

# Applying Network and Bibliometric Analyses to Mentions of Politicians in Plenary Speeches: Case ParliamentSampo – Parliament of Finland on the Semantic Web

Henna Poikkimäki<sup>1\*</sup>, Petri Leskinen<sup>2,1\*</sup> and Eero Hyvönen<sup>1,2\*</sup>

<sup>1\*</sup>Department of Computer Science, Aalto University, Maarintie 8, Espoo, 02150, Finland.

<sup>2\*</sup>Helsinki Centre for Digital Humanities (HELDIG), University of Helsinki, Helsinki, Finland.

\*Corresponding author(s). E-mail(s): [henna.poikkimaki@aalto.fi](mailto:henna.poikkimaki@aalto.fi); [petri.leskinen@aalto.fi](mailto:petri.leskinen@aalto.fi); [eero.hyvonen@aalto.fi](mailto:eero.hyvonen@aalto.fi);

## Abstract

Speakers mention each other for various reasons in parliamentary debate speeches. Recognizing person name mentions from textual speeches and linking them to corresponding speakers allows construction of mention networks of politicians and parties that can be analyzed using methods of network science and bibliometrics. This paper shows how networks based on person name mentions in speeches can be constructed and analyzed based on the speech corpus 2015–2022 of the Parliament of Finland. Our results give novel insights on how the party and its role (government, opposition) affect the way how the Members of Parliament (MP) reference each other in their speeches.

**Keywords:** parliamentary studies, network analysis, linked data, bibliometrics, digital humanities

## 1 Introduction

Openness and transparency of parliamentary work is a foundation of democracy. An important public part of parliamentary work are the plenary debates where the MPs

discuss and enact new laws, oversee the work of the government, and decide on the state budget. The minutes of the plenary sessions therefore provide interesting information about the state and functioning of democratic systems and enable study of political life, language, and culture [1].

The minutes of the plenary debates of various Parliaments have been forged into parliamentary corpora, see, e.g., [2]. Parliamentary discussions and other materials have also been transformed into Linked Data (LD) when, e.g., creating the LinkedEP [3] system based on the European Parliament’s data<sup>1</sup>, the LinkedSaeima for the Latvian parliament [4], and the Finnish ParliamentSampo [5–8] whose data is used in this paper. Transforming parliamentary data into LD provides well-defined semantics for representing and enriching knowledge<sup>2</sup> aggregated from heterogeneous data sources [3, 4], which makes it easier to query the data and to construct networks for data analyses.

Parliamentary corpora have been used mostly for linguistic analyses. For example, in [9] the contents of women’s parliamentary speeches in the British Parliament were analyzed, and thematic and conceptual analyses of the opinions and language were made in [10–14]. In earlier research, sociocentric and egocentric networks connecting the actors have been constructed from different texts based on, e.g., mentioned names, hypertext links, genealogical relations, or similarities in characteristics, such as lifetime events [15–17].

In this paper, the latter idea is applied to parliamentary speeches where speakers mention each other. This paper presents, how different networks of MPs and parties can be constructed based on person name mentions in parliamentary speeches, and how these networks can be analyzed using methods of network science and bibliometrics. As a case study, networks were created using the ParliamentSampo data and infrastructure [8], specifically the speech subcorpus 2015–2022 of the Parliament of Finland [5, 7] and the knowledge graph (KG) of the MPs [6]. Analyses of networks based on person name mentions reveal, does the party of the speakers (right-wing, left-wing, opposition, government) affect how the MPs are mentioned or who they mention, and who are the most central people in parliamentary discussions from a network analytic point of view. More detailed descriptions of the analyses presented can be found in [18]. How the parliament works in Finland is documented in [19].

## 2 Related Works

During the last two decades, network science [20–22] has become an increasingly significant field by successfully explaining phenomena and fundamental concepts in a wide array of systems from cellular biology to societies. In the context of parliamentary speech data, especially discourse network analysis (DNA) has been used. Discourse network analysis, combining network analysis and qualitative content analysis, can reveal connections between political actors at the discursive level [23].

For example, [24] clustered speeches given in the Parliament of New Zealand according to their topics and created a weighted network where MPs were connected

---

<sup>1</sup>The European Parliament publishes data about its work as LD online: <https://data.europarl.europa.eu/en/home>.

<sup>2</sup><https://www.w3.org/standards/semanticweb/>

when they had spoken about the same topics. Networks of MPs for different parliaments were analyzed using complex network techniques to study, e.g., community structure of networks. [25] constructed a network of German MPs based on how they agreed with each other in speeches and written explanations of votes related to the Greek crisis in 2010–2015. Bhattacharya studied how party unity held in discussions and decisions related to the Greek crisis. To the best of our knowledge, how MPs mention each other in their parliamentary speeches has not been studied before. Instead, the behavior of politicians on social media has been studied. For example, [26] studied how Dutch MPs interact on Twitter<sup>3</sup> using a mention function and concluded that mention networks based on tweets do not indicate political polarization.

The idea of our work is to construct different kind of networks based on person name mentions on parliamentary speeches and analyze these networks using methods from network analysis and bibliometrics in order to find out possibly interesting MPs or mention patterns in parliamentary discussions. Bibliometrics is a study of publications that is often based on citations between scientific publications [27]. Bibliometric studies can be done, e.g., on document, author or journal level in order to find central actors and possible clusters of publications or authors [28]. In our work parliamentary speeches correspond to the documents, MPs to authors, and parties or committees to journals. Bibliometric methods used in this work include citation analysis (CA), author co-citation analysis (ACA) and author bibliographical coupling analysis (ABCA).

CA can reveal the most significant documents or authors on the chosen field of analysis. It is assumed that a large number of mentions to a document or an author tells that the document or author in question is influential on its own field [29]. ACA is often used to study the intellectual structure of chosen disciplines by clustering authors based on co-citations whereas ABCA can also give hints of the future directions of the field [30].

Traditional methods of bibliometrics are purely citation-based and do not take into account the context of citations. Context has been taken into account by manually sorting citations into categories [31]. Manual sorting is not feasible when the number of citations to be studied is large. An other way for including citation context into analyses is to compare the location of citations in the articles or to calculate similarity of citation sentences [32]. Dynamics of citation-based networks can be taken into account by constructing networks for different time periods and comparing results [33].

### 3 Data

We used the ParliamentSampo data that contains all parliamentary debate speeches in Finland from 1907 as well as data about the MPs parliamentary organizations as Linked Open Data [8]. The ParliamentSampo data publication consists of two parts: 1) a knowledge graph (KG) of parliamentary debate speeches [5] and 2) a KG of the political actors [6]. In order to extract networks based on person name mentions, the original RDF speech graph was enriched with Natural Language Processing (NLP) methods, such as named entity recognition (NER) and linking (NEL) [34].

---

<sup>3</sup><https://twitter.com>

The named entities (NE) were extracted from the parliamentary speeches from end of April 2015 until end of January 2022, which limits the analyses for this particular time period and parliament. Recognized named entities, the mentioned people, places, groups, organizations, and their related information were linked internally to the ParliamentSampo KG of MPs. For a broader data enrichment, linkings to external data sources were created, including the Kanto<sup>4</sup> vocabulary for Finnish actors provided by the National Library, the YSO Places ontology<sup>5</sup> and PNR<sup>6</sup> gazetteer of Finnish place names by the National Survey. Named entity recognition and linking of person name mentions are explained in more detail in [35]. Since then the person name linkage have been improved by linking also family name mentions.

