1.61)
- Grün, Bettina & Hornik, Kurt, "Topicmodels: An R package for Fitting Topic Models", <https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf> s.2-3
- Hoppenbrouwers, Peter & Blockmans, Wim, Introduction to Medieval Europe 300-1500, Routledge, third edition, 2014.
- Johan Jarlbrink, "All the Work that Makes it Work: Digital Methods and Manual Labour", i Mats Fridlund m.fl. (red.), Digital Histories: Emergent Approaches within the New Digital History (Helsinki: Helsinki University Press, 2020), s. 113–126. <https://doi.org/10.33134/HUP-5>
- Kjær, Lars, Valdemar Sejr. Imperiebyggeren, Gads Forlag 1, udgave, 1. oplag, 2024.
- Kjær, Lars. Valdemar Den Store - Borgerkrigens barn. Gads Forlag, 1. udgave, 1. oplag, 2022.
- Lemercier, Claire og Claire Zalc. Quantitative Methods in the Humanities - an introduction. University of Virginia Press, 2019.
- "Lost the Plot," The Journal of the Social History Society 10 (2013): s. 9, downloaded 10. november 2024,

- Parland-von Essen, J. (2020). "Building historical knowledge byte by byte: Infrastructures and data management in modern scholarship." in M. Fridlund, M. Oiva, & P. Paju (Eds.), *Digital histories: Emergent approaches within the new digital history*. (s. 89-102). Helsinki: Helsinki University Press.
<http://doi.org/10.33134/HUP-5-5>
- Skyum-Nielsen, Niels, *Kirkekampen i Danmark 1241-1290*, Munksgaard, 1963.
- Tim Hitchcock, "Confronting the Digital, Cultural and Social History - Or How Academic History Writing
- Torgerson, W. Jesse. "Historical practice in the era of digital history", i *History and Theory* 61, no. 4(December 2022). S. 37-63. [DOI: 10.1111/hith.12276](https://doi.org/10.1111/hith.12276)

Internetressourcer

- Bibelselskabet. "Esajas' Bog Kapitel 56." Konsulteret 25. maj 2025.
<https://www.bibelselskabet.dk/brugbibelen/bibelenonline/es/56>
- *From historical source to network data*, Claire Lemercier, YouTube-video, udgivet af Historical Network Research Community, 8. juli 2021.
<https://www.youtube.com/watch?v=ZLgCSKNrTBY>
- Danmarks Leksikon. "Ingeborg: Prinsesse af Danmark og dronning af Frankrig, 1175-1237." Konsulteret 25. maj 2025.
https://danmarkshistorien.lex.dk/Ingeborg_Prinsesse_af_Danmark_og_dronning_af_Frankrig,_1175-1237
- Danmarks Leksikon. "Knud d. 6." Konsulteret 23. maj 2025.
https://lex.dk/Knud_6.?gad_source=1&gad_campaignid=17600838629&gbraid=0AAABeGnWrL5mNFil4KE6maOCKqa2PtA&gclid=CjwKCAjwruXBBhArEiwACB

[RtHTEhE6FYY8XwNpvhTbK288Cr99bZJfnMrhvPCgSe_6jy0aNhkOeK3hoCGPsQ
AvD_BwE](https://www.lex.dk/Knud_den_Store?gad_source=1&gad_campaignid=20505639983&gclid=Cj0KCQjwlrBBhDnARIsAHEQgOTnQknRPZtt02xZZ-eO3D5NMfRKXf0ZzYnk-XV3XbsjvgQel5jLwwIaAjFkEALw_wcB)

