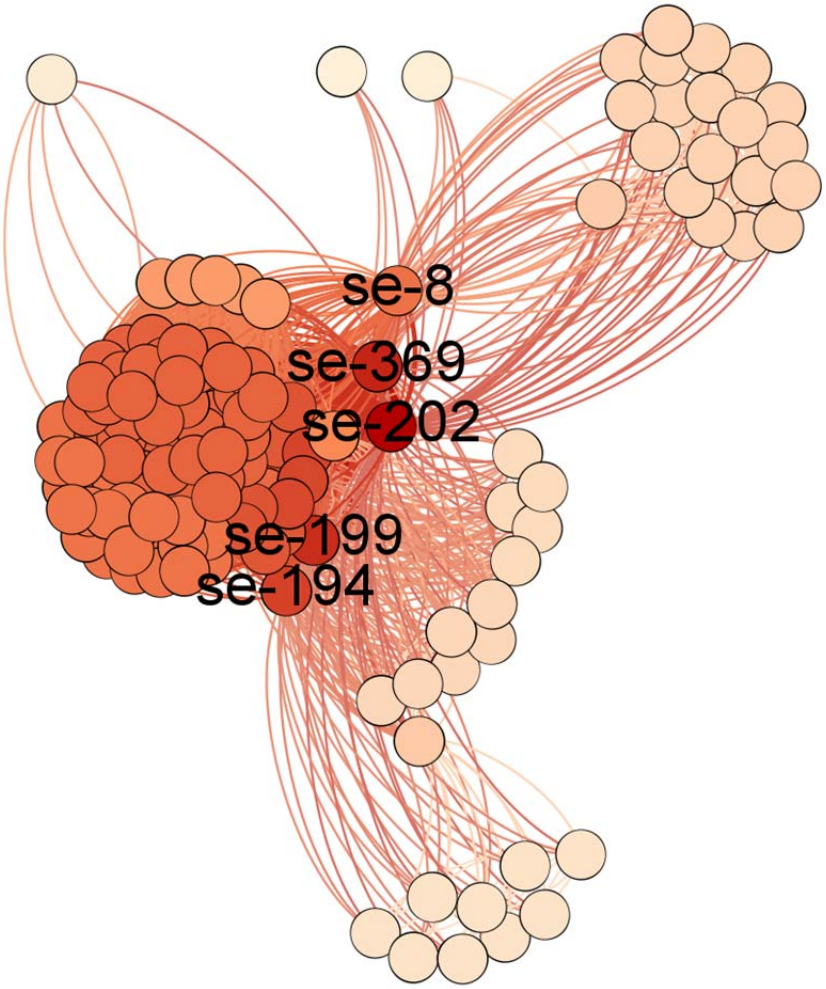
Manuscriptnetwerkvisualisatie en haar tekortkomingen

Een inleiding in de netwerkvisualisatie voor het onderzoek naar manuscripttransmissie



Celis Tittse

S1001174

Bachelorscriptie Geschiedenis

14 augustus 2022

Onder begeleiding van Shari Boodts

Inhoudsopgave

Inleiding

Manuscripten zijn in veel opzichten de toegangspoorten tot het middeleeuwse verleden. Het zijn de handgeschreven boeken (*codices*) die de middeleeuwse mens aan ons heeft overgedragen. Elke bladzijde moest met de hand worden overgeschreven van een ander manuscript.[1] Naast orale informatieoverdracht waren manuscripten een grote bron van kennis in de vroege en hoge middeleeuwen. Beide media gingen met de snelheid van paard en wagen, in tegenstelling tot de glasvezel van nu. Het in kaart brengen en analyseren van deze kennisstromen en de ontwikkeling van het manuscript vallen in het domein van de manuscriptstudies.[2] Doorgronden wat, hoe en waarom informatie werd doorgegeven, vertelt ons meer over de interpretatie van de manuscripten en hun inhoud in de toenmalige culturele en politieke situatie.

De studie naar manuscripten is al heel oud, maar de methodes om die te onderzoeken blijven veranderen. Zo is er in het afgelopen decennium een trend opgekomen die manuscripten onderzoekt door middel van netwerkanalyse. Deze methode is ontstaan in de wiskunde en toegepast op mensen in de sociologie waar het doorontwikkeld is. Het maakt gebruik van knopen en relaties om vanaf een afstand naar het netwerk te kijken en vanuit daar vragen over te stellen. Op deze manier kunnen de teksten in manuscripten ook in relatie tot elkaar worden onderzocht. Gustavo Fernandez Riva heeft in 2019 een belangrijk artikel gepubliceerd waarin hij de mogelijkheden van manuscriptstudies en netwerkvisualisatie uiteenzette.[3] Hij beschreef, middels beschrijvende informatie over manuscripten (metadata) uit de database *Handschriftencensus,* de relatie tussen middel-Duitse literatuurtradities. Hierin heeft hij de mogelijkheden van het vakgebied laten zien, maar hij is over de tekortkomingen en valkuilen minder spraakzaam. Aangezien dit een beginnend vakgebied is, is het nodig om niet alleen de mogelijkheden maar ook de potentiële risico's van manuscriptonderzoek met netwerkvisualisaties inzichtelijk te maken. Deze scriptie kan worden gezien als een toevoeging op het werk van Riva. Het dient als een nuancering op de potentie die Riva opwerpt, zodat we een vollediger beeld krijgen van netwerkvisualisaties in de manuscriptstudies wanneer er aan dit vakgebied wordt begonnen.

Daarom is voor deze scriptie één onderzoeksgebied binnen de handboekkunde van bijzonder belang. Het artikel van Riva gaat over de interactie van teksten en in welke hoedanigheid die tot ons zijn gekomen. Dit is het domein van de transmissiestudies en receptiestudies. De transmissie van teksten gaat over de reis die teksten maken voordat ze op hun huidige locatie belanden. Teksten

---

[1] E. Buringh, *Medieval manuscript production in the Latin West: explorations with a global database*. Global economic history series v. 6 (Leiden en Boston 2011). Dit boek geeft een goed overzicht van allemaal verschillende facetten van boekproductie in de middeleeuwen over heel Europa.; Albert Derolez, *The Palaeography of Gothic Manuscript Books: From the Twelfth to the Early Sixteenth Century* (Cambridge 2003).; Erik Kwakkel, 'Biting, Kissing and the Treatment of Feet: The Transitional Script of the Long Twelfth Century', in: Rosamond McKitterick e.a. ed., *Turning over a new leaf: change and development in the Medieval manuscript*. Studies in Medieval and Renaissance book culture (Leiden 2012) zijn twee goede boeken om de weg wijs te worden in paleografie of handschriftkunde vanaf de lange twaalfde-eeuwse renaissance.

[2] Erik Kwakkel, *Books before print*. Medieval media cultures (Leeds 2018) en Mathias Kluge, *Handschriften des Mittelalters: Grundwissen Kodikologie und Paläographie* (3e ed., Ostfildern 2019) zijn goede werken als introductie in de manuscriptstudies.

[3] Gustavo Fernandez Riva, 'Network Analysis of Medieval Manuscript Transmission', *Journal of Historical Network Research* 3 (2019) 30–49. Fiscarelli stelt dat Riva's artikel een van de trends is binnen de netwerkanalyseonderzoeken van digitale geesteswetenschappers. Antonio Maria Fiscarelli, 'Social network analysis for digital humanities', in: Andreas Fickers en Juliane Tatarinov ed., *Digital History and Hermeneutics: Between Theory and Practice* (Oldenbourg 2022) 23–42, aldaar 33–34.

werden overgeschreven in een bepaalde tijd en context.[4] Dit kan worden bestudeerd om het verondersteld belang en leven van een tekst te bekijken. Manuscripten zijn in deze context het vervoersmiddel van teksten. Echter, de teksten reisden vaak niet alleen. Ze werden gedeeltelijk of volledig samengevoegd met andere werken. Dit zegt doorgaans iets over de interpretatie van de tekst in de toenmalige culturele, politieke en religieuze situatie.[5] De analyse van reproductie, dus het kopiëren van de teksten, is besteed aan de transmissiestudies.

Doordat het ene manuscript van het andere werd overgeschreven, kunnen manuscripten in principe gezien worden als objecten die contact met elkaar hadden. De kopiist heeft een of meerdere manuscript(en) gebruikt om over te schrijven tot een nieuw boek. Dit zou kunnen gelden als contact of een verbinding. Daarom kan de gereedschapskist van de sociologen worden geraadpleegd om de transmissiepatronen te lijf te gaan. Dit kan bijvoorbeeld gedaan worden door manuscriptnetwerken als sociale netwerken te beschrijven. Dit is waar sociale netwerkanalyse (SNA) om de hoek komt kijken. SNA is een statistische benadering om de netwerken van mensen te analyseren.[6] Het gaat ervan uit dat mensen in een context opereren. Met netwerkanalyse ontstaat de mogelijkheid om uit te zoomen weg van het individu. Dit creëert overzicht en geeft sociologen de mogelijkheid om statistische vragen aan het netwerk te stellen. Deze manier van onderzoek kan worden toegepast op manuscripten.[7]

Sociologen hebben al veel werk zitten in het theoretiseren van hun actoren in een wiskundig beschreven netwerk, maar filologen nog niet. Daar wil deze scriptie een opstap voor zijn, om de mogelijkheden en valkuilen van netwerkvisualisatie in de transmissiestudies te onderzoeken. Dit doet het door eerst een overzicht te geven van de historiografie van netwerkanalyse in de manuscriptstudies. De studie is opgebouwd volgens de logica waarmee een onderzoek in dit domein kan worden opgestart. Het begint met een nieuwe manier van naar manuscripten kijken, namelijk als dataobject. Daarna kijkt het naar de verschillende vragen die gesteld kunnen worden aan de data en welk type methode hiervoor gebruikt kan worden. Daarna zal er dieper worden ingegaan op netwerkvisualisatie. Hiervoor zal veelal de manuscript- en inhoudsrelaties van de Augustinus' preek 202 (AU s 202 en se-202) worden gebruikt.[8] Preek 202 van Augustinus heeft namelijk een interessant karakter als het aankomt op transmissie. Het zit niet altijd in een grote collectie die altijd samen wordt gekopieerd, maar komt in meerdere collecties van preken voor en ook alleen. Daardoor is het een interessante casus om de problemen van netwerkenanalyse bij transmissienetwerken uit te lichten. Elke sectie heeft een theoretisch kader, potentiële valkuilen en een illustratie aan de hand van preek 202. Hieruit zal geconcludeerd worden dat er bij elke stap gereflecteerd moet worden op de tekortkomingen van de data, vraag en methode. Daarnaast zal blijken dat de statistische hulpmiddelen in de netwerkvisualisaties niet betrouwbaar zijn voor de manuscriptstudies.

---

[4] Ibid., 33.

[5] Maximilian Diesenberger en Yitzhak Hen, *Sermo doctorum: compilers, preachers, and their audiences in the early medieval West*. Sermo volume 9 (Turnhout 2013) 9.

[6] Martin Grandjean, 'Introduction to Social Network Analysis: Basics and Historical Specificities' (HNR+ ResHist Conference 2021) 1–22, aldaar 2–3.

[7] Mathias Blobel, Investigating Genre through Network Analysis of the Electronic Manuscript Catalogue (Master Thesis University of Iceland, Reykjavík 2015). Deze scriptie is een voorbeeld van manuscriptnetwerken waarin sages de verbindingen vormen.

[8] De metadata is onttrokken uit een metadatadocument opgesteld door Luc de Coninck, verder beschreven op pagina 10. De controle, evaluatie en discussie over de preek is gedaan door Gleb Schmidt. Ik ben erg dankbaar voor zijn tijdloze behulpzaamheid en kritische blik. Luc De Coninck, 'Sermones file 2017' (KU Leuven 2017).

Historiografie

Van grafentheorie naar manuscriptnetwerken

Voordat we in netwerkvisualisaties van manuscripten duiken, is het nuttig om te weten waaruit het vakgebied ontstaan is. Het ontleent allemaal verschillende elementen uit diverse disciplines. Eerst zal er een schets worden gegeven van de voorlopers van de manuscriptnetwerkanalyse. Vervolgens wordt er gekeken naar de twee scholen binnen manuscriptnetwerkanalyse.

De aanloop naar netwerkanalyse in manuscriptstudies kent grofweg drie ontwikkelingsfases. Het begon bij de uitvinding van de grafentheorie in de 18e eeuw door de wiskundige Euler.[9] Euler abstraheerde de verbindingen tussen de bruggen in knopen (*nodes*) en zijden (*edges*). Op deze manier kunnen er mathematische beschrijvingen worden gemaakt van relaties tussen punten. Dit zette de deur open voor het wiskundig analyseren van netwerken. De knopen en zijden werden in de 20ste eeuw individuen en hun relaties voor sociologen. Op deze manier kreeg grafentheorie een sociale toepassing als *Social Network Analysis* (SNA). Sociologen hebben decennialang gewerkt aan de theoretisering van deze toepassing om menselijke relaties te beschrijven. Er zijn nieuwe beschrijvingsmethodes ontwikkeld en er is een polemisch debat gevoerd over de mogelijkheden en tekortkomingen van deze methode.[10] Sinds de *digital turn* in de jaren ’90, hebben boekdeskundigen steeds meer data en bronnen om te analyseren.[11] Hierdoor gingen digitale geesteswetenschappers en historici meer statistische methodes gebruiken, waaronder netwerkanalyse.[12] Hiermee zijn in tweehonderd jaar de statistische methodes gekomen uit de wiskunde, het nadenken over sociale netwerken uit de sociologie en de bredere inzet van netwerkanalyse uit de Digital Humanities.

Verschillende manuscriptonderzoekers hebben ondertussen de weg naar het netwerkperspectief gevonden. Een analyse van het onderzoeksveld toont dat er grofweg twee scholen zijn. Dit onderscheid wordt in de literatuur niet gemaakt, maar de methodekritiek verschilt voor beide scholen, dus onderscheid ik ze in deze scriptie.[13] Er zijn boekdeskundigen die netwerken gebruiken om hun data te visualiseren en onderzoekers die statistische netwerkanalyse gebruiken om hun netwerken te toetsen. De methode in het artikel van Kapitan, Timothy en Wills is het meest exemplarisch voor de visualisatiecategorie.[14] Ze gebruikten metadata uit het The Skaldic Project om

---

[9] Gerald L. Alexanderson, ‘About the cover: Euler and Königsberg’s Bridges: A historical view’, *Bulletin of the American Mathematical Society* 43 (2006) 567–574, aldaar 567–568.

[10] Grandjean, ‘Introduction to Social Network Analysis’, 8–9; Mark Hoffman, *Methods for Network Analysis* <https://bookdown.org/markhoff/social_network_analysis/>; Jochem Tolsma, ‘Social Networks’, *Jochem Tolsma* <https://www.jochemtolsma.nl/courses/>. De laatste twee annotaties zijn online boeken die een sociologische introductie op SNA vormen. Kennis van de programmeertaal R is hiervoor noodzakelijk.

[11] Marilena Maniaci, *Trends in Statistical Codicology* (Berlijn 2021) 4–5.

[12] Matthew Hammond, ‘From Digital Prosopography to Social Network Analysis: Medieval People in the Twenty-First Century’, *Medieval People* 36 (2021). Dit artikel gaat in op de historiografie van een specifieke bron van middeleeuwse netwerkanalyse; briefwisselingen tussen personen.; Voorbeelden hiervan zijn: Ralph Kenna, Máirín MacCarron en Pádraig MacCarron ed., *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. Understanding Complex Systems (Cham 2017).; Anne S. Chao, Zhandong Liu en Qiwei Li, ‘Network of Words: A Co-Occurrence Analysis of Nation-Building Terms in the Writings of Liang Qichao and Chen Duxiu’, *Journal of Historical Network Research* 5 (2021).

[13] Grandjean, ‘Introduction to Social Network Analysis’, 10–11. Grandjean stelt hier dat er andere vormen zijn en dat netwerkvisualisatie geen resultaat op zichzelf is, maar een middel.

