

NETWORK ANALYSIS IN LAW: A LITERATURE OVERVIEW AND RESEARCH AGENDA

Tereza Novotná

PhD candidate, Masaryk University, Faculty of Law, Institute of Law and Technology
Veveří 158/70, 611 80 Brno, CZ
tereza.novotna@mail.muni.cz; <https://www.law.muni.cz>

Keywords: *Network Analysis, Literature Review, Citation Analysis, Data Collection*

Abstract: *Application of network analysis in legal domain is not a new approach, however it has noted an increased interest past two years. As the network analysis methods are widely adopted in legal domain, it is the aim of this review to provide an overview of the most current efforts in this field. This review methodologically and temporely builds on the previous review of Whalen's from 2016. Firstly, this article adresses data sample issue and the importance of complexity of datasets. In the second part it shows some repetitive trends appearing in current network analysis research results and its interpretation.*

1. Introduction

The great part of every lawyer's job is making a legal research, which means searching for applicable legal norm or relevant case law. With the growing number of legal documents of all the relevant types (codes, case law or literature), this task becomes challenging even for experts. However, legal and computer experts are trying to provide tools for processing large number of legal texts and extracting relevant information even for users without any technical background (for example [KUPPEVELT ET AL. 2017]). Network analysis is just one of the ways to handle such issue. It is the mathematical and statistical method based on graph theory and therefore offering the graphical output of data. The graphical output is what makes this method suitable for handling large amount of data and references among them (for instance individual judicial decisions, codes or even articles).

The network usually contains two types of data: *nodes* (or *vertices*) and *edges* (or *links*). Nodes are usually represented by legal sources (such as individual codes, laws, articles or judicial decisions) and edges are represented as relations among them (such as citation references or similarities). Network statistics then defines several descriptive variables based on these two types.

In this paper, I will provide a literature overview of current research in the field of legal network analysis and compare different approaches of legal scholars to application of network analysis in law. This way I will build on the last try to summarize works related to applying network analysis in law.

2. Framework for review: Whalen's Legal Networks

Last effort to compile a literature review on the network analysis in law is [WHALEN 2016] which describes network analysis methods and builds a strong theoretical background and summarizes the main efforts in applying such methods in different areas (namely case law analysis, patent citation analysis, statutory and regulatory analysis, social and organizational analysis). It shows a broad spectrum of possible use of network analysis in legal domain and advocates its potential use *pro futuro*. Subsequently, it shows different ways how to improve the methodology of network analysis to provide more accurate results. According to Whalen, the improvements lie mainly in data. He suggests to include more nuanced data and use techniques supporting more dimensions of data and its metadata to overcome its binary representation.

2.1. Data

Whalen argues, that the reduction of relationship among different legal sources into the binary links discards other contextual information. For this reason he suggests not to represent an edge of a network as a binary code but more as a multidimensional variable or simply a weighted edge.¹ That means to add another measurement to the links that shows the number of mutual citations – the weight. Subsequently, he suggests to use different types of relationships (edges) between nodes or vertices.² For example legal science usually distinguishes at least two types of citations between judgments – positive and negative. This distinction is highly relevant for citation analysis. As such these types do not usually have the same value and the same meaning. This important contextual characteristic is missing if the data are evaluated only as «existing» or «non-existing» links. On top of these two suggestions Whalen advocates using a multiplex network analysis to provide a multi-level overview of data.

2.2. Techniques

Firstly, Whalen describes different statistical models that are appropriate to use in legal network analysis. However he particularly focuses on two aspect of used technology – dynamics and theory. Network dynamics is described as a change of network over time as every legal system changes over time (with new documents and new links among them). Focus on static characteristics of legal networks can cause a loss of much of its context.³ From the theoretical point of view, Whalen emphasizes the need to cover also other levels of examined legal network than the structure and the centrality. These levels are simply summarized as «causes» of such a state of a network – simply why is the network structured this particular way. He proposes multilevel approach to legal networks – combining it with actor-level analysis and generally judicial behavioral analysis.⁴

3. Review methodology and included works

This review aims to cover papers published after Whalen’s review in [WHALEN 2016] – that means published from 2016 to 2018. Papers included are those appeared in online databases named in reference.⁵ Considering the fact that during past two years the number of published papers tackling this issue has significantly increased, this review concerns only the papers covering application of network analysis on legal sources as it has the biggest potential to be exploited in practice according to the author of this review. Most of included papers offer advanced approach to applying network analysis and build on the previous knowledge and efforts.

However, this overview will cover also one paper published in 2015 [DERLÉN ET AL. 2015] since this paper was not mentioned in Whalen’s review and it is very important and impactful work. Although the criteria were relatively narrow, every reviewed paper deals with different partial issue of network analysis in legal domain and aims to different results. In this part I will briefly introduce main focus of included works for the purpose of following review. There is no need, according to the author, to extend the theoretical part of network analysis, as the latest changes don’t really tackle the theoretical questions and the background remains the same.