- Danmarks Leksikon. "Knud den Store." Konsulteret d. 1. juni 2025.
https://lex.dk/Knud_den_Store?gad_source=1&gad_campaignid=20505639983&gclid=Cj0KCQjwlrBBhDnARIsAHEQgOTnQknRPZtt02xZZ-eO3D5NMfRKXf0ZzYnk-XV3XbsjvgQel5jLwwIaAjFkEALw_wcB
- Danmarks Nationalleksikon. "Eleonora af Aquitaine." Konsulteret d. 21. maj 2025.
https://lex.dk/Eleonora_af_Aquitaine
- Det Danske Sprog- og Litteraturselskab. "Danmarks Riges Breve." Konsulteret 7. april 2025. <https://dsl.dk/publication?id=122>
- Europarl.Europa. "Big data: Definition, fordele og udfordringer (infografik)." Konsulteret d. 9. april 2025.
<https://www.europarl.europa.eu/topics/da/article/20210211STO97614/big-data-definition-fordele-og-udfordringer-infografik>
- Github. "Danish Sentiment Lexicon." Konsulteret d. 1. juni 2025.
<https://github.com/dsldk/danish-sentiment-lexicon>
- Grün, Bettina & Hornik, Kurt, "Topicmodels: An R package for Fitting Topic Models", <https://cran.r-project.org/web/packages/topicmodels/vignettes/topicmodels.pdf> s.2-3
- Thomas Heebøll Holm og Lars Kjær, Mægtig Middelalder, Ingeborg, Dronning af Orléans I: Brylluppet i Amiens, [20:51–21:28], udgivet af Spotify, 13. september 2024, <https://open.spotify.com/episode/4UFHyVojXpX1ZKrTeRqC?si=995f472822784d44>

- Hvitfeldt Emil og Julia Silge. “Supervised Machine Learning for Text Analysis in R.” Konsulteret d. 12. maj 2025. <https://smltar.com>
- Klart som blæk. “Fra avissider til segmenterede artikler.” Konsulteret d. 1. juni 2025. https://hislab.quarto.pub/step_4.html
- Larsen, Morten & Peter Sestoft “Noter om R”,
<https://cran.rproject.org/doc/contrib/Larsen+Sestof-noter-om-R.pdf>
- Papal Encyclicals Online. “First Council of Lyons – 1245 A.D.” Konsulteret d. 1. juni 2025. <https://www.papalencyclicals.net/councils/ecum13.htm>
- Programming Historian. “Visualizing Urban and Demographic Data in R with ggplot2.” Konsulteret d. 1. juni 2025. <https://programminghistorian.org/en/lessons/urban-demographic-data-r-ggplot2#our-data-sister-cities-in-the-european-union>
- Programmering for historikere. “Kapitel 5: Visualisering 2: Kort.” Konsulteret d. 16. maj 2025. <https://johanheinsen.quarto.pub/programmering-for-historikere/kort.html>
- Text Mining with R: A Tidy Approach. “Kapitel 2: Sentiment analysis with tidy data.” Konsulteret d. 13. maj 2025. <https://www.tidytextmining.com/sentiment>

Kilder

- Anonymous, Powell, J. M. (2004). *The Deeds of Pope Innocent III*. Catholic University of America Press. <https://doi.org/10.2307/j.ctt284xfp>
- *Adam af Bremen om Menigheden i Norden*, oversat af P.W. Christensen, *Karl Schønbergs Forlag*, 1862

- Larry F. Field. *The Deeds of Philip Augustus: An English Translation of Rigord's "Gesta Philippi Augusti."* Edited by M. Cecilia Gaposchkin and Sean L. Field. Cornell University Press (2022).
- Scudder, Dutton Vida. Cathrine of Siena (1905) - TRANSLATED & EDITED WITH INTRODUCTION BY VIDA DUTTON SCUDDER London, New York: J.M. Dent and E.P. Dutton, 1905 <http://www.domcentral.org/trad/cathletters.htm>
- *Saxos Danmarkshistorie*, oversat af Peter Zeeberg, Det Danske Sprog- og Litteraturselskab & Gads Forlag, 2015. [//litteraturbanken.se/f%C3%B6rfattare/Saxo/titlar/SaxosDanmarksHistorie/sida/6/faksimil](http://litteraturbanken.se/f%C3%B6rfattare/Saxo/titlar/SaxosDanmarksHistorie/sida/6/faksimil)

Danmarks Riges Breve

- Christensen, C.A., og Herluf Nielsen. *Danmarks Riges Breve - 1. række 1. bind: 789-1052*. Det danske sprog- og litteraturselskab, 1975. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3001.pdf> Sidst tilgået d. 27. maj 2025.
- Christensen, C.A., Gustav Hermansen og Herluf Nielsen. *Danmarks Riges Breve - 1. række 2. bind: 1053-1169*. Det danske sprog- og litteraturselskab, 1975. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3002.pdf> Sidst tilgået d. 27. maj 2025.
- Christensen, C.A., og Herluf Nielsen. *Danmarks Riges Breve - 1. række 3. bind: 1170-1199*. Det danske sprog- og litteraturselskab, 1977. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3003.pdf> Sidst tilgået d. 27. maj 2025.
- Niels Skyum-Nielsen. *Danmarks Riges Breve - 1. række 4. bind: 1200-1210*. Det danske sprog- og litteraturselskab, 1958. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3004.pdf> Sidst tilgået d. 27. maj 2025.