[14] Katarzyna Anna Kapitan, Rowbotham Timothy en Tarrin Jon Wills, ‘Visualising genre relationships in Icelandic manuscripts’ (Digital humaniora i Norden 2017, Göteborg) 59–62.; Kapitan reageert later op dit Conferentiepaper door de digitale catalogi te bekritiseren en de problemen die dat opwerkt. Dit is een goede

te achterhalen of de IJslandse, mythische heldensage Fornaldarsögur hetzelfde genre was als de riddersaga Riddarasögur. Ze gebruikten de overeenkomst van andere sagen om de manuscripten te klusteren en keken of de twee sagen dichtbij elkaar lagen.[15] Visualisatie is hier de voornaamste vorm van analyse.

In de tweede categorie, de statistische, is het werk van Mac Carron en R. Kenna het meest typerende voorbeeld.[16] Zij vergelijken de transmissietraditie van verschillende Ierse mythen om de wiskundige eigenschappen van de transmissienetwerken uit te diepen. Hierbij ging het over de dichtheid van het netwerk van manuscripten met mythe 1 erin vergelijken met de manuscripten met mythe 2 erin. Dit deden ze niet door alle netwerken visueel met elkaar te vergelijken, maar door de wiskundige eigenschappen van de netwerken te bekijken. Denk hierbij aan de diameter van de netwerken, of hoeveel verbindingen een gemiddelde *node* heeft. Dit is dus een statistische aanpak om manuscriptnetwerken te analyseren.

Samengevat ligt de basis van de netwerkanalyse in de wiskunde. De wiskundige grafentheorie kreeg in de 20e eeuw een toepassing in de menswetenschappen, doordat sociologen het gebruikten om netwerken tussen mensen te beschrijven. Uiteindelijk werd het in de jaren '90 gebruikt door geesteswetenschappers onder de vleugels van de *digital turn* en de laatste tien jaar hebben boekdeskundigen het gebruikt om de relaties tussen handschriften te onderzoeken. Dat is het punt waar we nu zijn. Daarbinnen is er een onderscheid te maken tussen papers die netwerkvisualisatie gebruiken om tot resultaten te komen en papers die netwerkanalyse gebruiken om hun vragen te toetsen. Dit is de academische context waarbinnen Riva en deze scriptie opereert.

---

aansluiting bij het hoofdstuk *Data* van deze scriptie. Katarzyna Anna Kapitan, 'Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature', *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 6 (2021) 74–97.

[15] Blobel, Julien en Cheriet et. al. volgen hetzelfde patroon. Mathias Blobel, Investigating Genre through Network Analysis (2015).; Octave Julien, 'Délier, lire et relier. L'utilisation de l'analyse réseau pour construire une typologie de recueils manuscrits de la fin du Moyen Âge', *Hypothèses* 19 (2016) 211–224.; Mohamed Cheriet, Reza Farrahi Moghaddam en Rachid Hedjam, 'Visual language processing (VLP) of ancient manuscripts: Converting collections to windows on the past' (Conferentiepaper, 2013 7th IEEE GCC Conference and Exhibition) 407–412.

[16] Sandra D. Prado e.a., 'Temporal Network Analysis of Literary Texts', *Advances in Complex Systems* (2016).; van Mac Carron en R. Kenna, 'Network analysis of the Íslendinga sögur – the Sagas of Icelanders', *The European Physical Journal B* 86 (2013) 407.

Het manuscript als dataobject

De abstracte manuscriptbeschrijving

Manuscripten kunnen ons veel vertellen over het verleden. Maar ze kunnen ons niet alles vertellen en de kwaliteit van de metadata bepaalt welke antwoorden er kunnen worden gegeven. Als dateringen niet zijn geïncludeerd, kan de vraag naar een tijdsbepaling niet worden beantwoord. Maar ook de vorm van de data maakt uit. Wanneer we bijvoorbeeld een foto hebben van een bladzijde, kan er iets gezocht worden naar specifieke woorden, maar moeten de woorden eerst herkend worden door een computer. Stel we zouden vragen willen stellen waarbij netwerkanalyse gebruikt kan worden, dan moet de informatie uit een *codex* op een andere manier worden opgeschreven. Vele onderzoekers, onder wie Julien Octave en Dominique Stutzmann, gebruiken een dergelijke onderzoeksmethode, maar het theoretische raamwerk om een manuscript als dataobject te bekijken is nog niet expliciet onder woorden gebracht.[17] Daarom gaan we in op de theorie van een manuscriptanalyse, zodat een computer het kan begrijpen.

Computers denken op een andere manier dan mensen, dus moet er een vertaling worden gemaakt van een fysiek manuscript naar een digitaal dataobject. Stel we hebben een manuscript voor ons, dan zouden we die op de traditionele manier kunnen analyseren. Kijken we naar de buitenkant, lezen wat erin staat, en kijken we naar de individuele illustraties en andere details. In feite analyseren we een manuscript van groot naar klein. In principe zouden we elk aspect van het manuscript kunnen beschrijven op een kladblaadje van groot naar klein. Kleur van de kaft, de verluchting van Maria op de eerste pagina en lettervorm van de zestiende letter a op de 71^{e} pagina. Dit zou een traditionele manier zijn om een manuscript in groot detail te onderzoeken.

Als we dit zouden vertalen naar computerbeschrijvingen, ontstaat er een hiërarchische structuur van verschillende lagen die met elkaar verbonden zijn. Deze is vergelijkbaar met een huis. Aan de buitenkant is het een gebouw. Wanneer je binnenloopt zijn er verschillende kamers met verschillende functies en inrichtingen. Misschien zijn niet alle kamers tegelijk gebouwd of blijken er kamers verbouwd te zijn, maar dat is aan de buitenkant van het huis niet te zien. In iedere individuele kamer zijn meubels en objecten die op zichzelf ook kunnen worden bestudeerd. Voor manuscripten werkt het hetzelfde. De buitenkant is bijvoorbeeld gebonden in leer en heeft een opdruk van een zon. Sla je het manuscript open, dan zitten er allemaal boekdelen en katernen in die in verschillende tijden gemaakt kunnen zijn en verschillende teksten bevatten. Zo kan een manuscript op meerdere niveaus worden beschreven, zodat een computer het kan begrijpen.

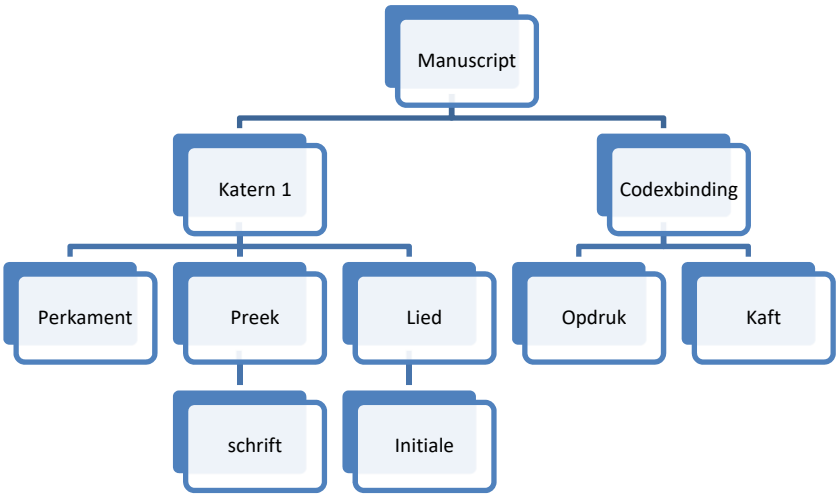
In deze benadering wordt een manuscript gezien als een hoofdobject met subobjecten die allemaal verschillende attributen hebben. Elke laag wordt gevormd omdat er subobjecten onder een hoofdobject zitten. Op een bladzijde (hoofdobject) staan verschillende illustraties (subobjecten). Deze objecten beschrijf je op een bepaalde manier. Deze beschrijving noemen we attributen. Een attribuut van een illustratie is bijvoorbeeld blauwe inkt. Naast deze beschrijvende attributen zijn er nog attributen die extra achtergrondinformatie geven over het object. Hierbij zou men kunnen

[17] Octave Julien, 'Construction et composition des recueils français du XVe siècle: apports de la codicologie quantitative', *Babel. Littératures plurielles* (2007) 13–30.; Dominique Stutzmann, TEI, Github, Zenodo, IIIF, Heurist, Docker: Looking back on Data Management, Tools, and Processes in the Funded Projects Oriflamms, Ecmen, Fama, Himanis, Horae, and Home (Ongepubliceerd Conferentiepaper, International Medieval Congress 2022, Life And Afterlife Of Digital Projects In Medieval Manuscript Studies II).

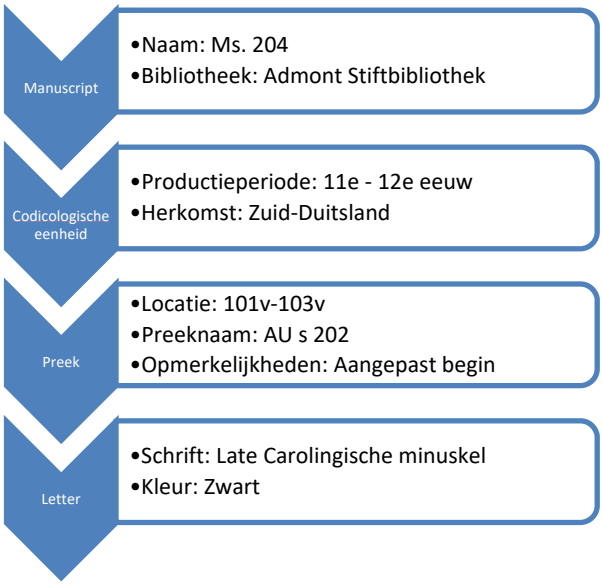
denken aan de productieregio of relevante secundaire literatuur. Dit geeft een abstracte benadering van een object waarin beschrijving en omschrijving gelijkwaardige attributen worden.

Om het minder abstract te maken passen we deze theorie toe op een voorbeeld. Een van de manuscripten waarin preek 202 staat is Admont, Stiftsbibliothek Ms.228.[18] De objecten en subobjecten kunnen los worden beschreven. De gebonden bladzijdes waarop preek 202 staat, heeft als attributen bijvoorbeeld *herkomst: Zuid-Duitsland* en *productieperiode: 1000-1199* (figuur 2). De binding kan bijvoorbeeld van een latere tijd zijn, waardoor het in het niveau van manuscript zou komen. Figuur 1 laat een fictief manuscript zien om te illustreren dat de subobjecten naast elkaar staan. Dus een subobject van een manuscript is bijvoorbeeld de kaft, maar ook zijn titel en de gebonden bladzijdes. Dit voorbeeld van manuscript 228 heeft een abstracter dataobject dan de gedigitaliseerde versie op de e-codiceswebsite.

Kortom, het onderzoeksonderwerp van de manuscriptstudies kan worden beschreven als een metadataobject. Het is namelijk een samenhangend geheel dat bestaat uit objecten met bijpassende beschrijvingen. De hoofdobjecten kunnen worden onderverdeeld in hiërarchische lagen. Deze lagen of onderdelen zijn te analyseren in de context van het manuscript, maar kunnen ook los worden geanalyseerd. Op deze manier is er ruimte voor nieuwe onderzoeksvragen en is er meer mogelijkheid om losse aspecten van manuscripten te analyseren. Deze manier van manuscripten beschrijven als dataobjecten is overigens noodzakelijk op het moment dat er kwantitatieve analyses, zoals netwerkanalyse, worden uitgevoerd. Een computer snapt namelijk de digitale foto's van een manuscript niet zoals wij dat doen, dus moet er een vertaalslag worden gemaakt.



Figuur 1: Manuscriptbeschrijving



Figuur 2: Manuscriptlagen met attributen

[18] Admont, Stiftsbibliothek, Ms.228, f. 101v-103v. Te raadplegen op: https://manuscripta.at/hs_detail.php?ID=26089.

De valkuilen van manuscriptmetadata

In de databeschrijving van Riva, drukt hij lezers op het hart om kritisch te blijven op de data.[19] Hierbij gaat hij verder niet in op de manier waarop je dat doet. Maar bronnenkritiek is een van de belangrijkste onderdelen van historisch onderzoek en met het gebruik van data komt er dus nog een extra laag bij. Er doen zich namelijk twee extra prominente problemen voor wanneer je werkt met datacorpora van manuscripten. Ten eerste kan het lijken alsof de aanwezige manuscriptbeschrijving (metadata) volledig is. Ten tweede kan het lijken alsof de data die beschikbaar zijn correct zijn. Beide zijn doorgaans onzeker en daar moet rekening mee gehouden worden.

Wanneer er rijen en rijen aan data onder elkaar staan, kan dit de illusie van objectiviteit, volledigheid en correctheid geven. Daarom benadrukken kwantitatieve historici met klem dat men altijd voorzichtig moet zijn en de kwaliteit van de gegevens niet moet overschatten.[20] Octave Julien benadrukt om altijd terug te gaan naar de manuscripten zelf om na te lopen wat er gebeurt in de databeschrijving.[21] Daar bovenop komt de schijn van volledigheid. Omdat alles in kolommen staat betekent het nog niet dat het een volledige dataset is. Dit klinkt als een open deur, maar dit wordt helaas nog weleens vergeten.

Op de eerste plaats heeft de statistische filoloog ook te kampen met het feit dat niet alle manuscripten zijn overgeleverd aan ons. De tand des tijds vergt zijn tol, waardoor bronnen uit vroegere tijden schaarser zijn dan bronnen uit nabijer gelegen tijden. Er moet over worden nagedacht hoe dit invloed heeft op de analyse. Een transmissiepatroon identificeren op basis van de 'hele geschiedenis' geeft dus een disproportioneel grote representatie van de latere eeuwen. Nu werd er ook meer geschreven, maar het is niet duidelijk wat het aandeel niet-overgeleverde bronnen is in een gemiddelde berekening.

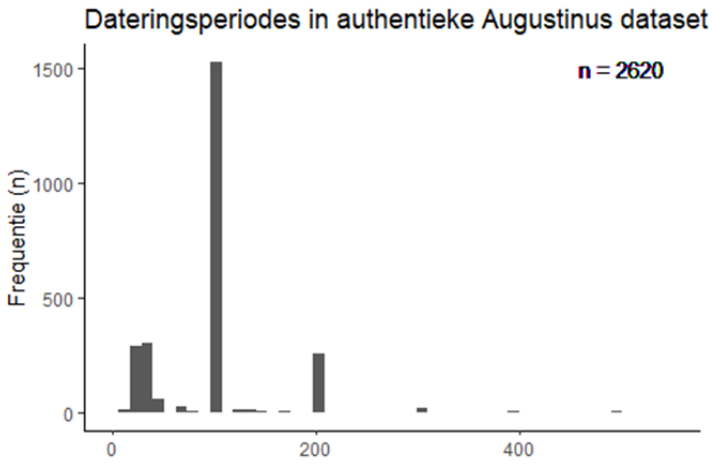
Daarnaast geeft Kapitan in haar paper uit 2021 aan dat een dataset niet volledig is en dat de beschrijving van de manuscripten nu, niets hoeft te zeggen over de productie toen.[22] Ze herinnert haar lezers eraan dat projecten die data beschikbaar stellen een afgebakende omvang hebben en dus niet alle manuscripten bevatten. De achterliggende doelstelling van het project kan een *bias* geven in de resultaten. Om een adequate beschrijving van de data te geven, is het nuttig om per beschrijvende categorie na te lopen waar de informatie vandaan komt en hoe het is opgesteld.

Aangenomen dat de informatie volledig is, dan zijn we er nog niet. Want in middeleeuwse studies is de onzekerheid van data dagelijkse kost. Dateringen zijn doorgaans aangeduid in periodes zoals 'ergens aan het einde van de 12e eeuw'. Er moet rekening mee gehouden worden dat men niet zomaar een gemiddelde van een waarschijnlijke periodisering kan nemen, als de verschillen tussen de potentiële productiedatum sterk uit elkaar lopen. Dit geldt voor de datering, maar ook voor alle andere databeschrijvingen is een kritische noot nodig.

