

Eduskunnan täysistuntojen puheenvuorojen henkilömainintoihin perustuvien verkostoiden analyysi

Henna Poikkimäki

Perustieteiden korkeakoulu

Diplomityö, joka on jätetty opinnäytteenä tarkastettavaksi
diplomi-insinöörin tutkintoa varten Espoossa 27.2.2023.

Työn valvoja

Prof. Eero Hyvönen

Työn ohjaaja

DI Petri Leskinen

Copyright © 2023 Henna Poikkimäki

Tekijä Henna Poikkimäki

Työn nimi Eduskunnan täysistuntojen puheenvuorojen henkilömainintoihin perustuvien verkostoiden analyysi

Koulutusohjelma Life Science Technologies

Pääaine Complex Systems

Pääaineen koodi SCI3060

Työn valvoja Prof. Eero Hyvönen

Työn ohjaaja DI Petri Leskinen

Päivämäärä 27.2.2023

Sivumäärä 59

Kieli Suomi

Tiivistelmä

Parlamentaarisen datan avoimuus on tärkeää muun muassa päätöksenteon läpinäkyvyyden ja parlamentaariseen keskusteluun liittyvän tutkimuksen kannalta. Semanttinen parlamentti konsortiohanke kokosi yhteen kaikki täysistuntojen puheenvuorot ja täysistunnoissa puhuneet henkilöt aina eduskunnan perustamisesta 1907 saakka luomalla linkitetyn avoimen datan infrastruktuurin, Parlamenttisammon.

Dataa on rikastettu esimerkiksi tunnistamalla puheenvuoroista henkilömainintoja, joiden pohjalta tässä työssä luodaan kansanedustajien ja puolueiden välisiä verkostoja. Työssä esitellään ensin verkosto- ja viiteanalyysi sekä käydään läpi niiden välisiä yhteyksiä. Työssä muodostetaan viiteanalyysistä tuttuja viittaus- ja yhteisviittausverkostoja sekä bibliografiseen kytkentään perustuvia verkostoja, joita analysoidaan viite- ja verkostanalyysin menetelmin. Verkostojen avulla pyritään tunnistamaan parlamentaarisisessa keskusteluissa aktiivisia ja keskeisiä henkilöitä sekä tutkimaan, onko puheenvuoroissa tehtyjen henkilömainintojen taustalla toistuvia kaavoja.

Viittausverkostojen avulla saatiin selville, että usein mainittuja kansanedustajia ovat ministereinä toimineet henkilöt ja itse paljon henkilömainintoja puheenvuoroissaan tekevät kansanedustajat. Paljon viittauksia tekevistä henkilöistä puolestaan kukaan ei toiminut ministerinä ja suurin osa heistä oli opposition edustajia. Yhteisviittausverkoston avulla saatiin selville henkilöt, jotka on mainittu usein samojen kansanedustajien toimesta kun taas bibliografiseen kytkentään perustuvan verkoston avulla selvitettiin, ketkä kansanedustajista ovat maininneet usein samoja henkilöitä.

Suurin vaikuttava tekijä siihen, ketkä henkilöön viittaavat tai keitä henkilö mainitsee puheenvuoroissaan vaikuttaa olevan henkilön puolueen parlamentaarinen rooli, hallitus tai oppositio. Selkeitä kansanedustajien ryhmittymiä ei kuitenkaan löydetty henkilömainintojen perusteella tässä työssä käytetyillä menetelmillä. Otettaessa huomioon puheenvuorojen lauseet, joissa henkilömaininnat tapahtuvat, opposition ja hallituksen poliitikkojen jako muuttuu selkeämmäksi. Lopussa työssä pohditaan, voisiko henkilömainintojen perusteella löytää selkeämpiä kansanedustajien ryhmittymiä ottamalla esimerkiksi paremmin huomioon henkilömainintojen kontekstit.

Avainsanat verkostanalyysi, viiteanalyysi, eduskunta, puheenvuoro, viittausverkosto, mainintaverkosto

Author Henna Poikkimäki

Title Analyses of networks based on person name mentions extracted from parliamentary speeches

Degree programme Life Science Technologies

Major Complex Systems

Code of major SCI3060

Supervisor Prof. Eero Hyvönen

Advisor MSc Petri Leskinen

Date 27.2.2023

Number of pages 59

Language Finnish

Abstract

Open parliamentary data enables research related to parliamentary debates and makes political decision making more transparent. Semantic Parliament consortium project has collected, harmonized and published all parliamentary debate speeches given in Parliament of Finland since 1907, related people and groups by creating linked open data infrastructure ParliamentSampo.

The data has been enriched by recognizing and linking person name mentions from debate speeches. In this work, based on person name mentions in parliamentary debate speeches different networks between members of Parliament (MPs) and parties are constructed. Networks are then analyzed using methods from network analysis and bibliometrics. The purpose of this work is to find out who are the most central MPs in parliamentary debates, how members of different parties mention each other in their speeches and if there are any patterns in the person name mentions.

Analyses of the citation network reveal most active debaters as well as most mentioned MPs. Most mentioned people are often ministers or people who have mentioned a lot of other MPs. Co-citation network reveals who are often mentioned together with other MPs and bibliographical coupling networks tell who have often mentioned same people.

The parliamentary role of the parties of politicians seem to be the most significant factor that affects who mentions some particular MP and whom some particular MP mentions. However, no clear clusters of MPs are found based on person name mentions using the methods described in this work. Finally additional strategies for finding possible patterns in person name mentions are discussed.

Keywords parliamentary debates, network analysis, bibliometrics, citation network, mention network

Esipuhe

Kiitos professori Eero Hyvöselle ja ohjaajalleni Petri Leskiselä työni ohjauksesta. Kiitos myös kaikille Parlamenttisarmon parissa työskenneille ja erityisesti Minna Tamperille, jonka työ henkilömainintojen tunnistamisen parissa mahdollisti tämän diplomityön. Kiitos Aalto-yliopiston tietotekniikan laitokselle ja Suomen Akatemialle diplomityöni rahoituksesta.

Otaniemi, 27.2.2023

Henna Poikkimäki

Sisällys

Tiivistelmä	3
Abstract	4
Esipuhe	5
Sisällys	6
Lyhenteet	7
1 Johdanto	8
2 Verkostoanalyysi	10
2.1 Verkostot	10
2.2 Verkostoanalyysijä parlamentaariselle datalle	11
3 Viiteanalyysi	14
3.1 Viiteanalyysin menetelmät verkostoanalyysin näkökulmasta	14
3.2 Viittausanalyysi	17
3.3 Kirjoittajien yhteisviittausanalyysi	18
3.4 Bibliografinen kytkentä	19
3.5 Kontekstin ja ajan huomiointi viiteanalyysissä	20
3.6 Kirjoittajien viiteanalyysin ja poliitikkojen mainintoihin perustuvien verkostojen analyysien eroavaisuudet	20
4 Parlamenttisanjo	24
4.1 Semanttinen parlamentti -hanke	24
4.2 Henkilömainintojen tunnistamien eduskunnan puheenvuoroista	25
5 Mainintaverkostojen tutkimusmenetelmät	27
5.1 Verkostojen luominen	27
5.2 Verkostometriikat	30
5.3 Perinteiset yhteisviittausanalyysin menetelmät	31
5.4 Muita verkostoanalyysin menetelmiä	33
6 Tulokset	34
6.1 Puolueiden välinen viittausverkosto	34
6.2 Valiokuntien välinen viittausverkosto	41
6.3 Kansanedustajien välinen viittausverkosto	43
6.4 Kansanedustajien yhteisviittausanalyysi	48
6.5 Kansanedustajien välisiin bibliografisiin kytkentöihin pohjautuva analyysi	50
7 Johtopäätökset	53

Lyhenteet

Puolueiden lyhenteet

KD	Suomen Kristillisdemokraatit
Kesk.	Suomen Keskusta
KL	Korjausliike
Kok.	Kansallinen Kokoomus
Liik.	Liike Nyt
PS	Perussuomalaiset
RKP	Suomen ruotsalainen kansanpuolue
SDP	Suomen Sosialidemokraattinen Puolue
TL	Seitsemän tähden liike
Vas.	Vasemmistoliitto
Vihr.	Vihreä liitto
VKK	Valta kuuluu kansalle

Menetelmiin liittyviä lyhenteitä

FA	faktorianalyysi (<i>factor analysis</i>)
MDS	moniulotteinen skaalaus (<i>multidimensional scaling</i>)
NEL	nimettyjen entiteettien linkitys (<i>named entity linking</i>)
NER	nimettyjen entiteettien tunnistus (<i>named entity recognition</i>)
NLP	luonnollisen kielen käsittely (<i>natural language processing</i>)

1 Johdanto

Parlamentaarisen datan, kuten eduskunnassa pidettyjen puheenvuoroihin ja puheenvuoron pitäjiin, yleensä kansanedustajiin, liittyvän tiedon helppo saatavuus ja käytettävyys on tärkeää päätöksenteon läpinäkyvyyden kannalta. Aalto-yliopiston, Turun yliopiston ja Helsingin yliopiston yhteinen Semanttinen parlamentti (SEM-PARL) konsortiohanke kokosi yhteen ja julkaisi kaikki Suomen eduskunnassa pidetyt puheenvuorot sekä puheenvuoroihin ja niiden pitäjiin liittyvää oleellista tietoa. Samalla julkaistiin myös verkkoportaali, jonka fasettihaku mahdollistaa puheenvuorojen ja niiden pitäjiin helpon hakemisen sekä hakutulosten rajaamisen.

Hankkeessa Eduskunnalta saatua dataa on rikastettu paitsi luomalla linkityksiä ulkoisiin tietolähteisiin mutta myös luonnollisen kielen käsittelyn (*natural language processing, NLP*) menetelmien avulla. Vaalikausien 2015–2018 ja 2019–2022 puheenvuoroista on tunnistettu nimettyjä entiteettejä, kuten henkilöitä, organisaatioita, paikkoja, ajan ilmaisuja ja lakeja. Tunnistetut nimetyt entiteetit on linkitetty vastaaviin tietolähteisiin sekä sisäisesti Parlamenttisammon dataan että ulkoisiin tietolähteisiin. Tunnistettujen ja linkitettyjen nimettyjen entiteettien avulla puheenvuorojen ja puheenvuorojen pitäjiin välillä voidaan havaita yhteyksiä esimerkiksi tutkimalla ketkä kansanedustajista mainitsevat samankaltaisia asioita tai henkilöitä puheenvuoroissaan.

Tässä työssä muodostetaan ja analysoidaan erilaisia verkostoja puheenvuoroissa esiintyvien henkilömainintojen perusteella. Henkilömaininta puheenvuorossa voi olla merkki debatista, tai vastaavasti puheenvuoron pitäjä voi kannattaa mainitun henkilön ajatuksia tai ehdotuksia. Henkilömainintojen tarkastelu antaa tietoa kansanedustajien välisestä vuorovaikutuksesta eduskunnassa. Tutkimalla henkilömainintojen avulla muodostettuja verkostoja verkostanalyysin ja viiteanalyysin menetelmin voidaan löytää vastauksia muun muassa seuraaviin kysymyksiin. Millaisiin henkilöihin kohdistuu paljon mainintoja? Ketkä tekevät paljon henkilömainintoja puheenvuoroissaan? Ketkä ovat aktiivisia keskustelijoita eduskunnassa tai merkittäviä henkilöitä henkilömainintojen näkökulmasta? Voiko kansanedustajia luokitella ryhmiin heidän tekemiensä tai heihin kohdistuvien henkilömainintojen pohjalta ja millaisia ryhmiä tällöin muodostuu? Mainitsevatko miehet ja naiset suhteessa yhtä paljon henkilöitä puheenvuoroissaan vai onko sukupuolten välillä eroavaisuuksia? Tapahtuuko viittausverkostoissa suuria muutoksia ajan kuluessa? Lisäksi yhdistelmällä henkilömainintoja puheenvuorojen pitäjiin ja mainittujen henkilöiden ryhmiin, kuten puolueisiin, voidaan tehdä johtopäätöksiä ryhmien välisistä suhteista.

Seuraavassa luvussa esitellään verkostanalyysiä yleisesti ja annetaan esimerkkejä verkostanalyysin soveltamisesta parlamentaariselle datalle. Luvussa 3 esitellään viiteanalyysin menetelmiä erityisesti verkostanalyysin näkökulmasta, vertaillaan perinteisen viittausdatan ja tämän työn henkilömainintojen eroavaisuuksia sekä pohditaan, miten ne näkyvät valituissa menetelmissä ja tuloksissa. Luvussa 4 esitellään hieman tarkemmin työssä käytettyä dataa ja luvussa 5 käydään läpi kuinka erilai-

sia verkostoja voidaan muodostaa puheenvuoroissa esiintyvien henkilömainintojen pohjalta sekä tarkastellaan hieman tarkemmin käytettyjä menetelmiä. Luvussa 6 esitellään verkosto- ja viiteanalyysin tulokset ja lopuksi käydään läpi tehtyjä johtopäätöksiä sekä pohditaan, mitä lisäanalyysejä henkilömainintojen pohjalta voisi tehdä.

2 Verkostoanalyysi

Verkostoanalyysin menetelmät mahdollistavat monimutkaistenkin ilmiöiden ja systeemien mallintamisen ja tutkimisen. Tämän luvun alussa käydään läpi verkostoihin liittyviä käsitteitä sekä esitellään tämän työn kannalta oleellisia verkostotyyppejä. Luvun lopussa näytetään esimerkkejä parlamentaariselle datalle tehdyistä verkostoanalyyseistä sekä poliitikkojen mainintaverkostoista Twitter-keskusteluiden pohjalta.

2.1 Verkostot

Verkostotiede (*network science*) on kasvanut merkittäväksi tieteenalaksi 2000-luvulla [1]. Ympäriällämme on paljon monimutkaisia järjestelmiä (*complex systems*), joiden osia sekä osien välisiä vuorovaikutuksia voidaan mallintaa verkostoilla [2]. Tällaisia verkostoja ovat esimerkiksi proteiinien väliset vuorovaikutusverkostot, ihmisten väliset sosiaaliset verkostot, julkisen liikenteen verkostot ja laitteiden väliset vuorovaikutusverkostot.

Verkosto $G = (V, E)$ on matemaattinen malli, joka koostuu kahdesta joukosta: solmuista $V = \{v_1, v_2, \dots, v_N\}$ ja linkeistä $E = \{e_1, e_2, \dots, e_m\}$. Solmujen määrää verkostossa merkitään kirjaimella N ja m on vastaavasti solmujen välisten linkkien määrä. Solmut kuvaavat tutkittavan järjestelmän tai systeemin osia, kuten tämän työn tapauksessa esimerkiksi kansanedustajia tai puolueita. Jokainen linkki on solmu pari $e_k = (v_i, v_j)$, $v_i, v_j \in V$ ja kuvaa solmujen välistä yhteyttä tai vuorovaikutusta. Tässä työssä linkit perustuvat eduskunnassa pidettyjen puheenvuorojen henkilömainintoihin. Kahden solmun sanotaan olevan toistensa naapureita, jos niiden välillä on linkki. Visualisoinneissa solmut kuvataan usein ympyröinä tai neliöinä ja linkit niiden välisinä viivoina.

Suuntaamattomissa verkostoissa solmujen välillä on yhteys, jolla ei ole suuntaa eli $e_k = (v_i, v_j) = (v_j, v_i)$. Tällöin kahden solmun välillä oleva linkki vihjaa näillä kahdella solmulla olevan jotakin yhteistä. Esimerkiksi solmujen ollessa kansanedustajia, solmujen välinen linkki voi kertoa kansanedustajien maininneen saman henkilön pitämässään puheenvuoroissa. Suunnatuissa verkostoissa linkeillä on suunta, jolloin verkostossa voi olla linkki solmusta v_i solmuun v_j ja erikseen linkki solmusta v_j solmuun v_i . Tämä voi kuvata esimerkiksi informaation virtausta lähdesolmusta kohdesolmuun tai tilannetta, jossa lähdesolmu viittaa kohdesolmuun. Verkostojen visualisoinneissa suunta näytetään merkitsemällä linkki nuolella. Tässä työssä hyödynnetään sekä suunnattuja että suuntaamattomia verkostoja.

Verkostot voivat olla joko painotettuja tai painottomia. Painotetuissa verkostoissa jokaisella linkillä $e_k = (v_i, v_j)$ on paino w_{ij} , joka voi kuvata esimerkiksi solmujen välisen yhteyden vahvuutta tai solmujen välillä kulkevan informaation määrää. Painottomilta verkostoilta tällainen ominaisuus puuttuu kokonaan. Tässä työssä hyödynnetään painotettuja verkostoja. Painotetut verkostot voidaan ilmaista $N \times N$ neliömatriisin, yhteysmatriisin (*adjacency matrix*) A avulla. Yhteysmatriisin alkiot

määritellään yhtälön (1) mukaan.

$$A_{ij} = \begin{cases} 0 & \text{jos solmujen } i \text{ ja } j \text{ välillä ei ole linkkiä} \\ w_{ij} & \text{jos solmujen } i \text{ ja } j \text{ välillä on linkki} \end{cases} \quad (1)$$

Yhtälössä (1) w_{ij} vastaa solmujen välisen linkin painoa. Painottomien verkostojen tapauksessa yhteysmatriisissa $w_{ij} = 1$, mikä ilmaisee vain linkin olemassaolon. Suuntaamattomien verkostojen tapauksissa yhteysmatriisi on symmetrinen. Suunnatuissa verkostoissa jokainen alkio A_{ij} kuvaa linkkiä solmusta i solmuun j . Solmusta voi olla linkki myös solmuun itseensä, jolloin $A_{ii} \neq 0$.

Yhteysmatriisin käyttö verkoston esittämiseen toimii erityisen hyvin tiheille ja suhteellisen pienille verkostoille. Suurien verkostojen tapauksessa yhteysmatriisin koko kasvaa, mikä vaatii paljon muistia. Harvojen verkostojen tapauksessa yhteysmatriisi sisältää paljon nollia, jolloin verkosto on huomattavasti tehokkaampaa ilmaista yhteyslistana. Yhteyslistassa jokaiselle solmulle on listattu solmun naapurit. Tämän työn tapauksessa solmujen määrää rajoittaa kansanedustajien määrä, jolloin verkostojen koot pysyvät suhteellisen pieninä ja yhteysmatriisin koosta ei muodostu ongelmaa. Yhteysmatriisin käytössä etuna on, että yhteysmatriisia voidaan käyttää laskuissa ja monimuuttuja-analyysissä.

Suunnattujen ja suuntaamattomien, painotettujen ja painottomien verkostojen lisäksi on olemassa muitakin erilaisia verkostotyyppisiä, joilla on erikoisominaisuuksia. Yksi tässä työssä hyödynnettävistä verkostotyypeistä on painotettu kaksijakoinen verkosto (*bipartite network*). Kaksijakoisen verkoston solmut voidaan jakaa kahteen ryhmään siten, ettei samaan ryhmään kuuluvien solmujen välillä ole lainkaan linkkejä. Kaksijakoisessa verkostossa solmujen välillä voi siis olla linkki vain jos solmut kuuluvat eri ryhmiin.

Verkostoanalyysin menetelmät mahdollistavat muun muassa verkoston solmujen klusteroinnin ryhmiin sekä verkoston kannalta keskeisten solmujen tunnistamisen. Tässä työssä käytettyjen verkostojen muodostaminen sekä niiden analysointiin käytetyt menetelmät esitellään luvussa 5.

2.2 Verkostoanalyysijä parlamentaariselle datalle

Esimerkkejä verkostoanalyysin menetelmien soveltamisesta parlamentaariselle datalle löytyy jonkin verran. Erityisesti verkosto- ja sisällönanalyysin yhdistävän diskursiivverkostoanalyysin (*discourse network analysis*) menetelmiä on sovellettu parlamentaarisiiin puheenvuoroihin [3]. Esimerkiksi Bhattacharya muodosti verkoston saksalaisista kansanedustajista, jossa kansanedustajat yhdistyivät sitä suuremmalla painoarvolla mitä enemmän he olivat asioista samaa mieltä puheenvuorojen ja äänestyksiin liittyvien selityksien perusteella [4]. Tämän verkoston avulla tutkittiin, rakoilivatko eri puoluiden rivit kun käsiteltiin Kreikan kriisiin liittyviä kysymyksiä

vuosina 2010–2015.

Curran ym. ryhmittelivät Uuden-Seelannin parlamentin puheenvuoroja aiheen mukaan ja muodostivat sen perusteella kaksijakoisen verkoston aiheiden ja niistä puhuneiden edustajien välille [5]. Tästä verkostosta projektiona saatiin edustajien verkosto, jossa samoista asioista puhuneet henkilöt yhdistyivät sitä suuremmalla painoarvolla mitä useammin he olivat puhuneet samoista aiheista. Vastaavasti kaksijakoisesta verkostosta pystyttiin muodostamaan aiheiden verkosto, jossa aiheet yhdistyivät aina saman henkilön puhuessa molemmista. Verkostojen avulla tutkittiin esimerkiksi edustajien ryhmittymiä, aktiivisuutta sekä miten eri aiheiden suosio on muuttunut eri puolueiden keskuudessa vuosien kuluessa. Kaksijakoista verkostoa ja siitä luotuja projektioita tarkastellaan tarkemmin luvussa 5.1.

Tämän työn pohjana on tieto siitä, kuinka usein ja keitä muita poliitikkoja kansanedustajat mainitsevat puheenvuoroissaan. Tällöin kansanedustajista voidaan muodostaa viittausverkosto, jossa poliitikot ovat solmuja ja linkit kulkevat puheenvuoroja pitäneistä poliitikoista puheenvuoroissa mainittuihin poliitikkoihin. Tällaisesta parlamentaarisiin puheenvuoroihin perustuvasta viittausverkostoanalyysistä ei juurikaan löydy aiempaa tutkimusta. Sen sijaan esimerkiksi Hollannin kansanedustajien poliittista polarisoitumista on tutkittu Twitter-keskusteluiden pohjalta muodostettujen mainintaverkostojen avulla, jotka perustuvat kansanedustajien välisiin mainintoihin Twitterissä [6].

Twitter¹ on sosiaalisen median alusta, jonka kautta käyttäjät voivat julkaista lyhyitä mikroblogeja eli twiittejä, jakaa muiden käyttäjien sisältöä ja keskustella vastaamalla muille käyttäjille tai mainitsemalla heitä. Käyttäjien mainintaverkostojen lisäksi muita tutkittuja Twitter-keskusteluihin pohjautuvia verkostoja ovat uudelleentwiittaus-verkosto ja vastausverkosto, jotka perustuvat siihen miten käyttäjät jakavat muiden käyttäjien sisältöä tai vastaavat muille käyttäjille [7]. Parlamentaarisisissa puheenvuoroissa esiintyvät henkilömaininnat voivat olla esimerkiksi vastauksia tai kysymyksiä mainitulle henkilölle. Vaihtoehtoisesti mainitun henkilön esityksiä ja ideoita voidaan nostaa esiin haastaen tai kannattaen mainittua henkilöä. Tällaisiin henkilömainintoihin perustuvan verkoston voidaan ajatella olevan eräänlainen sekoitus aiemmin mainittuja Twitter-keskusteluihin pohjautuvia verkostoja.

Cherepnalkoski ja Mozetic rajasivat uudelleentwiittaus-verkoston solmuiksi Euroopan parlamentin jäsenet ja totesivat jäsenien ryhmittyvän enimmäkseen niihin ryhmiin, joihin he kuuluvat Euroopan parlamentissa käyttämällä verkostojen yhteistötunnistamismenetelmiä [8]. Solmujen eli Twitter-käyttäjien ryhmittymisen lisäksi erilaisista verkostoista on tunnistettu merkittäviä käyttäjiä laskemalla solmuille erilaisia sosiaalisen verkostoanalyysin (*social network analysis, SNA*) metriikoita [9]. On huomattu, että käyttäjämainintoihin perustuvien mainintaverkostojen avulla ei voida tunnistaa esimerkiksi poliittista polarisaatiota, sillä käyttäjän mainitsevat usein

¹<https://twitter.com>

toisia käyttäjiä ikään kuin haastaen kun taas samaa mieltä olevien sisältöä jaetaan uudelleentwiittaus-ominaisuuden avulla, jolloin verkosto näyttää polarisoituneelta [6].

3 Viiteanalyysi

Tässä luvussa käydään läpi, mitä on viiteanalyysi ja mikä on sen suhde verkostoanalyysiin. Aluksi esitellään viiteanalyysin perusteita, jonka jälkeen annetaan esimerkkejä viiteanalyysin menetelmiä käyttäneistä tutkimuksista. Lopuksi pohditaan viiteanalyysin ongelmia sekä miten poliitikkojen viittausverkostojen tutkiminen viiteanalyysin menetelmin eroaa perinteisen viittausdatan analysoinnista.