- Niels Skyum-Nielsen. *Danmarks Riges Breve - 1. række 5. bind: 1211-1223*. Det danske sprog- og litteraturselskab, 1957. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3005.pdf> Sidst tilgået d. 27. maj 2025.
- Niels Skyum-Nielsen. *Danmarks Riges Breve - 1. række 6. bind: 1224-1237*. Det danske sprog- og litteraturselskab, 1979. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3006.pdf> Sidst tilgået d. 27. maj 2025.
- Skyum-Nielsen, Niels, og Herluf Nielsen. *Danmarks Riges Breve - 1. række 7. bind: 1238-1249* Det danske sprog- og litteraturselskab, 1990. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3007.pdf> Sidst tilgået d. 27. maj 2025.
- Gustav Hermansen. *Danmarks Riges Breve - 2. række 1. bind: 1250-1265*. Det danske sprog- og litteraturselskab, 1938. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3008.pdf> Sidst tilgået d. 27. maj 2025.
- Afzelius, Adam, og Gustav Hermansen. *Danmarks Riges Breve - 2. række 2. bind: 1266-1280*. Det danske sprog- og litteraturselskab, 1941. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3009.pdf> Sidst tilgået d. 27. maj 2025.
- Adam Afzelius. *Danmarks Riges Breve - 2. række 3. bind: 1281-1290*. Det danske sprog- og litteraturselskab, 1939. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3010.pdf> Sidst tilgået d. 27. maj 2025.
- C. A. Christensen. *Danmarks Riges Breve - 2. række 4. bind: 1291-1298*. Det danske sprog- og litteraturselskab, 1942. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3011.pdf> Sidst tilgået d. 27. maj 2025.
- C. A. Christensen. *Danmarks Riges Breve - 2. række 5. bind: 1299-1305*. Det danske sprog- og litteraturselskab, 1943. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3012.pdf> Sidst tilgået d. 27. maj 2025.

- Kåre Olsen. *Danmarks Riges Breve - 2. række 6. bind: 1306-1312*. Det danske sprog- og litteraturselskab, 1948. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3013.pdf> Sidst tilgået d. 27. maj 2025.
- Kåre Olsen. *Danmarks Riges Breve - 2. række 7. bind: 1313-1317*. Det danske sprog- og litteraturselskab, 1956. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3014.pdf> Sidst tilgået d. 27. maj 2025.
- Afzelius, Adam, Gustav Hermansen og Kåre Olsen. *Danmarks Riges Breve - 2. række 8. bind: 1318-1322*. Det danske sprog- og litteraturselskab, 1953. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3015.pdf> Sidst tilgået d. 27. maj 2025.
- Afzelius, Adam, og Gustav Hermansen. *Danmarks Riges Breve - 2. række 9. bind: 1323-1327*. Det danske sprog- og litteraturselskab, 1946. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3016.pdf> Sidst tilgået d. 27. maj 2025.
- Afzelius, Adam, og C. A. Christensen. *Danmarks Riges Breve - 2. række 10. bind: 1328-1332*. Det danske sprog- og litteraturselskab, 1948. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3017.pdf> Sidst tilgået d. 27. maj 2025.
- Afzelius, Adam, og C. A. Christensen. *Danmarks Riges Breve - 2. række 11. bind: 1333-1336*. Det danske sprog- og litteraturselskab, 1950. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3018.pdf> Sidst tilgået d. 30. maj 2025.
- C. A. Christensen. *Danmarks Riges Breve - 2. række 12. bind: 1337-1339*. Det danske sprog- og litteraturselskab, 1960. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3019.pdf> Sidst tilgået d. 30. maj 2025.