Een voorbeeld hiervan is de metadatabeschrijving van de authentieke Augustinus-preken in de PASSIM-database.[23] De metadata zijn afkomstig uit een document dat opgesteld is door Luc de Coninck. Het beschrijft legio manuscripten die preken bevatten van kerkvader Augustinus van Hippo (354 – 430). Er is een onderscheid te maken tussen preken die waarschijnlijk door Augustinus zijn

[19] Riva, 'Network Analysis of Medieval Manuscript Transmission', 35–36.
[20] Bruno Blondé e.a., *Trend en toeval: inleiding tot de kwantitatieve methoden voor historici* (Leuven 2012) 24.
[21] Julien, 'Délier, lire et relier', 10.
[22] Katarzyna Anna Kapitan, 'Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature', *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 6 (2021) 74–97, aldaar 82.
[23] Patristic Sermons in the Middle Ages, The dissemination, manipulation and interpretation of late-antique sermons in the medieval Latin West. https://passim.rich.ru.nl/

gemaakt, authentieke preken, en preken die aan Augustinus zijn toegeschreven, pseudo-Augustinus' preken. Het document benoemt alleen de eerst categorie. De Coninck heeft vanaf de jaren '90 verscheidene catalogi, secundaire literatuur en manuscripten ingezien ten einde een overzicht te maken van verschillende manuscripten. Naast de catalogi gebruikte hij de aantekeningen van Dom C. Lambot. Soms beschrijven verschillende catalogi hetzelfde manuscript en soms wordt een manuscript door een enkele catalogus beschreven. De Coninck raadpleegde tot 2017 in totaal 172 titels.



Grafiek 1: Verdeling van de periodegrootte in de dataset

In grafiek 1 zijn alle manuscripten in de dataset van De Coninck te zien. Hierbij zijn alle hypotheses voor de dateringen van een manuscript op een rij gezet. Elke datering bestaat uit een range, tussen 1000 en 1099 bijvoorbeeld. Hierbij zou de dateringsrange 99 zijn, omdat het verschil tussen het begin en einde van de periode 99 is. In grafiek 1 zijn alle ranges geteld en op de hoogte geeft aan hoe vaak die bepaalde range voorkomt. Dit heeft een wisselend beeld met grote en kleine ranges. Er is een piek te zien rond 99 te zien, dus veruit de meeste manuscripten worden aangeduid met de beschrijving: "dit is een 15e-eeuws manuscript" bijvoorbeeld.

De verscheidenheid aan dateringranges is deze dataset heeft te maken met een aantal oorzaken. Allereerst is de datering van manuscripten überhaupt geen exacte wetenschap. Het wordt gereconstrueerd uit de paleografische gegevens, fysieke kenmerken, wiskundige methoden of - met hoge uitzondering - een naam of een jaartal.[24] Dit resulteert doorgaans in een beargumenteerde hypothese van een periode en geen exact jaartal. Ten tweede zijn onderzoekers het niet altijd eens over de datering. Daarom worden er vaak twee tijdperken genomen waartussen de datering van de tekst ligt. Ten derde beschrijft deze dataset de datering van het manuscript als geheel. Dit is eigenlijk een zwaktebod daar een codex de binding van verschillende losse katernen is. Deze kunnen later bij elkaar gevoegd zijn. Als een katern uit de 13e eeuw komt en een andere uit de 16e eeuw, dan kan de periodisering van het manuscript 1200 tot 1599 aangeven. Deze drie aspecten dragen bij aan een zeer heterogene periodisering, waardoor ze niet allemaal over één kam kunnen worden geschoren.

Behalve de datering zijn wetenschappers het soms ook niet eens met de inhoud van de manuscripten. Sommige auteurs van catalogi identificeren bepaalde preken waar andere afwijkende of geen Augustinus-preken vinden. Dit gegeven kan worden doorgetrokken naar een algemene

[24] Enock Osoro Omayio, Sreedevi Indu en Jeebananda Panda, 'Historical manuscript dating: traditional and current trends', *Multimedia Tools and Applications* (2022) 1-30, aldaar 3–4.

onzekerheid over de inhoudsbepaling. Niet alle catalogen hebben alle manuscripten volledig beschreven, waardoor potentiële betwiste preken niet als zodanig staan aangegeven in de dataset. Deze onzekerheid dwingt tot voorzichtigheid in het formuleren van een conclusie.

Daarnaast zijn niet alle manuscripten van begin tot eind beschreven. In de dataset van De Coninck zijn bijvoorbeeld alleen authentieke Augustinus' preken opgenomen. Daardoor kunnen niet alle vragen beantwoord worden. Als je wilt weten hoeveel preken twee manuscripten gemeenschappelijk hebben, en alleen Augustinus' preken zijn beschreven, kan die vraag dus niet worden beantwoord. De inhoud en vooral de missende inhoud van een dataset is van groot belang om een goede vraagstelling en analyse te kunnen doen.[25]

Om alles nog onzekerder te maken, attendeert Kapitan ons erop dat de boekdelen of katernen door de tijd heen niet constant hoeven zijn gebleven.[26] Deze hoeven niet per se op hetzelfde moment geproduceerd te zijn. Het feit dat we een manuscript in deze vorm hebben, betekent niet dat dit niet veranderde over de eeuwen. Als we onderzoek zoals Riva willen starten waarbij verbindingen worden gemaakt tussen samen-aangetroffen teksten in een *codex,* moet daar rekening mee worden gehouden in het trekken van conclusies.

Kortom, wanneer je met data werkt moet je dus altijd kritisch blijven. De illusie van volledigheid, correctheid en objectiviteit ligt op de loer. Om dat het hoofd te kunnen bieden, is een goede data- en risicobeschrijving gepast. Hierin zijn de oorsprong, de accuraatheid en de tekortkomingen van de data van ruim belang. Een cijfer of een titel is niet zo zeker als het lijkt en de vraagstelling, analyse en conclusie moet daarop worden aangepast.

[25] Blondé e.a., *Trend en toeval*, 205–206.
[26] Kapitan, 'Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature', 82–84.

Netwerkperspectief

Er wordt een onderscheid gemaakt tussen twee categorieën van netwerkanalyse: netwerkvisualisatie en netwerkanalyse. Dit zijn twee kwantitatieve onderzoeksmethodes. Om de verschillen te zien, is het goed om een stap terug te nemen. Waar komt dit netwerkidee vandaag? Wat is de toegevoegde waarde van een netwerkperspectief, zoals Riva gebruikt? Om dit te onderzoeken kijken we naar het verschil tussen een netwerkperspectief, netwerkvisualisatie en netwerkanalyse. Dit zijn andere methodes die nuttig kunnen zijn om onderzoeksvragen te beantwoorden. Elke methode heeft voordelen om een bepaald soort vragen te beantwoorden en moet daarom zorgvuldig gekozen worden.

In de brede zin wordt bij een netwerkperspectief ervan uitgegaan dat je onderzoeksobject niet los staat van zijn omgeving, maar ingebed is in een context en ook bepaald wordt door de context.[27] Manuscripten kunnen daarom ook worden gezien als een geheel aan entiteiten die met elkaar interacteren en onderling vergeleken kunnen worden. Net als bij een mindmap stelt het de onderzoeker in staat om een stap terug te nemen en het geheel te beschouwen waarin het onderzoeksobject zich bevindt. Het netwerkperspectief beweegt zich weg van objecten geïsoleerd bestuderen en richting een interdependente manier van actoren analyseren.

In de context van manuscriptstudies betekent dit dat manuscripten altijd een netwerk vormen. Dit kan een conceptueel netwerk zijn van samen reizende teksten, een relationeel netwerk van gemeenschappelijke eigenschappen of een geografisch kennisnetwerk waarin manuscripten als de dragers van ideeën functioneren.[28] Die manier van kijken naar manuscripten noemen we het netwerkperspectief. In het eerste *Historical Network Research* boek wordt dit het "metaforische netwerk" genoemd.[29] De nadruk ligt hierbij op de afhankelijkheid van actoren ten opzichte van elkaar, ontleend uit de sociologie. Vanuit dat perspectief kan er op verschillende manieren nagedacht worden over aspecten van manuscripten.

Vervolgens kan er worden gekeken hoe men deze aspecten wil analyseren. Grofweg zijn hier drie vormen in: het metaforische netwerkperspectief, de netwerkvisualisatie en de netwerkanalyse. Het metaforische netwerkperspectief is een kwalitatieve aanpak waarin het perspectief op de positie van het onderzoeksobject in de omgeving ligt. Dit is de gebruikelijke methode in de manuscriptstudies en in de bredere geesteswetenschappen. Het gaat hier om vragen zoals: hoe verhouden teksten zich ten opzichte van elkaar over tijd en wat is de invloed de samenreizende teksten op andere teksten? Dit perspectief is veel besproken en ligt niet specifiek in de reikwijdte van deze scriptie.[30]

---

[27] Charles Wetherell, 'Historical Social Network Analysis', *International Review of Social History* 43 (1998) 125–144, aldaar 126.

[28] Riva's paper is het beste voorbeeld van samen reizende teksten. Voor een relationeel netwerk van geclusterde verluchtingen is het paper van Grana et. al. een voorbeeld: Costantino Grana, Daniele Borghesani en Rita Cucchiara, 'Automatic segmentation of digitalized historical manuscripts', *Multimedia Tools and Applications* 55 (2011) 483–506. En Stutzmann sprak in een conferentie over het idee om Cisterciënzer kloosters als netwerken te beschouwen en de relaties door middel van manuscripten in kaart te brengen: Dominique Stutzmann, 'Réseaux et politiques documentaires de l'Ordre cistercien aux XIIe et XIIIe siècles: des abbayes et des textes', in: *6e rencontre Res-Hist Réseaux bipartis en histoire* (Aix-en-Provence 2021).

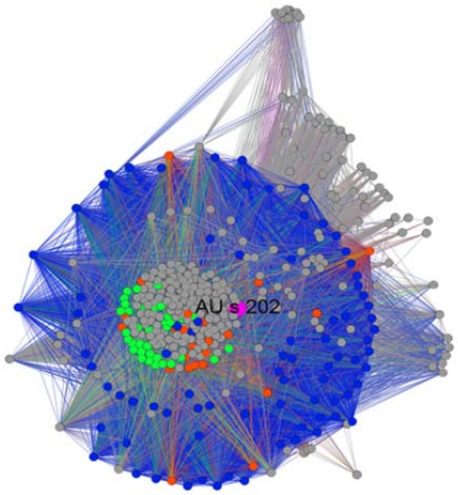
[29] Florian Kerschbaumer e.a. ed., *The Power of Networks: Prospects of Historical Network Research* (2020) 3.

[30] Voor de 'traditionele' manuscriptmethode zijn de volgende boeken goede aanknopingspunten: Joel Thomas Rosenthal ed., *Understanding medieval primary sources: using historical sources to discover medieval Europe*. Routledge guides to using historical sources (London en New York 2012); Erik Kwakkel en Rodney M Thomson,

Daar tegenover staat de kwantitatieve manier van netwerkonderzoek. Deze is onderverdeeld in netwerkvisualisatie en netwerkanalyse. Hierbij wordt een telbare manier gezocht om relaties te definiëren. Hierbij wordt een model gemaakt van de werkelijkheid door entiteiten en relaties die ze tot elkaar hebben te definiëren.[31] Dit zouden bijvoorbeeld Noord-Franse getijdenboeken kunnen zijn, die relationeel gedefinieerd worden op basis van de het aantal keren dat Maria verschijnt. Vervolgens moet de keuze worden gemaakt om een visuele representatie van het netwerk te maken, of statistisch uit te rekenen wat de relaties zijn. Welke aanpak het beste is om te gebruiken, hangt volledig af van de onderzoeksvraag.

Om dit te illustreren kan er gekeken worden naar de transmissietraditie van Augustinus' preek 202. We weten dat preken vaak in collecties samen reizen en zo aan ons zijn overgeleverd.[32] In dit geval zijn we geïnteresseerd in de collecties met Augustinus-preken die in de middeleeuwen zijn samengesteld. [33] Dat zijn *De verbis domini et apostoli, De diversis rebus, De lapsu mundi, Sancti Catholici Patres, Collectio Tripartita, Collectio Bruxellensis* en *Collectorium of Roberto de' Bardi*.[34] Hoe deze collecties tot stand zijn gekomen, maakt voor dit voorbeeld niet bijzonder veel uit. Het gaat hier om de manier waarop bepaalde vragen bepaalde methodes prefereren.

Voor een netwerkvisualisatie zouden we ons af kunnen vragen hoe de preken in middeleeuwse collecties waar sermoen 202 *geen* onderdeel in losse clusters samen reisden. Een cluster ontstaat wanneer bepaalde preken vaak met elkaar in manuscripten worden aangetroffen en minder met andere preken. Om deze visualisatie te maken, definiëren we de *edges* als samen reizende preken in een manuscript. De middeleeuwse collecties zonder 202 zijn de collecties *De verbis domini et apostoli, De diversis rebus* en *De lapsu mundi*. Hier weergegeven in respectievelijk blauw, groen en rood. In de visualisatie is te zien dat veel preken met elkaar verbonden zijn en dat de middeleeuwse collecties niet specifiek losse clusters vormen. Vervolgens kan worden gekeken naar de reden waarom dat is. Er is altijd onderbouwing of verder onderzoek nodig om de visualisatie te interpreteren. In het hoofdstuk *Netwerkvisualisatie* zal worden uitgediept waarom de visualisatie zich op deze manier aan ons ontvouwt.



Figuur 3: Netwerkvisualisatie van de preken in AU s 202 manuscripten met De verbis domini et apostoli (blauw), De diversis rebus (groen) en De lapsu mundi (rood) uitgelicht.

Netwerkanalyse daarentegen kan statistische antwoorden geven op gerichte vragen. Stel we zouden meer willen weten over de middeleeuwse collecties *Sancti Catholici Patres* en *Collectio Tripartita*. De eerste bevat onder meer 146 authentieke Augustinus' preken en de tweede 196. Zij

---

The European book in the twelfth century (2018).; L. D. Reynolds, N. G. Wilson en Nigel Guy Wilson, *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature* (Oxford 2013).

[31] Grandjean, 'Introduction to Social Network Analysis', 2.

[32] Anthony Dupont ed., *Preaching in the Patristic Era: sermons, preachers, and audiences in the Latin West*. A new history of the sermon 6 (Leiden en Boston 2018) 178.

[33] Hiermee excludeer ik de antieke collecties rond de tijd van Augustinus en de collecties van verzamelaar Caesarius van Arles (ca. 470 – 542).