The accuracy of the NER was estimated for 100 randomly selected mentions of people, places, organizations, and expressions of time. The precision was 97%, recall 77%, and F1-score 86%. Low recall is mostly due to problems in recognizing organizations, and it should not have too great effect on our work. The results for linking people were calculated for 100 randomly selected speeches that contained 100 entities in total. The precision was 98%, recall 96%, and F1-score 97%.

Table 1 contains the number of speeches and number of speeches in which at least one person name mention has been recognized and linked. In total NEL and NER have been done for about 120 000 speeches and 64 000 of them contain linked person name mentions. NEL and NER have not been done properly for speeches given in Swedish. The number of Swedish speeches is, however, quite low (1400) compared to the total number of speeches and leaving them out affects mostly the Swedish People’s Party as one third of the party’s speeches are given in Swedish. The number of speeches of the Speaker of the Parliament of Finland is higher on the later electoral term even if all the speeches of the electoral term are not in the data. This is due to differences on how those speeches are written down and it does not concern our work as the speeches of the Speaker are excluded from our analyses.

Electoral term	Speeches	Speeches with person name mentions
2015–2018	51807 (64934)	24428 (31075)
2019–2022*	33146 (55104)	16405 (31578)
Total	86820 (120038)	40833 (64221)

**Table 1:** Number of speeches and number of speeches that contain at least one person name mention. Numbers inside brackets tell the number of speeches when speeches of the Speaker of Parliament of Finland are also taken into account. \*Speeches from the electoral term 2019–2022 were available only until 6.5.2022.

---

<sup>4</sup><https://finto.fi/finaf/en/>

<sup>5</sup><https://finto.fi/ys0-paikat/en/?clang=en>

<sup>6</sup><http://www.ldf.fi/dataset/pnr>

## 4 Methods

### 4.1 Constructing Networks

Different networks based on mentions between MPs were created and analyzed using NetworkX [36]. For the networks where nodes are MPs, speeches from the electoral term 2015–2018 were used. Limiting speeches for one electoral term prevents situations where MPs who were chosen for both terms have had more opportunities to mention other MPs and to be mentioned by other MPs, which causes them to surface in the analyses. For the networks where nodes are parties, speeches from the first half of electoral term 2019–2022 were also used. Because many parties' parliamentary role, government or opposition, changed when the electoral term changed, including speeches from both electoral terms allows to study, if mentions received and made by the members of the party is affected by the party's parliamentary role. Speeches that do not contain any linked person names were excluded from the analysis. In addition, administrative speeches of the Speaker of the Parliament of Finland and speeches given in Swedish were not taken into account as person name recognition and linking is done using tools related to the Finnish language.

In citation (or mention) network of MPs the nodes are MPs. A directed link goes from one MP to other MP when an MP mentions another MP in his/her speech. The link weight is the number of speeches, where the first MP mentions the second MP at least once. Self-mentions are not taken into account, i.e., there is no links from a node to node itself. To obtain citation networks between parties, MPs were grouped by their parties and link weight summed. If an MP had changed party during an electoral term, mentions made and received by the MP were added to the party that the MP was member at the time of each speech where mentions were made.

Co-citation and bibliographical coupling based networks were also constructed for MPs. In co-citation networks a link was added between MPs when some other MP had mentioned both MPs in his/her speeches. When one MP has mentioned two other MPs in their speeches, we take the minimum of the number of times the MP has mentioned one or the other MP. By summing up minimums from all MPs we get the total co-citation count between two other MPs. In a network based on bibliographical coupling there is link between two MPs when they have mentioned at least one same person in their speeches. Bibliographical coupling strengths or link weights were calculated using similar idea as for co-mention graphs, but now summing up minimums of mentions made instead of received mentions.

To take mention context into account in co-citation and bibliographical coupling analyses, ideas presented by Jeong et al. [32] were followed with some modifications. They calculated cosine similarities for each citing sentence pair in the document and summed cosine similarities in all documents related to two authors to get co-citation count between those two authors. Networks presented in our work are based on MP level mentions instead of speech level mentions, and co-citation counts are then calculated differently. Sentences that contain mentions of MPs were first extracted from speeches and then lemmatized using Voikko linguistic software<sup>7</sup>. Cosine similarity for

---

<sup>7</sup><https://voikko.puimula.org/>

each sentence pair was calculated, where the MP mentions one MP in the first sentence and other one in the second sentence. Maximum cosine similarity was chosen and total co-citation count for the two mentioned MPs was acquired by summing all maximum cosine similarities. For content based bibliographical coupling analysis, a similar idea was used to get total bibliographical coupling strengths.

## 4.2 Bibliometric Methods

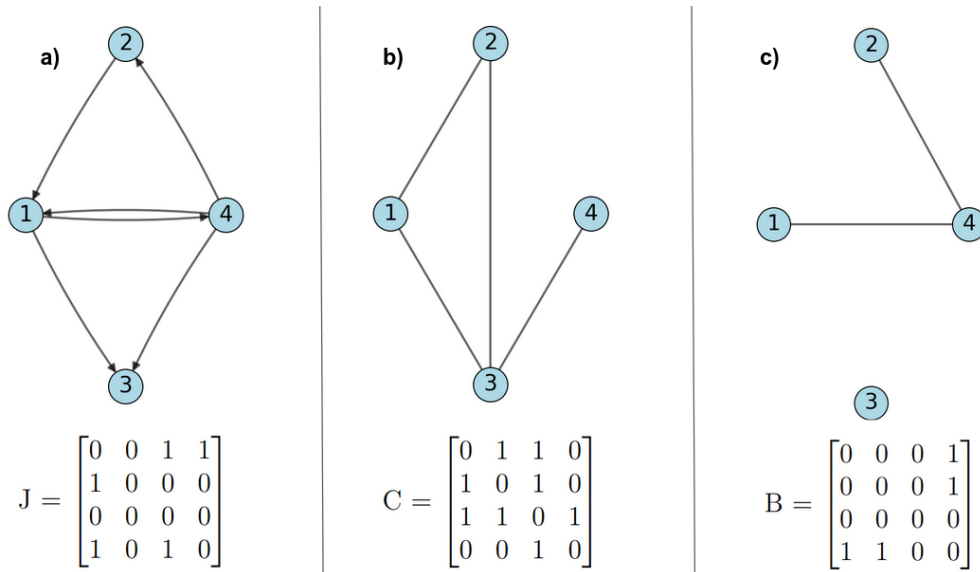
Traditional methods of bibliometrics are often based on multivariate analysis, such as factor analysis, clustering analysis, and multidimensional scaling of matrices based on citations, e.g., a co-citation matrix  $\mathbf{C}$ , citation matrix  $\mathbf{J}$ , or bibliographical coupling matrix  $\mathbf{B}$  [37–39]. In a citation matrix  $\mathbf{J}$  each element  $j_{ij}$  corresponds to the number of citations from actor  $i$  to actor  $j$  and self-citations can be ignored by assigning zero to corresponding elements in the matrix [40]. In case of our work, element  $j_{ij}$  corresponds to the number of times MP  $i$  has mentioned MP  $j$  or members of party  $i$  have mentioned members of other party  $j$ .