- Christensen, C.A., og K. Friis Johansen. *Danmarks Riges Breve - 3. række 1. bind: 1340-1343*. Det danske sprog- og litteraturselskab, 1958. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3020.pdf> Sidst tilgået d. 30. maj 2025.
- Christensen, C.A., og K. Friis Johansen. *Danmarks Riges Breve - 3. række 2. bind: 1344-1347*. Det danske sprog- og litteraturselskab, 1959. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3021.pdf> Sidst tilgået d. 30. maj 2025.
- Christensen, C.A., og K. Friis Johansen. *Danmarks Riges Breve - 3. række 3. bind: 1348-1352*. Det danske sprog- og litteraturselskab, 1963. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3022.pdf> Sidst tilgået d. 30. maj 2025.
- Christensen, C.A., og K. Friis Johansen. *Danmarks Riges Breve - 3. række 4. bind: 1353-1356*. Det danske sprog- og litteraturselskab, 1966. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3023.pdf> Sidst tilgået d. 30. maj 2025.
- Christensen, C.A., K. Friis Johansen og Herluf Nielsen. *Danmarks Riges Breve - 3. række 5. bind: 1357-1360*. Det danske sprog- og litteraturselskab, 1967. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3024.pdf> Sidst tilgået d. 30. maj 2025.
- Christensen, C.A., K. Friis Johansen og Herluf Nielsen. *Danmarks Riges Breve - 3. række 6. bind: 1361-1363*. Det danske sprog- og litteraturselskab, 1969. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3025.pdf> Sidst tilgået d. 30. maj 2025.
- Christensen, C.A., K. Friis Johansen og Herluf Nielsen. *Danmarks Riges Breve - 3. række 7. bind: 1364-1366*. Det danske sprog- og litteraturselskab, 1972. Digitaliseret

på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3026.pdf> Sidst tilgået d. 30. maj 2025.

- Christensen, C.A., K. Friis Johansen og Herluf Nielsen. *Danmarks Riges Breve - 3. række 8. bind: 1367-1370*. Det danske sprog- og litteraturselskab, 1980. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3027.pdf> Sidst tilgået d. 30. maj 2025.
- Christensen, C.A. og Herluf Nielsen. *Danmarks Riges Breve - 3. række 9. bind: 1371-1375*. Det danske sprog- og litteraturselskab, 1982. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3028.pdf> Sidst tilgået d. 30. maj 2025.
- Herluf Nielsen. *Danmarks Riges Breve - 4. række 1. bind: 1376-1379* Det danske sprog- og litteraturselskab, 1984. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3029.pdf> Sidst tilgået d. 30. maj 2025.
- Herluf Nielsen. *Danmarks Riges Breve - 4. række 2. bind: 1380-1385* Det danske sprog- og litteraturselskab, 1987. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3030.pdf> Sidst tilgået d. 30. maj 2025.
- Thomas Riis. *Danmarks Riges Breve - 4. række 3. bind: 1386-1388* Det danske sprog- og litteraturselskab, 1993. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3031.pdf> Sidst tilgået d. 30. maj 2025.
- Herluf Nielsen. *Danmarks Riges Breve - 4. række 4. bind: 1389-1392* Det danske sprog- og litteraturselskab, 1994. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3032.pdf> Sidst tilgået d. 30. maj 2025.
- Herluf Nielsen. *Danmarks Riges Breve - 4. række 5. bind: 1393-1395* Det danske sprog- og litteraturselskab, 1998. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3033.pdf> Sidst tilgået d. 30. maj 2025.

- Aage Andersen. *Danmarks Riges Breve - 4. række 6. bind: 1396-1398* Det danske sprog- og litteraturselskab, 1998. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3034.pdf> Sidst tilgået d. 30. maj 2025.
- Andersen, Aage, Russel Friedman og Herluf Nielsen. *Danmarks Riges Breve - 4. række 7. bind: 1399-1400* Det danske sprog- og litteraturselskab, 2000. Digitaliseret på Claus Rønlev Bibliotek. Hentet fra <https://www.ronlev.dk/bibliotek/3035.pdf> Sidst tilgået d. 30. maj 2025.