Legal Knowledge and Information Systems
A. Wyner and G. Casini (Eds.)

doi:10.3233/978-1-61499-838-9-95

Answering Legal Research Questions About Dutch Case Law with Network Analysis and Visualization

Dafne VAN KUPPEVELT [a], Gijs VAN DIJCK [b]
[a] *Netherlands eScience Center*
[b] *Maastricht University*

Abstract. The availability of large collections of digitalized legal texts raises an opportunity for new methodologies in legal scholarship. Analysis of citation networks of case law gives insight into which cases are related and to determine their relevance. Software tools that provide an graphical interface to case law networks are required in order to enable non-technical researches to use network analysis methodologies. In this study, we present open source software for the analysis and visualization of networks of Dutch case law, aimed for use by legal scholars. This technology assists in answering legal research questions, including determining relevant precedents, comparing the precedents with those identified in the literature, and determining clusters of related cases. The technology was used to analyze a network of cases related to employer liability.

Keywords. network analysis, case law, visual analytics

1. Introduction

Legal documents such as case law and legislation are increasingly made available to the public. In the Netherlands, the government provides a dataset with the most important case law as XML files on *www.rechtspraak.nl.* The LiDO data bank[1] (also provided by the Dutch government) offers a linked data platform linking different legal sources. It contains meta data of Dutch and European case law and legislation and, more recently, also computer-identified references in Dutch case law to legislation and other case law. Computer-processing of these datasets allows legal researchers to investigate a large number of cases, in contrast to traditional methods that focus on a few, allegedly relevant cases. One way to represent a collection of data with references is as a network. Since decisions of judged can form precedents for future cases, the references between cases represent how case law is made by judges. Thus a network representation reveals the structure in the data and provides insight into legal questions that are very difficult to answer by looking at cases individually. However, graphical interfaces to the underlying data structures are needed for non-technical legal scholars.

[1] http://linkeddata.overheid.nl/front/portal/lido

Several previous studies have applied network analysis on case law. Fowler et al. [1] apply network analysis on Supreme Court cases in the US to determine relevance based on network statistics. Winkels [2] applies network statistics on Dutch case law, Derlén and Lindholm [3] on European case law and Schaper [4] on European direct tax law. A few web-based legal network visualization exist, such as EUcaseNet [5] and LexMex [2]. There are also commercial tools that apply network analysis and visualization for legal practice, such as Ravellaw [3] for US case law, and Juribot [4] for Dutch legal data. Despite many experiments with network analysis on legal data, generic tools that empower a non-technical legal researcher to explore the structure of legal data are difficult to find. This is most likely due to the relative lack of publicly available APIs to several sources of law. Consequently, collecting data for network analysis is often time-consuming and requires technical expertise to transform the data into the proper format for visualization and subsequent analysis.

Our work takes the viewpoint of the legal scholar without a technical background, and provides a generic open source technology based on publicly available data. We evaluate the technology from the legal perspective, by studying an example network with the technology and showing how legal research question can be answered for this network.

2. Network Analysis on Case Law

The network approach views cases as *nodes* in the network, and references between cases as *edges*. Following previous related studies, the following network statistics have been defined as possibly meaningful for research in case law networks:

- *In-degree*: the number of incoming references. Considering the concept of precedent in case law, cases that are referenced frequently, are more likely to be important than cases that are not frequently cited.
- *Out-degree*: the number of outgoing references. Cases with a large out-degree can be considered well-grounded, since the decision is based on many sources [3].
- *HITS hubs and authorities*: the HITS algorithm [6] gives an authority score, meaning how much a node is cited by nodes that are 'hubs', and a hub score, meaning how much a node is cited by nodes that are 'authorities'. Nodes with a high hub-score thus represent cases that have many citations to authoritative cases.
- *Relative in-degree*: the number of incoming references, corrected for the number of cases that exist later in time. Introduced by Tarissan and Nollez-Goldbach [7], this metric attempts to account for the fact that early cases have a larger in-degree, simply because there is more opportunity to be cited by succeeding cases.
- *Betweenness centrality*: A measure for how many shortest paths go through a node, i.e. how important the node is to connect the network. A large betweenness centrality can indicate that a case connects several subareas of the network.
- *Pagerank*: The PageRank algorithm [8] assigns scores to nodes based on the scores of incoming references. Although it has been argued that PageRank is difficult to use for case law networks [2], Derlén and Lindholm [3] have used it to determine importance of cases.