3.1 Viiteanalyysin menetelmät verkostoanalyysin näkökulmasta

Viiteanalyysi on yksi tapa tehdä bibliometristä eli tieteellisiin julkaisuihin liittyvää tutkimusta. Tieteellisiä julkaisuja voidaan tutkia esimerkiksi dokumentti-, kirjoittaja- tai julkaisijatasolla. Erilaisilla menetelmillä pyritään esimerkiksi selvittämään, mitkä ovat tärkeitä julkaisuja tietyllä alalla tai ketkä kirjoittajat erityisesti vaikuttavat tietyn aihealueen kirjallisuudessa. Julkaisuja tutkimalla voidaan myös tehdä johtopäätöksiä tietyn tieteenalan rakenteesta, kuten siitä mitä aihepiirejä alaan liittyy ja ketkä kirjoittajat liittyvät mihinkin aihepiiriin sekä miten tieteenala on kehittynyt ja tulee kehittymään jatkossa. Viiteanalyysi voidaan jakaa lähdeanalyysiin ja viittausanalyysiin. Lähdeanalyysissä tutkitaan kirjoittajan tai työn käyttämiä lähteitä, viittausanalyysissä puolestaan työhön tai kirjoittajaan kohdistuvia viittauksia. Lähdeanalyysin yksi tärkeimmistä käsitteistä on bibliografinen kytkentä, joka kuvaa kahden dokumentin tai kirjoittajan lähdeluetteloiden samankaltaisuutta. Toinen tässä työssä käytetty tärkeä menetelmä on viittausanalyysin pohjalta kehitetty yhteisviittausanalyysi, jossa tarkastellaan, miten kirjoittajiin on viitattu yhdessä muiden kirjoittajien töissä. [10]

Monet perinteiset viiteanalyysin menetelmät perustuvat erilaisten viittausten tai lähteiden perusteella muodostettujen matriisien monimuuttuja-analyysiin. Tässä työssä bibliometriikan menetelmistä hyödynnetään erityisesti kirjoittajatason viiteanalyysiä, jolloin matriisien alkioden katsotaan kuvaavan kansanedustajien välisiä henkilömainintoja. Yhtälö (2) kuvaa viittausmatriisin J alkiota j_{ij} , joka vastaa viittausta henkilöltä i henkilöön j [11]. Jos $j_{ij} = 0$, henkilö i ei viittaa henkilöön j . Muulloin j_{ij} on positiivinen kokonaisluku w_{ij} , joka voi ilmaista viittauksen olemassa oloa tai esimerkiksi viittausten määrää. Vaikka analyysiä tehdään kirjoittajatasolla, kirjoittajien väliset viittaukset lasketaan kuitenkin dokumenttitasolla eli kirjoittaja viittaa toiseen kirjoittajaan, kun hän viittaa omassa työssään toisen kirjoittajan työhön.

$$j_{ij} = \begin{cases} 0 & \text{jos henkilö } i \text{ ei viittaa henkilöön } j \\ w_{ij} & \text{jos henkilö } i \text{ viittaa vähintään kerran henkilöön } j \end{cases} \quad (2)$$

Jos sama kirjoittaja esiintyy sekä viitattavien että viitattujen kirjoittajien joukossa, voi mahdollisia kirjoittajan viittauksia itseensä käsitellä eri tavoilla. Matriisissa

itseviittauksia vastaavina alkioina voi olla esimerkiksi itseviittausten määrä [12]. Vaihtoehtoisesti itseviittaukset voidaan asettaa nolleksiksi, jos kirjoittajien viittauksia itseensä ei haluta huomioida [11].

Viittausmatriisin pohjalta voidaan muodostaa kaksi symmetristä matriisia: bibliografiseen kytkentään perustuva matriisi B ja yhteisviittausmatriisi C . Yhtälössä (3) kuvatut bibliometriseen kytkentään perustuvan matriisin diagonaalien ulkopuoliset alkio b_{ij} kuvaavat, kuinka samankaltaisesti henkilöt i ja j ovat viittaneet muihin henkilöihin. Bibliografisen kytkennän suuruuden (w_{ij}) määrittämiseen on eri vaihtoehtoja. Ma ehdotti kolme erilaista menetelmää bibliografisen kytkennän suuruuden määrittämiseen kirjoittajien bibliografiseen kytkentään pohjautuvassa analyysissä, yksinkertainen menetelmä, minimimenetelmä sekä yhdistetty menetelmä [13]. Man tutkimuksessa minimimenetelmä tuotti parhaimmat tulokset. Yksinkertaisessa menetelmässä bibliografinen kytkentä kahden henkilön välillä on niiden henkilöiden määrä, joihin molemmat henkilöt ovat viittaneet. Tämä vastaa tilannetta, jossa vastaavan viittausmatriisin alkio saavat arvoiksi vain 0 tai 1 eli henkilöiden välisten viittausten määrää ei oteta huomioon. Minimimenetelmässä kahden henkilön välisen bibliografisen kytkennän voimakkuus saadaan valitsemalla minimi kolmannen henkilön ensimmäiseen ja toiseen kohdistamien viittausten määristä, käymällä läpi kaikki kolmannet henkilöt ja summaamalla minimi. Yhdistetyssä menetelmässä minimin ottamisen sijaan kolmansien henkilöiden viittausmäärät kerrotaan keskenään ja saadut tulot summataan yhteen.

Yhteisviittausmatriisin alkio on määritelty yhtälössä (4). Kirjoittajien yhteisviittausanalyysissä yhteisviittaukselle on useampi määritelmä johtuen enimmäkseen siitä, että kirjoittajien yhteisviittaukset lasketaan dokumenttitasolta, yhdellä dokumentilla voi olla useampi kirjoittaja ja johonkin kirjoittajaan voidaan viitata useasti samassa dokumentissa [14]. Kahden henkilön välisen yhteisviittauksista muodostuvan kokonaispainon w_{ij} laskemiseen voidaan käyttää samoja ideoita kuin bibliografisen kytkennän vahvuuden laskemiseen. Yksinkertaisessa menetelmässä kahden henkilön välinen yhteisviittauksen määrä vastaa niiden henkilöiden määrää, jotka ovat viittaneet kahteen ensimmäiseen henkilöön. Minimimenetelmä tuottaa taas painon, joka vastaa kirjoittajien yhteisviittausanalyysissä usein käytetyn yhteisviittauksen määritelmää.

Matriisien diagonaalit vaativat erikoiskäsittelyä, sillä yleensä ajatellaan ettei henkilö voi tulla yhteisviittatuksi itsensä kanssa eikä henkilön tekemiä viittauksia ole järkevää verrata itsensä kanssa. Yhteisviittausmatriisin ja bibliografiseen kytkentään perustuvan matriisin diagonaalien käsittelyyn ei ole olemassa yksiselitteistä ratkaisua. Yhteisviittausmatriisin tapauksessa diagonaalien lukuja on voitu käsitellä puuttuvina arvoina ja korvattu sarakekeskiarvoilla [15] tai ne korvattu esimerkiksi kirjoittajan töihin kohdistuvien viittausten yhteismäärällä [16]. Tässä työssä diagonaalien arvoja käsitellään puuttuvina arvoina.

$$b_{ij} = \begin{cases} 0 & \text{jos henkilöt } i \text{ ja } j \text{ eivät ole viittaneet samoihin henkilöihin, } i \neq j \\ w_{ij} & \text{jos } i \text{ ja } j \text{ ovat viittaneet vähintään yhteen samaan henkilöön, } i \neq j \end{cases} \quad (3)$$

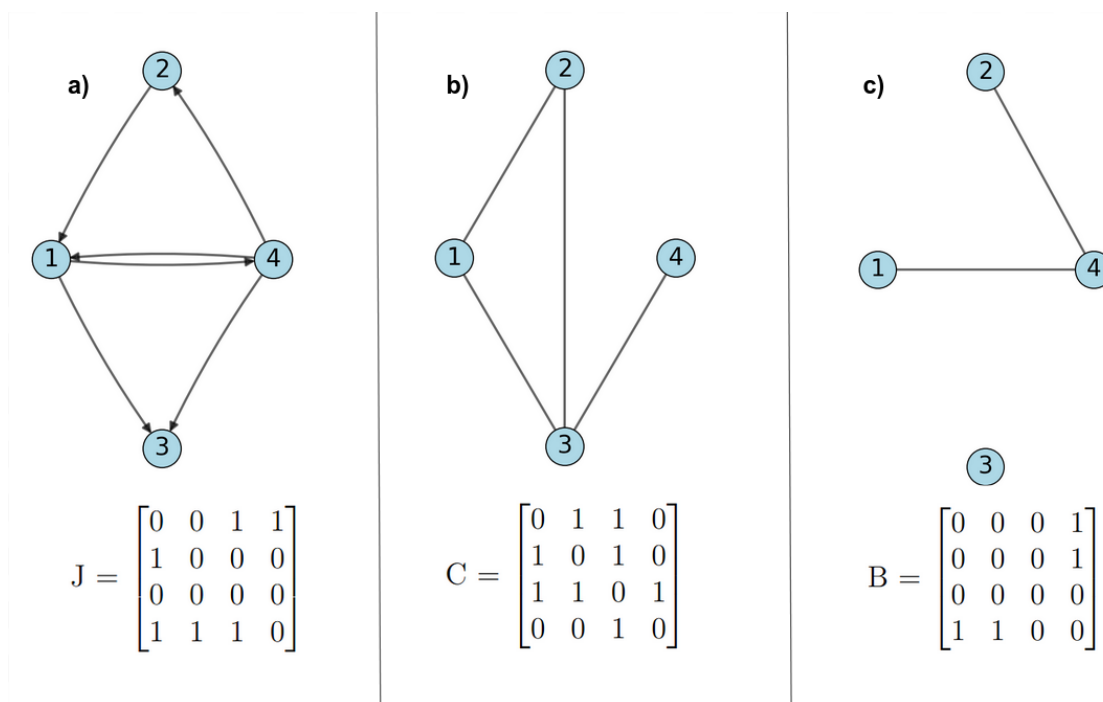
$$c_{ij} = \begin{cases} 0 & \text{jos henkilöihin } i \text{ ja } j \text{ ei ole viitattu saman tahon toimesta, } i \neq j \\ w_{ij} & \text{jos vähintään yksi henkilö on viitannut henkilöihin } i \text{ ja } j, i \neq j \end{cases} \quad (4)$$

Edellä mainitut matriisit voidaan ymmärtää verkostojen yhteysmatriiseina. Tällöin viittausmatriisi J vastaa viittausverkoston yhteysmatriisia. Tämän työn tapauksessa tarkastellaan dokumenteissa tapahtuvien viittausten sijaan eduskunnan puheenvuoroissa tehtyjä kansanedustajien välisiä henkilömainintoja, jolloin solmut ovat kansanedustajia. Kansanedustajasta toiseen menee suunnattu linkki, jos ensimmäinen edustaja on maininnut toisen edustajan puheenvuorossaan. Yhteisviittausmatriisin ja bibliografiseen kytkentään perustuva matriisin tapauksessa vastaavien yhteysmatriisien diagonaalien voidaan ajatella olevan nollia, jolloin vastaavissa verkostoissa yhdestäkään solmusta ei mene linkkiä solmuun itseensä. Yhteisviittausmatriisi ja bibliografiseen kytkentään pohjautuva matriisi ovat symmetrisiä, joten niiden pohjalta muodostuvat verkostot ovat suuntaamattomia. Yhteisviittausverkostossa linkit kansanedustajien välillä näyttävät, että joku tai jotkut ovat viittaneet heihin molempiin. Bibliografiseen kytkentään perustuvassa verkostossa linkki kansanedustajien välillä viittaa siihen, että edustajat ovat viittaneet samoihin, muihin edustajiin.

Kuvassa 1 näkyy a) yksinkertainen viittausverkosto ja sitä vastaava viittausmatriisi, b) viittausverkoston pohjalta muodostettu yhteisviittausverkosto ja yhteysmatriisi sekä c) bibliografiseen kytkentään perustuva verkosto ja vastaava yhteysmatriisi. Esimerkiksi viittausverkostossa solmu neljä mainitsee sekä solmun kaksi ja solmun kolme, joten niiden välissä on linkki yhteisviittausverkostossa. Vastaavasti sekä solmu 2 ja solmu 4 mainitsevat solmun 1 viittausverkostossa, jolloin solmujen 2 ja 4 välillä on linkki bibliografiseen kytkentään perustuvassa verkostossa. Yhteisviittausverkoston ja bibliografiseen kytkentään pohjautuvan verkoston yhteysmatriisit vastaavat diagonaalia lukuun ottamatta yhteisviittausmatriisia ja bibliografiseen kytkentään perustuvaa matriisia.

Viiteanalyysin voidaan siis katsoa liittyvän läheisesti myös verkostoanalyysiin ja verkostoanalyysin menetelmiä onkin hyödynnetty bibliometriikan tutkimuksissa muodostamalla erilaisia verkostoja viittautietojen pohjalta [17]. Viittauksiin perustuvien matriisien pohjalta muodostetut verkostot mahdollistavat keskeisten kirjoittajien tunnistamisen datasta sekä kirjoittajien klusteroinnin verkostoanalyysin menetelmin. Lisäksi verkostoanalyysin visualisointimenetelmät mahdollistavat viittausdatan esittämisen mielenkiintoisilla tavoilla.

Seuraavaksi esitellään esimerkkejä tutkimuksista, jotka ovat käyttäneet erilaisia viiteanalyysin menetelmiä. Tässä työssä käytetyt viite- ja verkostoanalyysin



Kuva 1: a) Esimerkki viittausverkostosta ja vastaavasta viittausmatriisista. b) Viittausverkoston pohjalta muodostettu yhteisviittausverkosto ja vastaava yhteysmatriisi. c) Viittausverkoston pohjalta muodostettu bibliografiseen kytkentään perustuva verkosto ja vastaava yhteysmatriisi.

menetelmät esitellään tarkemmin luvussa 5.

3.2 Viittausanalyysi

Viittausanalyysi, jossa tarkastellaan suoria töiden tai kirjoittajien välisiä viittauksia, paljastaa alan merkittävimpiä teoksia ja kirjailijoita. Suuren määrän viittauksia johonkin teokseen tai kirjoittajaan käsitetään kertovan, että teos tai kirjoittaja ovat vaikutusvaltaisia omalla alallaan [18]. Vastaavasti kansanedustajien viittausverkostossa paljon mainintoja keräävän edustajan voidaan olettaa olevan jollain tavalla erityisessä asemassa.

Esimerkiksi Backhaus ja muut tutkivat yritysmarkkinoinnin alan kehitystä viittaus- ja yhteisviittausanalyysin menetelmin [19]. Viittausanalyysit eri ajanjaksoille paljastivat, että vuosien aikana alaan liittyvien julkaisujen määrä ja sitä myöten myös keskiarvoinen viittausten määrä kasvoivat artikkelia kohti. Lisäksi kirjoittajien viittaukset kirjoittajien omiin artikkeleihin vähenivät, minkä tulkittiin kertovan alan vakiintumisesta. Laskemalla eri ajanjaksoina artikkelien saamia viittausten määriä saatiin selville kullakin ajanjaksolla eniten vaikuttaneet teokset. Laskemalla eri julkaisuiden saamia viittausten määriä saatiin selville alalle tärkeitä julkaisuita. Samalla huomattiin myös, että alalla psykologiaan liittyvien julkaisujen merkitys on pieni poiketen yleisestä markkinointitutkimuksesta. McLaren ja Bruner puolestaan esitteli-

vät, kuinka viittausverkostoanalyysiä voi käyttää liikuntapsykologian tarpeisiin ja nostivat esiin erilaisia viittausverkostovisualisointeja [20].

Suunnatun viittausverkoston ansiosta viittausanalyysissä voidaan myös huomioda paitsi dokumenttiin tai kirjoittajaan tehtyjä viittauksia mutta myös dokumentissa tehtyjä tai kirjoittajan tekemiä viittauksia. Esimerkiksi Calero-Medina ja Noyons muodostivat viittausverkoston tietyn käsitteen sisältävistä töistä, jossa työt ovat verkoston solmuja ja suunnatut linkit töiden välisiä viittauksia toisiinsa [21]. He laskivat verkoston solmuille keskus- ja auktoriteetti-arvot (*hub and authority weights*). Keskuksien ja auktoriteettien konsepti kehitettiin aluksi tärkeiden verkkosivujen tunnistamiseen [22]. Keskuksista ja auktoriteeteista kerrotaan tarkemmin luvussa 5.4. Oman työnsä tapauksessa Calero-Medina ja Noyons määrittelivät auktoriteetin olevan työ, johon on viitannut moni hyvä keskus ja keskuksen olevan työ, joka on viitannut moneen hyvään auktoriteettiin. Yleisesti ottaen keskuksset olivat paljon viittauksia tehneitä kirjallisuuskatsauksia ja auktoriteetit paljon viittauksia saaneita, vanhempia töitä. Joistain kirjallisuuskatsauksista tuli myös ajan myötä auktoriteetteja.

3.3 Kirjoittajien yhteisviittausanalyysi

Yhteisviittausanalyysi on tässä työssä käsitellyistä viiteanalyysin menetelmistä kenties käytetyin menetelmä. Yhteisviittausanalyysin avulla on perinteisesti pyritty tutkimaan eri tieteenalojen rakenteita selvittämällä, miten usein dokumentteihin, kirjoittajiin tai julkaisuihin on viitattu yhdessä kolmannen tahon toimesta [23]. Jos esimerkiksi kahteen kirjoittajaan on viitattu usein samojen kolmansien kirjoittajien toimesta, oletetaan heidän vaikuttavan samalla alalla. Etsimällä kirjoittajien ryhmitelmiä yhteisviittausten perusteella voidaan saada jonkinlainen käsitys tieteenalan sisäisestä rakenteesta. Kirjoittajien yhteisviittausanalyysi (*author co-citation analysis*), onkin yksi tärkeimmistä yhteisviittausanalyysin osa-alueista.

Parlamenttisammon tapauksessa kirjoittajien yhteisviittausanalyysin menetelmillä voidaan tutkia ryhmittyvätkö kansanedustajat tai puolueet jotenkin sen perusteella, miten kansanedustajat mainitsevat toisia kansanedustajia puheenvuoroissaan. Rajaamalla puheenvuoroja tietyille ajanjaksoille saadaan tietoa kyseisillä ajanjaksoilla vaikuttaneista kansanedustajista. Rajaamalla puheenvuoroja esimerkiksi aiheen mukaan voitaisiin mahdollisesti nähdä, ketkä edustajat, ja sitä myöten mitkä puolueet ja valiokunnat nousevat esiin käsiteltäessä eri aihepiireihin kuuluvia asioita. Esimerkkejä yhteisviittausanalyyseistä ja sen muunnelmista eri tieteenaloille löytyy paljon, kuten sosiaalisiin verkostoihin [24] sekä liiketoimintaan ja hallintoon [25] liittyen.

Monimuuttuja-analyysin menetelmistä kirjoittajien yhteisviittausanalyyseissä on käytetty esimerkiksi faktorianalyysiä (*factor analysis, FA*), moniulotteista skaalaus- ta (*multidimensional scaling, MDS*) ja hierarkkista agglomeratiivista klusterointia [26]. Esimerkiksi Andrews tutki lääketieteellisen informatiikan alaa käyttäen näitä

menetelmiä kirjoittajien ryhmittymien tunnistamiseksi [27] ja Zhao ja Strotmann hyödynsivät faktorianalyysiä informaatiotutkimuksen alan rakenteen selvittämiseksi [28].

Nerur ja muut tutkivat strategisen johtamisen alan rakennetta käyttäen yhteisviittausanalyysin menetelmiä [29]. He käyttivät faktorianalyysiä alan sisäisten suuntausten tunnistamiseen ja moniulotteista skaalausta kirjoittajien samankaltaisuuksien esittämiseen visuaalisesti. Lisäksi he käyttivät verkostanalyysin menetelmistä Pathfinder-analyysiä tunnistamaan erityisiä yhteyksiä kirjoittajien välillä sekä kirjoittajia, jotka ovat vaikuttaneet useampaan alan sisäiseen suuntaukseen. Solmujen ryhmittelyä voisi tehdä myös verkostanalyysin menetelmin muodostamalla Pathfinder-verkostoja. Pathfinder-verkostot ovat skaalattuja verkostoja, jotka pois karsittujen linkkien ansioista voivat paljastaa alkuperäisen verkoston tärkeimmän rakenteen [30]. Tästä kirjoittajan yhteisviittausanalyysin tapauksessa usein puumaisesta rakenteesta voidaan visuaalisesti tunnistaa ryhmittymiä, joskin ryhmittymien ymmärtäminen ja niiden oikeellisuuden arviointi vaatii datan erinomaista tuntemusta. Joissain tapauksissa Pathfinder-verkosto vastaa verkoston pienintä virittävää puuta (*minimum spanning tree*) [30].

Yhteisviittausanalyysiä on sovellettu muuhunkin kuin viittausdataan. Zhao ja Chen sovelsivat kirjoittajien yhteisviittausanalyysiä käyttäjien vuorovaikutusten analysointiin pikaviestintäryhmissä [31]. Käyttäjät laskettiin "yhteisviitatuksi", kun he osallistuivat keskusteluun samasta aiheesta. Myös järjestöjen ja erityisesti yritysten välisiä yhteyksiä on tutkittu yritysten välisillä yhteisviittausverkostoilla, joissa kaksi yritystä lasketaan yhteisviitatuksi niiden esiintyessä samassa uutisartikkelissa [32]. Työssä muodostetaan suuntaamaton verkosto, jossa solmut ovat yrityksiä ja linkit kuvaavat yritysten suhteita. Linkin paino vastaa niiden uutisten määrää, joissa yritykset esiintyvät yhdessä. Muodostetusta verkostosta voi selvittää keskeisiä yrityksiä laskemalla solmuille eli yrityksille keskeisyysmetriikoita.

3.4 Bibliografinen kytkentä

Bibliografiseen kytkentään (*bibliographical coupling*) perustuvia verkostoja on käytetty myös eri tieteenalojen kirjallisuuden tutkimiseen, mutta ei yhtä paljon kuin yhteisviittausanalyysiä. Bibliografiseen kytkentään pohjautuvassa verkostossa solmut yhdistyvät, jos ne jakavat samoja lähteitä. Siinä missä kirjoittajien yhteisviittausanalyysi tunnistaa ja ryhmittelee alaan erityisesti pidempään vaikuttaneita kirjoittajia, kirjoittajien bibliografisen kytkentä analyysi auttaa tunnistamaan alan tulevia suuntauksia [33]. Kirjoittajien bibliografisen kytkennän analyysissä ei tarvitse odottaa, että kirjoittajille kertyy tarpeeksi viittauksia, vaan voidaan selvittää, millaisiin julkaisuihin viime aikoina aktiivisena olleet kirjoittajat ovat viitanneet.

Ma testasi kirjoittajan bibliografisen kytkennän analyysiä kiinalaiseen akateemiseen tietokantaan pyrkien selvittämään, voiko kirjoittajan bibliografisen kytkennän

menetelmillä selvittää kiinalaisen kirjasto- ja informaatiotieteeseen liittyvien kirjoittajien ryhmittymiä heidän osaamisalojensa mukaan [13]. Hän ehdotti kolme erilaista periaatetta kirjoittajien bibliografisen kytkennän vahvuuden määrittämiseksi ja kokeili, mikä niistä mahdollistaa kirjoittajien selkeimmän ryhmittelyn. Hän muodosti Pathfinder-verkoston kirjoittajien bibliografisten kytkentöjen perusteella ja tunnisti verkoston visualisoinnista kirjoittajien ryhmittymiä. Zhao ja Strotmann vertailivat kirjoittajien bibliografisen kytkennän analyysin ja yhteisviittausanalyysin tuloksia informaatiotieteen kartoituksessa ja havaitsivat menetelmien täydentävän toisiaan sekä tuottavan yhdessä selkeämmän kuvan tieteenalojen rakenteesta [34].

3.5 Kontekstin ja ajan huomiointi viiteanalyysissä

Perinteiset viiteanalyysin menetelmät eivät huomioi viittausten kontekstia eli sitä, onko esimerkiksi viittaaja lähteen kanssa samaa mieltä, eri mieltä vai onko suhtautuminen neutraali. Viittausten laatua on myös pyritty ottamaan luokittelemalla viittauksia eri kategorioihin [35]. Tällöin ongelmana on mahdolliset virheet viittausten luokittelussa ja manuaalisen luokittelun hitaus, mikä rajoittaa tutkimukseen sisällytettävien viittausten määrää .

Yhteisviittausanalyysissä kontekstia on pyritty ottamaan huomioon muun muassa vertailemalla lauseita, joissa viittaajat viittaavat samaan kohteeseen ja tuloksia on pyritty myös parantamaan ottamalla huomioon, missä artikkelin osiossa viittaus tapahtuu [36]. Parlamentaariset puheenvuorot ovat useimmiten lyhyitä, eikä niitä ole jaoteltu selkeisiin osiin, jolloin henkilömaininnan paikan tunnistaminen puheenvuorosta ei tuo niin paljoa lisäarvoa.

Viittausverkostojen dynaamisuutta on taas huomioitu esimerkiksi muodostamalla viittausverkostoja eri ajanjaksoille, analysoimalla eri ajanjaksojen verkostoja ja vertailemalla tuloksia, jolloin pystytään tutkimaan, miten jokin tieteenala on kehittynyt [19]. Parlamentaarisessa keskustelussa mielenkiintoisia ajanjaksoja vertailtavaksi ovat esimerkiksi eri vaalikaudet tai valtiopäivät.

