Giving every case its (legal) due The contribution of citation networks and text similarity techniques to legal studies of European Union law

Yannis Panagis, Urska Sadl, Fabien Tarissan

Giving every case its (legal) due

The contribution of citation networks and text similarity techniques to legal studies of European Union law

Yannis PANAGIS [a,1], Urška ŠADL [a,b] and Fabien TARISSAN [c]
[a] *iCourts, Centre of Excellence for International Courts*
[b] *European University Institute*
[c] *Université Paris-Saclay, ISP, ENS Paris-Saclay, CNRS*

Abstract. In this article we propose a novel methodology, which uses text similarity techniques to infer precise citations from the judgments of the Court of Justice of the European Union (CJEU), including their content. We construct a complete network of citations to judgments on the level of singular text units or paragraphs. By contrast to previous literature, which takes into account only explicit citations of entire judgments, we also infer implicit citations, meaning the repetitions of legal arguments stemming from past judgments without explicit reference. On this basis we can differentiate between different categories and modes of citations. The latter is crucial for assessing the actual legal importance of judgments in the citation network. Our study is an important methodological step forward in integrating citation network analysis into legal studies, which significantly enhances our understanding of European Union law and the decision making of the CJEU.

Keywords. Network analysis, Citation networks, Text similarity, CJEU

1. Introduction

While citation network analysis has gained traction as an approach to understand law and courts, legal scholars remain reserved. Our study is motivated by this reticence, which can be summed up in three pertinent objections. We discuss them in turn. First, legal scholars see the approach as quantitative hence unfit for detailed qualitative investigations of legal rules and principles and their application to concrete cases. The latter are typically considered as the main purpose of a legal study. The position is reinforced by the existing use of the approach since, bar a few exceptions, the network approach has been paired with statistical quantitative analysis. The field is strongly focused on judicial behavior and judicial bias in legal decision making, using case citation networks to answer questions related to law rather than legal questions. This is particularly recognizable in research on the United States Supreme Court (USSC) but has been less true for studies of the CJEU [7]. Examples include inquiries into judicial activism, the rise of *stare decisis*, the depreciation of precedents in the USSC [4], citation strategies of international courts, and the (strategic) behavior of individual judges [8].

[1] Corresponding Author: University of Copenhagen, iCourts Centre of Excellence for International Courts, Karen Blixens plads 16, DK-2300, Copenhagen S, Denmark. email: ioannis.panagis@jur.ku.dk

Second, the network analysis approach treats all citations as equally important, and does not discriminate between different types of references. The judge might mention a case in passing or include it in a string citation; she might cite it to distinguish it from the case at hand or dismiss it as irrelevant, or because one of the parties to the case relied on it. She might, furthermore, cite the case as an example, to reason by analogy, rather than employ it as a binding, guiding, or even legally persuasive source of law that legally or *de facto* obliges her to reach a specific outcome. The information is crucial for the inquiry of what is valid law and which cases are the truly important reference points in a court's repository. The criticism is underscored by the fact that so far existing studies using case citations have in fact conceptualized citations as equal, and treated them as legally relevant. Moreover, the same studies have assumed that the "citation behavior of the Court provides information about which precedents serve important roles in the development of [...] law," [6]. The implication of this assumption would be that "[...] a judgment's value as a source of law is limited if it has never been cited by the Court of Justice." [7], which is something that most legal scholars would contest.

Third, because citation network analysis relies on explicit case citations *it can only be applied to courts with a developed and rigorous citation practice*. The criticism raises the question whether the method will yield inaccurate findings in the case of continental style courts like the CJEU, which tend to repeat the wording and the arguments established in past cases without citing the source (the so-called *implicit citations*). Albeit this is less true for the more recent judgments, the CJEU has especially in its earlier cases often resorted to such implicit citations.

To address the above challenges we combine the network approach with text analysis. First, we construct the citation network based on references to paragraphs of individual judgments as units rather than judgments as a whole (we use cases and judgments of the CJEU as synonyms). Namely, most cited paragraphs typically include a particular concept or a particular formulation, which is relevant in the process of construction of legal arguments[2]. By doing this we take a step further in identifying the aspects of cases that are legally important. We also acquire the information whether the case is relevant for one or several legal aspects. For instance, if only one paragraph of a case is repeatedly cited, the case is most likely important for resolving one legal issue. If, by contrast, there are several different paragraphs of one judgment that are cited, the case might be important for resolving more than one legal issue.