[2] http://www.lexmex.fr/
[3] http://ravellaw.com/
[4] https://referenties.semlab.nl/

Additionally, the Louvain community detection method [9] is used to identify clusters of nodes that form a community in the network. This can be used in the visualization to color the nodes by community, so that the parts of the network are visually separated. Tarissan et. al. [10] have shown that applying network statistics on a subnetwork around a certain topic can highlight landmark cases, in contrast to analysis on a complete network of a specific court. Therefore, we aim to develop an application for researchers to collect networks based on a specific set of cases related to a topic of interest.

3. Technology

The technology developed in this project consists of two web-based tools: the *caselawnet querier* [11] and the *case-law-app visualization* [12]. The querier application was built on top of the search API of rechtspraak.nl and used the link extractor API of LiDO. The querier allows users to search case law on keywords, and construct networks from a collection of cases. It is possible to include cases that were not in the original search result, but that are linked to one of the cases in the result. It is also possible to construct a network based on a user-defined set of cases, optionally including linked cases. The caselawnet querier is built in Python and Flask[5], and uses NetworkX[6] for calculating network statistics. The networks can be downloaded as csv-files of the nodes or links for further data analysis, or as JSON files to use in the visualization application.

Our visualization tool uses the Javascript library SigmaJS for graph rendering. The ForceAtlas2 layout algorithm [13] is used to position the nodes. In this algorithm, nodes that are connected through an edge are pulled closer together, resulting in a layout that emphasizes the structure of the network. The user can filter or change the appearance of the network, based on attributes of the nodes. These attributes include metadata of the cases, such as the court or year of the decision. It also includes network statistics, as described in the previous section.

4. Results in Legal Research

We discuss a number of research questions in legal research that can be answered with the technology presented in this paper. We illustrate these with an example network consisting of a set of 154 cases of Dutch supreme court, related to employer liability. This dataset was collected manually between 15 January and 5 April 2016, when computer-identified references were not available yet. Cases that were not directly about employer liability, but that were referenced by one of the employer liability cases were included as well. The citation network was enriched with meta data and network statistics using *caselawnet* and visualized with the *case-law-app*.

Which thematic subareas exist in the collection of cases? The Louvain method provides a starting point to define thematic subareas. In the employer liability network, 28 communities are detected by the Louvain method. Selecting only communities that are connected to the rest of the network, six communities are left. A qualitative analysis of

[5]http://flask.pocoo.org/
[6]https://networkx.github.io/

Table 1. Most cited decisions in the causality clusters

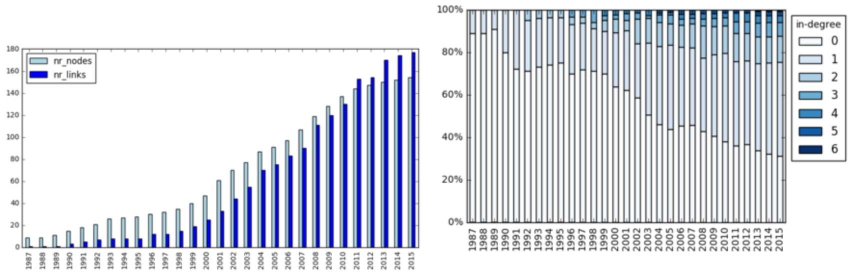
Name	ECLI	In-Degree	Authority	Relative In-Degree
Unilever/Dikmans	ECLI:NL:HR:2000:AA8369	6	0.165	0.055
Havermans/Luyckx	ECLI:NL:HR:2006:AW6166	4	0.138	0.069
Nefalit/Karamus	ECLI:NL:HR:2006:AU6092	3	0.087	0.051