3.6 Kirjoittajien viiteanalyysin ja poliitikkojen mainintoihin perustuvien verkostojen analyysien eroavaisuudet

Poliitikkojen puheenvuoroihin perustuvissa viittausverkostoissa on eroja perinteisiin, artikkeleihin ja niiden kirjoittajiin perustuviin viittausverkostoihin verrattuna. Tämän vuoksi jotkin viiteanalyysin ongelmat katoavat tai niiden merkitys pienenee ja vastaavasti syntyy toisenlaisia pulmia. Nämä erot täytyy myös ottaa huomioon päätettäessä käytettävistä menetelmistä sekä tulkitessa tuloksia.

Yksi perustavanlaatuisen ero viiteanalyysin ja poliitikkojen viittausverkostojen välillä on se, että vaikka viiteanalyysissä tutkimusta voidaan tehdä eri tasoilla, niiden

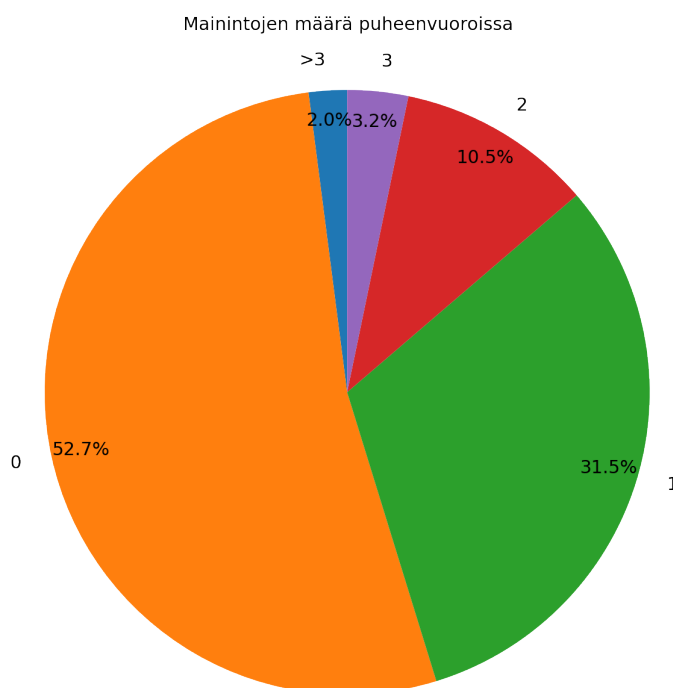
pohjalla ovat aina dokumenttien väliset viittaukset. Poliitikkojen viittausverkostossa maininnat puolestaan kohdistuvat henkilöihin, eivätkä välttämättä henkilön tiettyyn puheenvuoroon. Esimerkiksi vaalikaudella 2015–2018 noin 15% lauseista, jotka sisältävät henkilömaininnan sisältävät sanan "puheenvuoro" tai "puhe" jossain muodossa, mikä kertoo mahdollisesta viittauksesta toisen edustajan puheenvuoroon sen sijaan että viitattaisiin suoraan kansanedustajaan. Henkilömainintaverkosto ei siis perustu puheenvuorojen viittausverkostoon kuten kirjoittajien välinen viittausverkosto perustuu dokumenttien väliseen viittausverkostoon. Tästä syystä useat bibliometriikan tarpeisiin kehitetyt indeksit eivät ole laskettavissa tai ainakaan samalla tavalla tulkittavissa poliitikoille kuin kirjoittajille.

Toinen merkittävä ero viiteanalyysin ja poliitikkojen viittausverkostojen välillä on viittausten määrien ero dokumenteissa ja parlamentaarisisissa puheenvuoroissa. Dokumenteissa on yleensä useita, jopa kymmeniä lähteitä. Parlamentaarisisissa puheenvuoroissa ei niiden lyhyiden takia yleensä ehditä mainita kovin montaa henkilöä. Kuvassa 2 näkyy puheenvuorojen saman vaalikauden kansanedustajiin kohdistuvien henkilömainintojen määrät vaalikaudella 2015–2018. Vain hieman alle puolet puheenvuoroista sisältää henkilömaininnan vähintään yhteen saman vaalikauden kansanedustajaan. Vaalikaudella 2015–2018 suurin määrä mainittuja henkilöiden nimiä yhdessä puheenvuorossa on 22. Lähes kolmasosa puheenvuoroista sisältää maininnan yhteen henkilöön. Vain 15.7% puheenvuoroista sisältää maininnan useampaan kuin yhteen kansanedustajaan. Tästä syystä yhteisviittaus ja bibliografinen kytkentä määritellään tässä työssä henkilötasolla eikä dokumentti- tai puheenvuorotasolla.

Viiteanalyysissä kahden kirjoittajan välillä on yhteisviittaus, jos molempien kirjoittajien jotkin työt esiintyvät kolmannen kirjoittajan työn lähdeluettelossa. Tämän työn tapauksessa kahden poliitikon välillä on yhteisviittaus, jos joku kolmas poliitikko on maininnut kaksi edellistä puheenvuoroissaan eli mainintojen ei tarvitse tapahtua yhdessä ja samassa puheenvuorossa. Vastaavasti poliitikkojen välillä on bibliografinen kytkentä, jos he ovat maininneet jonkin kolmannen poliitikon nimen puheenvuoroissaan sen sijaan että he olisivat viittaneet kolmannen poliitikon johonkin tiettyyn puheenvuoroon.

Viittausanalyysissä ongelmana saattaa olla viitetietokantojen kattavuus [37]. Haettaessa esimerkiksi tiettyyn tieteenalaan liittyviä artikkeleita viitetietokannoista, joitain olennaisia kirjoittajia tai heidän artikkeleitaan voi jäädä puuttumaan datajoukosta. Tässä työssä on käytettävissä kaikki valittujen ajanjaksojen puheenvuorot ja niistä tunnistetut henkilömaininnat. Henkilömainintojen tunnistuksen tarkkuuden vaikutus viittausverkostojen laatuun täytyy kuitenkin ottaa huomioon. Toisaalta täytyy muistaa, että vaikka kaikki puheenvuorot ovatkin käytettävissä analyyseissä, puhemiehet jakavat sen rajallisen määrän puheenvuoroja mikä kuhunkin asiaan liittyen on käytettävissä [38]. Näin ollen kaikki halukkaat eivät välttämättä pääse ääneen.

Vaikka noin puolet puheenvuoroista eivät sisällä mainintoja saman vaalikauden kansanedustajia lainkaan, on henkilömaininnat eduskunnassa pidettyjen puheenvuo-



Kuva 2: Henkilömainintojen määrät puheenvuoroissa vaalikaudella 2015–2018.

rojen perusteella paljon keskustelunomaisempaa kuin viittaukset tieteellisten artikkelien kautta. Periaatteessa kirjoittajat voivat vastata artikkelien kautta toisilleen esimerkiksi jonkin menetelmän käytöstä tai toiminnasta johtuviin erimielisyyksiin liittyen. Eduskunnassa puheenvuorojen kautta vastaaminen on nopeampaa, vaikka puheenvuorojen määrä onkin rajallinen.

Edellä mainituista eroista huolimatta monia viiteanalyysin menetelmiä voidaan käyttää myös poliitikkojen viittausverkostoille. Tulosten tulkinnoissa on kuitenkin eroavaisuuksia. Esimerkiksi viiteanalyysissä kaksi kirjoittajaa voidaan käsittää samantapaiseksi eli he vaikkapa liittyvät samaan alaan, jos he ovat viitanneet paljon samoihin töihin ja heidän välillään on siten vahva bibliografinen kytkentä. Jos kahden poliitikon välillä on vahva bibliografinen kytkentä, he ovat maininneet usein samoja henkilöitä puheenvuoroissaan. Kirjoittajien tapauksessa tulokset kumpuavat dokumenttitasolta, kun taas tässä työssä poliitikkojen tapauksessa tulokset johtuvat suorista poliitikkojen välisistä maininnoista. Tämä voi hankaloittaa tulosten tulkintaa poliitikkojen viittausverkostojen tapauksessa.

Oletettavasti kirjallisissa töissä aihe on yleensä rajattu, jolloin kirjoittajan tekemien viittausten perusteella voi päätellä hänen mielenkiinnon kohteistaan ja samoihin teoksiin viittaavien kirjoittajien voidaan ajatella olevan kiinnostuneita samoista aiheista. Jos kaksi kirjoittajaa mainitsisivat kolmannen kirjoittajan vain nimeltä ilman kontekstia, olisi kahden kirjoittajan välisiä yhteyksiä vaikeampi selvittää. Kolmas

kirjoittaja voi esimerkiksi tuottaa kahteen hieman eri alaan liittyviä töitä, jolloin mahdollisesti ensimmäinen kirjoittaja on kiinnostunut ensimmäisestä alasta ja toinen toisesta. Tällöin pelkän nimimaininnan perusteella ei voida suoraan päätellä, ovatko kirjoittajat kiinnostuneita samoista asioista. Samoin jos kahdesta poliitikosta molemmat mainitsevat nimeltä jonkin kolmannen poliitikon, voi kahden poliitikon välisiä yhteyksiä olla vaikea selvittää, varsinkin jos maininnan kontekstia ei tunneta. Voi olla, että ensimmäinen poliitikko vastustaa ja toinen poliitikko puolustaa kolmannen poliitikon esitystä, jolloin saman henkilön nimen maininta ei edes viittaa mainitsevien poliitikkojen samankaltaisuuteen. Jos kaksi poliitikkoa mainitsevat kuitenkin usein samoja henkilöitä, on sen taustalla todennäköisesti jokin syy.

Parlamentarisessa keskustelussa esiin nousevat ministerit ja puolueiden puheenjohtajat. Esimerkiksi puolueeseen kohdistuvat asiat saatetaan kohdistaa puolueen puheenjohtajalle ja johonkin tiettyyn aihepiiriin liittyvissä asioissa saatetaan mainita asiasta vastaava ministeri. Tämän vuoksi olisikin syytä pohtia, pitäisikö ministereihin ja puoluejohtajiin kohdistuville nimimaininnoille antaa pienimpi painoarvo, jolloin "riivikansanedustajat" eivät jäisi niin selkeästi heidän varjoonsa. Viiteanalyysissä täytyy päättää, lasketaanko kaikki artikkeliin merkityt kirjoittajat artikkelin kirjoittajaksi vai vain esimerkiksi ensimmäinen kirjoittaja [39]. Eduskunnassa puheenvuoroja pitää yksi henkilö kerrallaan, joten tätä pohdintaa ei tarvitse käydä tämän työn osalta.

Kirjoittajien viittaus- ja yhteisviittausanalyyseissä ongelmana on, että uuden kirjoittajat jäävät yleensä varjoon, sillä heille ei ole vielä kertynyt yhtä paljon viittauksia kuin kauemmin alan julkaisuja tehneille kirjoittajille [37]. Sen sijaan jos viittaukset rajataan esimerkiksi tietylle vaalikaudella ja vain sillä kaudella toimineiden kansanedustajien välille, kaikilla kansanedustajilla on ollut sama aika kerryttää ja tehdä viittauksia muutamia poikkeustapauksia lukuun ottamatta. Esimerkiksi joku kansanedustaja on voinut siirtyä kesken vaalikauden muihin tehtäviin ja hänen paikalleen on noussut joku toinen varasijalta. Kaikille edustajille ei myöskään ole kertynyt koko vaalikauden aikana riittävän suurta määrää viittauksia, jotta heidät voitaisiin kartoittaa luotettavasti kirjoittajien yhteisviittausanalyysin menetelmillä, ja heidät joudutaan rajaamaan pois monimuuttuja-analyyseistä.

4 Parlamenttisampo

Tässä luvussa esitellään Parlamenttisampo ja sen sisältämä data, jota tässä työssä käytetään. Lisäksi luvussa käydään lyhyesti läpi, miten puheenvuoroista on tunnistettu henkilömainintoja, joihin tämän työn analyysit perustuvat.

4.1 Semanttinen parlamentti -hanke

Työssä käytetty data on peräisin on peräisin Parlamenttisammosta [40], joka on Semanttinen parlamentti (SEMPARL) konsortiohankkeen luoma linkitetyn avoimen datan (*linked open data, LOD*) palvelu. Hankkeessa rikastettiin ja julkaistiin eduskunnan aineistoja linkitettynä avoimena datana luomalla ontologia- ja tietoinfrastruktuuri, joka on tutkijoiden, kansalaisten, poliitikkojen, median ja sovelluskehittäjien käytettävissä. Lisäksi hanke tutki muun muassa poliittisen kulttuurin ja kielen kehittymistä Suomessa sekä kehitti ja julkaisi semanttisen portaalisovelluksen², joka mahdollistaa aineistojen selaamisen fasettihaun avulla. Hankkeessa oli mukana Helsingin yliopistosta digitaalisten ihmistieteiden keskus (HELDIG), Aalto-yliopistosta Semanttisen laskennan tutkimusryhmä (SeCo) sekä Turun yliopistosta Eduskuntatutkimuksen keskus.

Parlamenttisammon data on linkitettyä, RDF-muodossa³ (*Resource Description Framework*) olevaa dataa. SPARQL⁴ on semanttinen kyselykieli, jolla voidaan hakea ja muokata tietokannoissa olevaa RDF-muodossa olevaa dataa. Dataa voidaan hakea Parlamenttisammon palvelupisteestä SPARQL-kyselyillä esimerkiksi Python skriptien avulla, jonka jälkeen dataa voidaan analysoida hyödyntäen eri Python-kirjastoja.

Parlamenttisampo koostuu kahdesta tietograafista (*knowledge graph, KG*). Ensimmäinen tietograafi kuvaa eduskunnassa pidettyjä puheenvuoroja aina eduskunnan perustamisesta 1907 lähtien [41]. Puheenvuorot on jaoteltu datassa puhetyypin mukaan varsinaisiin puheenvuoroihin, vastauspuheenvuoroihin, puhemiehen puheenvuoroihin, esittelypuheenvuoroihin, ryhmäpuheenvuoroihin ja nopeatahtisiin puheenvuoroihin. Toinen tietograafi sisältää poliittisia toimijoita: kaikki suomalaiset kansanedustajat vuodesta 1907 lähtien, muita merkittäviä poliittisia henkilöitä sekä erilaisia organisaatioita, esimerkiksi puolueet, eduskuntaryhmät, hallitukset ja valiokunnat [42]. Tietograafit linkittyvät toisiinsa, jolloin puheenvuoroista on mahdollista saada selville paitsi esimerkiksi puheenvuoron tyyppi, kieli, sisältö ja päivämäärä, mutta myös puheenvuoron pitäjä ja sitä kautta henkilöihin liittyvää tietoa.

Tässä työssä käytetään vaalikausien 2015–2018 ja 2019–2022 puheenvuoroja, joita on saatavilla 6.5.2022 asti. Nimettyjä entiteettejä on toistaiseksi tunnistettu ja linkitetty vain näiden vaalikausien puheenvuoroista. Taulukossa 1 näkyy puheenvuorojen määrä sekä vähintään yhden henkilömaininnan sisältävien puheenvuorojen määrä

²<https://parlamenttisampo.fi/fi/>

³<https://www.w3.org/TR/PR-rdf-syntax/>

⁴<https://www.w3.org/TR/sparql11-overview/>

vaalikausittain. Suluissa oleva numero vastaa puheenvuorojen määrää, kun puhemiehen puheenvuoroja ei oteta huomioon. Tässä työssä puhemiehen puheenvuoroja ei huomioida analyyseissä, sillä ne eivät sisällä työn kannalta mielenkiintoista tietoa. Puhemiehen puheenvuorot eivät varsinaisesti sisällä mitään täysistunnon aiheisiin liittyvää, vaan ne liittyvät enemmänkin esimerkiksi puheenvuorojen jakamiseen.

Yhteensä puheenvuoroja on noin 120 000, joista reilu 33 000 on puhemiehen puheenvuoroja. Puheenvuoroja, joissa on mainittu vähintään yksi henkilö, on yli 64 000 eli noin joka toisessa puheenvuorossa on mainittu vähintään yksi henkilö. Jos puhemiehen puheenvuorot jätetään huomiotta, puheenvuoroja, joissa on henkilömainintoja, on hieman yli 40 000. Näistä puheenvuoroista noin kolmasosassa on mainittu vähintään yksi henkilö, joka on ollut puheenvuoron aikaan ministeri. Vaalikauden 2019–2022 puhemiehen puheenvuorojen määrä, joissa puhemies mainitsee henkilön, on suuri verrattuna vaalikauden 2015–2018 vastaavaan määrään. Tämä johtuu muutoksista eduskunnan pöytäkirjanpitokäytännöissä, eikä tällä ole vaikutusta tämän työn tuloksiin, sillä puhemiehen puheenvuoroja ei huomioida analyyseissä.

Kaikista puheenvuoroista 1374 on merkitty ruotsinkieliseksi ja 220 puheesta puuttuu tieto puheenvuoron kielestä. Analyyseissä otetaan huomioon vain suomenkieliset puheenvuorot, sillä vain suomenkielisistä puheenvuoroista on tunnistettu ja linkitetty henkilömainintoja. Kun puhemiehen puheenvuorot jätetään huomioimatta, tämä tarkoittaa, että yhteensä 1203 puheenvuoroa eli 1.4% mahdollisista puheenvuoroista jää analyyseiden puolelle. Näistä 1091 on Suomen ruotsalaisen kansanpuolueen edustajien pitämiä, mikä on noin kolmasosa puolueen jäsenien 22.4.2015–6.5.2022 välillä pitämistä puheenvuoroista.

Vaalikausi	puheenvuorot	puheenvuorot, joissa henkilömaininta
2015–2018	64934 (51807)	31075 (24428)
2019–2022*	55104 (33146)	31578 (16405)
Yhteensä	120038 (86820)	64221 (40833)

Taulukko 1: Puheenvuorojen sekä niissä esiintyvien henkilömainintojen määrät. Suluissa olevat luvut kertovat puheenvuorojen ja henkilömainintojen määrät, kun puhemiehen puheenvuoroja ei oteta huomioon. *Vaalikauden 2019–2022 puheenvuoroja on saatavilla 6.5.2022 saakka.

4.2 Henkilömainintojen tunnistamien eduskunnan puheenvuoroista

Minna Tamper ja muut ovat NLP-menetelmiä käyttäen tunnistanut tunnistanut puheenvuoroista mainintoja nimettyihin entiteetteihin eli Parlamenttisammon tapauksessa esimerkiksi henkilöihin, paikkoihin, organisaatioihin, ajan määreisiin ja lainsäädäntöihin. Lemmatisoiduista puheenvuoroista etsittiin maininnat käyttäen nimettyjen entiteettien tunnistamismenetelmiä (*named entity recognition, NER*) ja

FinBERT-mallia. Tunnistetut nimetyt entiteetit linkitettiin (*named entity linking, NEL*) sisäisesti Parlamenttissammon tietograafeihin sekä sopiviin ulkoisiin ontologioihin. [43]

NER-tarkkuutta arvioitiin sadalla satunnaisesti valitulla puheenvuorolla, joille sisäinen tarkkuus oli 97%, herkkyys 77% ja näiden harmoninen keskiarvo eli F1-arvo 86%. Suurin osa nimetyistä entiteeteistä on siis tunnistettu oikein, sillä sisäinen tarkkuus on suuri. Alhainen herkkyys puolestaan viittaa siihen, että monta nimettyä entiteettiä on jäänyt tunnistamatta. Alhainen herkkyys johtuu pääosin ongelmista organisaatioiden, kuten valiokuntien tunnistamisesta eikä sillä näin ollen ole suurta vaikutusta kansanedustajien välisen viittausverkoston laatuun.

Jotta tunnistettuja nimettyjä entiteettejä voidaan hyödyntää analyyseissä, ne täytyy linkittää oikeisiin resursseihin. Nimettyjen entiteettien linkittämisen tulokset esitetään NamedEntity-luokan avulla. Luokan kautta pääsee käsiksi niihin sanoihin, joista nimetty entiteetti on tunnistettu puheenvuorosta sekä nimetyt entiteetin esiintymismäärään puheenvuoroista. Luokan kautta saa myös selville nimetyt entiteetin tyypin, joka voi olla esimerkiksi henkilö, paikka tai järjestö. Luokka sisältää linkejä nimettyyn entiteettiin liittyviin ontologioihin. Nimetty entiteetti on voitu linkittää sisäisesti Parlamenttissammon toimija tietograafiin tai ulkoisiin tietolähteisiin, kuten YSO paikkaontologiaan⁵ tai Kansalliskirjaston Kanto-sanastoon⁶, joka sisältää suomalaisia kansallisia toimijoita.

Linkitysten tarkkuutta henkilömaininnoille arvioitiin 100:lla satunnaisesti valitulla puheenvuorolla. Sisäinen tarkkuus oli 98%, herkkyys 96% ja F1-arvo 97%. Suurimmat sisäistä tarkkuutta pienentävät ongelmat mainittujen henkilöiden linkittämisessä liittyvät henkilöihin, joilla on sama sukunimi. Esimerkiksi puheenvuoroista tunnistetut maininnat presidentti Niinistö, puolustusministeri Niinistö ja edustaja Niinistö linkittäminen oikein Sauli, Jussi tai Ville Niinistöön ei aina ole onnistunut. Joskus linkitys tehdään yhteen tai useampaan Niinistöön, jolla ei ole kyseistä titteliä. Samoin Krista Kiuru ja Pauli Kiuru on linkitetty yli sataan samaan puheenvuoroon, joissa monissa löytyy maininta pelkästään ministeri Kiuruun, jolla tarkoitetaan Krista Kiurua. Niiden henkilöiden kohdalla, joilla on sukunimikaimoja, täytyy siis olla erityisen tarkkana tuloksia tulkitessa. Henkilömainintojen tunnistamisen ja linkityksien tarkkuudet ovat kuitenkin riittävän hyvät analyysien tekoon sekä tuloksien tulkintaan.

⁵<https://finto.fi/yso-paikat/fi/>

⁶<https://finto.fi/finaf/fi/>

5 Mainintaverkostojen tutkimusmenetelmät

Tässä luvussa käydään ensin läpi, miten Parlamenttisarmon datasta muodostetaan erilaisia henkilömainintoihin liittyviä verkostoja. Luvun lopussa käydään tarkemmin läpi työssä käytettyjä viiteanalyysin ja verkostoanalyysin menetelmiä.

5.1 Verkostojen luominen

Parlamenttisarmonssa on vuonna 2015 alkaneesta vaalikaudesta eteenpäin tunnistettu ja linkitetty henkilöitä, paikkoja ja muita nimettyjä entiteettejä. Tässä työssä käytetään enimmäkseen vaalikauden 2015–2018 puheenvuoroja. Verkostoissa, joissa solmut ovat kansanedustajia, henkilömaininnat on rajattu vaalikaudelle 2015–2018, jolloin voidaan tarkastella kansanedustajien vuorovaikutuksia vaalikauden aikana. Tällöin maininnat seuraavalla vaalikaudella eivät nosta esiin lähinnä henkilöitä, jotka ovat tulleet uudelleen valituksi eduskuntaan ja pystyneet siten kerryttämään enemmän mainintoja. Vain samalle vaalikaudelle valittuihin kansanedustajiin kohdistuvat maininnat otetaan huomioon, eli esimerkiksi maininnat entisiin kansanedustajiin tai ulkomaisiin poliitikkoihin jätettiin pois analyysistä. Puolueiden viittausverkoston tapauksessa mukaan on otettu myös henkilömaininnat vaalikauden 2019–2022 alkupuoliskolta, jolloin voidaan tarkastella, miten puolueiden siirtyminen hallituksesta oppositioon tai päinvastoin vaikuttaa puolueen kansanedustajiin kohdistuviin tai heidän tekemiin mainintoihin. Puhemiehen puheenvuoroissa ja ruotsinkielisissä puheenvuoroissa tehdyt henkilömaininnat on jätetty huomiotta.

Viittausverkostot luotiin käyttäen Jupyter Notebook -palveluun perustuvaa Google Colaboratorya (Colab), joka mahdollistaa Python-koodin kirjoittamisen ja ajamisen selaimessa. Datan hakemiseen Parlamenttisarmon SPARQL-palvelupisteestä käytettiin SPARQLWrapper-pakettia⁷, joka mahdollistaa SPARQL-kyselyiden tekemisen ja tulosten käsittelyn Python-koodissa. Verkostojen analysointiin käytettiin NetworkX Python-pakettia [44]. Lisäksi viiteanalyysissä käytettiin Scikit-learn [45] ja FactorAnalyzer⁸ Python-moduuleja.

Viittausverkostojen luominen alkaa hakemalla linkit eli hakemalla SPARQL-kyselyllä vaalikauden 2015–2018 puheenvuoroista puheenvuoron pitäjät, mainitut kansanedustajat sekä puheenvuorojen määrän, joissa puheenvuoron pitäjä on maininnut kunkin toisen kansanedustajan. Puheenvuoroista suodatettiin pois puhemiehen puheenvuorot, ruotsinkieliset puheenvuorot sekä puheenvuorot, jotka eivät osu valitulle ajanjaksolle. Lisäksi henkilömaininnoista suodatettiin virkamiesten tekemät maininnat sekä maininnat, joissa mainittu henkilö ei ole kansanedustaja tai ministeri puheenpitohetkellä. Erillisellä SPARQL-kyselyllä hankitaan solmuihin liittyvää tietoa, kuten solmujen ollessa henkilöitä henkilön nimi ja puolue, johon henkilö kuuluu.