Second, we isolate the references that are directed to entire cases (global references) and use text similarity techniques to infer local references, references to particular parts of cases (paragraphs). The latter are called *implicit references* or *missing citations*, where the CJEU repeats the text of a particular paragraph / part of the judgment verbatim or with slight variations but does not cite it. Third, we assess the relevance of cited paragraphs from a legal perspective. We showcase our approach by evaluating the links to three of the best known cases in the CJEU doctrine: *Dassonville*, *Defrenne II* and *Francovich*[3].

[2]Typically, the judgments of the CJEU are separated into self-contained units or paragraphs, dealing with a particular point of law or fact. In the older judgments dating back to the 1970s the paragraphs are not numbered systematically. Later, in the 1980s, when the judgments became longer and the writing style of the CJEU more argumentative and informative, the CJEU began to number the paragraphs.

[3]Dassonville: case C-8/74, ECLI:EU:C:1974:82, Defrenne II: case C-43/75, ECLI:EU:C:1976:39 (also known as "Defrenne II") and Francovich: Case C-6/90, ECLI:EU:C:1991:428.

Table 1. The different types of references, illustrated on references to *Dassonville*.

Reference type	Paragraph
local	39. The prohibition of measures [...] (see, in particular, **Case 8/74 Dassonville [1974] ECR 837, *paragraph 5***; Case 178/84 Commission v Germany [1987] ECR 1227 ("Beer purity law"), paragraph 27; [...].
global	10 It must be recalled first of all that since its judgment in **Case 8/74 Procureur du Roi v Dassonville [1974] ECR 837**, the Court has consistently held [...]

To summarize, by leveraging the fine-grained paragraph data we: a) infer the missing citations and b) tease out and assess the potentially legally relevant parts of the cited case. Altogether, this information is crucial to evaluate the actual legal importance of a particular case and its influence on the development of legal doctrine.

The rest of the paper is organized as follows: in Section 2 we present the methods that we use, namely the missing link detection technique, lay out the assumptions and observations, on which we base our research strategy and explain the terminology (judicial formulas). In Section 3 we present a quantitative and qualitative evaluation of our the missing link detection technique and, finally, we conclude in Section 4.

2. Research Strategy, Method and Data

Our research strategy is based on a set of assumptions and observations about the CJEU and its style of decision writing.

2.1. Paragraphs, Networks, and Reference Types

Every judgment of the CJEU is divided into smaller text units or paragraphs. The paragraphs form the skeleton of the judgment and contain the legal arguments that the Court is communicating as well as references to previous cases (precedents), on which the CJEU relies in order to support these arguments.

We define a *judgment paragraph* as the part of a judgment that usually starts with an integer number, e.g. 10, and extends until the text paragraph starting with the next integer, i.e. 11, excluding perhaps quoted text. References (or citations) can be grouped into two categories: *local* references that precisely define which paragraph of the previous judgment is being cited with a number and *global* references where the entire judgment is cited without specifying the paragraph(s). Examples of both types are given in Table 1.

A citation network is defined as a pair $G = (V, E)$, where $V = V_{case} \cup V_{par}$ is the set of nodes, V_{case} is a set of cases that is referred to by means of global references, V_{par} is the set of judgment paragraphs that cite and get cited by means of local references, $E = E_{global} \cup E_{local}$ are the edges such that $E_{global} = \{(u,v) | u \in V_{par}, v \in V_{case}\}$ depict the global citations and $E_{local} = \{(u,v) | u,v \in V_{par}\}$ are the local ones. Lastly, we denote by $par(C)$ the set of all judgment paragraphs of a given case C.