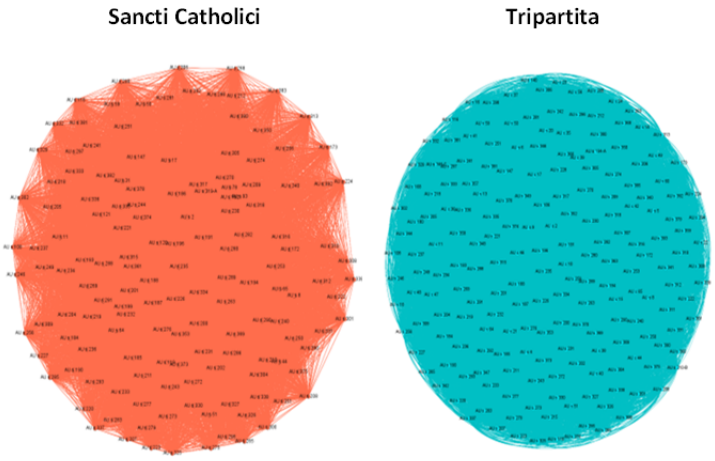
[34] Dupont ed., *Preaching in the Patristic Era*, 179.

hebben 132 preken gemeenschappelijk. Beide collecties komen we in hun (schijnbare) volledigheid tegen vanaf 1300, daarom beslaat deze analyse de periode tussen 1300 en 1600.[35] Hierbij formuleren we de vraag: herkennen we deze collecties geheel of gedeeltelijk in het corpus van manuscripten dat AU s 202 overlevert? Dit kunnen we achterhalen met netwerkanalyse. Het gaat namelijk over hoeveel relaties elke preek ten opzichte van elkaar heeft. Een relatie tussen twee preken wordt gevormd als ze samen in een manuscript worden aangetroffen. Als er in één manuscript tien preken zitten, zal elke preek in dat manuscript tien minus één (zichzelf) relaties hebben. Omdat twee preken maar éeen verbinding per manuscript kunnen hebben, wordt het totaal aantal verbindingen gedeeld door twee. Stel dat alle manuscripten de volledige hoeveelheid preken hebben in de collectie, dan zien we dat als de maximale score 100%. Door te meten hoeveel verbindingen er daadwerkelijk zijn, kunnen we zien hoeveel procent van de preken van een collectie in de overgeleverde se-202 manuscripten te vinden is. De berekening gaat als volgt:

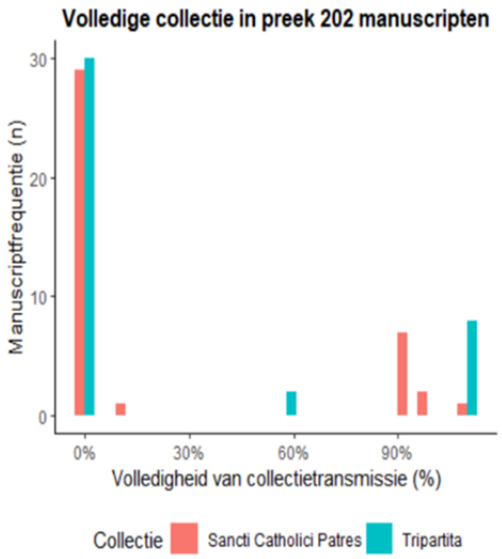
$$n = \ totaal\ aantal\ Augustinus'\ preken\ in\ een\ collectie$$

$$maximale\ hoeveelheid\ edges = \frac{n\ (n\ -1)}{2}\ x\ totaal\ aantal\ manuscripten$$

$$\%\ volledige\ collectie = \frac{gemeten\ edges}{maximale\ hoeveelheid\ edges}\ x\ 100\%$$



Figuur 4: Netwerkvisualisatie van de samenreisverbindingen tussen de preken van historische collecties Sancti Catholici (links) en Tripartita (rechts) vanuit AU s 202 manuscripten.



Grafiek 2: Histogram van AU s 202 manuscripten die een bepaald percentage van Sancti Catholici en Tripartitia in zich dragen.

[35] Sancti Catholici Patres was samengesteld door cisterciënzers in de 12e eeuw en alle delen van Tripartita werden in de 14e eeuw samengevoegd in een manuscript door Roberto di'Bardi. Richard W. Bishop, Johan Leemans en Hajnalka Tamas, *Preaching after Easter: Mid-Pentecost, Ascension, and Pentecost in Late Antiquity* (BRILL 2016) 315.

Hierbij komt gemiddeld de *Sancti Catholici Patres* op 23,8% volledige transmissie van de collectie in manuscripten van s. 202 uit, en *Tripartita* op 25,2%.[36] Om en nabij een vierde van de collectie wordt gemiddeld aangetroffen in manuscripten die preken van die collecties bevatten. Het verschil tussen beiden is niet bijzonder groot, daarom is het interessant om te kijken waar de verschillen zitten. Daarnaast overlapt een groot deel van de preken van beide collecties, dus misschien wordt dat in de resultaten gerepresenteerd. Om inzichtelijk te maken wat hier aan de hand is, kan er een netwerkvisualisatie worden gemaakt.

Echter, wanneer we naar de linker visualisaties kijken, worden we er niet wijzer van. Een netwerkvisualisatie is niet altijd handig. In dit geval is het veel handiger om een histogram te maken van de manuscripten die een bepaald percentage van de totale collectie in zich meedragen. Hierin zien we dat in de manuscripten die ook sermoen 202 in zich dragen, de collectie *Tripartita* vaker volledig wordt overgedragen dan *Sancti Catholici Patres*. Daarnaast zijn er veel manuscripten die maar een paar preken van beide collecties in zich dragen, maar die wel werden meegenomen in dat totale gemiddelde. Deze zijn zichtbaar in grafiek 2 rond de 0% van de x-as. De inzichtelijkheid van de statistische resultaten zit dus hier in nieuwe statistische vragen stellen en niet in netwerkvisualisaties.

Er kan dus op verschillende manieren worden gekeken naar manuscripten met een netwerkperspectief. De vraag die gesteld wordt, moet leidend zijn in de methode die gekozen wordt. Hierbij is het belangrijk om na te gaan of de data die je hebt volstaan om de vraag te beantwoorden. Wanneer dat het geval is en de methode succesvol is uitgevoerd, moeten de resultaten altijd worden geëvalueerd. Hierbij dient de methode zelf te worden geëvalueerd; welk aandeel heeft mijn methode in het formuleren van de resultaten? Vervolgens kan gekeken worden naar de historische verklaring voor de gevonden resultaten. Op deze manier is de traditionele 'wetenschappelijke' methode doorlopen, maar nu met vanuit een netwerkperspectief.[37]

[36] Tripartita bevat 196 authentieke Augustinus' preken, waarvan zonder excerpten en dubbelen, 183 unieke. Deze dataset beschrijft 41 manuscripten waarvan er 783.510 *edges* mogelijk waren en 197.238 bevonden waren. *Sancti Catholici Patres* bevat op zijn beurt 146 authentieke Augustinus' preken, waarvan zonder excerpten en dubbelen, 142 unieke. Deze dataset beschrijft 41 manuscripten waarvan er 433.985 *edges* mogelijk waren en 103.250 bevonden waren.

[37] Het ligt iets gecompliceerder dan enkel de wetenschappelijke methode. Hier wordt een toelichting gegeven aan de hand van de digitale hermeneutische cirkel. Andreas Fickers, Juliane Tatarinov en Tim van der Heijden, 'Digital history and hermeneutics–between theory and practice: An introduction', in: *Digital History and Hermeneutics: Between Theory and Practice*. Studies in Digital History and Hermeneutics (1ste druk; München 2022) 1–24.

Netwerkvisualisatie

Recapitulatie

De mogelijkheden van netwerkvisualisatie is de grote toegevoegde waarde van Riva's artikel. Zijn netwerkgrafieken lijken intuïtief, maar wanneer we inzoomen blijken er meer dingen te zijn om rekening mee te houden dan alleen de historische interpretatie van deze grafieken. Een paar van deze problemen hebben we al eerder langs zien komen. Het eerste is het verschil tussen netwerkvisualisatie en netwerkanalyse. Verder hebben we gezien dat een visualisatie een herordening van data of kennis is, dat gebruikt kan worden voor een overzicht of als index. Hierbij is een visuele representatie een relationeel model van data die mogelijk gebrekkig, inaccuraat en onvolledig is. Daardoor moet als altijd met grote voorzichtigheid en gepaste argwaan worden omgesprongen met dergelijke grafieken. Ook hebben we gezien dat een visualisatie "geen [op zichzelf staand] onderzoeksresultaat als zodanig is, maar een uitnodiging voor verder onderzoek."[38] Het moet daarom altijd verklaard en gecontextualiseerd worden op basis van de methode en de data. Dit is al een vlezige lijst van aspecten om rekening mee te houden.

In dit hoofdstuk zal gekeken worden naar de manier om van manuscripten data te maken, die geschikt zijn voor netwerkanalyse. Vervolgens zal gekeken worden naar de manier waarop data gevisualiseerd worden en hoe de vouwing van netwerkvisualisaties werkt. Tot slot zal kritisch gekeken worden naar de visuele, statistische weergavemogelijkheden binnen netwerkvisualisaties. Wanneer de totstandkoming inzichtelijk is, is de kans op inaccurate interpretaties beperkt en daarom is het goed om te weten wat er achter de schermen van netwerkprogramma's gebeurt.

Van manuscriptbeschrijving naar netwerkdata

Het grote voordeel van manuscripten als metadataobjecten te zien, is de isolatie van manuscriptaspecten, zoals productiedatum en inhoud bijvoorbeeld. Hierdoor kunnen objecten of attributen met elkaar worden vergeleken. Wanneer twee objecten iets gemeen hebben, zoals kleur of samen in een manuscript zitten kunnen ze een relatie krijgen. Dit opent de deur voor allerhande statistiek en netwerkanalyse. In het hoofdstuk *Netwerkperspectief* hebben we gezien dat preken van Augustinus genomen werden als actoren. In deze voorbeelden werd er van een relatie gesproken wanneer twee preken samen in één manuscript voorkwamen. Met andere woorden, de manuscripten zijn de *edges* en de preken de *nodes.*

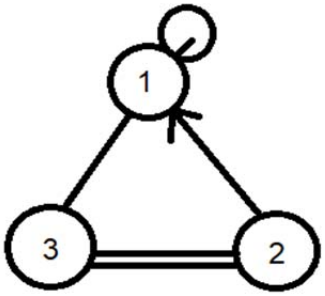
Deze *nodes* (knopen) en *edges* (verbindingen) kunnen op meerdere manieren worden weergegeven. Bijvoorbeeld door een lijst te maken van alle relaties. Ook kan er een matrix worden gemaakt waarin alle *nodes.* Hierbij staan alle *nodes* als de nummering van een schaakbord haaks op elkaar staan. De hoeveelheid relaties die zij delen worden dan weergegeven op het snijpunt (zie figuur 5). De eerste vorm noemen we een *edge list* en de tweede noemen we een sociomatrix of een *adjacency matrix*.[39] Dit is een datarepresentatie van de relaties. Stel preek 1 en preek 2 worden aangetroffen in drie manuscripten, dan worden er drie *edges* gevormd tussen preek 1 en 2. In de *edge list* zullen dit drie regels zijn met *edges* en in het sociomatrix staat op het kruispunt van preek 1 en 2, het nummer 3. Deze notatievorm voor *nodes* en *edges* zijn dus een geabstraheerd model van de onderzoeksobjecten.

---

[38] "[A complex netwerk visualisation] is not a research result as such, it's an exploration interface." Grandjean, 'Introduction to Social Network Analysis', 3.

[39] M. E. J. Newman, *Networks* (2e ed.; Oxford en New York 2018) 225–236. Dit geeft een solide beschrijving van de twee datavormen en hoe men ermee om kan springen.

Wanneer de notatievorm is geconstrueerd kan het worden omgezet naar een netwerk. Een netwerkanalist heeft de data om zijn netwerk op te bouwen, maar ook extra gereedschappen. Het eerste gereedschap is het gewicht van de relaties. De *edges* kunnen extra gewicht krijgen (*weighted*), dit betekent dat ze zwaarder worden meegenomen in de netwerkanalyseberekeningen. Als het gewicht van een *edge* van één naar twee wordt verhoogd, zal de visualisatie het als twee keer zo zwaar meenemen.[40] Het tweede stuk gereedschap is de gerichte relaties (*directed*). De visualisatie kan dit aangeven met een pijl. Dit betekent dat er een hoeveelheid inkomende en uitgaande relaties ontstaan. Voor drie *nodes* zijn de *edge list* en een sociogram beschreven om alle middelen onder elkaar te zetten:



Egde list:

{1:3} undirected
{3:2} undirected
{3:2} undirected
{2:1} directed
{1:1} directed

Sociogram	1	2	3
1	1	0	1
2	1	0	2
3	1	2	0

Figuur 5: Verschillende representaties van edge databeschrijvingen. Grafenfiguur (links), edge list (midden) en sociogram (rechts).

Dit is de manier waarop van een onderzoeksobject een datarepresentatie kan worden opgebouwd. Er is nog veel meer te zeggen over de technische aspecten van deze stap, maar die zijn voor deze scriptie niet bijzonder relevant. Wouter Nooy et. al. hebben het uitgebreid beschreven in een hoofdstuk over sociomatrices.[41] Belangrijk voor deze scriptie is de manier om metadata om te zetten naar *nodes* en *edges*, de datastructuren waarin ze gegoten kunnen worden en de extra soorten relaties die gebruikt kunnen worden om een adequate visualisatie te maken.

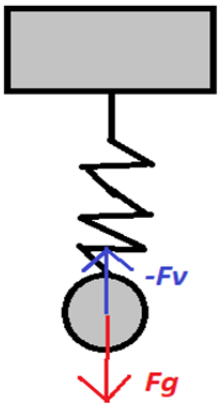
Lay-out van de visualisatie

Netwerkvisualisaties zijn in de basis wiskundige grafentheorielichamen die worden weergegeven in punten die verbonden zijn met lijnen. Een representatie op die manier heeft een aantal voordelen voor de geesteswetenschappen, constateert Venturini et. al.[42] Allereerst worden de punten gepositioneerd op basis van hun connectiviteit. Ten tweede kunnen punten visueel onderscheiden worden door aangegeven kleuren. Ten derde kunnen clusters makkelijk worden geïdentificeerd. Daarom staan zij voor het gebruik van netwerkvisualisatie als analysevorm.[43] Dit kan bijvoorbeeld worden gedaan middels een programma als Gephi, maar ook via modules in onder andere R en Python. Cherven heeft een uitgebreide handleiding gemaakt om Gephi te gebruiken, daarom zal deze scriptie hier verder niet op ingaan.[44] Het gebruik van dit soort programma's heeft de

[40] Wiskundig gezien is er geen verschil tussen twee *edges* of een *edge* met een gewicht van twee.
[41] Wouter de Nooy, Andrej Mrvar en Vladimir Batagelj, *Exploratory social network analysis with Pajek*. Structural analysis in the social sciences 46 (Cambridge en New York 2018) 315–348.
[42] Tommaso Venturini, Mathieu Jacomy en Pablo Jensen, 'What do we see when we look at networks: Visual network analysis, relational ambiguity, and force-directed layouts', *Big Data & Society* 8 (2021) 1-18, aldaar 3.
[43] Ibid., 2. Zij gebruiken de term *visual netwerk analysis* (VNA). Zij ageren tegen onderzoekers die VNA als gevaarlijke methode zien en het gebruik ervan willen minimaliseren.
[44] Ken Cherven, *Mastering Gephi network visualization: produce advanced network graphs in Gephi and gain valuable insights into your network datasets*. Open source: Community experience distilled (Birmingham

gebruiksvriendelijkheid van *social network analysis* voor wetenschappers zonder statistische achtergrond verhoogd. Dit betekent dat onderzoekers de berekeningen achter de visualisatie niet zelf doen, maar die wel gebruiken als aanknopingspunt voor hun analyses.

Een belangrijk aspect van de veronderstelde nuttigheid van deze programma's is de visuele ordening van de data. Dit wordt vaak de lay-out genoemd. Een grafenlichaam is een multidimensionaal figuur dat op een bepaalde manier gereduceerd moet worden tot een tweedimensionaal plaatje. Dit werkt op dezelfde manier als de afbeelding van een springveer waarbij schaduw en overlap wordt gebruikt om diepte te suggereren in een tweedimensionale representatie (zie figuur 6). Om dit te bewerkstelligen maken programma's gebruikt van een *force-directed layout*. Dit is voor SNA een goede optie omdat de verschillende clusters duidelijk worden onderscheiden.[45] Groepjes *nodes* die samen veel verbindingen hebben, worden bij elkaar gezet en weggeduwd van andere groepjes. Zo'n lay-out reduceert dus de originele complexiteit van een grafenlichaam. Om dit op een kritische manier te gebruiken, is het nuttig om te snappen hoe de visualisaties die we gebruiken tot stand komen en te weten wat er achter de schermen van het computerprogramma gebeurt.



Figuur 6: Krachten op een veer. Aantrekkende kracht en afduwende kracht.