⁷<https://github.com/RDFLib/sparqlwrapper.git>

⁸https://github.com/EducationalTestingService/factor_analyzer.git

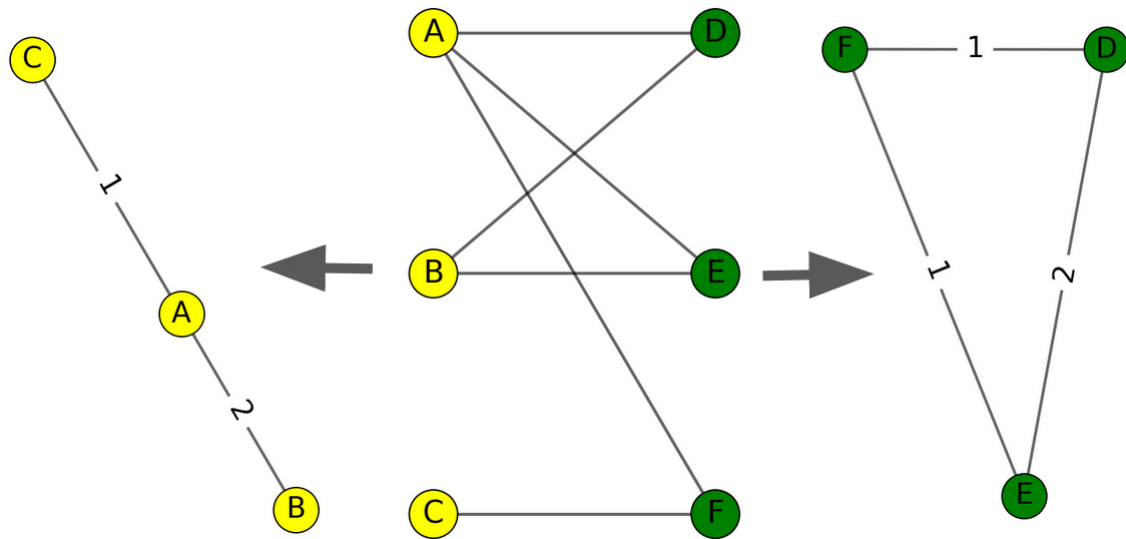
Linkitettyjen henkilömainintojen pohjalta voidaan muodostaa kansanedustajien välinen suunnattu viittausverkosto, jossa linkki menee puheenpitäjältä puheenvuorossa mainittuun henkilöön. Linkin paino vastaa niiden puheenvuorojen määrää, jossa ensimmäinen kansanedustaja mainitsee toisen kansanedustajan vähintään kerran. Ryhmittelemällä puhuja ja mainittu henkilö niihin puolueisiin tai valiokuntiin, joiden jäsenyys heillä on puheenpitohetkellä, saadaan puolueiden välinen ja valiokuntien välinen viittausverkosto. Tällöin solmuina toimivat puolueet tai valiokunnat, ja linkkien väliset painot vastaavat ensimmäiseen puolueeseen tai valiokuntiin kuuluvien kansanedustajien tekemien viittausten määrää toisen puolueen tai valiokunnan kansanedustajiin.

Yksi tapa muodostaa yhteisviittausverkosto ja bibliografiseen kytkentään perustuva verkosto on muodostaa ensin kaksijakoinen verkosto, jonka ensimmäisessä solmujoukossa on viittaavat solmut. Toisessa joukossa on puolestaan puheenvuoroissa mainitut solmut. Useat kansanedustajat esiintyivät molemmissa solmujoukossa, eli he ovat sekä maininneet muita kansanedustajia että tulleet muiden kansanedustajien mainitsemiksi. Viittaavan ja viitattun solmun välillä on suuntaamaton linkki, jos viittaavaan solmujoukkoon kuuluva on maininnut viitattuun solmujoukkoon kuuluvan kansanedustajan puheenvuorossaan.

Kaksijakoisesta verkostosta voidaan luoda kaksi erillistä verkostoa muodostamalla projektioita kuvan 3 mukaan. Ensimmäisessä projektiossa esiintyvät viittaavat solmut, joiden välillä on linkki heidän ovat viitattaessa vähintään yhteen samaan henkilöön. Tämä projektio vastaa bibliografiseen kytkentään perustuvaa verkostoa. Toiseen projektiin kuuluvat viittaavat solmut, joiden välillä on linkki, jos heidät on maininnut vähintään yksi sama henkilö. Toinen projektio vastaa yhteisviittausverkostoa.

Projektioissa kahden solmun välillä on siis linkki, jos niillä on yhteinen naapuri tai naapureita alkuperäisessä kaksijakoisessa verkostossa. Projektoiden linkit voivat olla painotettuja, jolloin painojen määrittämiseen on useita tapoja. Yksi tapa määrittellä kahden solmun välisen linkin paino projektiossa on solmujen yhteisten naapureiden määrä kaksijakoisessa verkostossa, kuten kuvan 3 tapauksessa. Tämä vastaa Man esittelemää yksinkertaista menetelmää, joka esiteltiin luvussa 3.1. Myös muita Man esittämiä menetelmiä voi hyödyntää linkkien painojen määrittämiseen. Lisäksi yhteisviittaukselle ja bibliografiselle kytkennälle voi vaatia aikaikkunan. Esimerkiksi kaksi kansanedustajaa lasketaan yhteisviitatuksi vain jos jokin kolmas kansanedustaja mainitsee heidät molemmat tietyn ajanjakson, kuten viikon aikana. Tässä työssä kokeiltiin erilaisia tapoja määrittellä yhteisviittausverkoston ja bibliografiseen kytkentään perustuvan verkoston linkkien painoja.

Yhteisviittausverkoston ja bibliografiseen kytkentään pohjautuvan verkoston voi myös muodostaa suunnatun viittausverkoston pohjalta. Ensin muodostetaan viittausverkosto, jonka yhteismatriisin avulla voidaan laskea yhteisviittausten ja bibliografisten kytkentöjen suuruudet solmujen välillä. Näin ollen voidaan muodostaa yhteisviittausmatriisi ja bibliografiseen kytkentään pohjautuva matriisi, jotka toimivat



Kuva 3: Kaksijakoisesta verkostosta (keskellä) projektiona muodostetut verkostot.

yhteysmatriiseina vastaavissa verkostoissa kun niiden diagonaalit asetetaan nollassi. NetworkX tarjoaa valmiin metodin, jolla voi muodostaa projektiota kaksijakoisen verkoston pohjalta. Metodille voi antaa parametriksi oman funktion, joka laskee linkkien painot projektiossa. Tämä tekee erilaisten painojen määrittelyiden testaamisesta selkeää ja helppoa, joten tässä työssä päädyttiin hyödyntämään kaksijakoisia verkostoja.

Sisältöön perustuvan yhteisviittausverkosto muodostettiin eri tavalla aiempaan verrattuna seuraten Jeongin ja muiden työtä [36]. Aluksi haettiin kaikki henkilömaininnat, niiden tekijät ja kohteet sekä se lause puheenvuorosta, jossa henkilömaininta tapahtuu. Aiempien verkostojen tapauksessa henkilö on laskettu kerran mainituksi, jos hänet on mainittu puheenvuorossa vähintään kerran. Sisältöön perustuvassa yhteisviittausanalyysissä päätettiin ottaa mukaan kaikki ne kerrat, joina henkilö mainitaan puheenvuorossa, sillä tällöin ei tarvitse arpoa mikä maininnan sisältävistä lauseista otetaan mukaan. Henkilömaininnan sisältävistä lauseista poistettiin kaikki välimerkit, tunnistetut henkilömaininnat, hukkas sanat sekä tittelit, kuten edustaja ja eri ministerien tittelit. Lauseet lemmatisoitiin eli jäljelle jääneet sanat muutettiin perusmuotoihin käyttämällä Voikko-kirjastoa⁹.

Kun kaksi kansanedustajaa tulee mainituksi kolmannen edustajan toimesta, saadaan heidän välisen yhteisviittauksen osapainon suuruus laskemalla ensimmäiseen kansanedustajaan kohdistuvan lauseen ja toisen kansanedustajaan kohdistuvan lauseen kosini-samankaltaisuus. Jos kolmas kansanedustaja on maininnut usein kaksi ensimmäistä kansanedustajaa, mikä on aika yleistä, lasketaan kaikkien mahdollisten lauseparien väliset etäisyydet ja valitaan niistä maksimi. Summaamalla yhteen kaikki osapainot saadaan lopullinen kansanedustajien välisen yhteisviittauksen suuruus, joka toimii solmujen välisenä painona yhteisviittausverkostossa. Suuri linkin paino

⁹<https://voikko.puimula.org/>

kertoo tässä tapauksessa, että kansanedustajat ovat tulleet mainituksi usein samojen henkilöiden toimesta ja samantapaisissa lauseissa. Samaa ideaa soveltaen voidaan muodostaa myös sisältöön ja bibliografiseen kytkentään perustuva verkosto, jossa henkilöiden välisen linkin suuruus on sitä suurempi mitä useampia samoja henkilöitä he ovat maininneet samankaltaisissa lauseissa.

5.2 Verkostometriikat

Viittausanalyysissä tutkitaan suunnattua viittausverkostoa, josta selvitetään verkoston yhtenäisyys, linkkien ja solmujen määrä, tiheys ja keskimääräinen aste. Yhtenäisessä verkostossa jokaisesta solmusta löytyy polku kaikkiin muihin solmuihin. Suunnattujen verkostojen tapauksessa verkosto on heikosti yhtenäinen, jos jokaisesta solmusta löytyy polku muihin solmuihin ottamatta huomioon linkkien suunnat. Jos linkkien suunnat otetaan huomioon ja jokaisesta solmusta löytyy polku muihin solmuihin, suunnattu verkosto on vahvasti yhtenäinen.

Verkoston tiheys on verkoston linkkien määrän suhde kaikkien mahdollisten linkkien määrään, jota verkostossa voisi olla. Täydellisen verkon, jossa kaikkien solmujen välillä on linkki, tiheys on 1.0. Jos verkossa on eristäytyneitä solmuja, jotka eivät yhdisty muihin solmuihin linkkien kautta, verkon tiheys on 0. Tiheys siis kuvaa, kuinka laajasti esimerkiksi kansanedustajat keskimäärin mainitsevat toisiaan.

Linkkien määrä vastaa solmuparien määrää, joista ensimmäinen solmu viittaa toisen solmun. Summaamalla yhteen linkkien painot, viittausverkoston tapauksessa saadaan viittausten kokonaismäärä. Keskimääräinen aste on viittausten kokonaismäärä jaettuna solmujen määrällä ja se kertoo, kuinka monta viittausta solmu kohti keskimäärin tulee ja kuinka monta viittausta solmu keskimäärin tekee muihin solmuihin.

Solmujen eli kansanedustajien, puolueiden ja valiokuntien roolia viittausverkostossa voidaan myös tutkia laskemalla erilaisia metriikoita. Solmujen keskeisyyttä selvittää laskemalla niiden astekeskeisyys (*degree centrality*) ja ominaisvektorikeskeisyys (*eigenvector centrality*). Eri keskeisyysmetriikat voivat merkitä eri asioita eri verkostojen kontekstissa, mikä täytyy ottaa aina huomioon keskeisyysmetriikoiden suuruuksia tutkittaessa.

Suuntaamattomassa verkostossa solmun astekeskeisyys vastaa solmuun tulevien linkkien painojen summaa tai jos linkkien painoja ei oteta huomioon, solmun naapureiden määrää. Suunnatun viittausverkon tapauksessa ulkoaste (*out-degree*) tarkoittaa niiden solmujen määrää, johon kyseinen solmu viittaa. Ottamalla huomioon linkkien painot, ulkoaste vastaa kaikkien solmun tekemien viittausten yhteismäärää. Vastavasti sisäaste (*in-degree*) vastaa niiden solmujen määrää, jotka viittaavat kyseiseen solmuun ja ottamalla huomioon linkkien painot sisäaste vastaa kyseiseen solmuun tehtyjen viittausten yhteismäärää. Sisä- ja ulkoasteita tutkimalla saadaan nopeasti

selville eniten viittauksia tehneet ja eniten mainitut kansanedustajat.

Ominaisvektorikeskeisyyden käsitetään yleensä kertovan solmun vaikutuksesta verkostoon. Ominaisvektorikeskeisyys riippuu solmujen naapureiden määrästä sekä naapureiden naapureiden määrästä. Suuntaamattomassa verkostossa solmulla on suuri ominaisvektorikeskeisyys, jos se on yhdistynyt linkillä useisiin muihin solmuihin, joilla puolestaan on useita naapureita. Suunnatun viittausverkoston tapauksessa solmun ominaisvektorikeskeisyys on korkea, jos solmuun viittaa usea solmu, joihin puolestaan viittaa moni muu solmu.

5.3 Perinteiset yhteisviittausanalyysin menetelmät

Perinteisesti kirjoittajien yhteisviittausanalyysin suorittaminen voidaan jakaa kuuheen askeleeseen: 1) kirjoittajien valinta, 2) yhteisviittausten määrien hakeminen, 3) yhteisviittausmatriisin muodostaminen, 4) yhteisviittausmatriisin normalisointi, 5) normalisoidun matriisin monimuuttuja-analyysi sekä 6) tulosten esittäminen ja validointi [26]. Monimuuttuja-analyysissä kirjoittajia pyritään ryhmittelemään tieteenalan rakenteen selvittämiseksi käyttäen yleensä klusteroinnin, faktorianalyysin ja moniulotteisen skaalauksen menetelmiä [27]. Monimuuttuja-analyysin menetelmiä voidaan käyttää myös viittausmatriisille ja bibliografiseen kytkentään perustuvalla matriisille.

Tässä työssä yhteisviittausmatriisi, bibliografiseen kytkentään perustuva matriisi ja viittausmatriisi saadaan muodostettujen verkostojen yhteismatriisina. Viittausmatriisin tapauksessa niin kutsutut itseviittaukset asetetaan nollassa eli niitä ei huomioida. Yhteisviittausmatriisin ja bibliografiseen kytkentään perustuvan matriisin diagonaalien alkioita kohdellaan puuttuvina arvoina ja ne korvataan sarakekeskiarvoilla.

Yhteisviittausdataa on normalisoitu muodostamalla jokin läheisyysmatriisi (*proximity matrix*) yhteisviittausmatriisin pohjalta [46]. Läheisyysmatriisilla tarkoitetaan symmetristä matriisia, jossa jokainen alkio kertoo alkion riviä ja saraketta vastaavien kohteiden välisestä etäisyydestä. Alkiot voivat siis olla etäisyyksiä tai samankaltaisuuksia. Korrelaatiomatriisin sijaan monimuuttuja-analyysissä voidaan käyttää myös jonkin muun etäisyys- tai samankaltaisuusmitan, kuten Jaccardin indeksin tai kosini-samankaltaisuuden, avulla muodostettua läheisyysmatriisia [16]. Siitä miten läheisyysmatriisi pitäisi muodostaa yhteisviittausmatriisista on ollut kiistaa, eikä yksimieliseen lopputulokseen ole päästy [47].

Bibliografiseen kytkentään perustuvien verkostojen analysointiin voidaan käyttää pitkälti samoja menetelmiä kuin kirjoittajan yhteisviittausanalyysissä. Yhteisviittausmatriisin sijaan muodostetaan bibliografiseen kytkentään perustuva matriisi, jota voidaan analysoida klusteroinnin, faktorianalyysin ja moniulotteisen skaalauksen menetelmillä [48]. Tulokset täytyy tällöin tulkita bibliografisen kytkennän näkökulmasta.

Bibliografisen kytkennän tapauksessa samoihin ryhmiin luokitellut solmut ovat viitanneet muihin solmuihin samankaltaisesti. Yhteisviittausanalyysin tapauksessa taas samaan ryhmän solmuihin on viitattu usein yhdessä joku samoissa töissä tai samojen kirjoittajien toimesta, jolloin heidän oletetaan liittyvään samaan tieteelliseen alaan.

Näillä menetelmillä pyritään ryhmittelemään dataa, jonka jälkeen tutkitaan löytyykö samaan ryhmään kuuluville yhdistäviä tekijöitä, joita ei löydy muista ryhmistä. Esimerkiksi tuottavatko tietyissä ryhmässä olevat henkilöt samaan aihealueeseen liittyviä tieteellisiä julkaisuja tai tässä tapauksessa kuuluvatko samassa ryhmässä olevat henkilöt samaan puolueeseen tai puhuvatko he samoista aiheista.

Solmujen, eli kansanedustajien tai puolueiden ryhmittelemiseen käytetään agglomeratiivista hierarkkista klusterointia, jossa jokainen solmu on aluksi oma klusterinsa. Kaksi klusteria yhdistetään vuoron perään käyttäen Wardin menetelmää eli yhdistetään aina kaksi klusteria, joiden sisäinen hajonta kasvaa vähiten yhdistämisen jälkeen [49]. Klustereita yhdistellään kunnes haluttu määrä klustereita tai yksi klusteri on jäljellä.

Faktorianalyysin menetelmistä tässä työssä käytetään pääfaktorianalyysiä. Komponenttien määrän valitsemiseen käytettiin Kaiserin sääntöä, jossa ensin suoritetaan pääfaktorianalyysi ilman rotaatiota siten, että komponenttien määrä vastaa muuttujien eli esimerkiksi tutkittavien kansanedustajien tai puolueiden määrää. Seuraavaksi tutkitaan, kuinka monen komponentin ominaisarvo on yli 1.0 ja näiden tätä määrää käytetään komponenttien määränä varsinaisessa faktorianalyysissä [50]. Faktorianalyysissä käytettiin tulosten selkeyttämiseksi Promax-rotaatiota. Rotaatio muuntaa pääfaktorianalyysin tuottamat komponentit helpommin tulkittaviksi, jolloin parhaassa tapauksessa jokainen komponentti vastaa yhtä piilevää muuttujaa [51]. Faktorianalyysi ei jaa muuttujia vain yhteen klusteriin, vaan yksi muuttuja voi latautua useampaan komponenttiin. Muuttujan suuri positiivinen tai negatiivinen latautumisen tiettyyn komponenttiin viittaa muuttujan ja komponentin välillä olevaan yhteyteen.

Moniulotteisessa skaalauksessa datapisteet pyritään esittämään alemmissa dimensioissa siten, että samankaltaiset datapisteet ovat lähellä toisiaan myös alempi dimensioisessa kartoituksessa [52]. Tällöin "keskeiset" datapisteet eli paljon muiden kanssa yhdessä mainitut kansanedustajat ovat keskeisillä koordinaateilla, kun taas heikommin linkitetty kansanedustajat eli vähemmän muiden kanssa mainitut edustajat ovat pistejoukon laidoilla moniulotteisessa avaruudessa [53]. Tässä tapauksessa paljon yhteisviitattut kansanedustajat eli kansanedustajat, jotka on mainittu usein samojen kolmansien kansanedustajien toimesta, ovat lähekkäin toisiaan moniulotteisessa avaruudessa. Moniulotteinen skaalaus ei suoraan ryhmittele datapisteitä, mutta sen tuloksena syntyneitä kuvaajaa voidaan käyttää klusteroinnin ja faktorianalyysin tulosten tukemiseen. Moniulotteisen skaalauksen merkitys visualisointimenetelmänä on pienentynyt verkostanalyysin visualisointimenetelmien kehityttyä [37].

5.4 Muita verkostanalyysin menetelmiä

Viittausanalyysiin liittyvissä esimerkeissä mainittiin HITS-algoritmi, jonka avulla voidaan laskea keskus- ja auktoriteettiarvoja suunnatun verkoston solmuille. Keskuksen ja auktoriteettien määrittely on iteratiivinen: hyvä keskus mainitsee monta hyvää auktoriteettia ja hyvä auktoriteetti tulee mainituksi monen hyvän keskuksen toimesta [22]. Viittausanalyysin tapauksessa korkean auktoriteettiarvon on ajateltu merkitsevän laadukasta artikkelia ja korkea keskusarvo viittaa artikkelin käyttävän laadukkaita lähteitä, jolloin ihanteellisella artikkelilla molemmat arvot ovat korkeita [21]. Tämän työn tapauksessa korkea auktoriteettiarvo ei tarkoita, että kansanedustajalla olisi esimerkiksi enemmän valtaa muihin nähden, vaan tällaisten henkilöiden voidaan olettaa olevan näkyviä henkilöitä parlamentaarisisessa keskustelussa ja he tulevat mainituksi usein puheenvuoroissa. Keskuksia taas mainitsevat usein tällaisia henkilöitä. Jos henkilöllä on sekä korkea auktoriteettiarvo että keskusarvo, on henkilö oletettavasti keskeinen keskustelija eduskunnassa.

Bibliometriikassa käytetyimmät algoritmit verkostojen solmujen sijaintien laskemiseen ovat Kamada-Kawai, joka pyrkii pitämään solmujen euklidisen etäisyyden visualisoinnissa samana kuin solmujen välinen lyhyin polku, ja Fruchterman-Reingold, joka pyrkii pitämään solmut sitä lähempänä toisiaan, mitä suurempi niiden välisen linkin paino on [37]. Joskus solmujen ryhmittymät voidaan nähdä suoraan verkostojen visualisoinneista. Tämän työn tapauksessa verkostojen, erityisesti yhteisviittausverkostojen ja bibliografiseen kytkentään perustuvien verkostojen, tiheyden todettiin olevan erittäin suuri. Tämä tarkoittaa, että jokaisesta solmusta menee linkki melkein kaikkiin loppuihin solmuihin ja tällöin solmujen väliset lyhyimmät polut ovat myös todella lyhyitä, jolloin Kamada-Kawai ei pystynyt juurikaan erottelemaan solmuja toisistaan. Lisäksi ryhmittymiä voi etsiä käyttämällä yhteisöjen etsimiseen tarkoitettuja algoritmeja [37]. Esimerkki tällaisesta algoritmista on Louvain-menetelmä [54]. Tässä työssä solmujen ryhmittelyyn käytetään kuitenkin vain perinteisempiä viiteanalyysin menetelmiä, jotka perustuvat viittausverkostojen yhteismatriisin käyttöön.

6 Tulokset

Tässä osiossa käydään läpi erilaisiin henkilömainintoihin pohjautuviin verkostoihin liittyviä analyysituloksia. Kaikissa analyyseissä henkilömaininnalla tarkoitetaan puhujan kanssa samalle vaalikaudelle valitun kansanedustajan mainitsemista vähintään kerran puheenvuorossa. Puhemiehen puheenvuoroja ja ruotsinkielisiä puheenvuoroja ei ole otettu huomioon, eikä myöskään viittauksia henkilöön itseensä tai puolueiden tapauksessa mainintoja puheenvuoron pitäjän oman puolueen jäseniin.

6.1 Puolueiden välinen viittausverkosto

Puolueille muodostettiin viittausverkostoja vaalikausien 2015–2018 ja 2019–2022 pidettyjen puheenvuorojen pohjalta. Vaalikaudella 2015–2018 hallituspuolueina aloittivat Suomen Keskusta (Kesk.), Kansallinen Kokoomus (Kok.) ja Perussuomalaiset (PS). Oppositiossa vaikuttivat Suomen Sosialidemokraattinen puolue (SDP), Vasemmistoliitto (Vas.), Vihreä Liitto (Vihr.), Suomen Kristillisdemokraatit (KD) sekä Suomen ruotsalainen kansanpuolue (RKP). Vuoden 2017 hallituskriisin seurauksena Korjausliike (KL) irtaantui Perussuomalaisista ja Perussuomalaiset siirtyivät oppositioon. Lisäksi vaalikauden lopulla syntyi kaksi "yhden miehen puoluetta", Liike Nyt (Liik.) ja Seitsemän tähden liike (TL). Puolueiden määrä kasvoi siis kahdeksasta yhteentoista vaalikauden aikana.

Tätä työtä tehdessä vaalikausi 2019–2022 on vielä kesken, joten kaikkia vaalikauden puheenvuoroja ei ole saatavilla. Vaalikaudella 2019–2022 hallituspuolueina toimivat SDP, Keskusta, Vasemmistoliitto, Vihreät ja RKP. Oppositiossa olivat Kokoomus, PS, Liike Nyt, KD sekä myöhemmin Perussuomalaisista erotetun Ano Turtiaisen Valta kuuluu kansalle -puolue (VKK). Vaalikauden ensimmäisellä puoliskolla yhteensä kymmenellä puolueella oli kansanedustaksi valittuja poliitikoita.

Puolueiden välisten mainintasuhteiden tutkimista varten muodostettiin kansanedustajien puheenvuoroissa tekemien henkilömainintojen pohjalta kolme viittausverkostoa. Kaksi viittausverkostoa muodostettiin vaalikauden 2015–2018 puheenvuorojen pohjalta jakamalla vaalikausi kahteen ajanjaksoon. Ensimmäinen ajanjakso on vaalikauden alusta hieman yli vaalikauden puolen välin aina 12.6.2017 saakka, jonka jälkeen Sininen tulevaisuus, nykyisin Korjausliike, erkaantui Perussuomalaisista. Datassa ensimmäinen Korjausliikkeen jäsenen puheenvuoro on päivältä 13.6.2017. Toinen viittausverkosto on muodostettu puheenvuorojen pohjalta, jotka on pidetty 13.6.2017 alkaen aina vaalikauden 2015–2018 loppuun asti.