2.2. Formulas

All language users depend on prefabricated phrases. That said, the language of courts is formalized to a much larger extent than natural language and is by far more repetitive. The language of the CJEU is particularly routinized, even when compared to the CJEU's

counterparts in France, Germany and the United Kindgom [1,14]. The CJEU makes use of a limited set of textual devices to construct its arguments [14]. These have been labeled judicial *formulas* in literature [1] and can be defined as legal phrases, which the CJEU repeats as self-standing statements of the law or in context with other prefabricated phrases. The formulas are not only rhetorical but simultaneously characterize the European Union legal order, establish its principles and fundamental concepts [1]. They speed up the process of judgment writing and make searching for relevant (legally similar) past cases more effective.[4]

With repetition the formulas detach from the judgments in which they were first pronounced (the original judgments) and acquire a broader relevance. They begin to function as abstract rules [11]. The modification of the content of the formulas reflects how the CJEU develops, elaborates, expands, or restricts legal concepts, and the reach of European Union law and how it adapts broad formulations to fit individual situations [11, 2]. This does not imply that the original judgment loses its legal relevance because it is not cited but rather that the legal relevance of the judgment becomes embedded, or implicitly acknowledged [10].

Among the best known examples is the so-called *Van Gend* formula, where the CJEU defined the Treaty as establishing *"a new legal order of international law"* and the formula in the *Grzelczyk* judgment[5], where the CJEU defined the concept of European Union citizenship, stating that *"Union citizenship is destined to be the fundamental status of nationals of the Member States."*. Both had far-reaching implications for the relationship between the European Union and the Member States and between the European Union and other international organizations.

2.3. Text Processing and Text Similarity

As already pointed out in the previous section the CJEU often paraphrases the wording of the original formula to express the same content. The new versions of the formula are thus not identical but similar to the original formula. To infer implicit citations (missing links) and local references we thus rely on text similarity.

We proceed by first applying a typical Natural Language Processing workflow which consists of the following steps: a) sentence and word segmentation, b) lowercasing, c) stemming, d) removal of stopwords, single letter words and numbers. For the last step we use the standard list of English stopwords.

Our purpose is to define a way to measure the formula similarity between any mutated paragraph a and the original paragraph b, and use this metric to detect the actual cited paragraphs in the case of global references. We therefore, use a special case of the Tversky index [15] (see Equation 1).

$$T(b,a) = \frac{|b \cap a|}{|b \cap a| + \alpha|b-a| + \beta|a-b|} \qquad \alpha,\beta \geq 0 \tag{1}$$

where $|\cdot|$ here denotes the number of words.

The *formula similarity index* between paragraphs b and a, $fsi(b,a)$, is merely the value of Tversky index we get by substituting $\alpha = 1$ and $\beta = 0$ and thus eliminating the influence of the mutant paragraph to the similarity score, which is desirable. Hence:

[4] Scholars have called these pre-fabricated phrases the *building blocks*, see e.g. [3]

[5] The corresponding ECLI numbers are ECLI:EU:C:1962:42 (Van Gend) ECLI:EU:C:2001:458 (Grzelzyk)

$$fsi(b,a) = \frac{|b \cap a|}{|b|} \tag{2}$$

Note that the above definition is not symmetric, i.e. $fsi(a,b) \neq fsi(b,a)$. An interesting property of the proposed similarity index is that it implies that the paragraphs are treated as *bags-of-words*, in the sense that the order of the words is not important. The latter property, together with stemming help us to partly overcome the effect of paraphrasing during link detection. In the context of inferring implicit links, a given paragraph a, refers to case B but not to a specific paragraph $b \in par(B)$. Hence, we will use *fsi* to infer which paragraph of B should have been the target of the implicit reference.

2.4. Dataset Construction

The dataset for this paper was first compiled by downloading from EUR-Lex[6], the texts of all judgments of the CJEU until the end of 2015. This yielded 10418 documents (judgments) in total. We then extracted all paragraphs in the *Grounds* section of the judgments[7]. We kept the English language versions of the judgments whenever available, and supplemented the dataset with the texts of the French language versions, yielding a total of around 445 000 paragraphs. Since the CJEU did not number judgment paragraphs systematically until the 1970s, we excluded all judgments that do not have numbered paragraphs from the extraction process.

Note that due to this technical issue, some older cases are left out in the present study. This includes some important landmark cases such as *Van Gend* and *Costa* for instance. However, we argue that this does not affect the way we assess the relevance of our methodology in Section 3, which is the core of the present study. This rather raises the question of completing the dataset which we let for a future work.