A co-citation matrix is based on co-citations, i.e., how often actors are cited together by a third actor [41]. Elements of a bibliographical coupling matrix are based on the similarity of reference lists of the documents or actors [39]. These matrices are symmetric similarity matrices. In multivariate analysis, the diagonals of these matrices can be handled as missing values and replaced by column means [42]. On author level analysis there are multiple ways to count co-citations and bibliographical couplings between authors [43, 44]. In the case of our work, two MPs are co-cited when a third MP has mentioned both of them in their speeches. The total co-citation count is obtained by choosing the minimum number of times a third MP has mentioned first or second MPs, and summing minimums from all third MPs. Similarly, there is a bibliographical coupling between two MPs when they have mentioned the same third MP in their speeches. The total strength of the bibliographical coupling is the minimum number of times the first or second MP have mentioned the third MP, and summing minimums for all third MPs.

The previous matrices can be seen as adjacency matrices to networks where nodes are documents, authors, or journals, and links are based on the citations between them. In our work, nodes are MPs or parties, and links are based on person name mentions in parliamentary speeches. The citation matrix gives the corresponding directed citation network [45]. The co-citation matrix and bibliographical coupling matrix give an undirected co-citation network and a bibliographical coupling network [28]. This allows usage of network analysis methods for finding communities and central nodes as well as for creating interesting visualizations [30].

Figure 1 shows a simple citation network and its adjacency matrix that corresponds to a citation matrix, and co-citation and bibliographical coupling networks based on citation network and their adjacency matrices. For example, node 2 has cited node 1 so there is a link from node 2 to node 1 and  $j_{21} = 1$ . Node 1 has cited nodes 3 and 4 so they are connected in a co-citation network. Nodes 2 and 4 have both cited node 1 in the original citation network and are then connected in a bibliographical coupling network. In the case of the co-citation and bibliographical networks, the diagonals of the adjacency matrices are set to zero as nodes do not have self-links. If adjacency

matrices are used in the analysis as co-citation and bibliographical coupling matrices, the diagonal values can be considered as missing values.



**Figure 1:** Example of a) simple citation network and its adjacency matrix or citation matrix b) co-citation network based on citation matrix and its adjacency matrix and c) bibliographical coupling network and corresponding adjacency matrix.

In co-citation analysis, multidimensional scaling (MDS), factor analysis (FA), and agglomerative hierarchical clustering (HC) are often used in order to find possible groups of authors and consequently possible subfields within the chosen research field. Same methods can be used for author bibliographical coupling analysis, in which case input matrix is a bibliographical coupling matrix instead of a co-citation matrix [39]. In the case of our work, it would be interesting to see if there is some pattern behind person name mentions made in parliamentary speeches, e.g., are some MPs clearly often mentioned together compared to other MP groups, or can MPs be divided into groups based on how they mention other MPs.

Multidimensional scaling is a group of visualization methods that try to represent data in lower dimensions while preserving the distances between the data points [46]. For example, in case of co-citation analysis, items that have been often cited together appear to be close to each other in visualization. MDS does not assign clusters to items but resulting visualization can be used as supporting evidence for HC and FA results [41]. For FA we used principal factor extraction with Promax rotation. An item was thought to belong to component (or cluster) when the corresponding loading was over 0.4 and one item might belong to multiple components.

There are some fundamental differences between citations in scientific publications and person name mentions in parliamentary speeches. Some of these differences remove common problems in bibliometrics, some differences bring new problems to be considered. In bibliometrics, citations, co-citations, and bibliographical couplings are always based on the document level counts even if the analyses are done, e.g., on an author or a journal level. In parliamentary discussions, person name mentions might be direct mentions to people rather than mentions to their speeches. For example, during the electoral term 2015–2018 about 15% of the sentences in speeches that contain person name mentions contain also the words ”puheenvuoro” or ”puhe” (speech) that indicates direct mention to mentioned person’s speech rather than mention to the person themselves. Person name mentions, co-mentions, and bibliographical couplings are then calculated on an MP level instead of speech level.

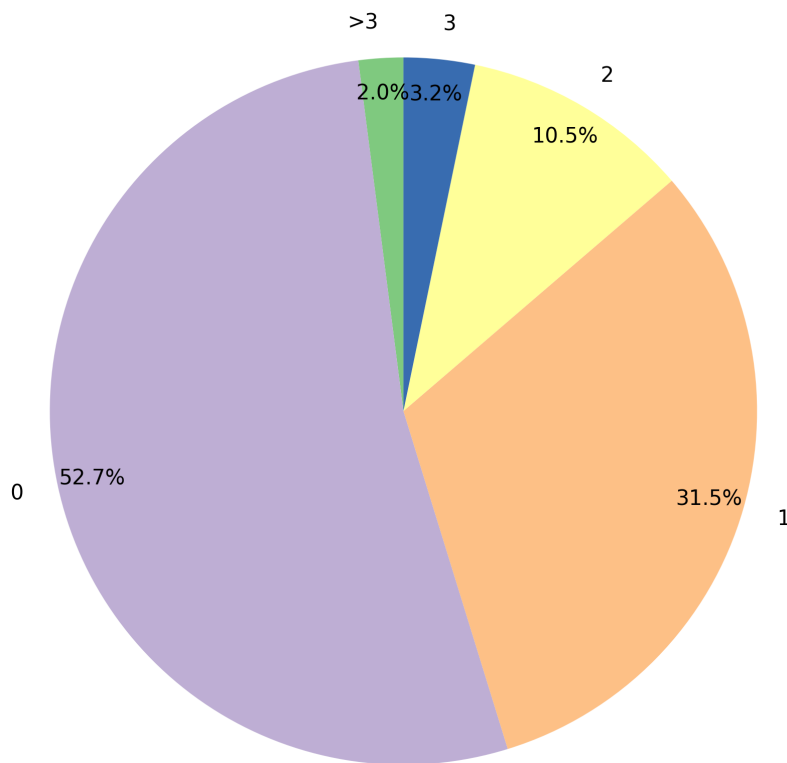
Parliamentary speeches in the Parliament of Finland are often quite short and there are not many person name mentions per speech. Speeches given during the electoral term 2015–2018 contained on the average 0.73 person name mentions to MPs chosen for the term when speeches of the Speaker of Parliament are not taken into account. Speeches that contained person name mentions had on the average 1.54 mentions. Figure 2 shows the proportion of speeches given during 2015–2018 term that contain mentions to certain number of different MPs and only 15.7% of speeches contain mentions to more than one people. For this reason, we ended up defining co-citation, or co-mention, between two MPs as a situation where the third MP mentions two MPs in their speeches but mentions do not have to happen in the same speech. Similarly bibliographical coupling happens when two MPs mention the same MP in some of their speeches.