Kolmas viittausverkosto on muodostettu vaalikauden 2019–2022 alkupuoliskon puheenvuoroista, jotka on pidetty 24.4.2019–2.7.2021 välillä eli vuoden 2021 kesätaukoon mennessä. Vaalikauden 2019–2022 toista puoliskoa ei oteta tässä työssä huomioon, sillä tätä työtä tehdessä osaa sen ajanjakson puheenvuoroja ei ole vielä saatu dataan mukaan eikä kaikkia vaalikauden puheenvuoroja ole edes pidetty. Viittausverkostoja analysoimalla pyritään esimerkiksi selvittämään, miten henkilömai-

ninnat jakautuvat puoluiden välille ja mainitaanko hallitus- ja oppositio puolueiden kansanedustajia eri tavalla. Koska vaalikauden vaihtuessa monet puolueet siirtyivät hallituksesta oppositioon tai oppositiosta hallitukseen, voidaan myös arvioida vaikuttaako puolueen rooli kansanedustajien tekemiin viittauksiin.

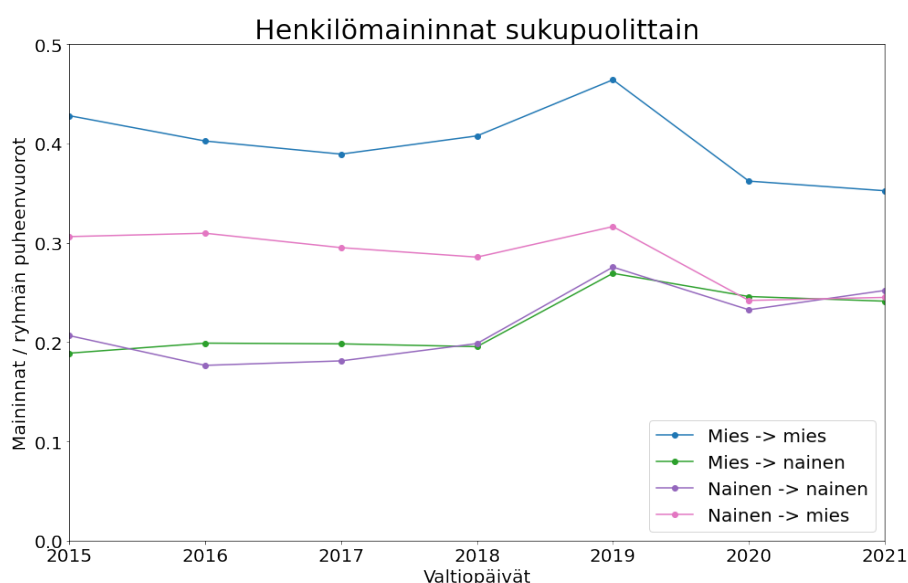
Taulukossa 2 on listattu valittuihin ajanjaksoihin ja niihin liittyviin viittausverkostoihin liittyvää tietoa. Niiden puheenvuorojen osuus, jotka sisältävät vähintään yhden henkilömaininnan saman vaalikauden kansanedustajiin on miltei sama jokaiselle ajanjaksolle, eli jokaisella ajanjaksolla noin 40% puheenvuoroista sisältävät vähintään yhden saman vaalikauden kansanedustajaan kohdistuvan henkilömaininnan. Ministerimainintojen osuus kaikista maininnoista, kun puhujan oman puolueen jäseniin kohdistuvia henkilömainintoja ei ole otettu huomioon, vaihtelee hieman. Vaalikaudella 2015–2018 Sipilän hallituksessa jaettiin aluksi 14 ministerisalkkua pääministerin pestin lisäksi. Huhtikuussa 2017 ministerien määrää lisättiin kolmella, mutta tämä ei kuitenkaan lisännyt ministereihin kohdistuvien henkilömainintojen osuutta vaalikauden 2015–2018 toiseen puoliskoon perustuvassa viittausverkostossa. Vaalikaudella 2019–2022 ministeriksi nimitettiin yhteensä 19 kansanedustajaa pääministeri mukaan lukien, ja ministereiden osuus henkilömaininnoista nousi, minkä saattaa osaltaan selittää ministerien suuri määrä. Kuten vaalikauden 2015–2018 tapauksessa nähtiin, ministerien määrä ei kuitenkaan välttämättä yksistään nosta ministereiden osuutta kaikista maininnoista.

Vaalikaudella 2019–2022 naisten osuus sekä ministereissä että puolueiden puheenjohtajissa nousi verrattuna edelliseen vaalikauteen, mikä näkyy selkeästi naiskansanedustajiin kohdistuvien viittausmäärissä kuvan 4 mukaan. Vastaavasti niiden puheenvuorojen osuus, joissa on mainittu vähintään yksi mieskansanedustaja on pienentynyt valtiopäivillä 2020 ja 2021 verrattuna aiempaan. Kuvassa on esitetty valtiopäivittäin niiden puheenvuorojen määrä, joissa mainitaan vähintään yksi mies- tai naiskansanedustajaa jaettuna puheenvuoron pitäjän sukupuolen valtiopäivien aikana pitämien puheenvuorojen määrällä. Miehet mainitsevat selkeästi naisia useammin puheenvuoroissaan henkilöitä. Sekä miehet että naiset ovat maininneet puheenvuoroissaan enemmän miehiä kuin naisia, mihin vaikuttaa myös se että mieskansanedustajia on valittu jonkin verran naiskansanedustajia enemmän eduskuntaan. Vaalikaudella 2019–2022 mainittujen naisten osuus on selvästi noussut ja valtiopäivillä 2021 naiset ovat maininneet useammassa puheenvuorossa naiskansanedustajia kuin mieskansanedustajia.

Verkostot ovat tiheitä, ja 22.4.2015–12.6.2017 aikana jokaisen puolueen välillä on maininta suuntaan ja toiseen. Kahdella muulla ajanjaksolla pienet, yhden tai muutaman kansanedustajan muodostamat puolueet pienentävät verkoston tiheyttä. Verkoston tiheydestä ja suurista viittausmääristä johtuen monimuuttuja-analyysissä on vaikeuksia erottaa puolueita toisistaan, sillä yhteisviittauslukemat puolueiden välillä kasvavat suuriksi. Tällöin suurin erottava tekijä on viittausmäärä, joka rajoittaa myös yhteisviittausmäärää ja yhteisviittausanalyysi lähinnä erottelee vähän viittauksia saaneet puolueet paljon viittauksia saaneista puolueista. Tämän

Ajanjakso	Päivät	Puolueet	Puheenvuorot	Maininnat	Puheenvuorot, joissa maininta	Ministerien osuus maininnoista	Maininta-verkoston tiheys
22.4.2015–12.6.2017	783	8	27467	14378	0.41	0.27	1.0
13.6.2017–16.4.2019	673	11	23634	12096	0.39	0.26	0.83
24.4.2019–2.7.2021	801	10	25292	13353	0.41	0.33	0.92

Taulukko 2: Päivien, puolueiden, puheenvuorojen ja mainintojen määrät valituilla ajanjaksoilla ja niiden puheenvuorojen osuus, joissa on vähintään yksi henkilömaininta muiden kuin puhujan puolueen jäseniin, ministerimainintojen osuus kaikista maininnoista (oman puolueen ministereihin kohdistettuja mainintoja ei ole otettu huomioon) sekä muodostettujen mainintaverkostojen tiheys.



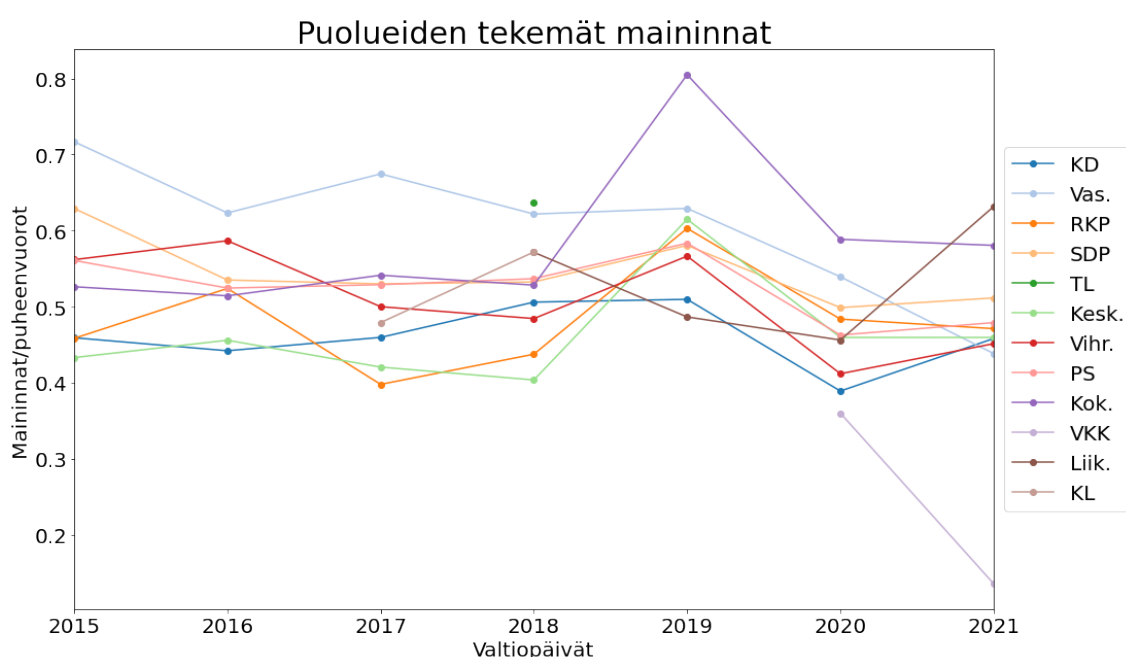
Kuva 4: Puheenvuorojen osuus valtiopäivittäin, joissa mieskansanedustaja mainitsee vähintään yhden nais- tai mieskansanedustajan miesten pitämässä puheenvuoroissa tai vastaavasti naiskansanedustaja mainitsee vähintään nais- tai mieskansanedustajan naisten pitämässä puheenvuoroissa.

vuoksi monimuuttuja-analyysissä käytetään puolueiden kohdalla viittausmatriisia yhteisviittausmatriisin sijaan Leydesdorffin ja Vaughanin suositusten mukaan [55].

Kuvassa 5 esitetään kuinka monta henkilömainintaa kunkin puolueen jäsenet ovat tehneet keskimäärin puheenvuoroissaan samalla vaalikaudella toimineisiin, muiden puolueiden kansanedustajiin valtiopäivittäin. Vaalikauden aloittavat valtiopäivillä, erityisesti valtiopäivillä 2019 henkilömainintojen suhteellinen osuus puheenvuoroissa on monella puolueella suurempi verrattuna muuhun vaalikauteen. Tähän saattaa vaikuttaa muun muassa hallitusohjelmaan liittyvä keskustelu.

Vasemmistoliiton ja Vihreän liiton osalta on havaittavissa, että henkilömainin-

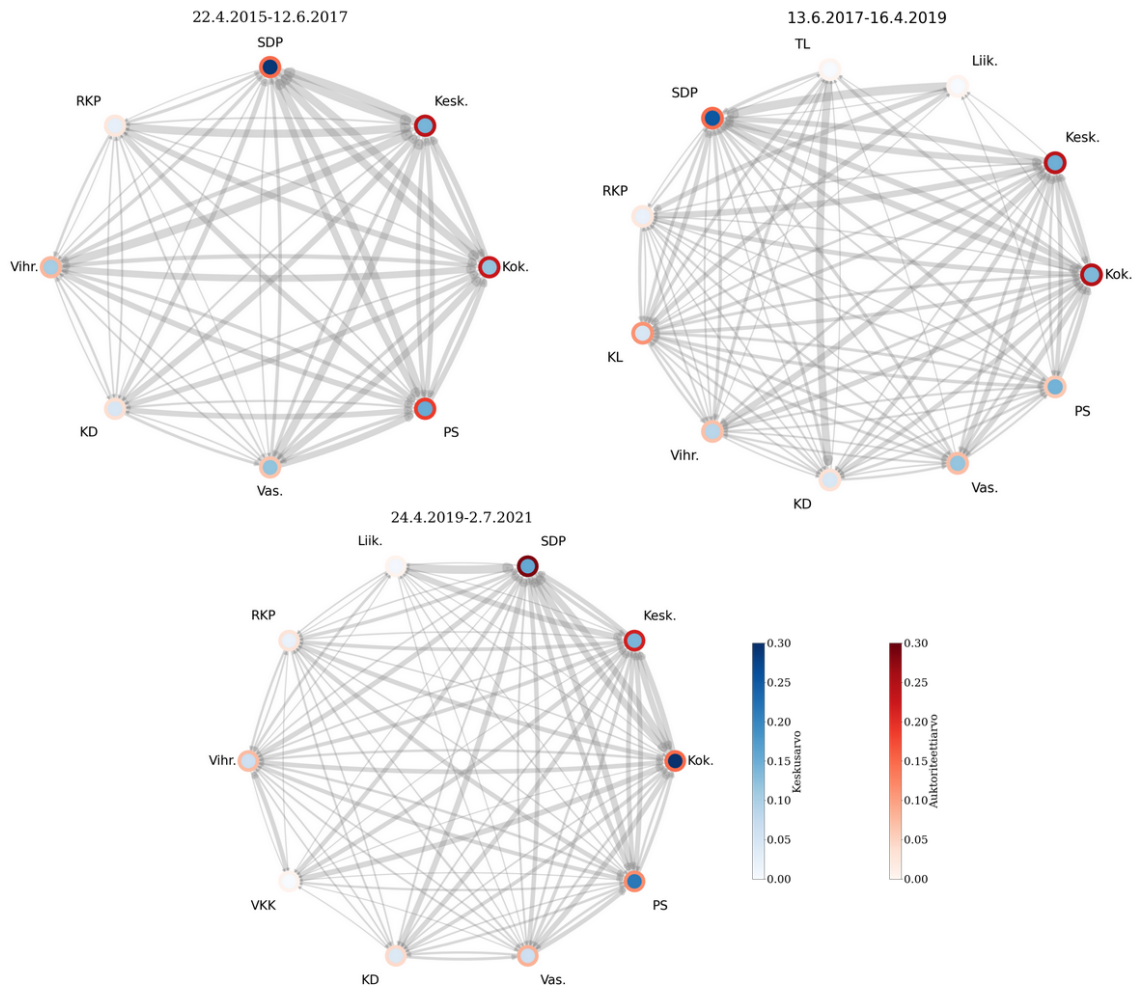
tojen suhteellinen määrä puheenvuoroissa on vähentynyt puolueiden siirryttyä oppositiosta hallitukseen vaalikauden 2015–2018 vaihduttua vaalikauteen 2019–2022. Vastaavasti Kokoomuksen kansanedustajilla henkilömainintojen suhteellinen osuus puheenvuoroissa on kasvanut puolueen siirryttyä oppositioon. Muutosten ei kuitenkaan voida tulkita johtuvan pelkästään opposition ja hallitusten välisestä erosta. Esimerkiksi Keskusta-puolueella henkilömainintojen osuus on myös kasvanut vaalikauden vaihtuessa, vaikka puolue pysyi hallituksessa. Toisaalta puolueen rooli vaihtui pääministeripuolueesta vaalitappion jälkeen pienemmäksi hallituspuolueeksi. Vastaavasti Perussuomalaisilla henkilömainintojen osuuksien pieneni vaalikauden vaihtuessa, vaikka puolue pysyi oppositiossa.



Kuva 5: Keskimääräinen henkilömainintojen määrä suomenkielisissä puheenvuoroissa toisten puolueiden kansanedustajiin valtiopäivittäin.

Muodostetut puolueiden väliset viittausverkostot on esitetty kuvassa 6. Puolueet on väritetty puolueiden keskus- ja auktoriteettiarvojen mukaan. Linkkien paksuus kertoo, kuinka suuri osa lähdepuolueen kaikista henkilömaininnoista kohdistuu kohdepuolueeseen. Esimerkiksi ensimmäisessä verkostossa suurin osa Vihreän liiton kansanedustajien tekemistä henkilömaininnoista kohdistuu Keskustan kansanedustajiin, mikä voi johtua esimerkiksi erimielisyyksistä ympäristöasioista. Suuri auktoriteettiarvo kertoo, että puolueeseen kohdistuu paljon mainintoja. Suuri keskusarvo kertoo taas, että puolueen kansanedustajat mainitsevat paljon muita henkilöitä puheenvuoroissaan. Puolueiden auktoriteetti- ja keskusarvoihin vaikuttaa erityisesti puolueen koko ja rooli. Suuremmilla puolueilla, kuten SDP:llä, Keskustalla, Kokoomuksella ja Perussuomalaisilla auktoriteetti- ja keskusarvot ovat lähtökohtaisesti suurempia, sillä näihin puolueisiin kuuluu enemmän kansanedustajia, jotka voivat tehdä mainintoja

ja joihin voidaan kohdistaa mainintoja.



Kuva 6: Puolueiden mainintaverkostot eri ajanjaksoina. Solmut on väritetty puolueiden keskus- ja auktoriteetti-arvojen mukaan. Linkkien paksuus vastaa kuinka suuri osuus lähdepuolueen kaikista tekemistä maininnoista kohdistuu kohdepuolueen kansanedustajiin.

Kuten aiemmin nähtiin, jokaisella valitulla ajanjaksolla yli neljäsosa kaikista henkilömaininnoista kohdistuu ministereihin. Tämä on päällimmäinen syy, joka nostaa hallituspuolueiden auktoriteetti-arvoa. Esimerkiksi ensimmäisessä mainintaverkostossa hallituspuolueilla Keskustalla, Kokoomuksella ja Perussuomalaisilla on suurin auktoriteetti-arvo. Perussuomalaisien hajotessa ja siirryttyä oppositioon puolueen auktoriteetti-arvo laski selvästi, mihin vaikuttaa myös puolueen jäsenmäärän pientyminen. Perussuomalaisista erkaantuneen hallituspuolueen Korjausliikkeen auktoriteetti-arvo on taas korkeampi kuin Perussuomalaisilla toisessa mainintaverkostossa, vaikka puolueiden kansanedustajien määrällä ei ollut suurta eroa. Hallitukseen siirtyminen näyttää myös laskevan puolueen keskusarvoa. Esimerkiksi viimeisessä

viittausverkostossa SDP:n ja Vasemmistoliiton keskusarvo on selkeästi laskenut puolueiden siirryttyä hallitukseen, kun taas Kokoomuksen keskusarvo on selvästi noussut oppositiossa.

Taulukossa 3 on esitetty kullekin ajanjaksolle puolueisiin liittyviä tilastoja ja metriikoita. Puolueen kansanedustajien määrä taulukossa vastaa kaikkien kyseisellä ajanjaksolla eduskunnassa puolueen nimissä vaikuttaneiden kansanedustajien määrää. Määrä saattaa siten erota hieman puolueen varsinaisten edustajapaikkojen määrästä, mikä itsessään voi vaihdella esimerkiksi kansanedustajan vaihtaessa puoluetta. Sisästeen eli puolueeseen kohdistuvien viittausten määrä näyttäisi olevan samassa linjassa auktoriteettiarvojen kanssa. Oppositiopuolueilla ulkoaste on usein sisäastetta suurempi, eli puolueiden kansanedustajat mainitsevat enemmän henkilöitä mitä kuin puolueen kansanedustajiin kohdistuu. Hallituspuolueilla tilanne on usein päinvastoin.

Puolueen pitämien puheenvuorojen määrä riippuu luonnollisesti puolueen kansanedustajien määrästä. RKP:n puheenvuorojen määrä on taulukon mukaan pieni, koska vain suomenkieliset puheenvuorot on otettu huomioon. Niiden puheenvuorojen osuudet, joissa on ainakin yksi henkilömaininta toisten puolueiden kansanedustajiin vaihtelevat puolueittain. Kuten aiemmin todettiin, esimerkiksi Kokoomuksella, Vihreällä liitolla ja Vasemmistoliitolla mainintojen osuus muuttui puolueiden vaihdettua roolia opposition ja hallituksen välillä. Tämä näkyy selkeästi myös taulukossa. Joillain puolueilla, kuten RKP:llä ja KD:llä puolueiden mainintoja sisältävien puheenvuorojen osuus näyttäisi pysyvän suhteellisen matalana muihin puolueisiin verrattuna. Pienempien puolueiden tapauksessa voi käydä niin, että puolueesta ei yksinkertaisesti tule valittua eduskuntaan kansanedustajaa, joka erityisen aktiivisesti mainitsee muita kansanedustajia osoittaen esimerkiksi tukea tai haastaen. Myös Keskustan puheenvuorojen osuus, joissa mainitaan muiden puolueiden kansanedustajia on matala. Tähän vaikuttaa puolueen rooli hallituksessa, mutta verrattuna muihin hallituspuolueisiin lukema on silti aika alhainen.

Monimuuttuja-analyysiä varten joukosta tiputettiin pois puolueet, jonka jäseniin oli valittuna ajanjaksona kohdistunut vain alle 50 henkilömainintaa. Kuten aiemmin mainittiin, analyyseissä päätettiin käyttää viittausmatriisia. Kuten aiemmissakin analyyseissä, henkilömainintoja puolueen omiin kansanedustajiin ei otettu huomioon ja vastaavat viittausmatriisin alkiot asetettiin nolaksi. Viittausdata normalisoitiin muuntamalla viittausmatriisi kosini-samankaltaisuuksien matriisiksi.

Kuvassa 7 näkyy kosini-samankaltaisuuksiin perustuvan matriisin pohjalta tehty MDS-kartoitus, jossa puolueet on väritetty hierarkkisen klusteroinnin tuottamien klustereiden mukaan. Näyttäisi siltä, että MDS tukee klusteroinnin tuottamia tuloksia, sillä samaan klusteriin kuuluvat puolueet ovat lähekkäin myös MDS-kartoituksissa. Tässä tapauksessa puolueiden ollessa lähekkäin MDS-kartoituksessa voidaan olettaa, että puolueisiin kohdistuu samankaltaisesti viittauksia muista puolueista. Esimerkiksi Vasemmistoliitto ja Vihreä liitto ovat lähekkäin jokaisessa MDS-kartoituksessa, mikä viittaa puolueiden kerryttävän henkilömainintoja muilta puolueilta samankaltaisilla

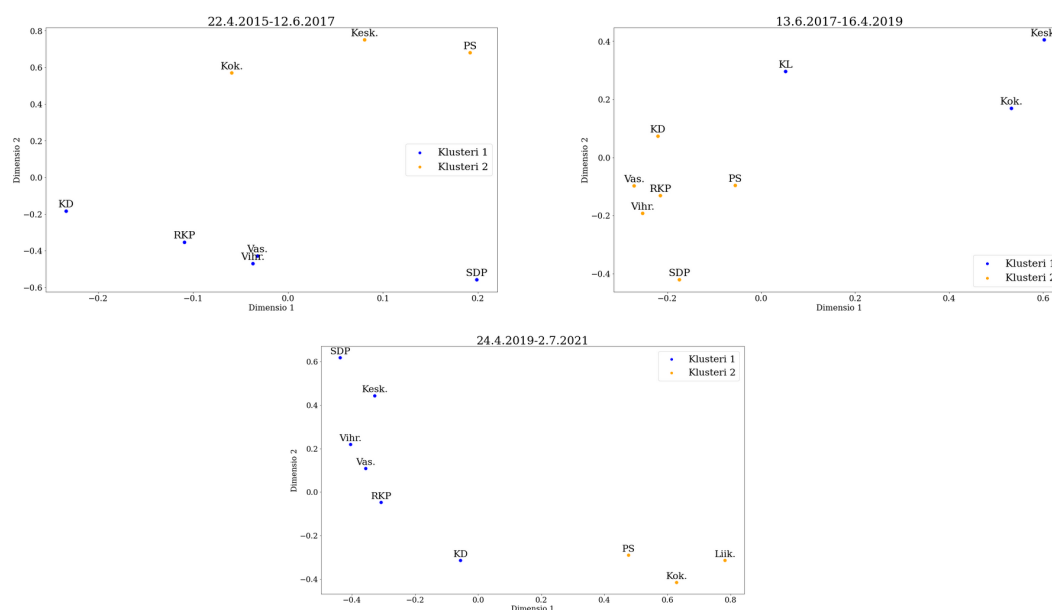
22.4.2015–12.6.2017						
Puolue	Edustajat	Puheenvuorot	Puheenvuorot, joissa henkilömaininta	Ulkoaste	Sisääste	Omainaisvektori-keskeisyys
SDP	36	6257	0.452	3533	2679	0.354
Kesk.	52	5809	0.329	2531	3145	0.354
PS	39	4765	0.389	2518	2450	0.354
Kok.	37	3934	0.407	2010	2905	0.354
Vihr.	15	2297	0.455	1298	1230	0.354
Vas.	12	2260	0.515	1503	1052	0.354
KD	5	1453	0.39	651	538	0.354
RKP	11	692	0.383	334	379	0.354
13.6.2017–16.4.2019						
Puolue	Edustajat	Puheenvuorot	Puheenvuorot, joissa henkilömaininta	Ulkoaste	Sisääste	Omainaisvektori-keskeisyys
Kesk.	52	5329	0.325	2214	2694	0.299
SDP	36	4781	0.426	2550	2094	0.342
Kok.	42	3639	0.417	1989	2694	0.322
PS	18	3104	0.39	1694	815	0.299
Vas.	12	1994	0.494	1269	935	0.299
Vihr.	15	1905	0.403	939	869	0.322
KL	21	1202	0.385	652	1258	0.342
KD	5	1128	0.384	557	418	0.322
RKP	10	534	0.341	221	295	0.318
TL	1	11	0.364	7	8	0.227
Liik.	2	7	0.571	4	16	0.187
24.4.2019–2.7.2021						
Puolue	Edustajat	Puheenvuorot	Puheenvuorot, joissa henkilömaininta	Ulkoaste	Sisääste	Omainaisvektori-keskeisyys
Kok.	38	5326	0.492	3446	2622	0.337
SDP	41	5163	0.401	2679	3335	0.337
PS	42	5126	0.391	2575	1899	0.337
Kesk.	32	4304	0.382	2089	2525	0.337
Vihr.	20	1823	0.331	830	870	0.337
Vas.	17	1583	0.399	870	1028	0.304
KD	5	1160	0.347	488	606	0.275
RKP	10	620	0.361	298	378	0.304
Liik.	1	125	0.392	64	56	0.275
VKK	1	62	0.194	14	34	0.308

Taulukko 3: Puolueiden välisten mainintaverkostojen ominaisuuksia. Edustajien määrä vastaa kyseisellä ajanjaksolla eduskunnassa vaikuttaneiden kansanedustajien määrää. Puheenvuorojen määrissä ei ole otettu huomioon puhemiehen puheenvuoroja. Puheenvuorot, joissa on maininta vastaa niiden puheenvuorojen osuutta kaikista puheenvuoroista, joissa edustaja on maininnut vähintään yhden muuhun puolueeseen kuuluvan kansanedustajan.

kaavoilla. Jokaisesta MDS-kartoituksesta puuttuu selkeästi keskeinen puolue, jonka bibliometriikassa tulkittaisiin olevan "monialainen". Lähimpänä tätä on ehkä toisessa viittausverkostossa (13.6.2017–16.4.2019) Perussuomalaiset.