Subsequently, we employed the core extraction methodology in [10], i.e. used GATE and a set of JAPE rules [5], to infer citations to paragraphs and to build a paragraph-to-paragraph citation network out of the entire paragraph dataset, including both *global* and *local* references. The main difference from the core methodology proposed in [10] is that in order to annotate the case names in the text, where possible, we preprocessed the paragraphs before passing them to GATE, instead of using a gazetteer.

Preprocessing was a necessary step to identify citations in the text that refer to the case by the name that it is commonly known by, in CJEU. For instance, the CJEU very often refers to a judgment without using case numbers, like in the following: *"37. The Court stated in **paragraph 16 of Keck and Mithouard**, cited above, that national provisions [...] within the meaning of the line of case-law initiated by **Dassonville**, cited above"*[8]. Text fragments like the previous, can however be annotated with the CELEX number of the case, by using a white-list of case names. The annotation can then be further used to identify every single case decided by the CJEU and stored in EUR-Lex. The use of gazetteer reaches the same final result in the general case, makes things more complicated, however, in the presence of ambiguous citations, e.g. *"Commision v. France"*.

The key figures of both the paragraph-to-paragraph and the corresponding case-to-case networks are summarized in Table 2.

[6] http://eur-lex.europa.eu

[7] The judgments of the CJEU are divided in sections. The section Grounds contains the statements of the CJEU about the legal arguments and is thus the part of the judgments that is most relevant for legal scholars.

[8] Karner, Case C-71/02, ECLI:EU:C:2004:181, par. 37

Table 2. The case-to-case and the paragraph-to-paragraph network of the CJEU. The numbers in parenthesis indicate, $|V_{case}|$ for nodes and $|E_{global}|$ for edges, respectively.

	case-to-case	paragraph-to-paragraph
Nodes	10418	74219 (4773)
Edges	49519	93713 (18778)

2.5. Predicting Local Links

The paragraph-to-paragraph network and the global references open the possibility to complement the network by predicting the actual target paragraphs from global citations on the basis of citing paragraphs alone. While legal experts who are familiar with the judgments of the CJEU and the relevant formulas would see this as an intuitive step in the legal analysis, the task is less straightforward for a computer program.

We nonetheless designed the following simple algorithm to overcome this difficulty: For every edge $(p,C) \in E_{global}$, run through the set $par(C)$ and for every paragraph $p_c \in par(C)$, compute $fsi(p_c,p)$ from Equation 2. We compute a candidate target paragraph p_t, taking $p_t = \max\{fsi(p_c,p) \forall p_c \in par(C)\}$ and then check if $fsi(p_t,p) \geq t$, for a specified threshold t, in which case we add the edge (p,p_t), which means that we infer that p should cite p_t among all paragraphs of C. If $fsi(p_t,p) < t$, no paragraph is predicted.

Computing the score $fsi(p_c,p)$, as above, implies that we consider p_c as the (candidate) source of law that paragraph p is citing, and *fsi* represents the percentage of p_c repeated by p. Another implication of the above approach is that we predict at most one edge for every (p,C)-pair even though in principle, a paragraph could refer to more than one paragraph of a cited judgment.

The selection of an appropriate value for t is not straightforward. In our case we worked with different values of t and observed that selecting $t \geq 0.5$ would exclude several true positive citations, when the formula that was reproduced in the global citation was a rather small fraction of the original formula. In fact we calculated the average $fsi_{avg} = \{fsi(v,u)/|E_{local}|, \forall (u,v) \in |E_{local}|\}$ and the result was $fsi_{avg} = 0.48$. Therefore, we tested several values of $t < 0.5$ and we ended with $t = 0.4$, which we will use for the rest of the paper. We omit the full results for a longer version of the paper.

3. Results and Interpretation

3.1. Predicting Local Citations

In order to assess the prediction method we evaluated the quality of the predictions by examining the predicted links towards three *landmark cases*[9] of the case-law of the CJEU, *Dassonville, Defrenne II* and *Francovich*. The reason for selecting those cases is on the one hand their qualitative characteristics, in particular their perceived doctrinal difference and versatility, and then their quantitative characteristics, see Table 3. As we see from Table 3, the three selected cases are cited more on average and vary greatly in the number of paragraphs of them that get cited.