First reviewed paper is [KUPPEVELT ET AL. 2017] on application of network analysis on case law and focusing on visualization of output. The aim of this work is the actual software tool developed to help in legal research. The data sample of case law used for the analysis is not complete, it is based on the dutch government-run website where only «the most important» case law is published.⁶ Moreover, the datasets are completed with a

¹ WHALEN 2016, p. 555.

² WHALEN 2016, p. 555–556.

³ WHALEN 2016, p. 558.

⁴ WHALEN 2016, p. 558.

⁵ Databases used for research: <https://dl.acm.org/>; <https://dblp.uni-trier.de/>; <https://www.ebsco.com/>; <https://www.scopus.com/>; <https://scholar.google.com> was used for searching through other databases.

⁶ <https://www.rechtspraak.nl>.

linked data platform, which is another task that usually legal or computer scholars need to deal with themselves. The article describes methodology very briefly – this part consists only from a few variables defined and used in following part and technology used for application of network analysis to data is defined only in the way of naming programming language or graph rendering tools. On the other hand it defines all possible options of using developed tool for legal queries on case law and it shows valuable example too. The queries are divided into three main groups – the clustering of case law regarding topics, the precedent value of different case law and the temporal analysis of a network.

In paper [WAGH ET AL. 2017] authors use two different network analysis methods to find case law similarity, which the authors find suitable for detecting relevant precedents in legal research. These two methods are found suitable since the possibility of weighting edges describing different case law relations. Data sample is incomplete similarly to the previous paper, however the data sample used in this work is based on commercially-run website⁷ publishing case law. Unlike in the previous case, in this one authors don't mention the content of the data sample except pointing out that data provided by mentioned website are completely unprocessed and raw. Methodology used in this paper is weighting edges of network using either cosine similarity matrix, which is being derived or Jaccard similarity measure is used and the result number is used as a weight of an edge. Weighted network is then built and variables are defined depending on network measures. The final comparison of variables show clearly that using of citation network analysis provides more valuable results.

[PANAGIS ET AL. 2017] is a paper from Panagis, Šadl and Tarissan. The authors perform network analysis on citations of case law but furthermore, they include implicit citations. Authors believe it is crucial to achieve more accurate understanding of the role of case law of Court of Justice in European Union. For this task they use text similarity techniques to extract the arguments from previous cases which are not explicitly marked. Such a method demands more text preprocessing before applying network analysis on data. Authors divide the network analysis as such into two parts, in the first part they build a standard network of explicit citation, however they only use the paragraph references but not references to a judgment as a whole. This particular approach focuses on the legally most important parts of judgments, also it shows whether the judgment deals with several legal aspects at the same time. As a second part, the global references (references to entire case) are added to the existing network. Authors define two types of these references – explicit and implicit references. Text analysis tools are used to extract implicit references from judgments. References are then compared with case law to obtain implicitly cited judgment and furthermore, its particular paragraph that is being cited. With the combination of these two steps authors offer a prediction technique to determine the particular paragraph cited from the knowledge of a global reference only, which is very interesting part of this work. Finally, authors assess the relevance of cited paragraphs from a legal perspective. This approach promises to bring out more accurate citation network image. The advantage of a data sample used in this work is that EUR-Lex website provides complete datasets of CJEU judgments⁸.

[BOULET ET AL. 2018] is the second part of network analysis of french legal codes. Shortly, the first part [BOULET ET AL. 2011] focuses on building a French legal codes citation network. Subsequently, the second part focuses on the role of weights in an existing network of French legal codes created in [BOULET ET AL. 2011]. This approach brings out some interesting findings. The first one is that the codes with the most links to other codes are connected to links with the strongest weights, this effect is also known as a power law effect. The second finding is that a density in links predicts higher weight measures of mutual links or that a density correlates with high weighted measures. As the additional information are added, the result in the form of weighted graph is found to be more accurate in the terms of legal research. It is also one of the two included works which use legal codes as a data sample (and not only the case law). In French legal system, there are

⁷ <https://indiankanoon.org>.

⁸ <https://eur-lex.europa.eu> Authors used mostly english versions of a text which was supported with french version in some cases.

52 codes and authors use all of them in a full text. In [BOULET ET AL. 2011], the «link» between two codes was stated if there was an explicit citation between these two codes; this link was considered undirected. In [BOULET ET AL. 2017], this link was extended with one more measure of «weight» as the number of citations between two codes (in both directions). Therefore, the data sample here means all the explicit citations in all of the french codes.