the legal content of the cases inside the communities shows thematic coherence within the communities. The themes of the six connected clusters are identified as follows: 1. Asbestos; 2. Duty of care (extent), recipients' liability, contributory negligence, gross negligence; 3. Causality; 4. Losses while carrying out work, duty to insure; 5. Reversal rule; 6. Experts, evidentiary burden. These themes partly overlap with the main subjects of employer liability mentioned in literature. For example, the reference work Asser [14] describes 22 subjects in the area of employer liability, among which are the subjects *duty of care*, *duty to insure*, *reversal rule*. Since most legal literature mentions many different subjects and doesn't attribute a subject to every case, it is difficult to make a quantitative comparison with the subjects from the clusters.

What are cases that have an important precedent value in the collection of case law? Network statistics can be used to identify important cases based on high relevance scores, both for the complete network and for each of the thematic clusters. The metrics in-degree, authority and relative in-degree all give an indication of the precedent value of a node in the network. We will thus study cases with high values for one or more of these measures. For the employer liability network, we will give an example for one of the *causality* cluster, but the same can be done for the rest of the network. The three most cited cases in the cluster are shown in Table 1. *Unilever/Dikmans* has the highest in-degree, but *Havermans/Luyckx* has the highest relative in-degree. Looking at the content, *Unilever/Dikmans* decides that the reversal rule is applicable, which forms an important precedent. *Havermans/Luyckx* builds on this decision by adding that it is the employee who needs to argue convincingly that he suffers from health problems that may be caused by the exposure to health hazards. In this way, *Havermans/Luyckx* takes over the role of precedent for cases that determine when the burden of evidence regarding causality is shifted from the employee to the employer. We also look at the case with the highest *betweenness centrality*, which is *Fransen/Stichting Pasteurziekenhuis* (ECLI:NL:HR:1999:AA3837). This case indeed connects clusters about asbestos, causality, the duty to insure and the reversal rule. The decision was a landmark case after the introduction of the 1992 Dutch Civil Code that gave direction as to how to interpret and apply the then new provision on employer liability.

How does the importance of cases change over time? The way in which judges apply the law depends on the context of the decision, such as the time period in which decision was taken. The network structure changes over time, and thus the network statistics such as in-degree and relative in-degree are not static. This can be explored by plotting the variables over time. The plots were created with Python using the *caselawnet* software. The code is available as iPython Notebook[7]. The size of the network (number of nodes and number of links) over time is plotted in Figure 1a. Naturally, the number of cases grows over time, as well as the number of links. The number of links grows faster than the number of nodes, which means that the network becomes 'denser' over time. By

[7] `https://github.com/caselawanalytics/CaseLawAnalytics/blob/master/notebooks/TimeAnalysis.ipynb`



(a) The size of the network over time.

(b) The distribution of in-degree over the years: the fraction of cases for each value of in-degree are shown.

Figure 1. The development of the network over time

looking at the distribution of in-degree varying over the years, as shown in Figure 1b, we see the distribution shifting so that relatively fewer cases have no incoming references. This indicates that over time, when the network becomes denser, the new links cite more *different* cases, instead of the cases that already had a large in-degree. Note that the shift might be partly caused by the way the network was constructed: cases that were referenced by one of the employer liability cases were included as well, so the selection of cases is based on the references in the most recent network. A solution for this would be to construct the networks in the exact same manner for each year, which for this network would require manual work that was outside of the scope of this paper.

5. Discussion and Conclusion

We have shown that the *caselawnet* and *caselawapp* prototype applications enable legal scholars to conduct empirical research using large number of decisions, by providing a graphical interface to the Dutch *rechtspraak.nl* and *LiDO* data sources. We used the technology for the analysis of a network about employer liability in Dutch Case law, which resulted in the identification of thematic sub-areas and important precedent cases in these sub-areas. It also provided insight in the evolution of the legal network over time.