[9] see e.g. [13] for the discussion on landmark cases

Table 3. Summary statistics for the number of citations of the selected cases compared with the network all paragraphs. The number of cited paragraphs for the entire network is the average.

	Dassonville	Defrenne II	Francovich	Network
Median	1.00	3.00	2.00	1.00
Mean	33.00	3.53	4.00	2.12
Number of cited paragraphs	3	19	27	5.23

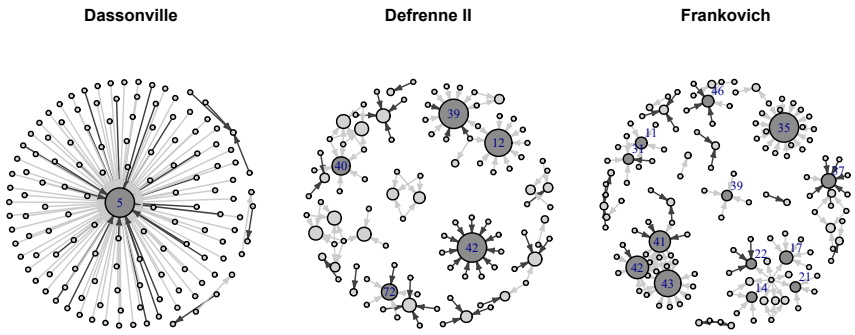

Figure 1. Examples or paragraph networks. Dark edges denote predicted links and numbered nodes correspond to highly cited paragraphs of each case.

Dassonville is prominent among legal scholars with regard to one legal aspect in a well defined area of EU law (free movement of goods). Defrenne II is typically considered by legal scholars in several areas of EU law, i.a. for its contribution to the general principle of non-discrimination on grounds of gender (the principle of equal pay for equal work), the limitation of temporal effects of judgments of the CJEU in exceptional circumstances, and horizontal direct effect of the Treaty as a fundamental characteristics of the EU legal order. Dassonville and Defrenne II are cases of creative judicial interpretation of the Treaty as the principal, written legal source of EU law. By contrast, Francovich is known as a judicial innovation. It establishes a new legal principle that does not originate from a written legal source of EU law and lays down the conditions under which it can be applied. This is reflected in the high number of different paragraphs that are cited in subsequent cases. The induced subgraphs of the paragraph networks of the above three cases are juxtaposed in Figure 1, where their differences in legal substance are very nicely represented by the fact that the Dassonville subnetwork consists almost entirely of one star, with incoming citations only to par. 5, whereas in Defrenne II and Francovich the citations to several legal aspects produce a number of smaller clusters.

This implies first, that Dassonville contains one formula, which is most often repeated in subsequent cases as a whole. By contrast, Defrenne II and Francovich contain more than one formulas. Second, since the Dassonville formula has a single and distinct meaning related to a particular legal problem, it is consistently repeated in one particular legal and several, factually similar contexts. Defrenne II and Francovich have a broader legal relevance because they concern the basic principles of the EU legal order that are applicable across subject areas. They can thus be repeated in more than one legal context and to factually distinct situations.

Altogether, we predicted 97 citations to all three cases while 33 citations remained unmatched. A legal expert validated the approach by reviewing the list of predicted citations and determining whether the citations were accurate. As *accurate citations*, we

Table 4. Method evaluation

Case	Recall	Precision	F1	Baseline Precision
Dassonville	87.1%	90.0%	88.5%	85.2%
Defrenne II	49.0%	66.7%	56.5%	35.6%
Francovich	85.7%	77.4%	81.4%	0.0%
TOTAL	69.4%	77.3%	73.2%	40.5%

considered citations that either matched the cited paragraph on legal language level (*text match*), meaning that they repeated the words of the formula, and on legal content level, meaning that they were predicted in a meaningful legal context and hence legally relevant (*content match*), or both (*full match*). Table 4 presents the results of this evaluation against a baseline approach, and with regard to *Precision*, *Recall* and *F1* measures, see [9]. Table 4 shows the performance both per case and in total (last row).

The baseline for our comparisons was to assign the local link to the most cited paragraph of the case. This method yielded a total precision of 40.5%. By way of comparison, the formula similarity index method was almost two times better with 77.3% precision with a satisfactory recall of 69.4% for all three cases.