These differences make, e.g., some bibliometrics indicators meaningless in the context of parliamentary speeches. In addition, results of CA, ACA, and ABCA have to be interpreted accordingly. For example, if two persons have cited the same document, it is more safe to assume that they are interested in the same topic than if they cite the same author but not necessarily the same document. The cited author could, for example, provide publications related to two different topics. The first citing author might be interested in the first topic and the second citing author in the other topic. Similarly in parliamentary debates when two persons mention the same person but we do not know if they refer to the same speech, we can not really know if mentioning MPs are interested in similar topics. Presumably person name mentions in parliamentary debate speeches also contain more often mentions that rise from disagreements than scientific citations.

Bibliographical studies rely often on different citation databases. The used database may not contain all works of all interesting authors, which may have some effect on the results [30]. In our work all the speeches from chosen time period are available, but mistakes in recognising and linking person name mentions may cause missing person name mentions or mentions are linked to wrong people. And even if all the speeches are in the data set, the number of speeches related to each topic are limited and all MPs may not have chance to give a speech [47]. In CA and ACA, older publications have had more time to accumulate citations, which makes them show up more prominent than newer publications [30]. If person name mentions in parliamentary speeches are



### Number of mentioned MPs in 2015–2018 speeches



**Figure 2:** The proportion of speeches given during electoral term 2015–2018 that contain mentions to 0, 1, 2, 3 or more different MPs chosen for 2015–2018 term. The speeches given by the Speaker of the Parliament are not taken into account.

limited to, e.g., some electoral term, all MPs have had same time to accumulate mentions. However, here are some exceptions when, for example, an MP has left the Parliament in the middle of the electoral term.

### 4.3 Network Analysis

For networks, some basic metrics like the number of nodes and edges and network density were calculated. Network density corresponds to the number of links in the network divided by the number of all possible links in the network. For example, in the case of the citation network, the density tells how comprehensively MPs mention each other. In-degree tells the number of citations an MP or a party has received from other MPs or parties. Out-degree is the number of mentions an MP or members of a party have made to other MPs or members of other parties. If link weights are not taken into account, in-degree corresponds to the number of MPs that one MP has mentioned and out-degree is the number of MPs that have mentioned that one MP.

For directed citation networks, hub and authority values were calculated for the nodes. In general, good hub links to several good authorities and good authority has links from good hubs [48]. In citation analysis, authorities are high quality sources and hubs cite many high quality sources, and ideal work has then high values in both [49]. MPs who have high authority values have probably been in charge of something significant and MPs who have high hub values have mentioned lot of other MPs. If an MP has a high hub and authority value, (s)he is probably central in parliamentary discussions in some way.

In addition, eigenvector centrality was calculated for the nodes. In the case of undirected networks, nodes with high eigenvector centrality have many neighbours and the neighbours in turn have many neighbours. In a directed citation network, a node has high eigenvector centrality if it has been cited often and those citing it have also been cited often.

## 5 Analyzing Networks of Politicians

### 5.1 Citation Networks of Parties

For parties, three citation networks for different time periods were constructed. The first citation network is based on speeches from the beginning of the electoral term 2015–2018 until June 12, 2017, when a group of politicians parted ways with the Finns party (PS) and formed the new party Finnish Reform Movement (KL), and the remaining MPs of the Finns party moved from government to opposition. Other parties that had MPs during the first time period were the National Coalition Party (Kok.), Centre Party (Kesk.), Social Democratic Party (SDP), Left Alliance (Vas.), Green League (Vihr.), Swedish People’s Party (RKP), and Christian Democrats (KD). The second network is based on speeches from the remaining period of the electoral term, and in addition to parties mentioned above the Movement Now (Liik.) and Seven Star Movement (TL) had both one MP. The third network is formed from speeches given during the first half of the electoral term 2019–2022. In addition to the parties in the first time period, the parties Liik. and Power Belongs to the People (VKK) had both one MP. Networks from different time periods allow studying how a party’s parliamentary role affects person name mentions made and referred to by the party members.

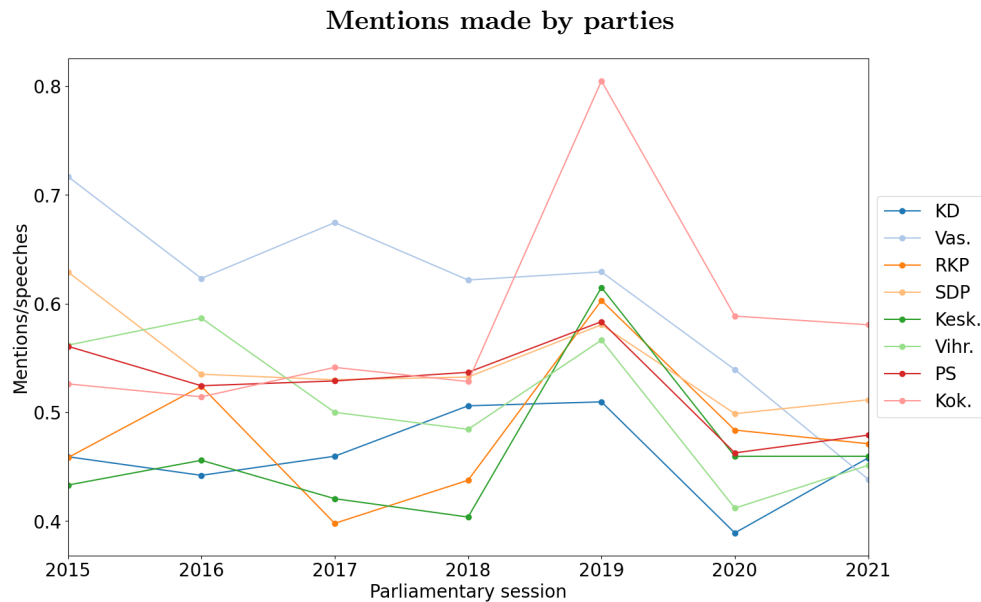
Table 2 provides statistics for the different time periods when speeches of speakers’ fellow party members are not taken into account. The proportion of speeches that contain person name mentions remains stable. Almost one third of these mentions are mentions of ministers. The resulting citation networks are dense and for the last two networks small ”one man parties” lower the density.

Figure 3 shows the proportion of the speeches containing at least one person name mention to an other MP given by party members during different parliamentary sessions. For some parties (Vas. and Vihr.) the proportion of speeches containing mentions seems to get smaller as the parties move from opposition to government from the parliamentary session 2019 onward. Similarly, the proportion of speeches that contain mentions to other MPs by National Coalition members rises as the party moves

from government to opposition when the electoral term changes. These observations, however, do not hold for all parties.

Time period	Days	Parties	Speeches	Mentions	Speeches with person name mention	Proportion of minister mentions	Density of network
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

**Table 2:** Number of days, parties, speeches, and person name mentions on chosen time periods. In addition, there are columns for the proportion of speeches where a speaker has mentioned at least one member of another party, proportion of minister mentions out of all person name mentions, and density of the corresponding citation network. Mentions to speakers' fellow party members are not taken into account.