Klusterointi jakaa puolueet hallitus- ja oppositiopuolueisiin. Vain kolmannen viittausverkoston tapauksessa KD on luokiteltu virheellisesti hallituspuolueeksi. MDS-kartoitus sijoittaa KD:n oppositio- ja hallituspuolueiden väliin. Tulokset viittaavat siis siihen, että hallituspuolueet erottuvat oppositiopuolueista sen perusteella miten puolueiden kansanedustajat mainitsevat toisten puolueiden kansanedustajia puheenvuoroissaan. Poistamalla ministereihin kohdistuvat maininnat tulokset pysyvät muu-

ten samana, mutta KL toisessa viittausverkostossa ryhmittyy oppositiopuolueiden klusteriin hallituksen sijaan. Ministereihin kohdistuvat maininnat eivät siis aiheuta ryhmittymistä hallitus- ja oppositiopuolueisiin. Puolueen roolilla näyttäisi siis olevan suurin vaikutus siihen, miten muut puolueet mainitsevat kyseisen puolueen kansanedustajia eikä esimerkiksi kyseisen puolueen arvoilla.



Kuva 7: Puolueiden MDS-karttoitus eri ajanjaksoilla viittausmatriisien perusteella. Puolueet on väritetty hierarkkisen klusteroinnin tuottamien tulosten mukaan.

Analyysit toistettiin myös normalisoimalla viittausdata muuttamalla viittausmatriisi Pearsonin korrelaatiomatriisiksi sekä käyttämällä yhteisviittausmatriisia viittausmatriisiin sijaan. Käyttämällä viittausmatriisia tulokset olivat pääosin samat normalisointitavasta riippumatta. Yhteisviittausmatriisin tapauksessa normalisointi kosini-samankaltaisuuden avulla tuotti selkeämmät tulokset ja puolueet ryhmittyivät pääosin oppositio- ja hallituspuolueisiin. Käyttämällä Pearsonin korrelaatioiden avulla normalisoitua yhteisviittausmatriisia joukosta erottui ensisijaisesti pienet puolueet, joihin kohdistuvien viittausten määrä oli vähäisin.

6.2 Valiokuntien välinen viittausverkosto

Työssä kokeiltiin myös, vaikuttaako kansanedustajan valiokuntajäsenyydet siihen, miltä tahoilta häneen tulee viittauksia käyttämällä vaalikauden 2015–2018 puheen- vuoroja. Jokainen kansanedustajan ministereitä ja puhemiestä lukuun ottamatta kuuluu keskimäärin kahteen valiokuntaan, mistä syystä ministerien ja puhemiehen tekemät sekä heihin kohdistuvat maininnat jätettiin pois. Myös edustajien itseensä tekemät maininnat jätettiin pois. Seuraavissa analyyseissä valiokuntiin kuuluviksi

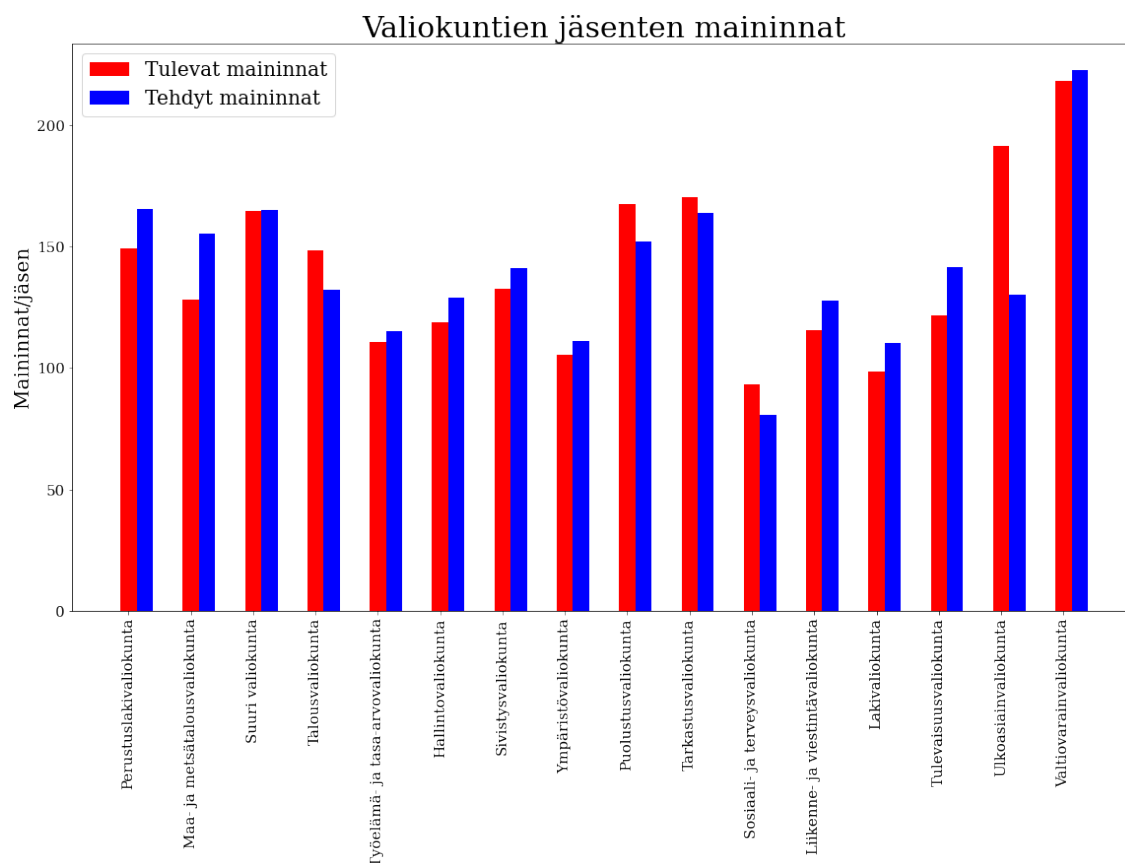
laskettiin vain varsinaiset jäsenet ja varajäsenet jätettiin huomioimatta. Samoin kuin puolueiden tapauksessa, kansanedustajien tekemät ja heihin kohdistuvat henkilömaininnat ryhmiteltiin niiden valiokuntien välille, joiden jäsenyydet kansanedustajilta löytyy henkilömaininnan aikaan.

Moni kansanedustaja on kuului useampaan valiokuntaan. Noin puolet vaalikauden 2015–2018 kansanedustajista kuuluivat yhteensä kolmeen tai neljän valiokuntaan vaalikauden aikana ja muutama kansanedustaja kuului jopa kuuteen valiokuntaan vaalikauden aikana. Valiokuntien kokoonpanoissa on siis oletettavasti tapahtunut muutoksia. Valiokuntien tapauksessa yksi kansanedustaja voi kuulua useampaan valiokuntaan ja valiokunnissa toimii sekä oppositio- että hallituspuolueiden kansanedustajia. Puolueiden tapauksessa nähtiin, kuinka hallituspuolueisiin ja oppositio- puolueisiin viitataan eri tavalla, jolloin ne voi erottaa toisistaan viittausten avulla. Koska valiokuntien tapauksessa hallitus- ja oppositiopuolueiden kansanedustajat ovat sekoittuneet, ei mitään selkeää valiokuntien ryhmittymistä ollut havaittavissa.

Kuvassa 8 on esitetty valiokuntaan tulevat ja valiokunnan tekemät henkilömaininnat jäsentä kohti. Ministereihin kohdistuvia sekä ministereiden tekemiä mainintoja ei ole otettu huomioon. Valiokunnan jäsenien määränä käytetään valiokunnan jäsenien määrää, joka vaihtelee valiokunnittain. Lisäksi mukaan lasketaan kunkin valiokunnan puheenjohtaja. Useimpien valiokuntien jäsenet tekevät suhteessa enemmän henkilömainintoja kuin valiokuntien jäseniin kohdistuu. Tehtyjen henkilömainintojen ja tulevien henkilömainintojen ero on kuitenkin usein pieni. Sosiaali- ja terveysvaliokunnan jäsenien kohdalla sekä jäseniin kohdistuvien mainintojen sekä jäsenien tekemien henkilömainintojen määrä on selkeästi alhaisempi muihin valiokuntiin verrattuna. Puolustus- ja ulkoasiainvaliokuntiin tulee selkeästi enemmän mainintoja kuin kyseisten valiokuntien jäsenet tekevät muiden valiokuntien jäseniin. Tehtyjen mainintojen määrä ei ole välttämättä pienempi jos otetaan huomioon myös ministereihin tehdyt maininnat.

Se, mihin valiokuntaan kansanedustajan kuuluu, vaikuttaa siihen, mistä asioista kansanedustajan on päättämässä. Jotkin asiat herättävät enemmän keskustelua kuin toiset, jolloin valiokuntatausta saattaa vaikuttaa kansanedustajaan kohdistuvien viittausten määrään. Kuten puolueisiin liittyvässä analyysissä nähtiin, oppositio- ja hallitusroolit vaikuttavat selkeästi siihen, miltä taholta ja missä määrään kansanedustajaan kohdistuu henkilömainintoja. Tämä viittaa siihen, että esimerkiksi kansanedustajan puolue vaikuttaa enemmän millaisia viittauksia kansanedustajaan kohdistuu kuin valiokuntatausta, sillä kansanedustajat ajavat usein puolueensa asiaa. Tällöin myös valiokunnan henkilömainintojen määrä riippuu enemmän siitä, keitä kansanedustajia valiokuntaan sattuu kuulumaan kuin siitä, mikä valiokunta on kyseessä. Esimerkiksi Antti Rinne, Matti Vanhanen, Ben Zyskowicz ja Jutta Urpilainen kuuluivat ulkoasiainvaliokuntaan ja heihin kohdistui yhteensä lähes 1000 henkilömainintaa enemmän kuin he tekivät kuuluessaan valiokuntaan. Tämä yksistään vastaa melkein eroa ulkoasiainvaliokuntaan tulevien ja ulkoasiainvaliokunnassa tehtyjen henkilömainintojen määrässä. Kaikki neljä ovat tunnettuja poliitikkoja, jolloin hei-

dän valiokuntataustansa ei yksinään selitä heihin kohdistuvien henkilömainintojen määrää. Esimerkiksi Rinne toimi SDP:n puheenjohtajana, minkä vuoksi häneen on voinut kohdistua enemmän mainintoja ja Zyskovicziin aktiivisena keskustelijana kohdistuu paljon henkilömainintoja.



Kuva 8: Valiokuntien jäseniin kohdistuvat ja heidän tekemät henkilömaininnat jäsentä kohti.

6.3 Kansanedustajien välinen viittausverkosto

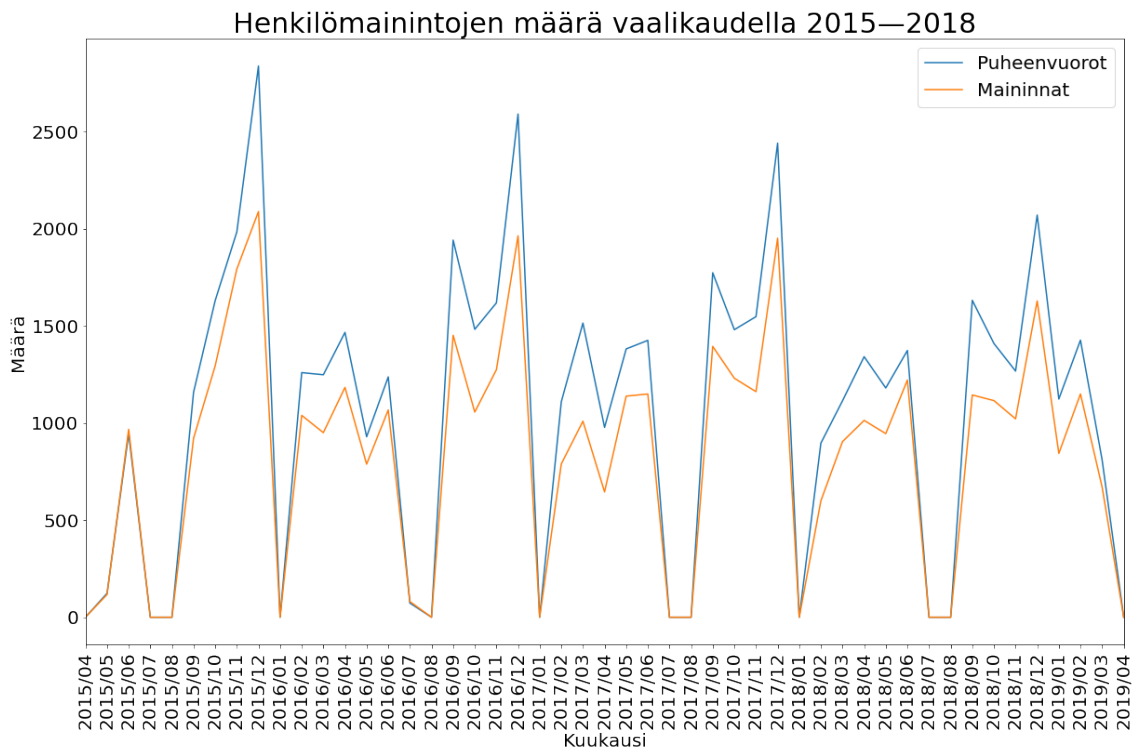
Kun tutkittiin kansanedustajien välisiä mainintoja, käytettiin vaalikauden 2015–2018 puheenvuoroja. Ottamatta huomioon puhemiehen puheenvuoroja, vaalikaudella 2015–2018 pidettiin yli 51 000 puheenvuoroa. Niistä yli 24 000:ssa mainittiin vähintään yksi samalla vaalikaudella kansanedustajana toiminut henkilö. Jokaista puheenvuoroa kohti tehtiin noin 0.73 henkilömainintaa. Puheenvuoroissa, joissa tehtiin henkilömaininta mainittiin keskimäärin 1.54 henkilöä. Vaalikaudella vaikutti 214 kansanedustajaa, joista 93 oli naisia. Hallituksella oli vaalikauden alussa 124 paikkaa kahdestasadasta ja vaalikauden lopussa 103 paikkaa. Saman vaalikauden kansanedustajiin tehdyistä henkilömaininnoista 24% kohdistui ministereihin.

Kansanedustajien välinen viittausverkosto on yhtenäinen verkosto, joka sisältää 213 solmua eli kansanedustajaa. Linkkejä verkostossa on noin 12 500 perustuen yli 37 000 henkilömainintaan. Kansanedustajapareja, joiden välillä on viittauksia on verkostossa 9406, ja niistä noin 40% välillä on vain yksi viittaus suuntaan tai toiseen. Henkilömainintoja vaalikauden 2015–2018 kansanedustajien välillä oli keskimäärin 175 kansanedustajaa kohti. Viittausverkoston tiheys on 0.29 eli varsin moni kansanedustaja on jättänyt suuren osan muista vaalikauden kansanedustajista mainitsematta puheenvuoroissaan.

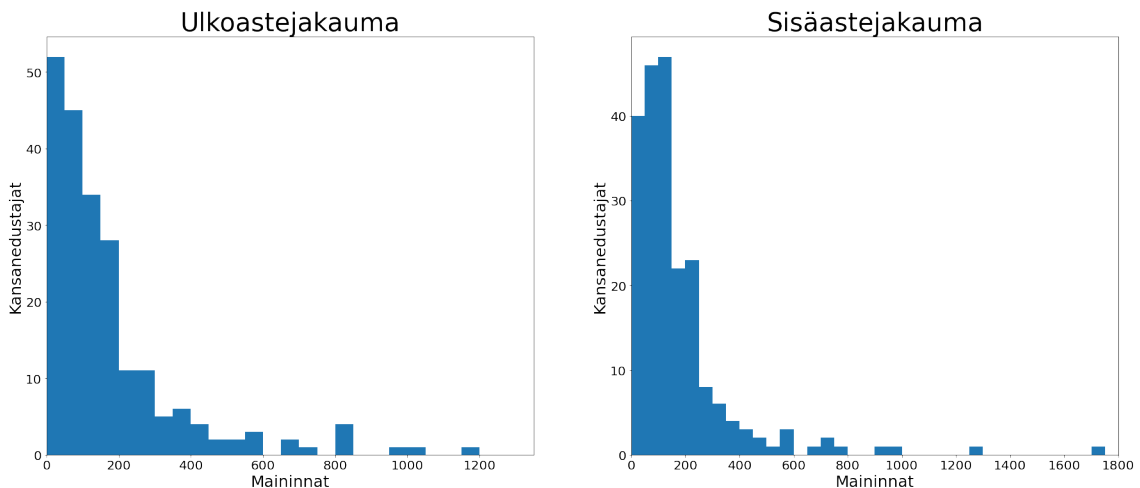
Kuvassa 9 on esitetty kuukausittain puheenvuorojen ja niistä tunnistettujen saman vaalikauden kansanedustajiin kohdistuvien henkilömainintojen määrä vaalikaudella 2015–2018. Mainintojen määrä seuraa puheenvuorojen määrää ja näyttääkin siltä, että mainintojen määrä suhteessa puheenvuorojen määrään pysyy varsin tasanaisena. Kesä- ja joulutauoista johtuen heinä-, elo- ja tammikuussa puheenvuorojen määrä putoaa lähelle nollaa. Syysistuntokaudella pidetään enemmän puheenvuoroja ja siten myös mainitaan enemmän henkilöitä verrattuna kevätistuntokauteen. Joulukuussa puheenvuorojen ja sitä myötä myös mainintojen määrä on aina valtiopäivien korkein. Joulukuulle ajoittuu erilaisia vuosikatsauksia, kuten oikeusasiamiehen kertomus edelliseltä vuodelta. Lisäksi joulukuussa käsitellään hallituksen esitys eduskunnalle valtion talousarvioksi seuraavalle vuodelle, mikä herättää paljon keskustelua.

Verkostolle laskettiin erilaisia perusmetriikoita keskeisten kansanedustajien tunnistamiseksi. Lasketut metriikat ovat ominaisvektorikeskeisyys, henkilömainintojen määrää kuvaavat, sisä- ja ulkoasteet sekä auktoriteetti- ja keskusarvot. Ottaen huomioon linkkien painot, suurin muihin kansanedustajiin lähtevien viittausten määrä eli suurin ulkoaste on SDP:n Pia Viitasella, joka on maininnut muita kansanedustajia yhteensä 1183 kertaa puheenvuoroissaan. Vastaavasti suurin sisäaste eli suurin määrä muilta kansanedustajilta saatuja viittauksia on pääministeri Juha Sipilällä (Kesk.), joka on mainittu 1725 puheenvuorossa. Sen lisäksi että Sipilään pääministerinä kohdistuu paljon mainintoja, jotkin häneen linkitetyistä henkilömaininnoista ovat itseasiassa mainintoja Sipilän hallitukseen eikä Juha Sipilään itseensä. Jos linkkien painoja ei oteta huomioon, sisäaste vastaa henkilöiden määrää, jotka ovat maininneet kohdesolmun ainakin yhdessä puheenvuorossa. Tällöinkin suurin sisäaste on Juha Sipilällä, jonka on maininnut 177 eri kansanedustajaa vaalikauden aikana. Vastaavasti ottamatta huomioon linkkien painot, suurin ulkoaste on Timo Heinosella, joka on maininnut 173 eri kansanedustajaa puheenvuoroissaan. Kansanedustajista SDP:n Pia Viitaseen ja Eero Heinäluoman välillä on 153 mainintaa, mikä on eniten mainintoja kansanedustajaparin välillä. Näistä maininnoista 105 on Viitaseen Heinäluomaan kohdistamia mainintoja.

Kuvassa 10 näkyvät verkoston ulko- ja sisäastejakaumat eroavat hieman toisistaan. Ulkoastejakauma on enemmän oikealle vino ja sillä on lyhyempi häntä kuin sisäastejakaumalla. Ulkoastejakauman alakvartiili, mediaani ja yläkvartiili ovat 55, 115 ja 203. Vastaavasti sisäastejakauman alakvartiili, mediaani ja yläkvartiili ovat 65, 121 ja 212. Henkilömaininnat kohdistuvat siis jossain määrin tiettyihin henkilöihin.



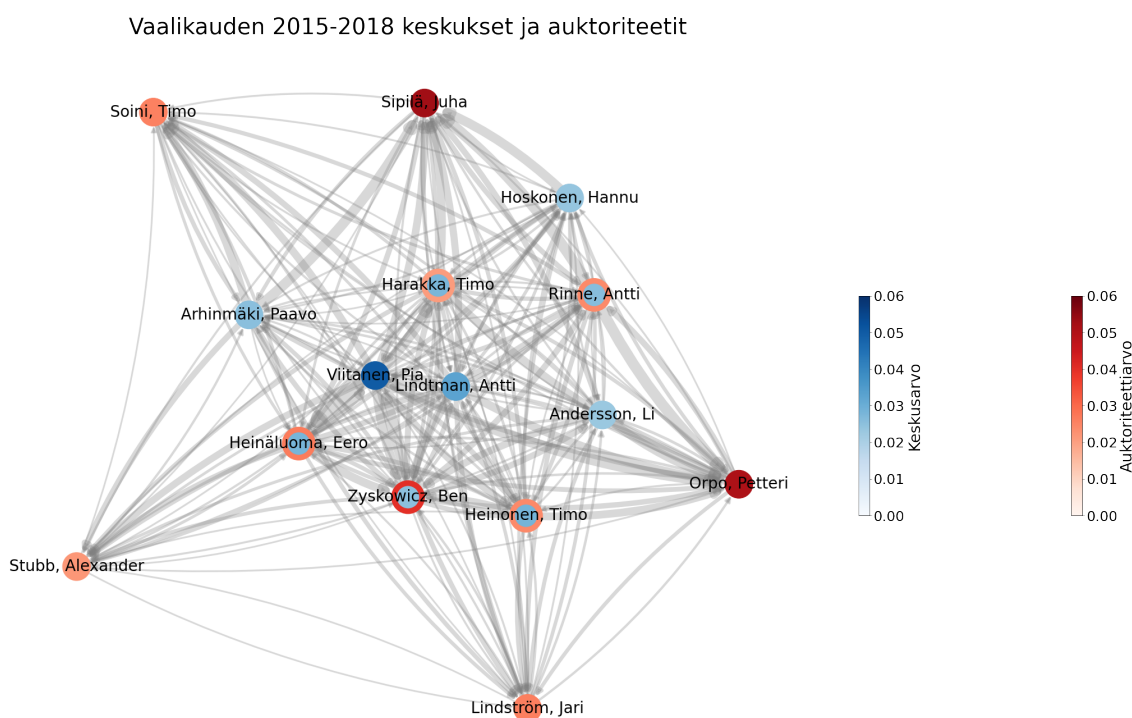
Kuva 9: Puheenvuorojen ja henkilömainintojen määrä kuukausittain vaalikaudella 2015–2018.



Kuva 10: Kansanedustajien viittausverkoston ulko- ja sisäastejakaumat perustuen vaalikauden 2015–2018 kansanedustajien välisiin viittauksiin.