Network visualization can assist in learning about a specific area of law, not only in research but also in education and legal practice. To familiarize with a new legal area, the tool can be used to answer questions about the data, such as: What are the most-cited cases about? What thematic subareas exist in this collection? In what time period do many (important) cases appear? Do different courts, in particular the Supreme court, refer to different cases? How dense and how connected is the network? In other words, how many citations do cases have on average and how much interaction is there between different subareas?

Since network analysis is a novel methodology for legal researchers, guidance and education is required to get the legal community familiarized with network analysis. The network statistics, other than in-degree and out-degree, have no simple meaning and require thorough interpretation. Users have to be careful not to make false assumptions based on the technology, such as spatial closeness in the ForceAtlas2 layout. In addition, the manner in which the network was constructed can influence the results of the network analysis, as we have shown in the analysis of the distribution of in-degree over time. It is therefore important that the network construction is done carefully and documented well.

The data collection tool should be further improved, by combining keyword search with network exploration and filter options to assemble data collections in an interactive manner, possibly also including European legal sources (HUDOC and EurLex). Natural Language Processing techniques could assist in identifying themes of communities in the network. Research about developments of case law networks over time, as presented in this paper, could benefit from further visualizations that show the temporal aspect. Examples of such visualizations are animations of the network changing over time and interactive graphs that show statistics such as network size and in-degree varying in time.

The methodology can be evaluated further in a more quantitative manner, by asking a group of subjects to perform a series of tasks using the software and measuring their time use and satisfaction. The Louvain method could be further evaluated by manually annotating a data set with themes and validating the communities against the annotations. Lastly, it would be valuable to apply the network methodology to many different legal domains and compare the analyses across the networks. The methodology will eventually prove most value if it leads to novel research results in different legal domains.

References

[1] J. H. Fowler, T. R. Johnson, J. F. Spriggs, S. Jeon, and P. J. Wahlbeck, "Network analysis and the law: Measuring the legal importance of precedents at the US supreme court," *Political Analysis*, vol. 15, no. 3, pp. 324–346, 2007.

[2] R. Winkels, J. d. Ruyter, and H. Kroese, "Determining authority of Dutch case law," *Legal Knowledge and Information Systems*, vol. 235, pp. 103–112, 2011.

[3] M. Derlén and J. Lindholm, "Goodbye van Gend en Loos, Hello Bosman? using network analysis to measure the importance of individual cjeu judgments," *European Law Journal*, vol. 20, no. 5, pp. 667–687, 2014.

[4] M. G. Schaper, "A computational legal analysis of Acte clair rules of eu law in the field of direct taxes," *World Tax Journal*, vol. 6, no. 1, 2014.

[5] N. Lettieri, A. Altamura, A. Faggiano, and D. Malandrino, "A computational approach for the experimental study of EU case law: analysis and implementation," vol. 6, 12 2016.

[6] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[7] F. Tarissan and R. Nollez-Goldbach, "Temporal properties of legal decision networks: a case study from the international criminal court," in *28th International Conference on Legal Knowledge and Information Systems (JURIX'15)*, 2015.

[8] L. Page, S. Brin, R. Motwani, and T. Winograd, "The PageRank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[9] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of statistical mechanics: theory and experiment*, vol. 2008, no. 10, p. P10008, 2008.

[10] F. Tarissan, Y. Panagis, and U. Šadl, "Selecting the cases that defined Europe: complementary metrics for a network analysis," in *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on*. IEEE, 2016, pp. 661–668.

[11] D. van Kuppevelt, "caselawanalytics/caselawanalytics," Aug. 2017, doi: 10.5281/zenodo.854541. [Online]. Available: https://github.com/caselawanalytics/CaseLawAnalytics

[12] ——, "Nlesc/case-law-app," Jul. 2017, doi: 10.5281/zenodo.596839. [Online]. Available: https://github.com/NLeSC/case-law-app

[13] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software," *PLOS ONE*, vol. 9, no. 6, pp. 1–12, 06 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0098679

[14] G. H. van Voss, *Asser 7-V Arbeidsovereenkomst*. Kluwer, 2012.