**Figure 3:** Proportion of biggest parties' speeches that contain at least one mention to MPs from other parties during different parliamentary sessions. Speeches of the Speaker of Parliament and speeches given in Swedish are not taken into account.

In the case of parties, a citation matrix normalized by Pearson correlations was used as basis for HC and MDS and parties whose members had been mentioned

under 50 times were left out. Figure 4 shows results of MDS where nodes are colored according to HC results. Based on how the members of other parties have mentioned parties' politicians, parties can be divided into government and opposition parties. In other words, members of opposition parties and members of government parties are mentioned with different patterns by other parties' members. The results do not change even if mentions to ministers are left out.

## 5.2 Citation Network of MPs

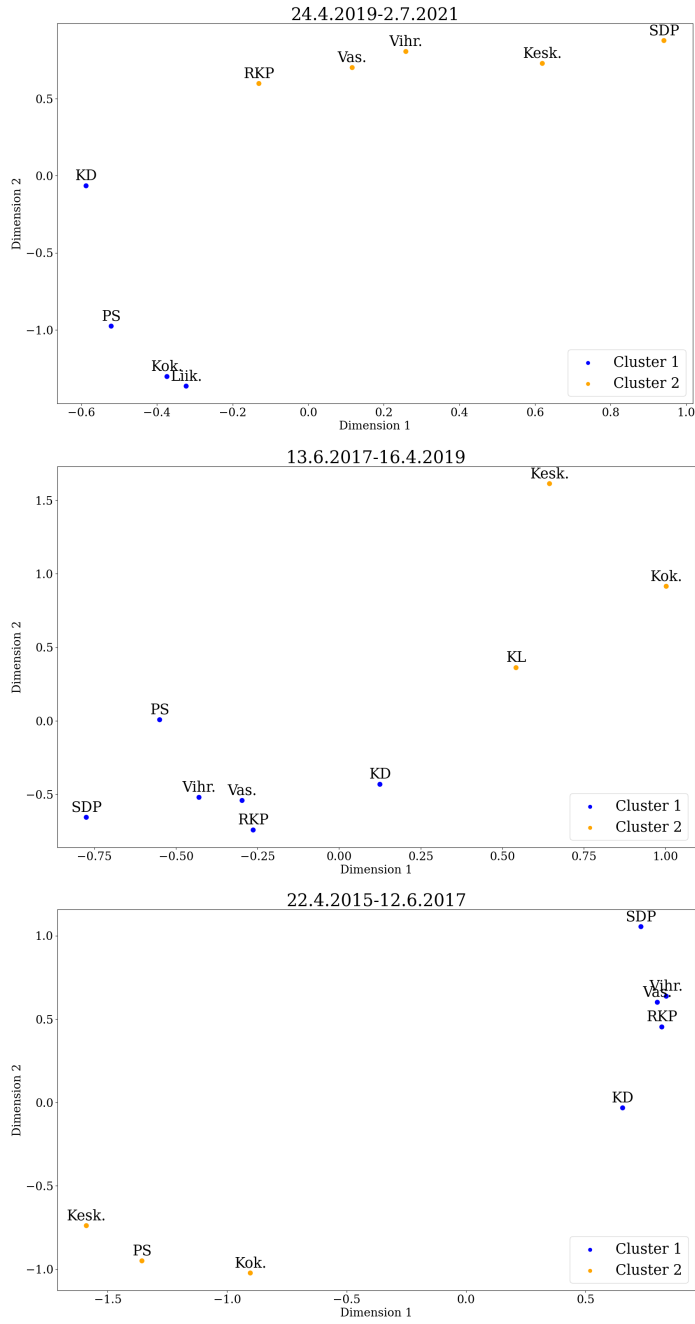
The citation network of MPs was constructed based on over 51 000 speeches given during the electoral term 2015–2018 of which over 24 000 contained at least one mention to an other MP chosen for the term. About 24% of these mentions were mentions of ministers. In total 214 different MPs were active during the electoral term.

To study and visualize part of the citation network, hub and authority values were calculated for nodes using the HITS algorithm [50]. In case of our work, MPs with high authority values have been mentioned often by good hubs, and MPs with high hub values have made mentions often especially of good authorities. Ten MPs with highest authority values and ten MPs with highest hub values are shown in Fig. 5. Table 3 contains MPs with highest hub and/or authority values. High in-degree, when the number of speeches where an MP has been mentioned is high, leads to high authority and eigenvector centrality values. Similarly, high out-degree, when the number of speeches where an MP has mentioned some other MP, results in high hub values. The most often mentioned MP was prime minister Juha Sipilä. Mrs. Pia Viitanen from opposition had mentioned other MPs most often, as well as given the largest number of speeches.

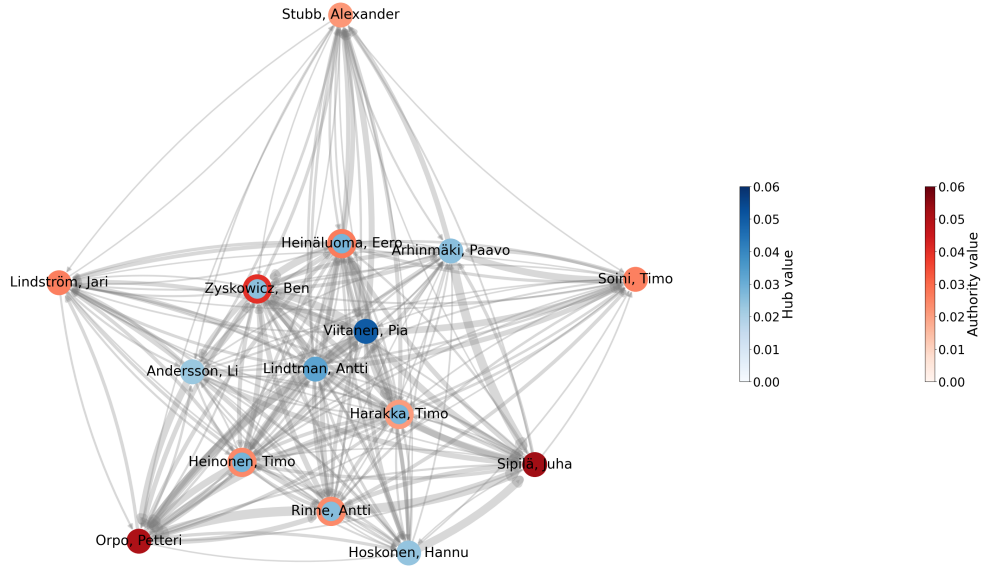
Nodes with highest authority values contain ministers and/or party leaders (Sipilä, Soini, Orpo, Stubb, Lindström, and Rinne) and MPs who have also highest hub values (Rinne, Zyskowicz, Heinonen, Heinäluoma, and Harakka). Ministers and party leaders presumably get mentions when discussing matters related to the minister's area of responsibility or issues related to parties. Nodes that have both high authority and hub values can be considered to be central people in parliamentary discussions; they mention other MPs often and they are also getting answers by receiving person name mentions. Other hubs include Heinonen, Arhinmäki, Anderson, Viitanen, and Lindtman. Hubs have given a lot of speeches and were mentioned many other MPs. Seven of the top ten authorities are government politicians, and seven of the top hubs are members of opposition parties.