Kuvassa 11 on esitetty verkoston kymmenen suurimman auktoriteetti- ja keskusarvon omaavat solmut eli kansanedustajat. Jotkut kansanedustajista ovat sekä suurimpien auktoriteettiarvojen että keskusarvojen omaavien solmujen joukossa. Kor-

keiden auktoriteettiarvojen joukossa on ministereinä toimineita henkilöitä (Sipilä, Orpo, Soini, Stubb, Lindström). Loput kuvassa nähtävissä olevista auktoriteeteista ovat myös keskuksia eli he mainitsivat paljon muita henkilöitä, jolloin on luonnollista että heidätkin mainitaan usein. Heistä Heinonen ja Zyskowitz edustavat Kokoomusta, joka oli hallituspuolue. Rinne, Harakka ja Heinäluoma edustavat SDP:tä eli he olivat mukana oppositiossa. Rinne toimi myös SDP:n puheenjohtajana, mikä saattoi nostaa häneen kohdistuvien henkilömainintojen määrää. Myös Sipilä, Soini ja Stubb, jonka myöhemmin korvasi Orpo, toimivat puolueidensa puheenjohtajina vaalikauden aikana. Loput keskussolmuista, jotka eivät kuulu auktoriteetteihin, ovat Viitanen (SDP), Lindtman (SDP), Arhinmäki (Vas.), Andersson (Vas.) ja Hoskonen (Kesk.). Seitsemän kymmenestä solmusta, joilla on suurin auktoriteettiarvo, ovat hallituspuolueen edustajia ja loput kolme oppositiopuolueiden edustajia. Keskussolmujen kohdalla tilanne on päinvastoin.



Kuva 11: Viittausverkoston kymmenen suurinta keskusta ja auktoriteettia. Linkkien paksuus riippuu mainintojen määrästä.

Taulukossa 4 on esitetty puolue, puheenvuorojen määrä, sisä- ja ulkoaste sekä ominaisvektorikeskeisyys kuvassa 11 esiintyvillä kansanedustajille. Kymmenestä mainituimmista henkilöistä taulukosta puuttuu Keskustan Kimmo Tiilikainen, joka toimi aluksi Maatalous- ja ympäristöministerinä ja myöhemmin Asunto-, energia- ja ympäristöministerinä. Hän oli vaalikauden 2015–2018 kymmenenneksi mainituin henkilö 577 maininnalla. Kymmenestä mainituimmista henkilöstä seitsemän on hallituspuolueiden edustajia, joista Sipilä, Orpo, Lindström, Soini ja Tiilikainen toimivat

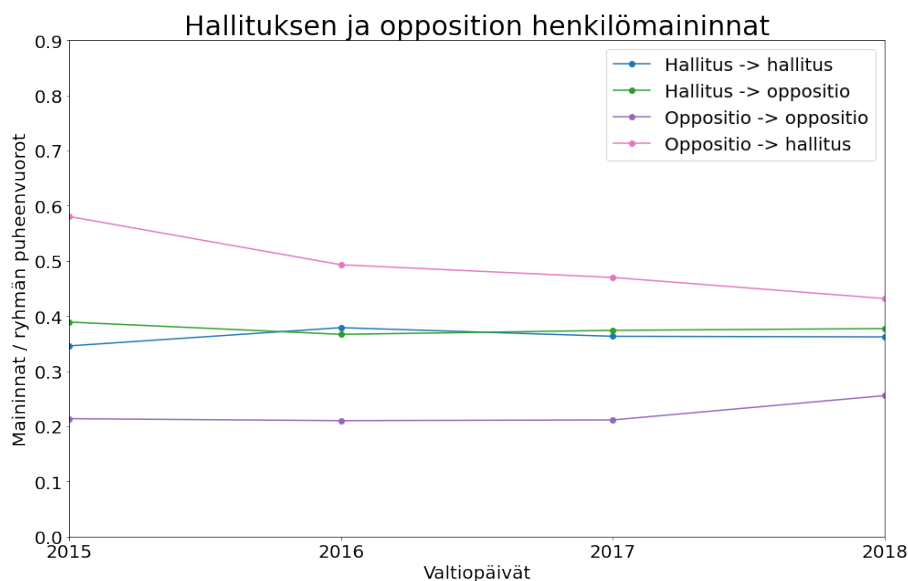
ministereinä. Mainituimmat henkilöt, jotka eivät toimineet ministereinä, ovat Antti Rinnettä lukuun ottamatta myös sekä kymmenen eniten puheenvuoroja pitäneiden ja kymmenen eniten muita kansanedustajia maininneiden listalla. Tämä viittaa siihen, että he ovat osallistuneet aktiivisesti keskusteluun, jolloin heidät on myös mainittu usein. Antti Rinne puolestaan toimi SDP:n puheenjohtajana, mikä voi vaikuttaa häneen kohdistuvien mainintojen määrään. Suuri sisäaste näyttäisi myös johtavan verkostossa suureen ominaisvektorikeskeisyyteen.

Kymmenestä aktiivisimmasta henkilömainintojen tekijöistä taulukosta puuttuu Jari Myllykoski (Vas., 844 mainintaa), Leena Meri (PS, 692 mainintaa) ja Simon Elo (PS/KL, 657 mainintaa). Kukaan kymmenestä eniten henkilömainintoja tehneestä henkilöstä ei toiminut ministerinä tai puolueensa puheenjohtajana vaalikauden aikana. Heistä kuusi on oppositiopuolueiden kansanedustajia mukaan lukien Leena Meri, joka siirtyi Perussuomalaisten mukana hallituksesta oppositioon. Eniten mainintoja puheenvuoroa kohden taulukon 4 kansanedustajista ovat tehneet Ben Zyskowitz (1.40), Antti Lindtman (1.36) ja Paavo Arhinmäki (1.36). Kaikista kansanedustajista eniten henkilömainintoja puheenvuoroja kohden on tehnyt puolustusministerinä toiminut Jussi Niinistö, joka mainitsi keskimäärin 3.78 henkilöä puheenvuoroa kohden. Pidettyjä puheenvuoroja hänelle kertyi 115 vaalikauden aikana.

Edustaja	Puolue	Puheenvuorot	Sisäaste	Ulkoaste	Ominaisvektori-keskeisyys
Sipilä, Juha	Kesk.	672 (8)	1725 (1)	406 (21)	0.153 (1)
Orpo, Petteri	Kok.	619 (12)	1293 (2)	363 (25)	0.138 (3)
Zyskowitz, Ben	Kok.	690 (7)	1001 (3)	963 (3)	0.128 (5)
Lindström, Jari	PS/KL	529 (19)	907 (4)	541 (14)	0.139 (2)
Heinonen, Timo	Kok.	929 (3)	799 (5)	1045 (2)	0.137 (4)
Soini, Timo	PS/KL	326 (56)	742 (6)	357 (26)	0.125 (6)
Rinne, Antti	SDP	492 (24)	733 (7)	454 (17)	0.112 (13)
Heinäluoma, Eero	SDP	727 (6)	695 (8)	804 (7)	0.120 (8)
Viitanen, Pia	SDP	1051 (1)	579 (9)	1183 (1)	0.107 (14)
Harakka, Timo	SDP	652 (9)	560 (11)	700 (8)	0.123 (7)
Hoskonen, Hannu	Kesk.	970 (2)	539 (12)	831 (5)	0.117 (12)
Stubb, Alexander	Kok.	205 (99)	469 (13)	129 (98)	0.099 (24)
Lindtman, Antti	SDP	603 (14)	384 (18)	821 (6)	0.101 (22)
Arhinmäki, Paavo	Vas.	439 (32)	371 (20)	595 (11)	0.098 (28)
Andersson, Li	Vas.	743 (5)	369 (21)	570 (12)	0.106 (15)

Taulukko 4: Kansanedustajat, jotka ovat kymmenen suurimman auktoriteetti- tai keskusarvon omaavien solmujen joukossa sekä heidän puheenvuorojen määrä, sisä- ja ulkoaste ja ominaisvektorikeskeisyys. Suluissa oleva lukuarvo kertoo, kuinka monenneksi suurin arvo kansanedustajalla on kaikista kansanedustajista.

Kuvassa 12 on esitetty, kuinka suuressa osassa hallituksen puheenvuoroja mainittiin vähintään yksi hallitus- tai oppositiopuolueiden kansanedustaja ja vastaavasti kuinka suuressa osassa oppositiopuolueiden puheenvuoroja mainittiin vähintään yksi hallitus- tai oppositiopuolueiden kansanedustaja. Hallituspuolueiden edustajat mainitsevat suunnilleen yhtä monessa puheenvuorossa sekä hallituspuolueiden ja oppositiopuolueiden edustajia. Oppositiopuolueiden edustajat mainitsevat taas selkeästi suuremmassa osassa puheenvuorojaan vähintään yhden hallituspuolueen edustajan



Kuva 12: Hallituspuolueiden ja oppositiopuolueiden kansanedustajien hallitus- ja oppositiopoliitikkoihin kohdistamien henkilömainintojen määrät valtiopäivittäin jaetuna vastaavasti hallituksen ja opposition puheenvuorojen määrällä.

kuin oppositiopuolueen edustajan, mikä nousee oppositiopuolueiden roolista haastaa hallituksen tekemiä päätöksiä.

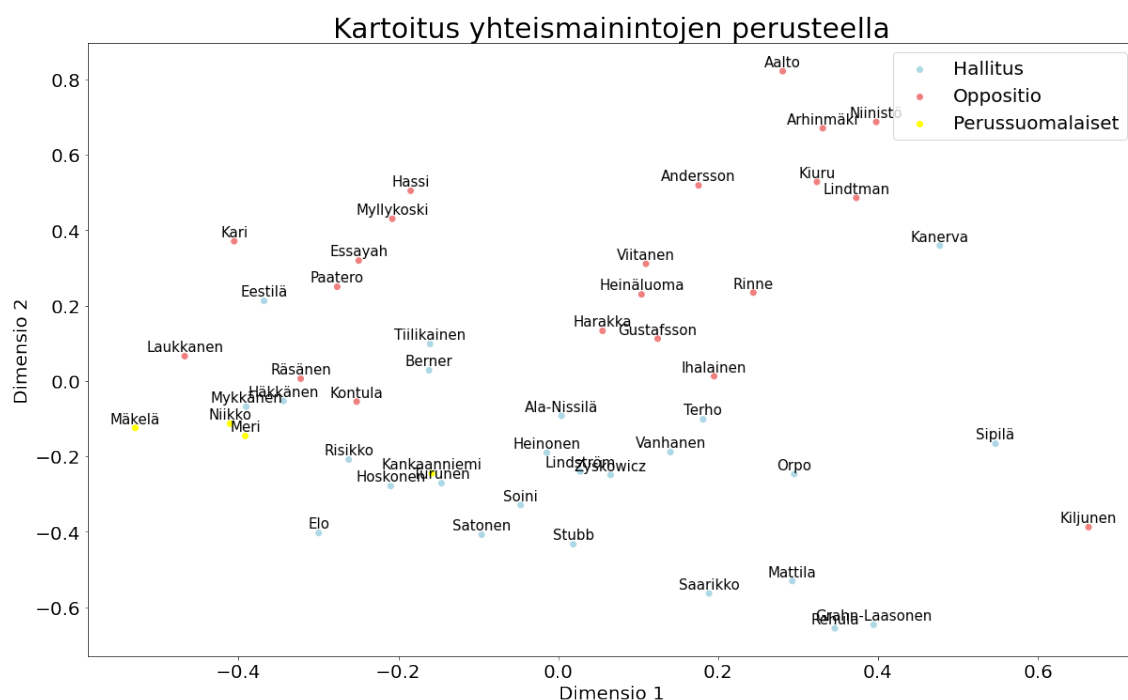
6.4 Kansanedustajien yhteisviittausanalyysi

Jotta saataisiin selvitettyä, miten samankaltaisesti kansanedustajiin viitataan muodostettiin kansanedustajien välinen yhteisviittausverkosto. Yhteisviittauksia kokeiltiin määritellä useammalla eri tavoilla, jotka on esitelty luvuissa 3.1 ja 5.1 sekä asettamalla 30 päivän aikaikkuna viittauksille, joiden perusteella henkilö voidaan laskea yhteisviitatuksi. Joka tapauksessa verkoston tiheys oli todella lähellä lukua 1, eli melkein kaikkien mahdollisten solmuparien välillä oli linkki. Klusterointimenetelmät ja faktorianalyysi eivät tuottaneet selkeästi tulkittavissa olevia klustereita. Seuraavissa esiteltävissä tuloksissa on käytetty niin kutsutulla minimimenetelmällä laskettuja yhteisviittauksia, eli henkilöiden välinen yhteisviittaus on sitä suurempi, mitä useammin sama kansanedustaja on maininnut molemmat henkilöt vaalikauden aikana.

Rajaamalla kansanedustajien määrä 50 eniten mainittuun, klusterointimenetelmät ja MDS pystyivät jossain määrin jakamaan poliitikkoja oppositio- ja hallituspuolueiden kansanedustajiin. Esimerkiksi kuvassa 13 on minimimenetelmän avulla muodostetun yhteisviittausverkoston 50 viitatuimman solmun MDS-kartoitus, jossa kansanedustajien pitäisi olla sitä lähempänä toisiaan, mitä samankaltaisemmin muut kansanedustajat ovat maininneet heitä. Kuvassa alalaidasta löytyy hallituspuolueiden kansanedustajia, kun taas kuvan ylälaidasta löytyy oppositiopuolueiden edustajia.

Oppositiopuolueiden edustajat näyttäisivät jakautuneen kahteen ryhmään. Ylhäällä keskellä ja oikealla on pelkästään SDP:n, Vasemmistoliiton ja Vihreän liiton kansanedustajia, jotka käsitetään perinteisesti vasemmistolaisiksi puolueiksi. Ylhäällä vasemmalla on muutamia henkilöitä näistä puolueista ja lisäksi kristillisdemokraatteja sekä hallituksesta oppositioon siirtyneitä perussuomalaisia, jotka taas ovat arvoiltaan enemmän oikealla. Voi olla, että henkilön puolueen parlamentaarisen roolin lisäksi itse puolue voi jossain määrin vaikuttaa siihen, miten henkilöön viitataan.

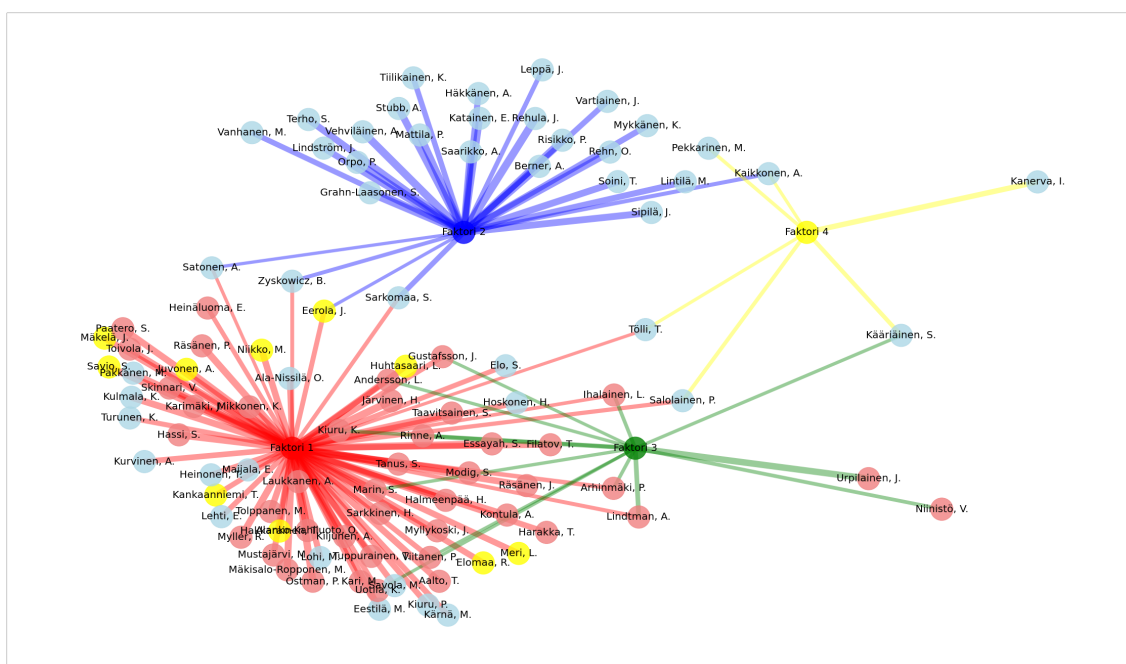
Ne 10 henkilöä, joilla on verkostossa suurin astekeskeisyys ovat Hoskosta mukaanottamatta myös viittausverkoston 10 suurinta auktoriteettiarvon omaavaa kansanedustajaa. Yhteisviittausverkoston ja viittausverkoston keskeisimmät henkilöt ovat osittain samoja, koska henkilöön kohdistuvien viittausten määrä rajoittaa sitä, kuinka usein henkilö voi tulla yhteisviitatuksi muiden henkilöiden kanssa. Suurin linkin paino on Juha Sipilän ja Petteri Orpon välillä, eli moni kansanedustaja on maininnut usein heidät molemmat.



Kuva 13: MDS-karttoitus vaalikauden 2015–2018 50 viitatuimmasta kansanedustajasta yhteisviittausmatriisin pohjalta. Yhteisviittausmatriisi on normalisoitu muodostamalla Pearsonin korrelaatiomatriisi yhteisviittausdatan pohjalta.

Ottamalla huomioon lauseet, joissa henkilömaininnat tapahtuivat hallituksen ja opposition välinen ero selkiytyi hieman. Kuvassa näkyy sadan mainituimman kansanedustajien latautumisen neljään faktoriin. Kansanedustajasta menee linkki faktoriin, jos kansanedustajan lataus faktoriin on vähintään 0.4. Oppositioapuolueiden edustajat on merkitty punaisella, hallituspuolueiden edustajat sinisellä ja perussuomalaiset

harmaalla. Ensimmäiseen faktoriin kuuluu sekä oppositio- että hallituspuolueiden jäseniä sekä kaikki perussuomalaiset. Toiseen ja neljanteen faktoriin kuuluu vain hallituspuolueiden edustajia ja kolmanteen faktoriin lähinnä muutamia oppositio- puolueiden edustajia. Kymmenen suurinta astetta omaavat kansanedustajat olivat samoja kuin tavallisessa yhteisviittausverkostossakin, mutta suurin aste oli Pette- ri Orpolla Juha Sipilän sijaan. Lisäksi koska ottamalla huomioon lauseet, joissa henkilömaininta tapahtui, otettiin mukaan kaikki samassa puheenvuorossa samaan henkilöön kohdistuneet maininnat mainintojen määrä nousi yli 3000 maininnalla. Mainintoja löytyi nyt yli 40 000.



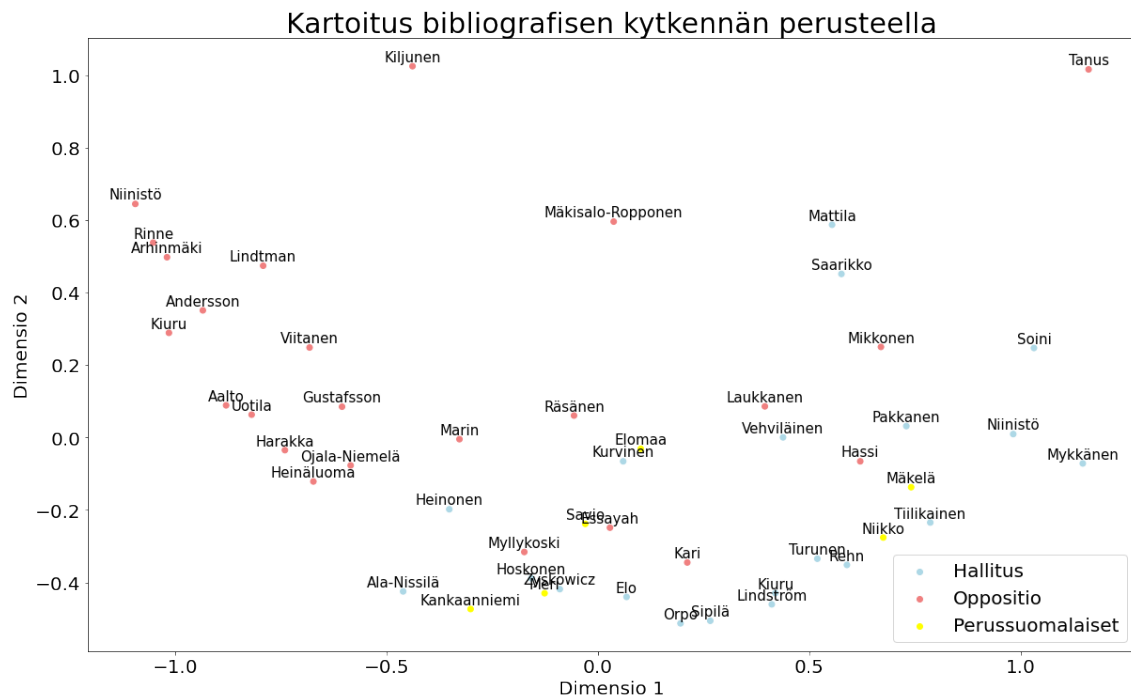
Kuva 14: 100 viitatuimman kansanedustajan latautuminen faktoreihin yhteisviittaus- ten pohjalta konteksti huomioiden. Jussi Niinistö ei kuulunut mihinkään faktoriin vähintään 0.4 suuruudella. Neljä faktoria selittivät 89% varianssista.

6.5 Kansanedustajien välisiin bibliografisiin kytkentöihin pohjautuva analyysi

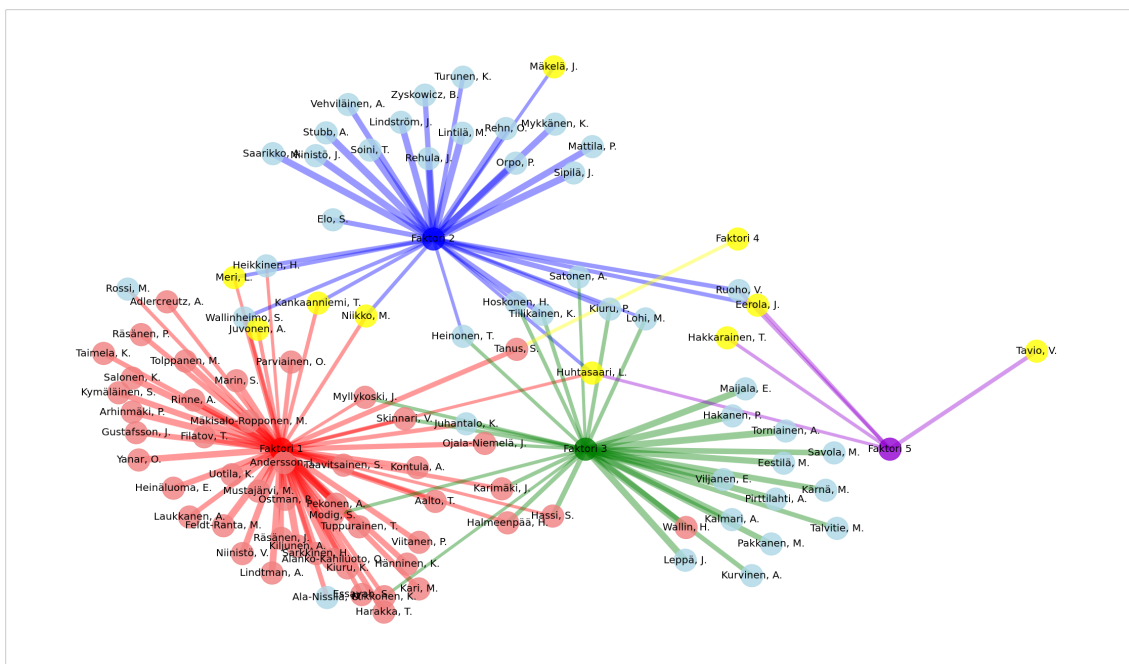
Myös bibliografiseen kytkentään pohjautuvat kansanedustajien väliset verkostot oli- vat hyvin tiheitä ja selkeiden klustereiden löytäminen verkostoista ei onnistunut. Kuvassa 15 esitetään 50 eniten henkilömainintoja tehneiden kansanedustajien MDS- kartoitus, kun bibliografisten kytkentöjen suuruudet on laskettu minimenetelmää käyttäen. Kartoituksessa lähekkäin olevat henkilöt ovat maininneet usein samoja henkilöitä puheenvuoroissaan. Kuvan vasemmassa laidassa on oppositiopoliitikkojen ryhmittymä, ja hallituspuolueiden edustajat sijoittuvat pääosin kuvan oikeaan alalai- taan. Selkeitä ryhmittymiä kuvassa ei kuitenkaan ole. 50 eniten viittauksia tehneiden kansanedustajien joukossa on enemmän oppositiopuolueiden edustajia kuin 50 eniten

viittauksia saaneiden kansanedustajien joukossa. Verkostossa kymmenen kansanedustajaa, joilla on suurin aste, ovat myös 10 eniten mainintoja tehnyttä kansanedustajaa ja heistä seitsemän kuuluu viittausverkoston suurimpiin keskuksiin.