Het principe van *force-directed layouts* is dat er vanuit *nodes* een aantrekkende en een afstotende werking bestaat. Deze is vergelijkbaar met de krachten op een veer (zie figuur 6).[46] Er is een kracht die aantrekt en een kracht die afstoot. De mate waarin lay-out algoritmes dat doen varieert, maar ze bouwen allemaal voort op datzelfde principe. De aantrekkende kracht heeft een lineair verband tot de hoeveelheid relaties die twee objecten delen. Hoe meer relaties er zijn, des te sterker de kracht tot aantrekking wordt. De afstotende kracht is geïnspireerd op de formule voor de veerconstante en is niet lineair. Deze kracht wordt groter als de *nodes* in kwestie überhaupt veel verbindingen hebben, maar weinig verbindingen met elkaar delen.[47] Hierbij is de constante een waarde die bepaald wordt door de instellingen. De instellingen en de formules kunnen variëren bij verschillende lay-outs, maar het komt allemaal neer op de afstotende en aantrekkende werking van twee punten, gebaseerd op de veerconstanteformule. Dit heeft als effect dat *nodes* met veel relaties van elkaar worden weggeduwd. Hierdoor zijn clusters makkelijker te herkennen.

$$F(aantrekkend) = \ hoeveelheid\ edges\ die\ node\ 1\ en\ 2\ delen$$

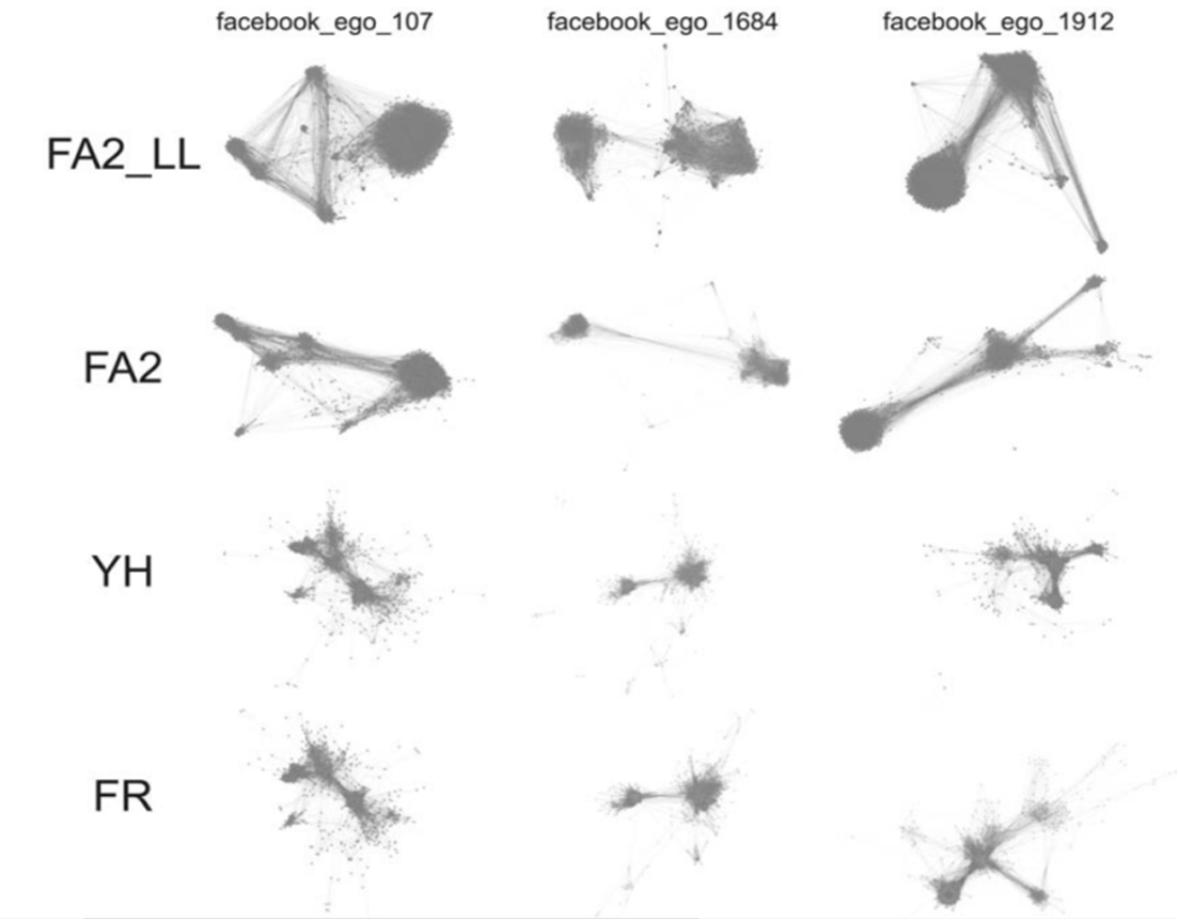
$$F(afstotend) = -\ constante\ \ x\ \frac{(totale\ aantal\ connecties\ node\ 1)\ x\ (totale\ aantal\ connecties\ node\ 2)}{aantal\ edges\ die\ node\ 1\ en\ 2\ delen}$$

Mumbai 2015). Cherven heeft echt een handleiding geschreven waarin in grote lijnen wordt uitgelegd wat functies doen en hoe ze gebruikt moeten worden. Hij laat de technische aspecten hierin achterwege.

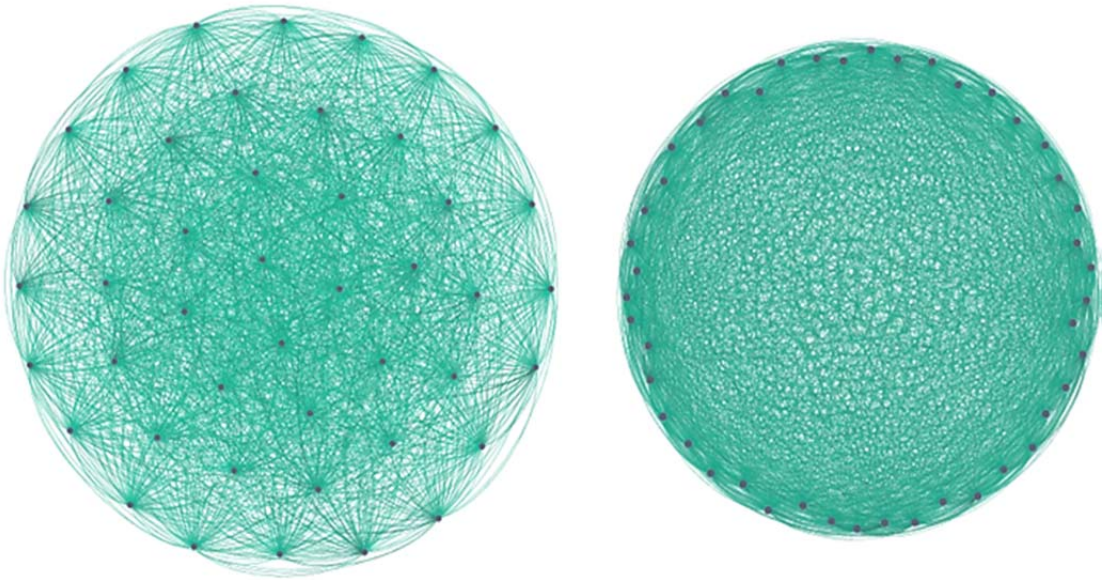
[45] Mathieu Jacomy e.a., 'ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software', *PLOS ONE* 9 (2014) 1–11, aldaar 1–2.; Nooy, Mrvar en Batagelj, *Exploratory social network analysis with Pajek*, 18.

[46] Jacomy e.a., 'ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software', 2.

[47] De positieve en negatieve getallen die uit de formules komen, staan voor de richting waarin ze zich bewegen. Het negatieve getal duwt *nodes* uit elkaar en het positieve getal trekt ze naar elkaar toe.



Figuur 7: Vier lay-outs op drie facebookdatasets. Uit: Mathieu Jacomy e.a., 'ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software', p. 9.



Figuur 8: Netwerkvisualisatie van een manuscriptcluster die met evenveel edges zijn verbonden. Links is gevisualiseerd met ForceAtlas 2 (scaling = 100). Rechts is gevisualiseerd met Yifan Hu Proportional (scaling = 10000).

De variatie binnen de *force-directed* formule resulteert in andere visualisaties met andere lay-outs. Het zijn allemaal andere representaties van dezelfde data. In figuur 7 zijn vier verschillende *force-directed* lay-outs te zien, toegepast op drie datasets van Facebookrelaties.[48] Ze lijken zo verschillend doordat een multidimensionaal figuur moet worden gereduceerd tot een tweedimensionale ruimte. Hiervoor gebruiken ze de veerconstanteformule die de weergave van de data stevig kan beïnvloeden. Het is daarom wijselijk om goed te weten wat de lay-out doet, zodat er begrepen kan worden waarom de data zich op een bepaalde manier ontvouwen.

Er schuilt echter ook gevaar in deze manier van visualiseren. Door de verschillende manieren waarop data gevouwen kunnen worden, kunnen er verkeerde interpretaties ontstaan. *Nodes* die dichtbij elkaar liggen, hoeven in de multidimensionale ruimte helemaal niet dichtbij elkaar te liggen. In figuur 8 zijn twee visualisaties te zien van dezelfde data. Het is een netwerk van manuscripten die allemaal met evenveel *edges* met elkaar verbonden zijn. In dit grafenlichaam staan alle punten even ver uit elkaar omdat ze allemaal even veel verbindingen hebben. Maar wanneer we het gaan visualiseren met een twee *force-layouts*, ontstaan er andere resultaten. Zonder kennis van de lay-outs zou de interpretatie bij de linker visualisatie van figuur 8 kunnen zijn dat de punten niet allemaal even ver uit elkaar liggen. De vouwing van grafiek kan dus de interpretatie beïnvloeden.

Om dit soort valkuilen tegen te gaan is het belangrijk om te weten wat er achter de visualisatie zit. Het maken van een hypothese over de vouwing van het grafenlichaam kan hier een uitweg in bieden. Hierbij is bovengenoemde kennis nodig over de lay-out en begrip van data die erachter zitten. Weten hoe de data verzameld is en een inkijkje in de platte sociomatrix kunnen hierbij helpen. Degelijke visualisaties werken uitstekend om overzicht te creëren in de data, maar moeten met gepaste voorzichtigheid worden ingezet voor analytische doeleinden.

Het gevaar van netwerkvisualisatie met statistische toevoegingen

Door de gunstige lay-outs kunnen netwerken met het blote oog worden geïnterpreteerd. Waarmee kan het een goede bron van informatie zijn om nieuwe verbanden te leggen tussen datapunten. Het geeft overzicht in een soms warrige, complexe situatie. Dit soort netwerken dienen als datavisualisatie.[49] Dit kan vergelijkbaar zijn met een familieboom of een grafiek. Het is daarom het overwegen waard om naar andere vormen van datavisualisatie te kijken wanneer er overzicht moet worden gecreëerd in de gegeven dataset. Er zijn legio voorbeelden van alternatieve vormen van visualisatie, naast de bekende bolletjes-met-streepjesvorm.[50] Met andere datavisualisaties kunnen kleine netwerken ook goed worden geïnventariseerd.

Echter, de netwerken waarmee gewerkt worden in de manuscriptstudies zijn meestal niet klein. Hierbij spreken we over grote of complexe netwerken. Het is hierin minder makkelijk om op basis van een visualisatie een betekenisvolle interpretatie te geven. Het is daarom beter om statistische hulpmiddelen te gebruiken om orde in de chaos te scheppen. Er zijn verschillende methodes om dit te doen en daarvan zou ik de *centrality measurements* verder willen toelichten. De *centrality measurements* worden in de regel gebruikt om vanuit chaotische netwerken, nieuwe

[48] Jacomy e.a., 'ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software', 9.
[49] Grandjean, 'Introduction to Social Network Analysis', 11.
[50] Deze site geeft allemaal verschillende vormen van netwerkvisualisaties. Nathan Yau, 'Network Visualization | FlowingData' <https://flowingdata.com/category/visualization/network-visualization/>; Deze website geeft een overzicht van diverse vormen van datavisualisaties. 'The Data Visualisation Catalogue' <https://datavizcatalogue.com/>.

onderzoeksobjecten te vinden om verder te analyseren.[51] Het is een bekend doel in de *Digital Humanities* als geheel om met nieuwe methodes onderbelichte onderzoeksobjecten te vinden.[52] Daarnaast kan het worden gebruikt om de structuur van het netwerk beter te kunnen begrijpen.

Deze *centrality measurements* zijn standaard netwerkstatistieken die kunnen worden toegepast om de relaties tussen *nodes* met elkaar te vergelijken. Dit kan gaan over individuele *nodes*, maar ook over volledige netwerken. Bij de individuele *nodes* beantwoordt het vragen zoals: wat zijn de verbindingspunten tussen clusters en welke *nodes* zijn met veel andere *nodes* verbonden? Dit zijn de *closeness centrality* vraagstukken.[53] Op dat soort vragen kan een berekening worden uitgevoerd om tot inzichtelijke antwoorden te komen. Vaak kan met een druk op de knop de *centrality* van alle *nodes* worden berekend en gevisualiseerd.[54]

Er zit echter een risico aan deze manier van netwerkvisualisaties gebruiken, constateert Joe Chick.[55] Hij verwijderde een deel van zijn dataset om het dataverlies van vroegmoderne bronnen te simuleren. Hierbij kwam hij tot de conclusie dat de gemiddelde scores van de *centrality measurements* aangaande het volledige netwerk niet bijzonder veel werden aangetast, maar dat de uitlichting van 'bijzondere' *nodes* wel fragiel werden. Dit betekent dat resultaten die gegenereerd worden door *centrality measurements,* die individuele *nodes* moeten doen oplichten, niet even betrouwbaar zijn.

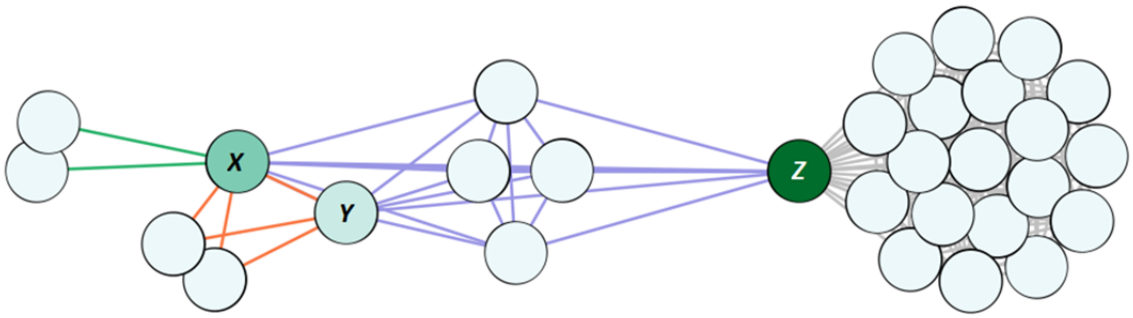
Sébastien de Valeriola ging zelfs een stap verder door te stellen dat de *centrality* het minst betrouwbaar is van de statistische toetsen wanneer er veel data ontbreken.[56] Hij simuleerde de verwijdering van data op een manier die vergelijkbaar is met historische vernietiging van documenten. Vervolgens keek hij in verschillende gradaties naar de invloed van de erosie van data op drie datasets. Hieruit concludeerde hij dat specifieke statistische modellen nodig zijn voor specifieke data met specifieke vragen.[57] Het is dus niet betrouwbaar om de standaard calculaties op een dataset toe te passen.

De Valeriola zegt dit niet zonder reden. Netwerkanalyse is immers een stevige tak van de statistische sociologische discipline. Die heeft een heel scala aan potentiële fouten en vaak voorkomende problemen met hun data opgesteld. Zij hebben oplossingen ontwikkeld om die problemen te overbruggen. Deze oplossingen zijn in de regel statistische operaties die kunnen worden ingezet, wanneer het type data dat toelaat.[58] Echter, deze statistische foutencompensaties zijn nog niet doorgedrongen tot de historische disciplines en niet worden toegepast op de visualisaties van Riva bijvoorbeeld. Dat maakt dat in visualisaties op twee manieren kritisch naar de gegenereerde resultaten moet worden gekeken. Allereerst moet er gekeken worden naar de manier waarop de data zijn geconstrueerd. Ten tweede moet het duidelijk zijn wat de analysemethode wiskundig, dus in de formule, doet. Daarna kunnen de resultaten pas, met de nuance van verloren

[51] Riva, 'Network Analysis of Medieval Manuscript Transmission', 45–46.
[52] Adam James Bradley e.a., 'Visualization and the Digital Humanities:', *IEEE Computer Graphics and Applications* 38 (2018) 26–38, aldaar 31.
[53] Hoffman, 'Centrality'; Cherven, *Mastering Gephi network visualization*, 14–17.
[54] Riva, 'Network Analysis of Medieval Manuscript Transmission', 47.
[55] Joe Chick, 'The impact of missing data on network centrality measures' (University of Warwick 2021) 1-5, aldaar 4.
[56] Sébastien de Valeriola, 'Can historians trust centrality? Historical network analysis and centrality metrics robustness', *Journal of Historical Network Research* 6 (2021) 85-125, aldaar 106, 120-121. Hij toetste dataverlies van 1%, 2%, 5%, 10%, 20% en 40% van de data.
[57] Ibid., 121.
[58] Newman, *Networks*, 275–301.

data, historisch worden geïnterpreteerd. Er komt dus nog een extra statistische interpretatie bij die extra methode en bronnenkritiek nodig heeft.