## 5.3 Co-citation and Bibliographical Coupling Networks of MPs

Co-citation and bibliographical coupling networks were also constructed for MPs based on the speeches of the electoral term 2015–2018. The resulting networks have densities close to 1.0, i.e., almost every pair of MPs have been mentioned by at least one same MP and almost every pair of MPs have mentioned at least one same person. We tried to look for clusters of MPs based on how often they have been co-mentioned and how often they have mentioned same MPs.



**Figure 4:** MDS and HC results for parties during three different time periods. Parties are colored based on HC results. Parties colored with yellow are also government parties whereas parties colored with blue are opposition parties.

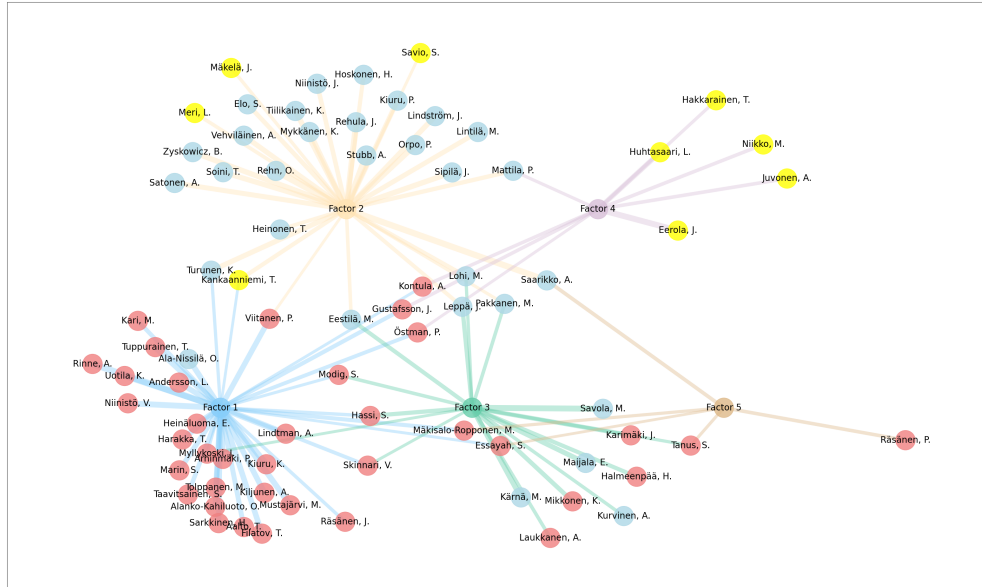


**Figure 5:** Ten MPs with highest hub and authority values based on the HITS algorithm. The darker red, the larger authority value, and the darker blue, the larger hub value.

MP	Party	Speeches	In-degree	Out-degree	Eigenvector centrality
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)
Heinenon, 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)
Heinälüoma, 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)
Hoskohen, 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)

**Table 3:** MPs with ten highest hub and/or authority values, their parties, number of given speeches, in-degree (mentions received), out-degree (mentions made) and eigenvector centrality. Number inside brackets tell how MPs high rank out of all MPs active during electoral term 2015–2018.





**Figure 7:** FA with Promax rotation results based on bibliographical couplings for 100 MPs that have mentioned most often other MPs. MP is connected to the factor if they have loading of at least 0.4 to the factor. Five factors explain 81% of the variance. Members of opposition parties are colored with red, members of government parties with blue, and members of the Finns party that belonged first to government and then to opposition during chosen time period are colored with yellow.

The results of our case study suggest that the methods used can reveal potentially interesting phenomena in parliamentary discussions. For example, the parliamentary role, government or opposition, of parties seems to have big effect on who mentions whom. Reference networks of MPs reveal the most active debaters as well as biggest authorities. However, interpreting results more profoundly requires domain knowledge and close reading of the related speeches. The ParliamentSampo semantic portal<sup>8</sup> can support both distant and close reading tasks by integrating semantic faceted search and browsing facilities with data-analytic tools. Faceted search can be used to filtering subsets of speeches to be analyzed, or to finding certain speeches when interpreting the results [51].

Our results suggest that taking the context into account, the distinction based on person name mentions between opposition and government politicians becomes more clear. In future analyses, it would be interesting to include the context of person name mentions in the analyses by, for example, taking into account the topics of the speeches. Alternatively, the speeches could be limited to a certain topic in order to study behaviour of parties and MPs related to certain topics.

<sup>8</sup>Portal in use at: <https://parlamentisampo.fi>; project homepage: <https://seco.cs.aalto.fi/projects/sem parl/en/>



## 7 List of Abbreviations

### Abbreviations of Political Parties in Finland

KD	Christian Democrats ( <i>Suomen Kristillisdemokraatit</i> )
Kesk.	Centre Party ( <i>Suomen Keskusta</i> )
KL	Finnish Reform Party ( <i>Korjausliike</i> )
Kok.	National Coalition Party ( <i>Kansallinen Kokoomus</i> )
Liik.	Movement Now ( <i>Liike Nyt</i> )
PS	Finns Party ( <i>Perussuomalaiset</i> )
RKP	Swedish People’s Party ( <i>Suomen ruotsalainen kansanpuolue</i> )
SDP	Social Democratic Party ( <i>Suomen Sosialidemokraattinen Puolue</i> )
TL	Seven Star Movement ( <i>Seitsemän tähden liike</i> )
Vas.	Left Alliance ( <i>Vasemmistoliitto</i> )
Vihr.	Green League ( <i>Vihreä liitto</i> )
VKK	Power Belongs to People ( <i>Valta kuuluu kansalle</i> )

### Other Abbreviations

ABCA	author bibliographical coupling analysis
ACA	author co-citation analysis
CA	citation analysis
FA	factor analysis
HC	agglomerative hierarchical clustering
MDS	multidimensional scaling
MP	member of Parliament
NEL	named entity linking
NER	named entity recognition
NLP	natural language processing

**Acknowledgments.** We thank Minna Tamper for her work with named entity recognition and linking related to the ParliamentSampo system, and other members of the larger project team. Thanks to Ari Apiola, Sari Wilenius, Päivikki Karhula, and other collaborators at the Parliament of Finland for help regarding provision of the data. Our work was funded by the Research Council of Finland (formerly Academy of Finland) and is related to the EU project InTaVia<sup>9</sup>, the EU COST action Nexus Linguarum<sup>10</sup>, and the Eminentia Grant by the Finnish Cultural Foundation to the third author. Computing resources of the CSC – IT Center for Science were used.

**Data Availability.** The data can be downloaded at <https://doi.org/10.5281/zenodo.7636420>. The dataset is also available on a public SPARQL endpoint <http://ldf.fi/semparl/sparql> and data can be browsed on the ParliamentSampo semantic portal <https://parlamenttisampo.fi/fi/> (in Finnish).

**Code Availability.** Google Colab Notebooks, modified data and Python script for data modification are available at <https://doi.org/10.5281/zenodo.8143795>.