3.2. Various Types of Citations

The least satisfactory results concern cases, in which the formulas are – in part or whole – repeated several times within a single document, as in Francovich, pars. 28, 37 and 46. For instance, a formulation that the Member States are "*obliged to make good loss and damage suffered by individuals as a result of the failure to transpose...*", appears in all three paragraphs, however, for very different reasons: in par. 28 as a reproduction of the question of the national court ("*28. In the second part of the first question the national court seeks to determine whether a Member State is obliged to make good loss and damage suffered by individuals as a result of the failure to transpose Directive 80/987*"), in par. 37 as a genuine statement of the law by the CJEU ("*37. It follows from all the foregoing that it is a principle of Community law that the Member States are obliged to make good loss and damage caused to individuals by breaches of Community law for which they can be held responsible.*") and in par. 46 as the reply of the CJEU to the national court ("*46 The answer to be given to the national court must therefore be that a Member State is required to make good loss and damage caused to individuals by failure to transpose Directive 80/987.*"). While the reference to par. 46 is recalled with a 100% precision, the reference to par. 28 is recalled with 0% precision and the reference to par. 37 is recalled with 87% precision[10]. In cases of high inter-textual similarity such as in our example the first occurrence would most likely be a reformulation of a question, while the last occurrence would most likely be an answer. The central – both legally and with regard to the position in the text – would be the middle occurrence.

Generally speaking, our predictions failed mostly with regard to linguistically too indistinct formulations and longer formulations, which repeated the arguments of the parties, or the questions of the national courts that also repeated the formulas taken from past cases, often tying them to the particular facts of the case, or in combination with either national or EU secondary legislation. The confusion arose because the CJEU refers to these arguments or preliminary questions in the Grounds of the judgment, to indicate

[10]Due to space constraints we have omitted the detailed results from this paper.

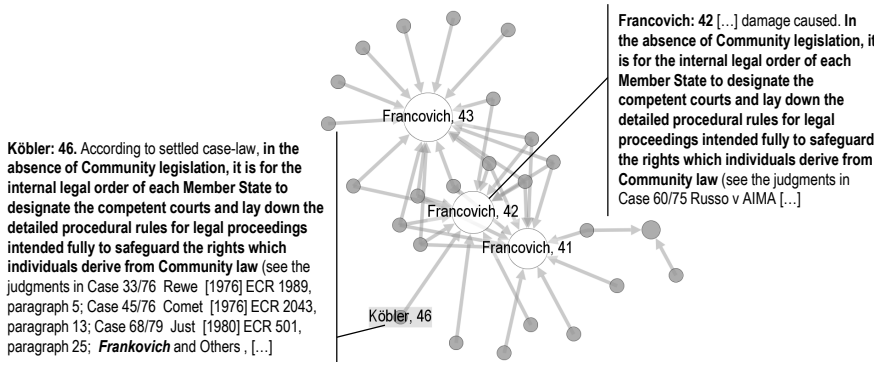

Figure 2. A predicted link from Köbler, 46 to Francovich, 42. The repeated formula is shown with bold letters.

which one of the several legal issues it is dealing with. The latter is especially common in longer and legally more complex judgments, where the national court formulates several preliminary questions.

3.3. Revealing How a Case Is Used

The most interesting findings concern individual predictions. For instance, a legally novel development of the so-called *Francovich principle* occurred first in *Brasserie du Pecheur*, and later in *Köbler*. While the workflow outlined in Section 2.4 does not detect a citation from *Köbler*, par. 46, to *Francovich* par. 42, due to a typo, a misspelling of "Francovich", as shown in Figure 2, the citation is detected on the basis of text similarity. This is not an isolated occurrence hence it can be argued that the text similarity techniques are indispensable for obtaining a more accurate picture of case citations and case centrality.

The final example demonstrates the contribution of our approach to the study of legal development, in particular legal change. Namely, the approach, which the CJEU created with *Dassonville* with regard to the national measures restricting trade, was importantly narrowed down in *Keck*. The central paragraph, in which this occurs, is *Keck*, par. 16, which refers to *Dassonville*, par. 5. Our method successfully detects this reference by link prediction on the basis of text similarity. The analysis on document level, without the use of text similarity would not detect this reference, which is crucial for legal scholars.