Om dit te illustreren nemen we de inhoud van drie manuscripten die enigszins gelijke inhoud hebben.[59] De *nodes* zijn gedefinieerd als de authentieke Augustinus' preken in de manuscripten en de relaties zijn gevormd doordat twee preken in een manuscript samen voorkomen. Stel dat er vier preken in een manuscript zitten, dan zullen er vier *nodes* zijn die allemaal met elkaar verbonden zijn. Stel dat er een ander manuscript is dat één preek gemeenschappelijk heeft met het eerdergenoemde manuscript, dan de gemeenschappelijke preek de verbindende *node* zijn tussen twee clusters (zie figuur 9, *node* Z). De verwachting is dat er dus drie clusters ontstaan die onderling *nodes* delen.



Figuur 9: Preekoverlapvisualisatie van drie manuscripten, aangegeven met gekleurde edges, en een fictief manuscript (grijs) waarbij de betweenness centrality weergegeven is in groentinten.

Stel we zouden willen weten welke preek als contactpunt tussen deze clusters zit. Hiervoor kunnen we de *betweenness centrality* gebruiken. "[Met de *betweenness centrality* berekening] zullen we *nodes* vinden die zeer invloedrijk zijn in het verbinden van verder afgelegen gebieden van een grafiek [...]. Deze knooppunten vormen een brug tussen delen van de grafiek en spelen dus een sleutelrol in het verkleinen van de afstanden van de paden bij het doorkruisen van de grafiek". [60] Riva stelt dat "het [een] uitstekende maatstaf [is] om te bepalen welke teksten als knooppunt tussen verschillende clusters van het netwerk dienen."[61] De berekening houdt in dat er elke keer twee willekeurige *nodes* in het netwerk worden gekozen en dat er wordt gekeken of het punt tussen die twee punten ligt. Een punt dat tussen twee clusters in zit zal daarom een hoge *centrality* score krijgen ten opzichte van *nodes* aan de buitenkant van een cluster.[62] Deze score geeft dus aan dat het een belangrijke verbindende *node* is tussen clusters. De formuleberekening hiervan is als volgt:

[59] Dit zijn Benevento, Biblioteca Capitolare, cod.13.; Benevento, Biblioteca Capitolare Cod.8.; Augsburg, Staats- und Stadtbibliothek, 2Â° Cod.184. Daarnaast is een fictieve 'manuscript' toegevoegd met twintig preken waarvan er een gedeeld wordt met 2Â° Cod.184 (*node* Z).

[60] Cherven, *Mastering Gephi network visualization*, 14. "[W]e will find nodes that are highly influential in connecting otherwise remote regions of a graph, even though these nodes might have low influence as measured by other centrality measures. These nodes form a bridge between parts of the graph and thus play a key role in reducing path distances when traversing the graph."

[61] Riva, 'Network Analysis of Medieval Manuscript Transmission', 46. "[B]etweenness centrality is an excellent measure of what texts serve as nexus between different clusters of the network." In SNA worden dit de 'brokers' genoemd, waarover Burt een overzichtsartikel heeft geschreven: Ronald S. Burt, 'The Network Structure Of Social Capital', *Research in Organizational Behavior* 22 (2000) 345–423.

[62] M. E. J. Newman, *Networks* (2e ed.; Oxford en New York 2018) 175.

$$Betweenness\ score\ (node\ X) = \sum_{k=1}^{n} \frac{aantal\ kortste\ routes\ van\ node\ A\ tot\ B\ dat\ door\ node\ X\ lopen}{aantal\ kortste\ routes\ van\ node\ A\ tot\ B}$$

Hierbij zijn *nodes* x, a en b dus willekeurig gekozen *nodes*. Deze berekening wordt uitgevoerd tot er geen knoopparen meer over zijn. De resultaten van de berekeningen zijn aangegeven in de groene tinten van figuur 9.[63] Deze berekening zou volgens de uitleg van Cherven in staat moeten zien om de belangrijkste knooppunten te onderscheiden.

Echter, er schijnt een onverwachte uitkomst te zijn in de uitwerking van deze berekening wanneer de wiskundige resultaten wordt vergeleken met de tekstuele uitleg. Deze uitleg was meer gefocust op de bruggen tussen clusters, maar de wiskundige definitie concentreert zich op de hoeveelheid paden van A naar B. De grootte van de clusters, gevormd door de hoeveelheid teksten in een manuscript, maakt uit voor de score. Dit is te zien in figuur 9, waar een fictief manuscript is toegevoegd met 20 preken erin, waarbij er één preek verbonden is aan een ander manuscript (*node* Z). De preek die de meeste clusters deelt is hier *node* X, maar die heeft een lagere *betweenness centrality* score dan *node* Z*,* omdat die tussen grotere clusters zit. Dus de *betweenness score* zegt ook iets over de grootte van de clusters en niet alleen iets over de hoeveelheden clusters. Dit kom je alleen te weten wanneer je de wiskundige formule kent en van tevoren bedenkt wat voor effect die gaat hebben op de data.

In het voorbeeld van figuur 9 is de onderzoeksvraag naar significante 'bruggen' tussen clusters niet te beantwoorden, aangezien grote clusters per definitie gevormd worden door grote manuscripten. Wanneer de wiskunde achter statistische berekeningen niet bekend is bij een onderzoeker kan dit tot interpretatiefouten leiden. Wanneer dit een complex netwerk was geweest waarin de structuur moeilijk te onderscheiden was, dan was het goed mogelijk geweest dat ik meteen een historische verklaring zou zoeken voor de populariteit van *node Z*, terwijl een wiskundige verklaring meer op zijn plek is.

Het is aanlokkelijk om statistische methodes toe te passen op complexe netwerkvisualisatie, omdat dit orde in de chaos schept. Er moet wel rekening gehouden worden met twee potentiële problemen. De eerste schuilt in de onvolledigheid van de data die we hebben, waardoor de visualisatie geen goede representatie van de historische werkelijkheid is. De onvolledigheid is voor de wiskundige beschrijving van een volledig netwerk minder relevant, maar voor het uitlichten van individuele datapunten wel, omdat daar vaak een foute uitkomst uit komt. Het tweede gevaar komt om de hoek kijken wanneer er een historische vertaling wordt gegeven van wat een statistische methode doet in de plaats van wiskundig begrijpen wat de methode doet. Wanneer wiskundige methodes worden gebruikt, moet zowel methodologische kritiek als bronkritiek worden toegepast om tot betekenisvolle conclusies te komen.

[63] De *betweenness centrality scores* van *nodes* Z, X, Y, en de rest uit figuur 9 zijn respectievelijk 190, 78, 24 en 0.

Case study

Data-schaarste en de belangrijkste knooppuntenberekening

We hebben gezien dat het gevaarlijk is om individuele datapunten uit te lichten met statistiek. Dit komt doordat er veel manuscripten verloren zijn gegaan. Om dit te laten zien, doe ik een *case study* met de dataset van De Coninck en de statistische operatie *betweenness centrality.* Hierbij is de centrale vraag: hoe groot is de kans dat de vier belangrijkste knoopuntpreken dezelfde zijn in de kleinere subset als in totale dataset?

In deze *case study* bevraag ik het prekennetwerk van de manuscripten die preek 202 in zich dragen. Hierbij inventariseer ik welke vier preken als belangrijkste knooppunten uit de bus komen wanneer er data wordt verwijderd. Om dit te doen gebruik ik een aangepaste versie van de experimenten die De Valeriola gebruikte.[64] Deze methodes heb ik toegespitst op De Coninck's dataset. Deze dataset is vrij groot voor filologen en vergelijkbaar met wat Riva's dataset was.

Hiermee doorloop ik de stappen, datakritiek, vraagstelling en methodekritiek, die zijn besproken in deze scriptie. Daarmee dient deze *case study* een dubbel doel: ten eerste om de onbetrouwbaarheid van *betweenness centrality* te illustreren als de onderzoeksvraag over individuele preken gaat, en ten tweede om een voorbeeld te geven van de praktische uitvoering van genoemde methodologische obstakels van netwerkanalysetoepassingen in de manuscriptstudies.

Data

Zoals gezegd maakt deze *case study* gebruik van de dataset die is opgesteld door Luc de Coninck. Deze is in detail beschreven in hoofdstuk *De valkuilen van manuscriptmetadata*. Uit die beschrijving zijn een paar belangrijke dingen te destilleren. Allereerst is de dataset niet 100% betrouwbaar en niet volledig. We weten simpelweg niet of het een goede representatie van de werkelijkheid is. Er zitten 2643 manuscripten die authentieke Augustinus' preken bevatten in deze database. Dat lijkt misschien veel, maar is waarschijnlijk een fractie van het totaal dat ooit geproduceerd is. Dat maakt deze dataset op het niveau van manuscriptobjecten onvolledig.

Ten tweede is de identificatie van de preken niet altijd accuraat. Verschillende auteurs hebben de manuscripten beschrijven en zijn het soms niet eens over welke preken erin staan. In deze analyse zal alleen worden gekeken naar de preken waarover men het eens is, om een zo accuraat mogelijke benadering te krijgen. Echter, er zijn ook manuscripten die maar door één auteur zijn beschreven. Daarbij is het niet duidelijk of de preken die benoemd worden ook daadwerkelijk in het manuscript staan. Wanneer er twee onderzoekers hetzelfde manuscript beschrijven, is de kans groter dat de preken waar ze het over eens zijn daadwerkelijk in het manuscript staan. Er zijn 1075 van de 2643 manuscripten die door een auteur zijn beschreven. Hierbij zal dus een benadering van de hoeveelheid mogelijke foute beschrijvingen moeten worden gevonden.

Ten derde is de beschrijving van een preeknaam niet alleen de naam van de preek zelf, maar soms ook uit welke paragrafen die bestaat. Het zou kunnen zijn dat preek 202 in twee delen in het manuscript staat en dat het twee keer in de inhoudsbeschrijving staat. Bijvoorbeeld AU s 202,1-2 en later AU s 202,3-5. Dit zullen twee verschillende *nodes* worden in de berekening. Wanneer we de paragraafaanduiding weg zouden halen, zal het voor dat manuscript dubbele *edges* creëren, omdat het er twee keer in staat.

[64] Valeriola, 'Can historians trust centrality?', 100–105.

Bij elkaar is deze dataset onvolledig, inaccuraat en niet direct met een netwerkperspectief te verenigen. Het is een incomplete dataset door erosie van manuscriptaantallen door de tand des tijds. Daarnaast wordt er door verschillende auteurs gedebatteerd over sommige preken en kunnen er onbenoemde, betwiste preken zitten in de manuscripten die door een auteur zijn beschreven. Als laatste zal er een keuze moeten worden gemaakt over het omgaan met de partiële preken. De gevolgen van deze onzekerheden kunnen worden gesimuleerd en getoetst. We toetsen hiermee of we dezelfde resultaten bij de subset krijgen als bij de totale dataset. Dit is een simulatie van de totale populatie manuscripten t.o.v. de dataset die we nog over hebben.

Netwerkperspectiefmogelijkheden

De vraag naar de validiteit van resultaten is een methodologische vraag, maar het perspectief dat gebruikt wordt ligt in het verlengde van het onderzoek dat Riva uitvoerde. Hierbij ging het om de transmissie van teksten en onder andere welke teksten er zich tussen clusters bevonden. In deze context fungeren de teksten als *nodes* en de manuscripten als *edges*. Wanneer teksten met elkaar samen worden aangetroffen in een manuscript, zullen zij één *edge* met elkaar delen. Daarbij zullen dichte clusters gevormd worden door grotere manuscripten, omdat elke extra tekst met alle andere teksten in een manuscript via een *edge* wordt verbonden. Een manuscript met vier teksten zal weergegeven wordt door vier *nodes* die samen zes *edges* delen, waar een manuscript met acht teksten, acht *nodes* en samen 28 *edges* krijgen.[65] Daarnaast zullen teksten die vaker samen voorkomen ook dichtere verbindingen krijgen. Wanneer we dit vertalen naar De Coninck's dataset, zijn de preken de *nodes* en de manuscripten de *edges.* De benadering van preektransmissie op deze manier maakt de weg vrij om netwerkvisualisaties en netwerkanalyse te gebruiken.

De beschikbare data maken het mogelijk om een kwantitatieve vraag te stellen. In deze *case study* zijn we benieuwd naar de consistentie van de belangrijkste knooppunten wanneer er data verwijderd worden. De statistische methode *betweenness centrality* kan hier antwoord op geven. Dat betekent dat statistische netwerkanalyse gebruikt gaat worden om de vraag te beantwoorden. Hierbij gaan we van een extreem voorbeeld uit, waarbij de onbetwiste preken worden aangehouden en de verliesratio van Kestemont als leidraad geldt. Daarnaast, is het prettig om visueel overzicht te hebben over wat er ongeveer gebeurt en daar kan een netwerkvisualisatie voor gebruikt worden.

Methode

Om de consistentie van de *betweenness centrality* in individuele punten te onderzoeken doen we alsof De Coninck's dataset de daadwerkelijke totale hoeveelheid manuscripten is met authentieke Augustinus manuscripten erin. In deze dataset zitten betwiste preken die eruit gefilterd worden om een zo goed mogelijke representatie van de 'historische werkelijkheid' te simuleren. De manuscripten in De Coninck's dataset fungeren dus als de volledige scala aan manuscripten die geproduceerd zijn in de middeleeuwen.

Om tot de overgeleverde hoeveelheid manuscripten te komen, dient er een subset te worden gemaakt van het totaal om de verloren manuscripten te simuleren. Naar verwachting is er voor allerhande manuscripten in het Heilige Roomse Rijk ongeveer 7% de totale hoeveelheid manuscripten uit de middeleeuwen aan ons overgeleverd.[66] Kestemont et. al. komt tot een gemiddelde van 9% door kwantitatieve methodes om aantallen uitgestorven diersoorten te

[65] $\text{Aantal edges} = \text{aantal teksten in een manuscript x (aantal teksten in een manuscript -1)2}$

[66] Mike Kestemont e.a., 'Forgotten books: The application of unseen species models to the survival of culture', *Science* 375 (2022) 765–769, aldaar 765 en 769.

benaderen en toe te passen op West-Europese manuscripten die geschreven zijn in de volkstaal. Het is ingewikkeld om een precies percentage op te drukken, zeker omdat preken een lagere kans hadden om te overleven vergeleken met literaire manuscripten.[67] Echter, het kiezen van een percentage is noodzakelijk voor de berekeningen en zal dus, semi-arbitrair, op het gemiddelde van 7% en 9% worden gezet. Uit programmatische overwegingen is het gemiddelde van beide genomen: acht procent. De selectieprocedure volgt de methode van Valeriola door willekeurig een manuscript te kiezen en die op te nemen in de subset tot acht procent van het geheel gehaald is.[68] Er wordt dus willekeurig van 2643 De Coninck's manuscriptbeschrijvingen 211 uitgekozen. Dit is de subset van manuscripten die zogenaamd aan ons zijn overgeleverd.