---

<sup>9</sup><https://intavia.eu>

<sup>10</sup><https://nexuslinguarum.eu>

## References

- [1] Benoît, C., Rozenberg, O. (eds.): Handbook of Parliamentary Studies: Interdisciplinary Approaches to Legislatures. Edward Elgar Publishing, UK (2020). <https://doi.org/10.4337/9781789906516>
- [2] Laponi, E., Søyland, M.G., Vellidal, E., Oepen, S.: The Talk of Norway: a richly annotated corpus of the Norwegian parliament, 1998–2016. *Lang Resources & Evaluation* **52**, 873–893 (2018) <https://doi.org/10.1007/s10579-018-9411-5>
- [3] Van Aggelen, A., Hollink, L., Kemman, M., Kleppe, M., Beunders, H.: The debates of the European Parliament as Linked Open Data. *Semantic Web – Interoperability, Usability, Applicability* **8**(2), 271–281 (2017) <https://doi.org/10.1007/s42001-019-00060-w>
- [4] Bojārs, U., Dargis, R., Lavrinovičs, U., Paikens, P.: LinkedSaeima: A linked open dataset of Latvia’s parliamentary debates. In: *Semantic Systems. The Power of AI and Knowledge Graphs. SEMANTiCS 2019*, pp. 50–56. Springer, Berlin, Heidelberg (2019). [https://doi.org/10.1007/978-3-030-33220-4\\_4](https://doi.org/10.1007/978-3-030-33220-4_4)
- [5] Sinikallio, L., Drobac, S., Tamper, M., Leal, R., Koho, M., Tuominen, J., Mela, M.L., Hyvönen, E.: Plenary debates of the Parliament of Finland as linked open data and in Parla-CLARIN markup. In: *3rd Conference on Language, Data and Knowledge, LDK 2021*, pp. 1–17. Schloss Dagstuhl- Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Germany (2021). <https://drops.dagstuhl.de/opus/volltexte/2021/14544/pdf/OASiCS-LDK-2021-8.pdf>
- [6] Leskinen, P., Hyvönen, E., Tuominen, J.: Members of Parliament in Finland knowledge graph and its linked open data service. In: *International Conference on Semantic Systems, 6-9 September 2021, Amsterdam, The Netherlands*, pp. 255–269 (2021). <https://doi.org/10.3233/SSW210049>
- [7] Hyvönen, E., Sinikallio, L., Leskinen, P., Mela, M.L., Tuominen, J., Elo, K., Drobac, S., Koho, M., Ikkala, E., Tamper, M., Leal, R., Kesäniemi, J.: Finnish parliament on the semantic web: Using ParliamentSampo data service and semantic portal for studying political culture and language. In: *DIPADA, Digital Parliamentary Data in Action. CEUR Workshop Proceedings, Vol. 3133, Aachen* (2022). <http://ceur-ws.org/Vol-3133/paper05.pdf>
- [8] Hyvönen, E., Sinikallio, L., Leskinen, P., Drobac, S., Leal, R., La Mela, M., Tuominen, J., Poikkimäki, H., Rantala, H.: Plenary speeches of the parliament of finland as linked open data and data services. In: *Proceedings of the International Workshop of Knowledge Generation from Text (TEX2KG), Co-located with ESWC 2023. CEUR Workshop Proceedings, Aachen* (2023). In press, preprint: <https://seco.cs.aalto.fi/publications/2023/hyvonen-et-al-ps-text2kg-2023.pdf>
- [9] Blaxill, L., Beelen, K.: A feminized language of democracy? the representation

- of women at Westminster since 1945. *Twentieth Century British History* **27**(3), 412–449 (2016) <https://doi.org/10.1093/tcbh/hww028>
- [10] Quinn, K.M., Monroe, B.L., Colaresi, M., Crespin, M.H., Radev, D.R.: How to analyze political attention with minimal assumptions and costs. *American Journal of Political Science* **54**, 209–228 (2010) <https://doi.org/10.1111/j.1540-5907.2009.00427.x>
- [11] Beelen, K., Thijm, T.A., Cochrane, C., Halvemaan, K., Hirst, G., Kimmins, M., Lijbrink, S., Marx, M., Naderi, N., Rheault, L.: Digitization of the Canadian parliamentary debates. In: Säily, T., Nurmi, A., Palander-Collin, M., Auer, A. (eds.) *Exploring Future Paths for Historical Sociolinguistics*, pp. 83–107. John Benjamins, Amsterdam (2017). <https://doi.org/10.1017/S0008423916001165>
- [12] Guldi, J.: Parliament’s debates about infrastructure: An exercise in using dynamic topic models to synthesize historical change. *Technology and Culture* **60**(1), 1–33 (2019) <https://doi.org/10.1353/tech.2019.0000>
- [13] Ihalainen, P., Sahala, A.: Evolving conceptualisations of internationalism in the UK parliament: Collocation analyses from the League to Brexit. In: Fridlund, M., M., Oiva, Paju, P. (eds.) *Digital Histories: Emergent Approaches Within the New Digital History*, pp. 199–219. Helsinki University Press, Helsinki (2020). <https://doi.org/10.33134/HUP-5-12>
- [14] Kettunen, K., La Mela, M.: Semantic tagging and the Nordic tradition of everyman’s rights. *Digital Scholarship in the Humanities* (2021) <https://doi.org/10.1093/dsch/fqab052>
- [15] Tamper, M., Leskinen, P., Hyvönen, E., Valjus, R., Keravuori, K.: Analyzing biography collection historiographically as linked data: Case National Biography of Finland. *Semantic Web – Interoperability, Usability, Applicability* **14**(2), 385–419 (2023) <https://doi.org/10.3233/SW-222887>
- [16] Leskinen, P., Rantala, H., Hyvönen, E.: Analyzing the Lives of Finnish Academic People 1640–1899 in Nordic and Baltic Countries: AcademySampo Data Service and Portal. In: 6th Digital Humanities in Nordic and Baltic Countries Conference, Long Paper. CEUR Workshop Proceedings, Vol. 3232, Aachen (2022). <https://ceur-ws.org/Vol-3232/>
- [17] Elson, D.K., McKeown, K., Dames, N.J.: Extracting social networks from literary fiction. *aclweb.org*. <https://www.aclweb.org/anthology/P10-1015.pdf>
- [18] Poikkimäki, H.: Eduskunnan Täysistuntojen Puheenvuorojen Henkilömainintoihin Perustuvien Verkostoiden Analyysi. MSc Thesis. <https://seco.cs.aalto.fi/publications/2023/poikkimaki-msc-2023.pdf>
- [19] Hidén, M., Honka-Hallila, H.: Miten Eduskunta Toimii (How Parliament of