[KONARIIS ET AL. 2017] is an exhausting work on the network analysis of all of the European Union legal sources. The aim of this work is to provide a complete system of European legal sources – a model. This aim is to be achieved through analysis including different levels of relationship between legal documents. Among others, this work confirms the power law effect in legal sources. It explicitly shows that a high number of legal sources depends on a small number of the high rated sources. On the legislation level, that means a change or a novelisation of a prescription can lead to an unexpected consequences in the whole legislation network. On the other hand the results of temporal analysis are not surprising. The number of edges grows faster than number the number of nodes and that means that a density is growing over time. In this case, data sample is a complete dataset of European Union legal sources. These are all of the treaties, international agreements, legislation, complementary legislation, preparatory acts and jurisprudence that are available to download at the EUR-Lex website. These data are complemented with metadata – date of effect and date of expiry. The complete dataset of legal sources allows to build a multilayer network which is even changing in time.

[DERLÉN ET AL. 2015] uses network analysis to adress legal question on the actual role of precedent of European Court of Justice in the legal system of European Union – it uses a mathematical methods to answer a fundamental legal question. It uses a distinction of three basic legal characteristics (action, actors, legal area) to define an overall impact on a judgment and its legal power and it shows that there is almost none actual impact on a precedential power of a judgment. However, this study contains some new findings. First of all it shows that preliminary actions are used as precedents more often than direct actions. And subsequently unsuccessful direct actions are used as precedents more often than succesful actions. Furthermore, this paper defines legal areas with the most important role of case law – mainly those related to internal market (competition law and fundamental freedoms). Concerning the actors, it shows that number of observations submitted by different Member States has an impact on the power of a judgment *pro futuro* – «*persuasive power*» but also on the precedential power of the judgment. This study confirms a power law effect in case law of CJEU as well as a correlation of a persuasive and precedential power variables. Authors here used whole dataset of case law provided on EUR-Lex website.

[DERLÉN ET AL. 2017] from same authors thematically follows the findings of their previous work and examines just a small sample of case law on internal market. It examines precedential and persuasive power as variables defined in previous work. They generally compare two types of measurements (HITS and PageRank) and highlight that the power depended on these measurements is changing over time. For that analysis authors use only a part of CJEU case law related to internal market.

4. Remarks on methodology of reviewed works

In the last part of the literature review I would like to follow up on the Whalen's review and point out some remarks on the methodology applied in reviewed papers. I will also try to compare the results of included works with the expectations of Whalen to what will actually bring out his methodological advices *pro futuro*. This way I would like to evaluate if the conclusions he provided were followed and if so, did it really improved the performance of network analysis methods. For that reasons I gathered them into two main thematical parts – issues connected to data sample and technology (to follow up Whalen's conclusions) and issues connected to the interpretation of results obtained by applying network analysis methods.

4.1. Data sample and technology

Network analysis is a method combining two different mathematical branches – statistics and graph theory. Thanks to this combination, it can easily process a large amount of data and display them graphically. Therefore network analysis is very common method used for different legal sources analysis as number of sources of law is constantly growing. This type of analysis then provides tools to process these data into more comprehensible graphical outputs.

For the reason that this method is partly based on statistics, the issue of a data sample and its characteristics is crucial. Furthermore, as this review is focusing on the role of network analysis in legal research, in order to retrieve complete and undistorted result the sample of sources ought to be complete. The conclusion is based purely on the practical aspects of the matter – non-complete sample of legal sources cannot guarantee the right result. Concerning the data sample of legal sources, another important question arises – the possibility to obtain complete dataset of digitalised legal sources itself. As this review shows next, unfortunately this issue is not the one that author can easily influence.

In spite of that, the complete data sample was not used in all reviewed works – it was used namely in [PANAGIS ET AL. 2017], [BOULET ET AL. 2018], [KONIARIS ET AL. 2017], [DERLÉN ET AL. 2015] with the temporal limitation and in [DERLÉN ET AL. 2017]. It is significant that works including complete datasets of legal sources analysed are mainly those analyzing legal sources of European Union. European Union is providing all the datasets within the Eur-Lex website database.⁹ Such a practice is not common in India where the analysis in the paper [WAGH ET AL. 2017] is held. Data analysed in this work come from private website providing indian legal sources to public, neither the paper nor the website provide information about the complexity or incompleteness of sources published.¹⁰ Paper number [KUPPEVELT ET AL. 2017] is analysing dutch case law, which is made public by government-run website, although the website does include only the most important case law.¹¹ Such a database is deprived of the «non-important cases» and that automatically arises a question: what the criteria for importance is and who is in charge to make such decision. Nevertheless, answers for this questions are beyond the scope of this review. Although it seems network analysis can successfully run on incomplete data samples too, for the purpose of legal research it should be complete data sample providing accurate representation of legal sources concerning legal queries.