In deze subset zijn alleen Augustinus' preken beschreven en de vraagstelling gaat daarom ook alleen over het netwerk van authentieke Augustinus' preken. Het gaat om de vraag welke preken met elkaar samen reisden. Hierbij wordt er uitgegaan van het idee van een preek in zijn volledigheid, en maken we geen onderscheid tussen de verschillende paragrafen van een preek als losse ideeën.[69] Door deze aanname, werkt deze *case study* met manuscripten die ideeën van preken in zich dragen.

De data-technische consequentie hiervan, houdt in dat *nodes* volledige preken zijn en dat de uittreksels en aantekeningen verwijderd worden. Daarbij komt het voor dat preken twee keer voorkomen in een manuscript, of wel in hun geheel of in delen. Aangezien het hier gaat over een manuscript dat ideeën van preken meedraagt, wordt die gesimplificeerd naar alle unieke preken. De excerpten worden vertaald naar volledige preken, AU s 201,1-3 wordt AU s 201, en dubbele preekaanduidingen in een manuscript worden er een. Dus een manuscript heeft alleen maar unieke hele preken in deze analyse.

Vervolgens moeten de betwiste preekbeschrijvingen worden gesimuleerd. Auteurs die manuscripten beschrijven zijn het niet altijd met elkaar eens, maar dit is niet te zien bij de manuscripten die maar door een persoon worden beschreven. 40,7% van alle manuscripten in de dataset is beschreven door één auteur. Gebaseerd op Valeriola's aanpak, wordt er gekeken naar hoeveel manuscripten, beschreven door meerdere auteurs, een betwiste preek hebben.[70] Om vervolgens te kijken wat het percentage betwiste preken is in de manuscriptbeschrijvingen zonder consensus. Op deze manier kan het percentage betwiste preekbeschrijvingen in kaart worden gebracht op meerdere niveaus.

Om dat te simuleren in de subset, wordt de overige 59,3% manuscriptbeschrijvingen geconstateerd dat 11% (166/1568 manuscripten) één of meerdere betwiste preken bevat. In deze manuscripten blijkt dat de kans op zo'n preek 1/33 (gemiddelde = 0.03, standaarddeviatie = 0.14). Om de enkel-auteurs manuscriptbeschrijvingen dezelfde percentage foutmarge toe te bedelen worden de bovengenoemde percentages willekeurig verwijderd. Eerst, wordt er 11% van de manuscripten willekeurig geselecteerd en vervolgens wordt er één op de 33 keer, een preek verwijderd uit de beschrijving. Op die manier wordt de inaccurate data gesimuleerd in de subset.

Deze subset is een simulatie van de manuscripten die we overhebben. Het is een vergelijking van de soort data dat De Coninck's dataset is, t.o.v. de historische gehele dataset. Van deze subset

[67] Peter Marshall, 'Julia Crick and Alexandra Walsham, eds. The Uses of Script and Print, 1300–1700. New York: Cambridge University Press, 2004. Pp. xiv+298. $70.00 (cloth). ISBN 0-521-81063-9.', *Journal of British Studies* 44 (2005) 637–638, aldaar 638.
[68] Valeriola, 'Can historians trust centrality?', 102.
[69] Riva, 'Network Analysis of Medieval Manuscript Transmission', 36.
[70] Valeriola, 'Can historians trust centrality?', 103–104.

worden alle manuscripten met preek AU s 202 erin geselecteerd en wordt er een *node* en *edge list* van gemaakt door voor elk manuscript met dezelfde preek erin, een *edge* te maken. De *nodes* zijn de preken. Op deze manier is er een netwerkdataset gemaakt van de metadataset. Als dat gemaakt is, kan worden uitgerekend met *betweenness centrality* welke vier preken de hoogste waarde hebben. Dit proces wordt vervolgens duizend keer herhaald om de consistentie van de uitkomsten te achterhalen.[71] Hierbij wordt de dataset met alle manuscripten die se-202 bevatten vergeleken met een dataset waarvan 8% is overgebleven en sommige preken fout zijn geïdentificeerd. Hiermee kunnen we een inschatting maken van de kans dat de specifieke preken worden uitgelicht als belangrijkste knooppunten in het netwerk van preken die samen met preek 202 reizen.

Resultaten en discussie

Om wat intuïtie te krijgen voor wat er achter de berekeningen gebeurt, is er een visualisatie gemaakt van de originele dataset en drie willekeurige subsets na het doorlopen van het bovengenoemde proces van uitdunning. Hierbij is de *betweenness centrality* zichtbaar in groen en zijn de vier preken met de hoogste waarde van namen voorzien. De data, scripts voor de calculaties en de grafieken zijn te vinden op Github.[72]

Betwn. Centr. Scr.	Totale dataset	Subset 1	Subset 2	Subset 3
1ste	AU s 202	AU s 202	AU s 202	AU s 202
2de	AU s 200	AU s 369	AU s 201	AU s 194
3de	AU s 369	AU s 199	AU s 200	AU s 192
4de	AU s 201	AU s 8	AU s 194	AU s 201

Tabel 1: Vier hoogste betweenness centrality knopen van de totale De Coninck's dataset en drie willekeurige subsets.

In de drie verschillende visualisatie uit figuur 10 en tabel 1 zijn drie dingen opvallend. Ze hebben net als de totale dataset een groot cluster en allemaal kleine clusters daaraan vast. Verder is AU s 202 als verbindende factor. Daarnaast lijken de drie andere 'belangrijkste verbinders' tussen de clusters lichtelijk te variëren. Dit laatste is nog beter te zien in grafiek 3 op de pagina 29. Hier staan alle preken, die als belangrijkste vier verbinders werden berekend in de duizend verschillende subsets die willekeurig geselecteerd werden. Elke preek die in de top vier van belangrijkste verbinders is gekomen in een van de subsets staat in de grafiek. Hierbij is gekleurd op welke 'plaats' ze terecht zijn gekomen. De meeste preken hebben dus een van de vier kleuren.

Hierin is te zien dat er een verscheidenheid aan preken als belangrijkste verbinders naar boven komen in de verschillende berekeningen. Sommige meer dan andere, maar in zijn totaliteit een flinke verscheidenheid. De dataset die historici in handen krijgen, is één van deze duizend berekeningen en zou technisch gezien een *outlier* kunnen zijn. Maar dat een pessimistische insteek. Wanneer we kijken wat de gemiddelde kans is dat de resultaten gelijk zijn, kunnen we afleiden of het

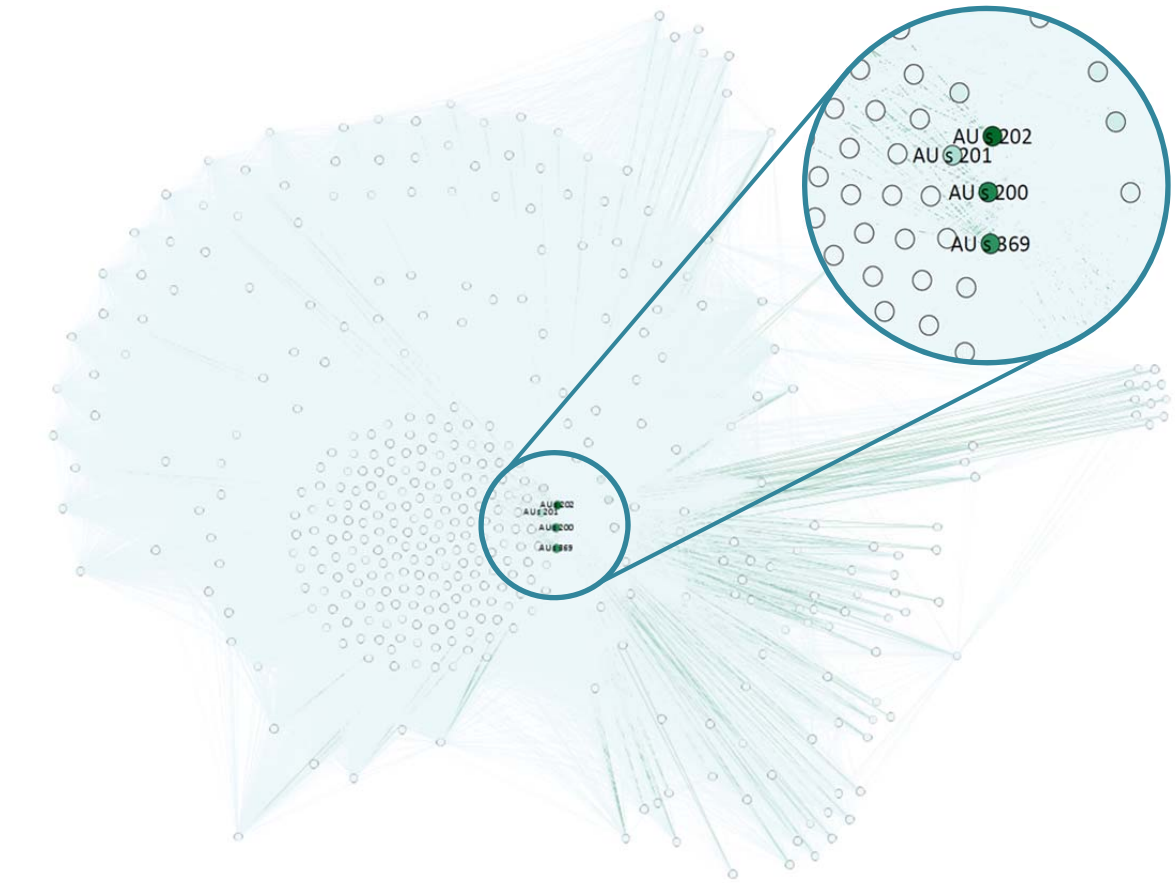
Betwn. Centr. Totale set	Kans gelijke positie subset
AU s 202	66,40%
AU s 200	12,30%
AU s 369	5,40%
AU s 201	13,10%

Tabel 2: Kans dat de preek van de totale set op dezelfde koppelingsplaats terecht komt in de subset.

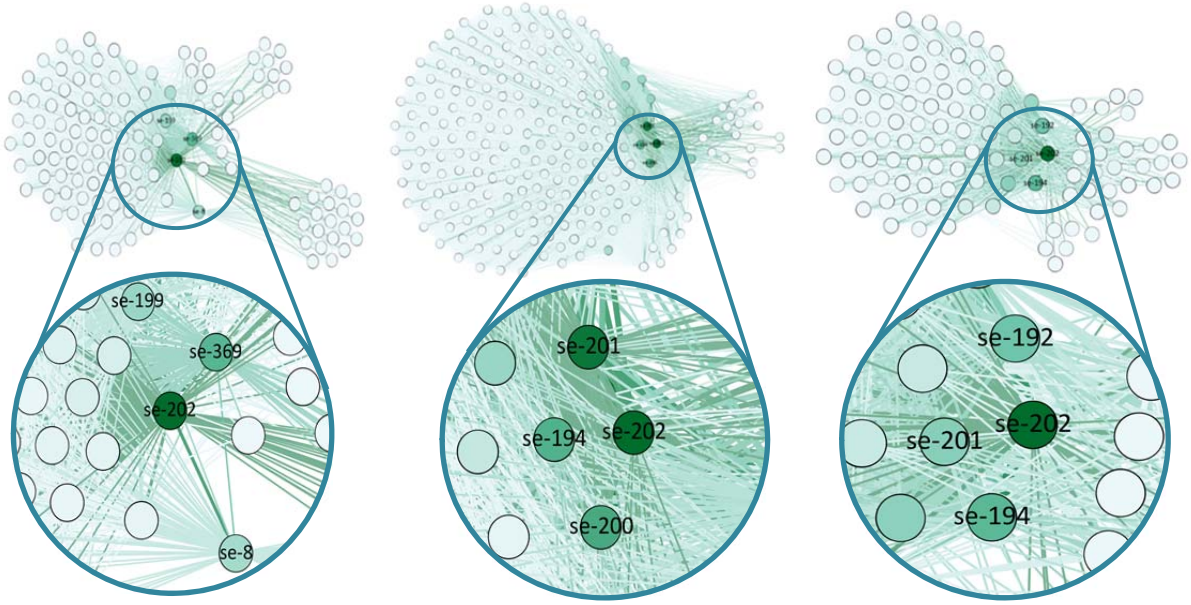
uitlichten van individuele preken met *betweenness centrality* een betrouwbare aanpak is.

Het is alleen niet verrassend dat AU s 202 als consistent wordt gezien in deze berekening. Dat is niet inherent aan preek 202, maar aan de manier waarop de dataset is gemaakt. Het bevat uitsluitend manuscripten die preek 202 bevatten en een cluster wordt gevormd door de inhoud van een manuscript. De verbindende positie bij een dergelijke dataset is dan logischerwijs preek 202. Hierbij is dus een data-technische uitleg gepaster dan een historische uitleg. Daarmee is de *betweenness centrality* als extra visualisatie optie om vanuit daar analyse te beginnen, historisch gezien een minder betrouwbare optie.

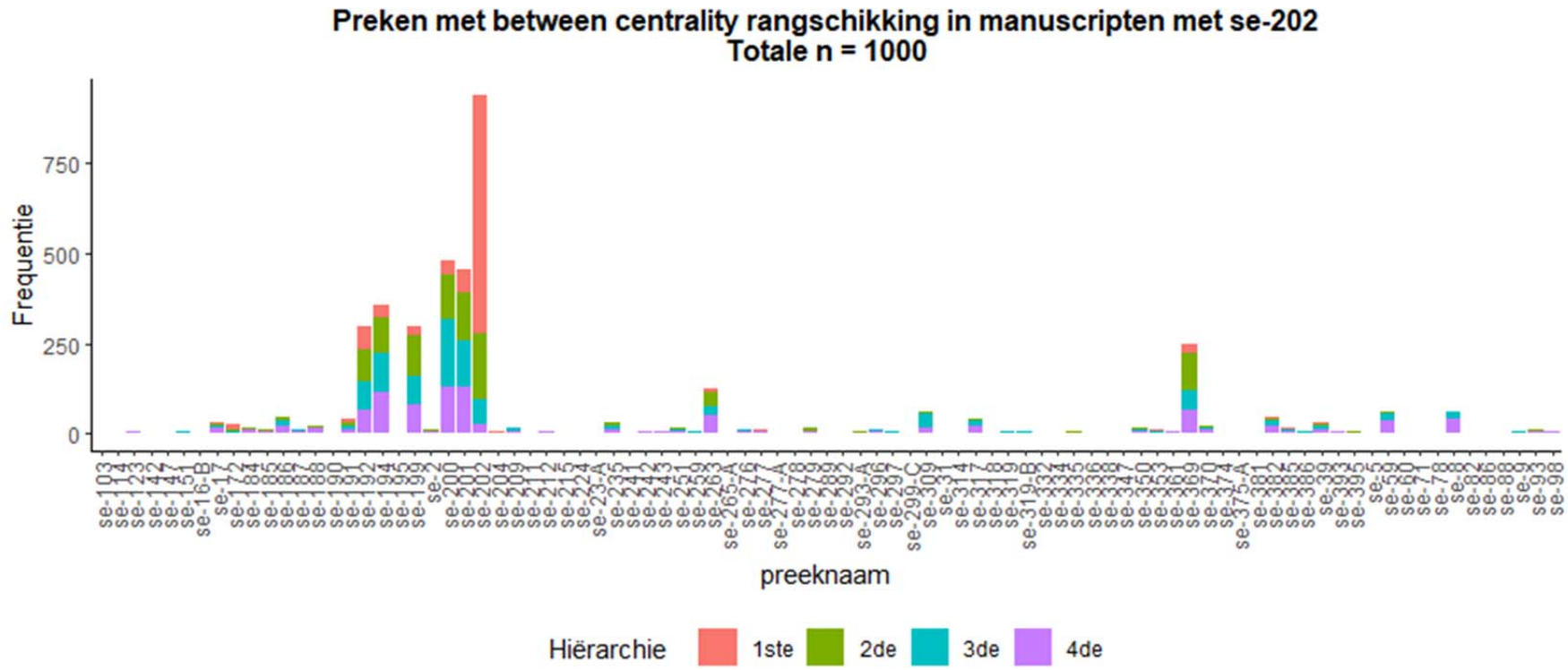
Concluderend, wanneer de onvolledigheid en onzekerheid van de datasets waarmee historici en filologen werken wordt nagebootst, zien we verschillen in de belangrijkste knooppunten tussen clusters. De belangrijkst *node* in de totale dataset kwam vaak als belangrijkste naar voren in de subsets. Dat is echter te wijten aan het feit dat de dataset was gekozen op preken die samen met deze ene preek meereisden. Logischerwijs komt deze *node* dan als verbindende factor naar boven. Voor de andere langrijke verbinders, ontstond er een verscheidenheid aan preken die naar voren komen en is de 'ware' belangrijkste verbinders niet gegarandeerd van een stabiele positie. Daarmee is het punt van Valeriola geverifieerd en kunnen historici *betweenness centrality* niet vertrouwen om individuele punten uit te lichten. Laten we netwerkanalyse gebruiken waarvoor het bedoeld is, namelijk het statistisch beschrijven van een netwerk en niet om individuele punten te onderzoeken.