4. Conclusions

In this article we constructed a network of individual text units or paragraphs of the judgments of the CJEU and used text similarity techniques to obtain a complete information about the content of case citations. This level of granularity enabled us to draw a more complete picture of implicit citations, meaning the repetitions of legal arguments stemming from past judgments without explicit references to those judgments. The implicit citations provided the missing data about the actual use of a case by the CJEU. On the basis of the precise information about the content of citations we were furthermore able to differentiate between various types of citations. Together, this information is important to empirically determine to what extent a specific case has influenced the law, and thus giving every case its doctrinal due.

Our findings show first, that algorithms can correctly predict the local links (citations to specific paragraphs), in cases where the formulations are limited to a one-sentence linguistically characteristic original sequence. By contrast, the results were not as convincing in the case of very short or very long linguistically indistinct formulations with broad legal application (for instance, a reference to "*legal certainty*").

Second, the findings reveal that it is possible to tease out legal development by a more detailed categorization of predicted links. A content match would often indicate a mutation of formulas, which is often an indication of legal change or important legal innovation (this was the example of Keck, par. 16, citing Dassonville, par. 5, only to reverse the course of the law established by Dassonville).

Our study is, to the best of our knowledge, the first to demonstrate that by combining the citation network approach with text similarity detection techniques, we can access the legal content behind citations. Thereby, our research opens avenues for original research, which by further improving and fine tuning the basic approach, can detect the doctrinal origin of legal formulas, their modifications in the judgments of the CJEU over time, as well as, their generalization across different areas of law. Finally, we believe that the findings of this paper will allow us to develop machine learning approaches to the problem of detecting legal formulas, in a spirit similar to recent developments, e.g. [12].

References

[1] Loïc Azoulai. La fabrication de la jurisprudence communautaire. *Dans la Fabrique du Droit Européen, Brussels: Bruylant*, pages 153–170, 2009.

[2] Loïc Azoulai. The retained powers' formula in the case law of the european court of justice: EU Law as Total Law. *Eur. J. Legal Stud.*, 4:178, 2011.

[3] Gunnar Beck. *The Legal Reasoning of the Court of Justice of the EU*. Bloomsbury Publishing, 2013.

[4] Ryan C Black and James F Spriggs. The citation and depreciation of US Supreme Court precedent. *Journal of Empirical Legal Studies*, 10(2):325–358, 2013.

[5] Hamish Cunningham. GATE, a general architecture for text engineering. *Comput Humanities*, 36(2):223–254, 2002.

[6] James H Fowler and Sangick Jeon. The authority of Supreme Court precedent. *Soc. networks*, 30(1):16–30, 2008.

[7] Johan Lindholm and Mattias Derlén. The court of justice and the Ankara agreement: Exploring the empirical approach. *Europarättslig tidskrift*, (3):462–481, 2012.

[8] Yonatan Lupu and James H Fowler. Strategic citations to precedent on the us supreme court. *The Journal of Legal Studies*, 42(1):151–186, 2013.

[9] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.

[10] Yannis Panagis and Urška Šadl. The force of EU case law: A multi-dimensional study of case citations. In *JURIX*, pages 71–80, 2015.

[11] Urška Šadl. Case–Case-Law–Law: Ruiz Zambrano as an illustration of how the court of justice of the European Union constructs its legal arguments. *Eur Const Law Rev*, 9(2):205–229, 2013.

[12] Olga Shulayeva, Advaith Siddharthan, and Adam Wyner. Recognizing cited facts and principles in legal judgements. *Artificial Intelligence and Law*, 25(1):107–126, Mar 2017.

[13] Fabien Tarissan, Yannis Panagis, and Urška Šadl. Selecting the cases that defined Europe: complementary metrics for a network analysis. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2016)*, pages 661–668. IEEE, 2016.

[14] Aleksandar Trklja. A corpus investigation of formulaicity and hybridity in legal language: a case of EU case law texts: A case of EU case law texts. In S.G. Roszkowski and G. Pontrandolfo, editors, *Phraseology in Legal and Institutional Settings: A Corpus-based Interdisciplinary Perspective*. Routledge, 2017.

[15] Amos Tversky. Features of similarity. *Psychological review*, 84(4):327, 1977.