Secondly, I would like to follow Whalen's proposition on more detailed data. We can observe that applying multidimensional approach to data actually leads to finer results as it is mainly in papers [PANAGIS ET AL. 2017], [BOULET ET AL. 2018], [KONIARIS ET AL. 2017] and [DERLÉN ET AL. 2015]. These examples show that despite its possible technical difficulty, this approach provides more information and even more potential to be used in practice even for public. In the [DERLÉN ET AL. 2015] case we can even see that purely mathematical method can be relevant in a legal-theoretical discussion on relevance of CJEU case law. Efforts including also temporal analysis show in [KUPPEVELT ET AL. 2017], [KONIARIS ET AL. 2017] and partly [DERLÉN ET AL. 2015], on the other hand, that a dynamical model of legal sources better follows its characteristics.

4.2. Results and interpretation

When taking a closer look at the results of reviewed works, an observer can spot a few repetitive tendencies. It is not really possible, from the methodological point of view, to compare the results as even if the network analysis method is moreless similar, the data sample is different in every paper. Notwithstanding, the aim of this part of review is to point out the patterns that some of the results are following.

⁹ <https://eur-lex.europa.eu>.

¹⁰ <https://indiankanoon.org>.

¹¹ <https://www.rechtspraak.nl>.

First of these trends shows that in all of the legal sources used in the reviewed papers there are few of them with a huge impact and on the other hand a vast majority with very low impact (this is in more precise version called power law). This trend is repeated in the case law analysis in [KUPPEVELT ET AL. 2017], [PANAGIS ET AL. 2017], [DERLÉN ET AL. 2015] and [DERLÉN ET AL. 2017] and even in legal codes and other sources analysis (as shown in [BOULET ET AL. 2018] and [KONARIS ET AL. 2017]). In addition, the interesting finding provides [PANAGIS ET AL. 2017] and [BOULET ET AL. 2018] because the complementing the network analysis with other methods is even strengthening this trend.

Another interesting observation concerns the temporal analysis in [KUPPEVELT ET AL. 2017] and [KONARIS ET AL. 2017] – both of these papers show that with the time the density of network is increasing and the number of edges among sources is growing faster than number of nodes (legal documents).

Every one of the papers aims to complete the work with a visualisation and graphical interface, but only in [KUPPEVELT ET AL. 2017] the software tool was actually developed.

5. Conclusion

This review follows Whalen's review in introducing different approaches to network analysis in legal domain in papers published within past two years. It aims to provide an overview for every other researcher who wants to follow up on applying network analysis methods on different legal sources in order to provide a better and more comprehensive view on different legal sources.

It specifically focuses on two issues concerning network analysis in law: the data collection or data sample issue and the results and its interpretation issue. It confirms some of Whalen's conclusions on data samples and their use in analysis – that more nuanced data and technology leads to more accurate results in network analysis.

6. References

- BOULET, OMAIN/MAZZEGA, PIERRE/BOURCIER, DANIELE, A network approach to the French system of legal codes—part I: analysis of a dense network, *Artificial Intelligence and Law*, 2011.
- BOULET, ROMAIN/MAZZEGA, PIERRE/BOURCIER, DANIELE, Network approach to the French system of legal codes part II: the role of the weights in a network, *Artificial Intelligence and Law*, 2018.
- DERLÉN, MATTIAS/LINDHOLM, JOHAN, Characteristics of Precedent: The Case Law of the European Court of Justice in Three Dimensions, *German Law Journal*, 2015.
- DERLÉN, MATTIAS/LINDHOLM, JOHAN, Is it Good Law? Network Analysis and the CJEU's Internal Market Jurisprudence, *Journal of International Economic Law*, 2017.
- KONARIS, MARIOS/ANAGNOSTOPOULOS, IOANNIS/VASSILIOU, YANNIS, Network Analysis in the Legal Domain: A complex model for European Union legal sources, *Journal of Complex Networks*, 2017.
- KUPPEVELT, DAFNE VAN/DIJK, GIJS VAN, Answering Legal Research Questions About Dutch Case Law with Network Analysis and Visualization, *Legal Knowledge and Information Systems: JURIX 2017: The Thirtieth Annual Conference*, B.m.: IOS Press, 2017.
- PANAGIS, YANNIS/SADL, URSKA/TARISSAN FABIEN, Giving every case its (legal) due: The contribution of citation networks and text similarity techniques to legal studies of European Union law, *30th International Conference on Legal Knowledge and Information Systems (JURIX'17)*, B.m.: IOS Press, 2017.
- WAGH, RUPALI/ANAND DEEPA, Application of citation network analysis for improved similarity index estimation of legal case documents000202f: A study, *2017 IEEE International Conference on Current Trends in Advanced Computing (ICCTAC)*, 2017.
- WHALEN, RYAN, *Legal Networks: The Promises and Challenges of Legal Network Analysis*, Michigan State Law Review, 2016.