Myös bibliografisen kytkentään perustuvan verkoston tapauksessa ero hallituksen ja opposition selkeytyi, kun viittauslauseet otettiin huomioon. Kuvassa 16 näkyy 100 eniten viittauksia tehneiden kansanedustajien latautuminen faktoreihin. Faktoriin yksi kuuluu lähinnä oppositiopuolueiden edustajia ja faktoreihin kaksi ja kolme hallituspuolueiden edustajia. Faktoriin neljä kuuluu vain kristillisdemokraattien Sari Tanus. Faktoriin viisi kuuluu perussuomalaisia sekä Veera Ruoho, joka loikkasi Perussuomalaisista Kokoomukseen. Tässäkin tapauksessa henkilö lasketaan kuuluvaksi faktoriin, jos hänen latauksensa faktoriin on vähintään 0.4.



Kuva 15: MDS-kartoitus vaalikauden 2015–2018 50 eniten viittauksia tehneestä kansanedustajasta bibliografisten kytkentöjen pohjalta. Yhteisviittausmatriisi on normalisoitu muodostamalla Pearsonin korrelaatiomatriisi yhteisviittausdatan pohjalta.



Kuva 16: 100 eniten viittauksia tehneiden kansanedustajien latautuminen faktoreihin yhteisviittausten pohjalta konteksti huomioiden. Sami Savio ja Ritva Elomaa eivät latautuneet mihinkään faktoriin vähintään 0.4 suuruudella ja puuttuvat siten kuvasta. Viisi faktoria selittivät 85% varianssista.

7 Johtopäätökset

Tässä työssä esiteltiin, kuinka eduskunnassa pidetyissä puheenvuoroissa esiintyvien henkilömainintojen pohjalta voi muodostaa erilaisia kansanedustajien ja puolueiden välisiä verkostoja sekä kuinka näitä verkostoja voidaan analysoida viiteanalyysin ja verkostoanalyysin menetelmin. Työssä käytiin myös läpi, miten viiteanalyysi ja verkostoanalyysi liittyvät toisiinsa. Verkostoista pyrittiin löytämään verkoston kannalta keskeiset solmut, jotka ovat aktiivisesti mukana puheenvuorojen kautta käydyissä vuorovaikutuksissa sekä tunnistamaan mahdollisia kaavoja viittausten taustalla.

Työssä muodostettiin puolueiden välisiä viittausverkostoja kolmelle eri ajankaksolle ja tutkittiin miten puolueen parlamentaarinen rooli, hallitus tai oppositio, vaikuttaa puolueen jäsenien tekemiin ja vastaanottamiin viittauksiin. Puolueita pystyttiin niiden jäseniin kohdistuvien viittausten perusteella erottelemaan hallitus- ja oppositiopuolueisiin.

Vaalikauden 2015–2018 puheenvuoroissa esiintyvien henkilömainintojen pohjalta muodostettiin kansanedustajien välinen viittausverkosto, yhteisviittausverkosto ja bibliografiseen kytkentään pohjautuva verkosto. Yhteisviittausverkostosta luotiin myös versio, jossa viittausten määrien lisäksi otettiin huomioon viittauslauseet. Viittausverkoston kymmenen mainituinta henkilöä olivat merkittäviä solmuja myös yhteisviittausverkostossa ja vastaavasti kymmenen eniten mainintoja tehneitä henkilöä olivat keskeisiä solmuja bibliografiseen kytkentään pohjautuvassa verkostossa. Eniten viittauksia kohdistui henkilöihin, jotka toimivat ministereinä ja henkilöihin, jotka mainitsivat itse paljon muita kansanedustajia puheissaan.

Kansanedustajia yritettiin ryhmitellä sen perusteella, kuinka usein heidät on mainittu samojen kansanedustajien toimesta sekä kuinka samankaltaisesti he ovat maininneet muita kansanedustajia. Ryhmittelyt eivät tuottaneet selkeitä klustereita, mutta jonkinlaista jakoa hallituksen ja opposition välillä oli kuitenkin havaittavissa. Jako oppositioon ja hallitukseen on luonnollinen, sillä oppositiopoliitikkojen tehtävä on haastaa hallitusta. Ottamalla huomioon lauseet, joissa viittaukset tapahtuvat, ryhmittäminen oppositioon ja hallitukseen henkilömainintojen perusteella selkiytyi.

Koska yhteisviittausverkostoista ja bibliografiseen kytkentään pohjautuvat verkostot olivat hyvin tiheitä, ainoa tapa erottaa klustereita on tutkia solmujen välisten linkkien suuruuksia. Viittausten määrät eivät tuoneet riittävän suuria eroja selkeiden klustereiden syntymiseksi. Linkkejä voisi karsia verkostoanalyysin menetelmin, kuten muodostaen Pathfinder-verkostoja alkuperäisistä verkostoista. Vaihtoehtoisesti yhteisviittaukselle ja bibliografiselle kytkennälle voisi vielä kokeilla tiukempia määritelmiä esimerkiksi siten, että kansanedustajat lasketaan yhteisviitatuksi vain, jos heidät on mainittu samassa puheenvuorossa, täysistunnossa tai esimerkiksi samaan aiheeseen liittyvässä keskustelussa. Vastaavasti kansanedustajan voidaan katsoa mainitsevan toisen kansanedustajan vain silloin, kun hän on maininnut toisen kansanedustajan vähintään jonkin tietyn määrän. On myös mahdollista, että henkilömainintojen poh-

jalta kansanedustajia ei voida jakaa selkeämpiin ryhmiin eli henkilömaininnoilla ei ole selkeitä kaavoja esimerkiksi puolueiden välillä.

Mahdolliset virheet henkilömainintojen tunnistamisessa puheenvuoroista ja linkityksessä oikeisiin tietolähteisiin voivat vaikuttaa siihen, että yksittäisille kansanedustajille lasketut metriikat voivat hieman vaihdella todellisesta erityisesti niiden kansanedustajien kohdalla, joilla on sukunimikaimoja. Oletettavasti henkilömainintojen NER- ja NEL-tarkkuudet ovat kuitenkin sen verran hyviä, että tutkittaessa puolueiden välisiä mainintoja tai ryhmitellessä mainituimpia ja eniten mainintoja tehneitä kansanedustajia tulokset vastaavat aika hyvin todellisuutta.

Jos puheenvuoroja pystytään hakemaan aiheen mukaan, voisi verkostoja muodostaa eri aiheiden pohjalta, jolloin saadaan selville kuhunkin aiheeseen liittyvät keskeiset kansanedustajat. Parlamenttisammossa puheenvuoroihin on linkitetty puheenvuoroihin liittyviä asiasanoja, joiden kautta puheenvuorojen aiheita voitaisiin kokeilla rajata. Lupaava suunta kansanedustajien ryhmittymisien löytämiseen onkin henkilömainintojen kontekstin ottaminen huomioon. Pelkän viittauslauseen sijaan voitaisiin kokeilla ottaa huomioon koko puheenvuoro. Edellä mainittujen tapojen lisäksi kansanedustajien klustereiden etsimiseen voisi kokeilla yhteysmatriisin pohjautuvan monimuuttuja-analyysin sijaan verkostoanalyysin yhteisöjen löytämiseen tarkoitettuja algoritmeja.

Tulevaisuudessa verkostoja voisi myös luoda eri ajanjaksoille, kuten puolueiden tapauksessa on tehty, ja tutkia ketkä kansanedustajat ovat keskeisessä asemassa minäkin ajanjaksona. Samalla voidaan tutkia, tapahtuuko viittauksiin perustuvissa verkostoissa suuria muutoksia ajan kuluessa sekä voisiko verkostoista tällöin tunnistaa erilaisia tapahtumia, kuten Perussuomalaisen hajoamisen vaalikaudella 2015–2017.

Tässä työssä etsittiin ja analysoitiin parlamentaaristen keskusteluiden keskeisiä henkilöitä. Parlamenttisammossa sisältämä data mahdollistaa monenlaisten kansanedustajien verkostojen luomisen, esimerkiksi kansanedustajien mainitsemien paikkojen perusteella tai organisaatioyhteyksien perusteella. Näin ollen avoin parlamentaarinen data mahdollistaa kansanedustajien keskeisyyden mittaamisen eri aihepiireissä, mikä on tärkeää päätöksenteon läpinäkyvyyden kannalta.

Viitteet

- [1] R. Molontay ja M. Nagy. ”Two decades of network science”. Teoksessa: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ASONAM ’19: International Conference on Advances in Social Networks Analysis and Mining. IEEE, 2019, s. 586–583. ISBN: 1-4503-6868-9. DOI: [10.1145/3341161.3343685](https://doi.org/10.1145/3341161.3343685).
- [2] M. Barabási A.-L. and Pósfai. *Network science*. Cambridge: Cambridge University Press, 2016. ISBN: 9781107076266 1107076269. URL: <http://barabasi.com/networksciencebook/>.
- [3] P. Leifeld. ”Policy debates and discourse network analysis: A research agenda”. *Politics and Governance* 8.2 (2020), s. 180–183. ISSN: 2183-2463. DOI: [10.17645/pag.v8i2.3249](https://doi.org/10.17645/pag.v8i2.3249).
- [4] C. Bhattacharya. ”Gatekeeping the plenary floor: Discourse network analysis as a novel approach to party control”. *Politics and Governance* 8.2 (2020), s. 229–242. ISSN: 2183-2463. DOI: [10.17645/pag.v8i2.2611](https://doi.org/10.17645/pag.v8i2.2611).
- [5] B. Curran et al. ”Look who’s talking: Two-mode networks as representations of a topic model of New Zealand parliamentary speeches”. *PloS one* 13.6 (2018), e0199072. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0199072](https://doi.org/10.1371/journal.pone.0199072).
- [6] M. Esteve Del Valle, M. Broersma ja A. Ponsioen. ”Political interaction beyond party lines: Communication ties and party polarization in parliamentary Twitter networks”. *Social Science Computer Review* 40.3 (2022), s. 736–755. ISSN: 0894-4393. DOI: [10.1177/0894439320987569](https://doi.org/10.1177/0894439320987569).
- [7] M. Weber D. and Nasim, L. Mitchell ja L. Falzon. ”Exploring the effect of streamed social media data variations on social network analysis”. *Social network analysis and mining* 11.1 (2021). ISSN: 1869-5450. DOI: [10.1007/s13278-021-00770-y](https://doi.org/10.1007/s13278-021-00770-y).
- [8] D. Cherepnalkoski ja I. Mozetic. ”A retweet network analysis of the European Parliament”. Teoksessa: *Signal-Image Technology & Internet-Based Systems (SITIS), 2015 11th International Conference on*. 2015 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). IEEE, 2015, s. 1–357. ISBN: 1-4673-9722-9. DOI: [10.1109/SITIS.2015.8](https://doi.org/10.1109/SITIS.2015.8).
- [9] A. Said et al. ”Mining network-level properties of Twitter altmetrics data”. *Scientometrics* 120.1 (2019-7), s. 217–235. ISSN: 0138-9130.
- [10] M. Forsman. *Julkaisut ja tieteen mittaaminen : Bibliometriikan käännekohtia*. 2. painos. Enostone kustannus, 2022. ISBN: 9789515150622. DOI: [10.31885/9789515150622](https://doi.org/10.31885/9789515150622).
- [11] J.D. West et al. ”Author-level Eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community”. *Journal of the American Society for Information Science and Technology* 64.4 (2013), s. 787–801. ISSN: 1532-2882. DOI: [10.1002/asi.22790](https://doi.org/10.1002/asi.22790).

- [12] W. Glänzel ja H.F. Moed. "Journal impact measures in bibliometric research". *Scientometrics* 53.2 (2002), s. 171–193. ISSN: 0138-9130. DOI: [10.1023/A:1014848323806](https://doi.org/10.1023/A:1014848323806).
- [13] R. Ma. "Author bibliographic coupling analysis: A test based on a Chinese academic database". *Journal of informetrics* 6.4 (2012), s. 532–542. ISSN: 1751-1577. DOI: [10.1016/j.joi.2012.04.006](https://doi.org/10.1016/j.joi.2012.04.006).
- [14] R. Rousseau ja A. Zuccala. "A classification of author co-citations: Definitions and search strategies". *Journal of the American Society for Information Science and Technology* 55.6 (2004), s. 513–529. ISSN: 1532-2882. DOI: [10.1002/asi.10401](https://doi.org/10.1002/asi.10401).
- [15] Y.-S. Hyun, D. Cho ja S.W. Yoon. "Landscape of human resource development research in Korea: Results from author co-citation network analysis". *Human Resource Development International* 18.5 (2015), s. 446–463. ISSN: 1367-8868. DOI: [10.1080/13678868.2015.1079295](https://doi.org/10.1080/13678868.2015.1079295).
- [16] L. Leydesdorff. "On the normalization and visualization of author co-citation data: Salton's Cosine versus the Jaccard index". *Journal of the American Society for Information Science and Technology* 59.1 (2008), s. 77–85. ISSN: 1532-2882. DOI: [10.1002/asi.20732](https://doi.org/10.1002/asi.20732).
- [17] E. Yan ja Y. Ding. "Scholarly network similarities: How bibliographic coupling networks, citation networks, cocitation networks, topical networks, coauthorship networks, and cword networks relate to each other". *Journal of the American Society for Information Science and Technology*. 63.7 (2012), s. 1313–1326. ISSN: 1532-2882. DOI: [10.1002/asi.22680](https://doi.org/10.1002/asi.22680).
- [18] M.J. Culnan. "Mapping the intellectual structure of MIS, 1980-1985: A co-citation analysis". *MIS Quarterly : Management Information Systems* 11.3 (1987), s. 341–350. ISSN: 0276-7783. DOI: [10.2307/248680](https://doi.org/10.2307/248680).
- [19] K. Backhaus, K. Lügger ja M. Koch. "The structure and evolution of business-to-business marketing: A citation and co-citation analysis". *Industrial Marketing Management* 40.6 (2011), s. 940–951. ISSN: 0019-8501. DOI: [10.1016/j.indmarman.2011.06.024](https://doi.org/10.1016/j.indmarman.2011.06.024).
- [20] C.D. McLaren ja M.W. Bruner. "Citation network analysis". *International Review of Sport and Exercise Psychology* 15.1 (joulukuu 2022), s. 179–198. ISSN: 1750-984X. DOI: [10.1080/1750984X.2021.1989705](https://doi.org/10.1080/1750984X.2021.1989705).
- [21] C. Calero-Medina ja E.C.M. Noyons. "Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field". *Journal of Informetrics* 2.4 (2008), s. 272–279. ISSN: 1751-1577. DOI: [10.1016/j.joi.2008.09.005](https://doi.org/10.1016/j.joi.2008.09.005).
- [22] J.M. Kleinberg. "Authoritative sources in a hyperlinked environment". *Journal of the Association for Computing Machinery* 46.5 (1999), s. 604–632. ISSN: 0004-5411. DOI: [10.1145/324133.324140](https://doi.org/10.1145/324133.324140).

- [23] J. Yun, S. Ahn ja J.Y. Lee. "Return to basics: Clustering of scientific literature using structural information". *Journal of Informetrics* 14.4 (2020), s. 101099. ISSN: 1751-1577. DOI: [10.1016/j.joi.2020.101099](https://doi.org/10.1016/j.joi.2020.101099).
- [24] W.-L. Shiau, Y.K. Dwivedi ja H.S. Yang. "Co-citation and cluster analyses of extant literature on social networks". *International Journal of Information Management* 37.5 (2017), s. 390–399. ISSN: 0268-4012. DOI: [10.1016/j.ijinfomgt.2017.04.007](https://doi.org/10.1016/j.ijinfomgt.2017.04.007).
- [25] K.-C. Yang. "Intellectual structure of trust in business and management: A co-citation analysis". *The Electronic Library* 34.3 (2016), s. 358–370. ISSN: 0264-0473. DOI: [10.1108/EL-07-2014-0121](https://doi.org/10.1108/EL-07-2014-0121).
- [26] K.W. McCain. "Mapping authors in intellectual space: A technical overview". *Journal of the American Society for Information Science*. 41.6 (1990), s. 433–443. ISSN: 0002-8231. DOI: [10.1002/\(SICI\)1097-4571\(199009\)41:6<433::AID-ASI11>3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<433::AID-ASI11>3.0.CO;2-Q).
- [27] J.E. Andrews. "An author co-citation analysis of medical informatics". *Journal of the Medical Library Association : JMLA* 91.1 (2003), s. 47–56. ISSN: 1536-5050.
- [28] D. Zhao ja A. Strotmann. "Intellectual structure of information science 2011–2020: An author co-citation analysis". *Journal of Documentation* 78.3 (2022), s. 728–744. ISSN: 0022-0418. DOI: [10.1108/JD-06-2021-0119](https://doi.org/10.1108/JD-06-2021-0119).
- [29] S.P. Nerur, A.A. Rasheed ja V. Natarajan. "The intellectual structure of the strategic management field: An author co-citation analysis". *Strategic Management Journal* 29.3 (2008), s. 319–336. ISSN: 0143-2095. DOI: [10.1002/smj.659](https://doi.org/10.1002/smj.659).
- [30] A. Arleo, O.-H. Kwon ja K.-L. Ma. "GraphRay: Distributed pathfinder network scaling". Teoksessa: *2017 IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV)*. Vol. 2017-December. 2017 IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV). IEEE, 2017, s. 7–83. ISBN: 1-5386-0618-6. DOI: [10.1109/LDAV.2017.8231853](https://doi.org/10.1109/LDAV.2017.8231853).
- [31] R. Zhao ja B. Chen. "Applying author co-citation analysis to user interaction analysis: A case study on instant messaging groups". *Scientometrics* 101.2 (2014), s. 985–997. ISSN: 0138-9130. DOI: [10.1007/s11192-014-1314-7](https://doi.org/10.1007/s11192-014-1314-7).
- [32] S.P. Sidorov et al. *QAP analysis of company co-mention network*. 1st ed. 2018. Vol. 10836. Algorithms and Models for the Web Graph. Cham : Springer International Publishing : 2018, s. 83–98. ISBN: 3-319-92870-8. DOI: [10.1007/978-3-319-92871-5_{_}7](https://doi.org/10.1007/978-3-319-92871-5_{_}7).
- [33] J.-P. Qiu, K. Dong ja H.-Q. Yu. "Comparative study on structure and correlation among author co-occurrence networks in bibliometrics". *Scientometrics* 101.2 (2014), s. 1345–1360. ISSN: 0138-9130. DOI: [10.1007/s11192-014-1315-6](https://doi.org/10.1007/s11192-014-1315-6).

- [34] D. Zhao ja A. Strotmann. "Author bibliographic coupling: Another approach to citation-based author knowledge network analysis". *Proceedings of the American Society for Information Science and Technology* 45.1 (2008), s. 1–10. ISSN: 2373-9231. DOI: [10.1002/meet.2008.1450450292](https://doi.org/10.1002/meet.2008.1450450292).
- [35] D. Zhao, A. Cappello ja L. Johnston. "Functions of uni- and multi-citations: Implications for weighted citation analysis". *Journal of Data and Information Science* 2.1 (2017), s. 51–69. ISSN: 2096-157X. DOI: [10.1515/jdis-2017-0003](https://doi.org/10.1515/jdis-2017-0003).
- [36] Y.K. Jeong, M. Song ja Y. Ding. "Content-based author co-citation analysis". *Journal of Informetrics* 8.1 (2014), s. 197–211. ISSN: 1751-1577. DOI: [10.1016/j.joi.2013.12.001](https://doi.org/10.1016/j.joi.2013.12.001).
- [37] I. Zupic ja T. Čater. "Bibliometric methods in management and organization". *Organizational Research Methods* 18.3 (2015), s. 429–472. ISSN: 1094-4281. DOI: [10.1177/1094428114562629](https://doi.org/10.1177/1094428114562629).
- [38] K. Makkonen ja P. Loukasmäki. "Eduskunnan täysistunnon puheenaiheet 1999–2014: Miten käsitellä LDA-aihemalleja?" *Politiikka* 61.2 (heinäkuu 2019), s. 127–159. URL: <https://journal.fi/politiikka/article/view/77163>.
- [39] D. Zhao. "Towards all-author co-citation analysis". *Information Processing & Management* 42.6 (2006), s. 1578–1591. ISSN: 0306-4573. DOI: [10.1016/j.ipm.2006.03.022](https://doi.org/10.1016/j.ipm.2006.03.022).
- [40] E. Hyvönen. "Finnish Parliament on the Semantic Web: Using Parliament-Sampo data service and semantic portal for studying political culture and language". *CEUR workshop proceedings* 3133 (2022), s. 69–85. ISSN: 1613-0073.
- [41] L. Sinikallio et al. "Plenary debates of the Parliament of Finland as linked open data and in Parla-CLARIN Markup". Teoksessa: *3rd Conference on Language, Data and Knowledge, LDK 2021*. Vol. 93. Open Access Series in Informatics (OASICs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, elokuu 2021, 8:1–8:17. DOI: [10.4230/OASICs.LDK.2021.8](https://doi.org/10.4230/OASICs.LDK.2021.8). URL: <https://drops.dagstuhl.de/opus/volltexte/2021/14544/pdf/OASICs-LDK-2021-8.pdf>.
- [42] Petri Leskinen, Eero Hyvönen ja Jouni Tuominen. "Members of Parliament in Finland knowledge graph and its linked open data service". Teoksessa: *Further with Knowledge Graphs. Proceedings of the 17th International Conference on Semantic Systems, 6-9 September 2021, Amsterdam, The Netherlands*. IOS Press, helmikuu 2021, s. 255–269. DOI: [10.3233/SSW210049](https://doi.org/10.3233/SSW210049). URL: <https://ebooks.iospress.nl/volumearticle/57420>.
- [43] Minna Tamper et al. "Extracting knowledge from parliamentary debates for studying political culture and language". Teoksessa: *Proceedings of the 1st International Workshop on Knowledge Graph Generation From Text and the 1st International Workshop on Modular Knowledge co-located with 19th Extended Semantic Conference (ESWC 2022)*. Toim. S. Tiwari et al. Vol. 3184. CEUR WS, toukokuu 2022, s. 70–79. URL: http://ceur-ws.org/Vol-3184/TEXT2KG_Paper_5.pdf.

- [44] A.A. Hagberg, D.A. Schult ja P.J. Swart. "Exploring network structure, dynamics, and function using NetworkX". Teoksessa: *Proceedings of the 7th Python in Science Conference*. Toim. G. Varoquaux, T. Vaught ja J. Millman. Pasadena, CA USA, 2008, s. 11–15.
- [45] F. Pedregosa. "Scikit-learn: Machine learning in Python". *Journal of machine learning research : JMLR* 12 (2011), s. 2825–2830. ISSN: 1532-4435.
- [46] N.J. van Eck ja L. Waltman. "How to normalize cooccurrence data? An analysis of some well-known similarity measures". *Journal of the American Society for Information Science and Technology* 60.8 (2009), s. 1635–1651. ISSN: 1532-2882. DOI: [10.1002/asi.21075](https://doi.org/10.1002/asi.21075).
- [47] E. Mênigbêto. "Controversies arising from which similarity measures can be used in co-citation analysis". *Malaysian Journal of Library & Information Science* 18.2 (2013), s. 25–31. ISSN: 1394-6234.
- [48] D. Zhao ja A. Strotmann. "Evolution of research activities and intellectual influences in information science 1996-2005: Introducing author bibliographic-coupling analysis". *Journal of the American Society for Information Science and Technology*. 59.13 (2008), s. 2070–2086. ISSN: 1532-2882. DOI: [10.1002/asi.20910](https://doi.org/10.1002/asi.20910).
- [49] G.J. Szekely ja M.L. Rizzo. "Hierarchical clustering via joint between-within distances: Extending Ward's minimum variance method". *Journal of Classification* 22.2 (2005), s. 151–183. ISSN: 0176-4268. DOI: [10.1007/s00357-005-0012-9](https://doi.org/10.1007/s00357-005-0012-9).
- [50] W.R. Zwick ja W.F. Velicer. "Comparison of five rules for determining the number of components to retain". *Psychological Bulletin* 99.3 (1986), s. 432–442. ISSN: 0033-2909. DOI: [10.1037/0033-2909.99.3.432](https://doi.org/10.1037/0033-2909.99.3.432).
- [51] J. Dien. "Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations". *Psychophysiology* 47.1 (2010), s. 170–183. ISSN: 0048-5772. DOI: [10.1111/j.1469-8986.2009.00885.x](https://doi.org/10.1111/j.1469-8986.2009.00885.x).
- [52] N.J. Van Eck et al. "A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS". *Journal of the American Society for Information Science and Technology* 61.12 (2010), s. 2405–2416. DOI: [10.1002/asi.21421](https://doi.org/10.1002/asi.21421).
- [53] Y. He ja S. Cheung Hui. "Mining a Web citation database for author co-citation analysis". *Information Processing & Management* 38.4 (2002), s. 491–508. ISSN: 0306-4573. DOI: [10.1016/S0306-4573\(01\)00046-2](https://doi.org/10.1016/S0306-4573(01)00046-2).
- [54] V.D. Blondel et al. "Fast unfolding of communities in large networks". *Journal of Statistical Mechanics : Theory and Experiment* 2008.10 (2008), P10008. ISSN: 1742-5468. DOI: [10.1088/1742-5468/2008/10/P10008](https://doi.org/10.1088/1742-5468/2008/10/P10008).
- [55] L. Leydesdorff ja L. Vaughan. "Co-occurrence matrices and their applications in information science: Extending ACA to the Web environment". *Journal of the American Society for Information Science and Technology* 57.12 (2006), s. 1616–1628. ISSN: 1532-2882. DOI: [10.1002/asi.20335](https://doi.org/10.1002/asi.20335).