- Finland Works). Edita Publishing, Helsinki, ??? (2006)
- [20] Vespignani, A.: Twenty years of network science. *Nature* **558**(7711), 528–529 (2018) <https://doi.org/10.1038/d41586-018-05444-y>
- [21] Watts, D.J., Strogatz, S.H.: Collective dynamics of ‘small-world’ networks. *Nature* **393**(6684), 440–442 (1998) <https://doi.org/10.1038/30918>
- [22] Saramäki, J., Moro, E.: From seconds to months: an overview of multi-scale dynamics of mobile telephone calls. *The European Physical Journal B* **88**(6) (2015) <https://doi.org/10.1140/epjb/e2015-60106-6>
- [23] Leifeld, P.: Policy debates and discourse network analysis: A research agenda. *Politics and Governance* **8**(2), 180–183 (2020) <https://doi.org/10.17645/pag.v8i2.3249>
- [24] Curran, B., Higham, K., Ortiz, E., Vasques Filho, D., Gargiulo, F.: Look who’s talking: Two-mode networks as representations of a topic model of new zealand parliamentary speeches. *PloS one* **13**(6), 0199072 (2018) <https://doi.org/10.1371/journal.pone.0199072>
- [25] Bhattacharya, C.: Gatekeeping the plenary floor: Discourse network analysis as a novel approach to party control. *Politics and Governance* **8**(2), 229–242 (2020) <https://doi.org/10.17645/pag.v8i2.2611>
- [26] Esteve Del Valle, M., Broersma, M., Ponsioen, A.: Political interaction beyond party lines: Communication ties and party polarization in parliamentary twitter networks. *Social Science Computer Review* **40**(3), 736–755 (2022) <https://doi.org/10.1177/0894439320987569>
- [27] Forsman, M.: *Julkaisut Ja Tieteen Mittaaminen : Bibliometriikan Käännökohtia*, 2nd edn. Enostone kustannus, Vantaa, Finland (2022). <https://doi.org/10.31885/9789515150622>
- [28] Yan, E., Ding, Y.: 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), 1313–1326 (2012) <https://doi.org/10.1002/asi.22680>
- [29] Culnan, M.J.: Mapping the intellectual structure of MIS, 1980-1985: A co-citation analysis. *MIS Quarterly: Management Information Systems* **11**(3), 341–350 (1987) <https://doi.org/10.2307/248680>
- [30] Zupic, I., Čater, T.: Bibliometric methods in management and organization. *Organizational Research Methods* **18**(3), 429–472 (2015) <https://doi.org/10.1177/1094428114562629>

- [31] Zhao, D., Cappello, A., Johnston, L.: Functions of uni- and multi-citations: Implications for weighted citation analysis. *Journal of Data and Information Science* **2**(1), 51–69 (2017) <https://doi.org/10.1515/jdis-2017-0003>
- [32] Jeong, Y.K., Song, M., Ding, Y.: Content-based author co-citation analysis. *Journal of Informetrics* **8**(1), 197–211 (2014) <https://doi.org/10.1016/j.joi.2013.12.001>
- [33] Backhaus, K., Lügger, K., Koch, M.: The structure and evolution of business-to-business marketing: A citation and co-citation analysis. *Industrial Marketing Management* **40**(6), 940–951 (2011) <https://doi.org/10.1016/j.indmarman.2011.06.024>
- [34] Tamper, M., Leal, R., Sinikallio, L., Leskinen, P., Tuominen, J., Hyvönen, E.: Extracting knowledge from parliamentary debates for studying political culture and language. In: Tiwari, S., Mihindukulasooriya, N., Osborne, F., Kontokostas, D., D’Souza, J., Kejriwal, M. (eds.) *TEXT2KG 2022 & MK 2022. First International Workshop on Knowledge Graph Generation From Text and First International Workshop on Modular Knowledge*, pp. 70–79. *CEUR Workshop Proceedings, Vol. 3184*, Aachen (2022). [http://ceur-ws.org/Vol-3184/TEXT2KG\\_Paper\\_5.pdf](http://ceur-ws.org/Vol-3184/TEXT2KG_Paper_5.pdf)
- [35] Poikkimäki, H., Leskinen, P., Tamper, M., Hyvönen, E.: Analyses of networks of politicians based on linked data: Case ParliamentSampo – Parliament of Finland on the Semantic Web. In: *Semantic Web and Ontology Design for Cultural Heritage (SWODCH 2022)*, Turin, Italy, Proceedings. *CEUR WS Proceedings*, Aachen (2022). Accepted
- [36] Hagberg, A.A., Schult, D.A., Swart, P.J.: Exploring network structure, dynamics, and function using networkx. In: Varoquaux, G., Vaught, T., Millman, J. (eds.) *Proceedings of the 7th Python in Science Conference*, Pasadena, CA USA, pp. 11–15 (2008)
- [37] McCain, K.W.: Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*. **41**(6), 433–443 (1990) [https://doi.org/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)
- [38] Leydesdorff, L., Vaughan, L.: 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), 1616–1628 (2006) <https://doi.org/10.1002/asi.20335>
- [39] Zhao, D., Strotmann, A.: 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), 2070–2086 (2008) <https://doi.org/10.1002/asi.20910>

- [40] West, J.D., Jensen, M.C., Dandrea, R.J., Gordon, G.J., Bergstrom, C.T.: 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), 787–801 (2013) <https://doi.org/10.1002/asi.22790>
- [41] Andrews, J.E.: An author co-citation analysis of medical informatics. *Journal of the Medical Library Association : JMLA* **91**(1), 47–56 (2003)
- [42] Hyun, Y.-S., Cho, D., Yoon, S.W.: Landscape of human resource development research in korea: Results from author co-citation network analysis. *Human Resource Development International* **18**(5), 446–463 (2015) <https://doi.org/10.1080/13678868.2015.1079295>
- [43] Rousseau, R., Zuccala, A.: A classification of author co-citations: Definitions and search strategies. *Journal of the American Society for Information Science and Technology* **55**(6), 513–529 (2004) <https://doi.org/10.1002/asi.10401>
- [44] Ma, R.: Author bibliographic coupling analysis: A test based on a chinese academic database. *Journal of informetrics* **6**(4), 532–542 (2012) <https://doi.org/10.1016/j.joi.2012.04.006>
- [45] McLaren, C.D., Bruner, M.W.: Citation network analysis. *International Review of Sport and Exercise Psychology* **15**(1), 179–198 (2022) <https://doi.org/10.1080/1750984X.2021.1989705>
- [46] Van Eck, N.J., Waltman, L., Dekker, R., Van Den Berg, J.: A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology* **61**(12), 2405–2416 (2010) <https://doi.org/10.1002/asi.21421>
- [47] Makkonen, K., Loukasmäki, P.: Eduskunnan täysistunnon puheenaiheet 1999–2014: Miten käsitellä lda-aihemalleja? *Politiikka* **61**(2), 127–159 (2019). Chap. Artikkelit
- [48] Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery* **46**(5), 604–632 (1999) <https://doi.org/10.1145/324133.324140>
- [49] Calero-Medina, C., Noyons, E.C.M.: 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), 272–279 (2008) <https://doi.org/10.1016/j.joi.2008.09.005>
- [50] Kleinberg, J.M.: Authoritative sources in a hyperlinked environment. *Journal of the Association for Computing Machinery* **46**, 604–632 (1999) <https://doi.org/10.1145/324133.324140>

- [51] Hyvönen, E., Leskinen, P., Rantala, H.: Integrating faceted search with data analytic tools in the user interface of ParliamentSampo - Parliament of Finland on the Semantic Web. In: Proceedings of ESWC 2023, Poster and Demo Papers. Springer, Cham (2023). In press