Figuur 9: Netwerkvisualisatie van AU s 202 manuscripten uit de volledige dataset gekleurd op between centrality



Figuur 10: Netwerkvisualisatie van drie willekeurige subsets van De Coninck's dataset van manuscripten die preek 202 bevatten



Grafiek 3: Preken die als belangrijkste vier verbinders werden berekend in de subsets van manuscripten die preek 202 bevatten. De ordening van de vier is aangegeven in kleur.

Conclusie

Riva publiceerde een invloedrijk artikel dat de deuren opende voor manuscriptonderzoek met netwerkvisualisaties. Hij gebruikte hiervoor een traditie die vanuit de grafentheorie van de sociologen is ontleend om met een andere blik naar manuscripten te kijken. Hierbij heeft hij allerlei opties laten zien waarop de transmissiepatronen van manuscripten konden worden onderzocht met netwerkvisualisatie. In deze scriptie heb ik de obstakels behandeld die Riva onderbelicht houdt. Er is gekeken naar een andere manier om manuscripten te beschrijven zodat het voor netwerkanalyse geschikt is. Manuscripten worden hierbij gezien als een dataobject dat kan worden beschreven in verschillende lagen. Elke laag is een stukje dieper in het manuscript en heeft attributen die het beschrijven. Op deze manier wordt het manuscript een abstract dataobject.

Daarbij kwam boven water dat de data niet altijd kloppen en volledig zijn. Het blijkt dat er onder de beschrijvende data onzekerheden en hiaten zitten, waarmee rekening moet worden gehouden, zeker wanneer er grootschalige vragen over de verbindingen tussen manuscripten worden gesteld. Deze vragen vallen uiteen in drie perspectieven: het metaforische netwerkperspectief, de netwerkvisualisatie en de netwerkanalyse. Deze perspectieven hebben hun eigen soort vragen en kunnen op een andere manier antwoord hierop geven.

Als een vraag beantwoord kan worden door een netwerkvisualisatie moet er kritisch naar de mechaniek van de visualisatie worden gekeken. Een bepaalde vouwing van de netwerken kan namelijk tot verschillende interpretaties leiden. Dit komt vaak door de manier waarop de data zijn geconstrueerd. Daarom is een goed inzicht in de dataopbouw en achterliggende algoritmes noodzakelijk. Dit geldt evenzeer voor de statistische methodes die gebruikt kunnen worden in visualisaties. Een historische vertaling van wat een formule doet is niet afdoende om tot een accurate verklaring te komen, hiervoor is kennis over de wiskundige formule nodig. Als er geen verklaring op basis van de data en de wiskunde kan worden gegeven, kan er een historische interpretatie worden gegeven.

Het grote nadeel van statistiek gebruiken op historische data is dat de representativiteit van de data vaak in het geding is. Mediëvisten hebben te kampen met bronnen die maar in kleine getalen, en soms met inaccurate beschrijvingen, aan hen zijn overgeleverd. Dit zorgt ervoor dat de betrouwbaarheid van statistische methodes, die individuele punten uitlichten, afneemt. Echter, dit is een aanlokkelijke manier om resultaten uit de visualisatie te halen. Aangezien de ontdekking van individuele actoren een populaire vraag is binnen de Digital Humanities, moet er in de toekomst gewaakt worden voor het gebruik van netwerkvisualisaties om dat te doen. Het is aan te raden om een simulatie te maken van de potentiële toets die er gedaan gaat worden met een kleinere dataset om te kijken of de foutmarge niet te groot is, op basis van de kwaliteit van de dataset.

Dit betekent niet dat het hier ophoudt. De netwerkvisualisatie is voor sociologen een beginstadium bedoeld om intuïtie voor hun data te krijgen. Zij theoretiseren al een eeuw de mogelijkheden en onmogelijkheden van hun data en methoden. Hierbij hebben ze een scala aan uitgebreide statistische vernuftigheden bedacht om hun potentiële fouten te omzeilen en vervolgens een benadering te geven van hun missende data. Voor historici en filologen ligt er nog een wereld open om lering te trekken uit het werk van sociologen en hun methodes. In het laatste decennium heeft het startschot geklonken voor de manuscriptnetwerken en hun relaties. Daarin zijn er vele moeilijkheden om te overbruggen en onzekerheden om rekening mee te houden. Maar we zijn pas aan het begin van deze reis en kunnen nog veel leren van andere disciplines die al eerder een vergelijkbare reis hebben afgelegd.

Referenties

Alexanderson, Gerald L., 'About the cover: Euler and Königsberg's Bridges: A historical view', *Bulletin of the American Mathematical Society* 43 (2006) 567–574 <doi:10.1090/S0273-0979-06-01130-X>.

Andersen, Eva, e.a., *Digital History and Hermeneutics: Between Theory and Practice*. Studies in Digital History and Hermeneutics (München 2022) <doi:10.1515/9783110723991>.

Andrews, Tara L., en Caroline Macé ed., *Analysis of ancient and medieval texts and manuscripts: digital approaches*. Lectio: studies in the transmission of texts & ideas 1 (Turnhout 2014).

Bishop, Richard W., Johan Leemans en Hajnalka Tamas, *Preaching after Easter: Mid-Pentecost, Ascension, and Pentecost in Late Antiquity* (BRILL 2016).

Blobel, Mathias, Investigating Genre through Network Analysis of the Electronic Manuscript Catalogue (Master Thesis, University of Iceland, Reykjavík 2015).

Blondé, Bruno, e.a., *Trend en toeval: inleiding tot de kwantitatieve methoden voor historici* (Leuven 2012).

Bradley, Adam James, e.a., 'Visualization and the Digital Humanities:', *IEEE Computer Graphics and Applications* 38 (2018) 26–38 <doi:10.1109/MCG.2018.2878900>.

Burt, Ronald S., 'The Network Structure Of Social Capital', *Research in Organizational Behavior* 22 (2000) 345–423 <doi:10.1016/S0191-3085(00)22009-1>.

Chao, Anne S., Zhandong Liu en Qiwei Li, 'Network of Words: A Co-Occurrence Analysis of Nation-Building Terms in the Writings of Liang Qichao and Chen Duxiu', *Journal of Historical Network Research* 5 (2021) <doi:10.25517/jhnr.v5i1.122>.

Cheriet, Mohamed, Reza Farrahi Moghaddam en Rachid Hedjam, 'Visual language processing (VLP) of ancient manuscripts: Converting collections to windows on the past' (Conferentiepaper, *2013 7th IEEE GCC Conference and Exhibition*) 407–412 <doi:10.1109/IEEEGCC.2013.6705813>.

Cherven, Ken, *Mastering Gephi network visualization: produce advanced network graphs in Gephi and gain valuable insights into your network datasets* (Birmingham 2015).

Chick, Joe, 'The impact of missing data on network centrality measures' (University of Warwick 2021) 1 - 5.

Crymble, Adam, *Technology and the historian transformations in the digital age*. Topics in the digital humanities 2021:1 (Urbana 2021).

Da Rold, Orietta, en Elaine Treharne ed., *The Cambridge Companion to Medieval British Manuscripts*. Cambridge Companions to Literature (Cambridge 2020) <doi:10.1017/9781316182659>.

De Coninck, Luc, 'Sermones file 2017' (KU Leuven 2017).

Dupont, Anthony ed., *Preaching in the Patristic Era: sermons, preachers, and audiences in the Latin West*. A new history of the sermon 6 (Leiden en Boston 2018).

Fickers, Andreas, Juliane Tatarinov en Tim van der Heijden, 'Digital history and hermeneutics–between theory and practice: An introduction', in: *Digital History and Hermeneutics:*

Between Theory and Practice. Studies in Digital History and Hermeneutics (1ste druk; München 2022) 1–24.

Field, Andy P, Jeremy Miles en Zoo Field, *Discovering statistics using R* (London 2012).

Fiscarelli, Antonio Maria, 'Social network analysis for digital humanities', in: Andreas Fickers en Juliane Tatarinov eds., *Digital History and Hermeneutics: Between Theory and Practice* (München 2022) 23–42 <doi:10.1515/9783110723991-002>.

Golbeck, Jennifer, 'Analyzing networks', in: *Introduction to Social Media Investigation* (Elsevier 2015) 221–235 <doi:10.1016/B978-0-12-801656-5.00021-4>.

Grana, Costantino, Daniele Borghesani en Rita Cucchiara, 'Automatic segmentation of digitalized historical manuscripts', *Multimedia Tools and Applications* 55 (2011) 483–506 <doi:10.1007/s11042-010-0561-8>.

Grandjean, Martin, 'Introduction to Social Network Analysis: Basics and Historical Specificities' (Zenodo 2021) 1–22 <doi:10.5281/ZENODO.5083036>.

Hammond, Matthew, 'From Digital Prosopography to Social Network Analysis: Medieval People in the Twenty-First Century', *Medieval People* 36 (2021) 1–10.

---, 'From Digital Prosopography to Social Network Analysis: Medieval People in the Twenty-First Century', *Medieval People* 36 (2021).

Hering, Katharina, 'Technology and the Historian: Transformations in the Digital Age by Adam Crymble (review)', *Information & Culture* 57 (2022) 222–224.

Hoffman, Mark, 'Centrality', in: *Methods for Network Analysis* (2021).

Jacomy, Mathieu, e.a., 'ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software', *PLOS ONE* 9 (2014) 1–11 <doi:10.1371/journal.pone.0098679>.

Julien, Octave, 'Construction et composition des recueils français du XVe siècle: apports de la codicologie quantitative', *Babel. Littératures plurielles* (2007) 13–30 <doi:10.4000/babel.686>.

---, 'Délier, lire et relier. L'utilisation de l'analyse réseau pour construire une typologie de recueils manuscrits de la fin du Moyen Âge', *Hypothèses* 19 (2016) 211–224 <doi:10.3917/hyp.151.0211>.

Kapitan, Katarzyna Anna, 'Perspectives on Digital Catalogs and Textual Networks of Old Norse Literature', *Manuscript Studies: A Journal of the Schoenberg Institute for Manuscript Studies* 6 (2021) 74–97 <doi:10.1353/mns.2021.0002>.

---, Rowbotham Timothy en Tarrin Jon Wills, 'Visualising genre relationships in Icelandic manuscripts' (Göteborg 2017) 59–62.

Kenna, Ralph, Máirín MacCarron en Pádraig MacCarron eds., *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. Understanding Complex Systems (Cham 2017) <doi:10.1007/978-3-319-39445-9>.

Kerschbaumer, Florian, e.a. ed., *The Power of Networks: Prospects of Historical Network Research* (1ste druk; Routledge 2020).

Kestemont, Mike, e.a., 'Forgotten books: The application of unseen species models to the survival of culture', *Science* 375 (2022) 765–769 <doi:10.1126/science.abl7655>.

Kienle, Miriam, 'Between Nodes and Edges: Possibilities and Limits of Network Analysis in Art History', *Artl@s Bulletin* 6 (2017) <https://docs.lib.purdue.edu/artlas/vol6/iss3/1>.

Kluge, Mathias, *Handschriften des Mittelalters: Grundwissen Kodikologie und Paläographie* (3e ed., Ostfildern 2019).

Kwakkel, Erik, *Books before print*. Medieval media cultures (Leeds 2018).

---, en Rodney M Thomson, *The European book in the twelfth century* (2018).

Mac Carron, P., en R. Kenna, 'Network analysis of the Íslendinga sögur – the Sagas of Icelanders', *The European Physical Journal B* 86 (2013) 407 <doi:10.1140/epjb/e2013-40583-3>.

Maniaci, Marilena ed., *Trends in Statistical Codicology* (De Gruyter 2021).

Marshall, Peter, 'Julia Crick and Alexandra Walsham, eds. The Uses of Script and Print, 1300–1700. New York: Cambridge University Press, 2004. Pp. xiv+298. $70.00 (cloth). ISBN 0-521-81063-9.', *Journal of British Studies* 44 (2005) 637–638 <doi:10.1086/432211>.

McGee, F., e.a., 'The State of the Art in Multilayer Network Visualization', *Computer Graphics Forum* 38 (2019) 125–149 <doi:10.1111/cgf.13610>.

Newman, M. E. J., *Networks* (2e ed.; Oxford en New York 2018).

Nooy, Wouter de, Andrej Mrvar en Vladimir Batagelj, *Exploratory social network analysis with Pajek*. Structural analysis in the social sciences 46 (2e [rev.] ed.; Cambridge en New York 2018).

Omayio, Enock Osoro, Sreedevi Indu en Jeebananda Panda, 'Historical manuscript dating: traditional and current trends', *Multimedia Tools and Applications* (2022) 1–30 <doi:10.1007/s11042-022-12927-8>.

Pierazzo, Elena, *Digital scholarly editing: theories, models and methods* (London en New York 2016).

Reynolds, L. D., N. G. Wilson en Nigel Guy Wilson, *Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature* (OUP Oxford 2013).

Riva, Gustavo Fernandez, 'Network Analysis of Medieval Manuscript Transmission', *Journal of Historical Network Research* 3 (2019) 30–49 <doi:10.25517/jhnr.v3i1.61>.

Rosenthal, Joel Thomas eds., *Understanding medieval primary sources: using historical sources to discover medieval Europe*. Routledge guides to using historical sources (London en New York 2012).

Schober, Regina, 'America as Network: Notions of Interconnectedness in American Transcendentalism and Pragmatism', *Amerikastudien / American Studies* 60 (2015) 97–119.

Stutzmann, Dominique, ‘Réseaux et politiques documentaires de l’Ordre cistercien aux XIIe et XIIIe siècles: des abbayes et des textes’, in: *6e rencontre Res-Hist Réseaux bipartis en histoire* (Aix-en-Provence 2021) <https://hal.archives-ouvertes.fr/hal-03503301>

---, ‘TEI, Github, Zenodo, IIIF, Heurist, Docker: Looking back on Data Management, Tools, and Processes in the Funded Projects Oriflamms, Ecmen, Fama, Himanis, Horae, and Home’, (Ongepubliceerd Conferentiepaper, International Medieval Congress, Life And Afterlife Of Digital Projects In Medieval Manuscript Studies II, Leeds 2022).

‘The Data Visualisation Catalogue’ <https://datavizcatalogue.com/>.

Tolsma, Jochem, ‘Social Networks’, *Jochem Tolsma* <https://www.jochemtolsma.nl/courses/>.

Valeriola, Sébastien de, ‘Can historians trust centrality? Historical network analysis and centrality metrics robustness’, *Journal of Historical Network Research* 6 (2021) 85–125 <doi:10.25517/jhnr.v6i1.105>.

Venturini, Tommaso, Mathieu Jacomy en Pablo Jensen, ‘What do we see when we look at networks: Visual network analysis, relational ambiguity, and force-directed layouts’, *Big Data & Society* 8 (2021) 1–18 <doi:10.1177/20539517211018488>.

Wetherell, Charles, ‘Historical Social Network Analysis’, *International Review of Social History* 43 (1998) 125–144.

Wilmart, André, ‘Easter Sermons Of St Augustine: General Evidence’, *The Journal of Theological Studies* 28 (1927) 113–144.

Yau, Nathan, ‘Network Visualization | FlowingData’ <https://flowingdata.com/category/visualization/network-visualization/>.