Identification of Case Content with Quantitative Network Analysis:
an Example from the ECtHR

Martin Lolle CHRISTENSEN a,1, Henrik Palmer OLSEN a and Fabien TARISSAN b
a iCourts – Centre of Excellence for International Courts, University of Copenhagen
b ISP – ENS Cachan & CNRS, University Paris-Saclay, France

Abstract. What is a case decided by the European Court of Human Rights about? The Courts own case database, HUDOC, lists all the articles mentioned in a specific case in their metadata. They also supply a number of keywords, but these keywords for the most part are reduced to repeating phrases from the relevant articles. In order to enhance information retrieval about case content, without relying on manual labor and subjective judgment, we propose in this paper a quantitative method that gives a better indication of case content in terms of which articles a given case is more closely associated with. To do so, we rely on the network structure induced by existing case-to-case and case-to-article citations and propose two computational approaches (referred to as MAININ and MAINOUT) which result in assigning one representative article to each case. We validate the approach by selecting a sample of important cases and comparing manual investigation of real content of those cases with the MAININ and MAINOUT articles. Results show that MAININ in particular is able to infer correctly the real content in most of the cases.

Keywords. Network science, Legal studies, European Court of Human Rights, Complex Networks, Directed acyclic graphs, Bipartite graphs

1. Introduction

As many real-world networks, corpus of legal decisions lend themselves to the use of graphs to analyze their structure and identify pertinent properties. In such a context, the nodes of the network usually stand for judgments and a link between two judgments A and B exists if A cites B to formally ground its decision or in some other way support the legal reasoning leading to its decision. This approach has been used for decades in many judicial contexts, ranging from judgments of the American Supreme Court [10,4,5] to European jurisdiction [1,14,3,2,8,12]. Those studies showed that network science is useful in particular to identify key important judgments in a corpus. However little has been done to exploit the network structure induced by the citations in order to automatically infer the content of a judgment. This is what this paper intends to do by proposing a new method that takes advantage of the network structure to identify the main subject of a judgment.

1 Corresponding Author
More precisely, we propose to rely on two elements of the network structure: 1) the citations towards former judgments and 2) references to articles. This is driven by the assumption that, given a case C, its content can be revealed through, not only what articles C cites, but also through what articles are cited by the cases that cite or are cited by C. Relying on this hypothesis, we derived two computations (referred to as MAININ and MAINOUT) aiming at automatically identifying the main legal subject matter of cases.

In order to evaluate this approach, we applied the proposed techniques to a subset of judgments of the European Court of the Human Rights (ECtHR) and qualitatively validated the results by reading through a selection of cases in the network, chosen among the most cited ones. Results show that 1) although the MAINOUT approach does not correctly identify the most dominant content in each and every case in the network, it overall seems convincing and 2) that the MAININ approach succeeds in most of the cases to infer the real content of a judgment.

The remaining of the paper is organized as follow: we first give a description of the corpus of judgments studied in this paper and the way the information is represented as a network (Section 2). Then we present the main results (Section 3) before discussing the possible improvements as well as the limitations of the proposed techniques (Section 4).

2. Background

In this section, we introduce the required background for the remainder of the paper. First, we present the jurisdiction under study before defining the formalism used for the network analysis and providing the experimental setting we used to conduct the study.

2.1. The European Court of the Human Rights

The European Court of Human Rights is one of the most active international court in the world. It has handed down over 18,500 judgments since it was inaugurated in 1959. It is established as part of the European Convention on Human Rights that was enacted a few years after the end of second world war. Although almost inactive for the first years of its existence, the Court today – after some changes made to the Convention that established a direct access to the Court for individuals who think that their human rights have been violated by a member state – deals with thousands of cases every year. Its jurisdiction spans from Iceland and Finland in the north, to Spain and Greece in the south and from the United Kingdom in the west to Russia and Turkey in the east. Many of the Court’s cases involve controversial and sensitive political questions, including issues such as the rights of sexual minorities, freedom to exercise religious practices, immigration issues (through the right to family life and respect for privacy), etc.

In this paper we have selected a sample of cases for the purposes of assessing the relevance of our method. We have selected all cases that are listed in HUDOC (the Court’s own database) as cases that cite Art. 9 (the right to freedom of religion). Two reasons motivate this choice. First, Art. 9 cases often involve some other rights issue. A typical example is that of a case where a group or a person has made derogatory expressions about a religious figure. These cases are most often decided as freedom of expression cases (Art. 10), but often also cite Art. 9 because of the protection that is offered through
this article to religious groups. Without having any prior information about the content of the case, it will not be possible to know whether the reasoning in the judgment concerns Art. 9 or 10, and we claim that our method can help to solve this issue. The second reason is related to the validation perspective: since the complete set of cases that cite article 9 is relatively small (compared for instance to the number of cases that cite article 6), it makes our qualitative test more representative of the overall results, regarding the cases related to Art. 9.

2.2. The network science approach

2.2.1. Definitions

As pointed out in the introduction, it is common to represent the network of citations among legal documents as a graph. Usually one considers directed graphs where the nodes stand for the judgments – identified by their HUDOC number when the network consists of cases from the ECtHR – and a link between nodes \( u \) and \( v \) exists if the decision \( u \) cites the decision \( v \). This representation turned out to be useful to compute standard properties of the network and assess the relative importance of the nodes. In this context, scholars usually assume that the more important a decision is, the more it will attract citations and/or that important decisions will be those decisions in the network that have the highest hub scores \([7,5]\). Thus, one often considers the in-degree of a node (number of inward links) or it’s hub score as a good proxy for the importance of a judgments. But this formalism does not account for indirect relations that judgments might have in terms of content. Such a property is better captured if one looks at which articles a judgment refers to. This remark has led the community to use bipartite networks instead, in which the set of nodes are partitioned into two distinct sets – here the set of judgments and the set of articles – and the links rely the judgments to the articles they mention.

The two structures mentioned above (directed and bipartite graphs) are well established frameworks in the literature and have already proved to be useful in the context of analyzing a corpus of judgments (see \([10,4,5,1,14,3,2,8,11]\) for instance). However, they have almost always been studied separately. In this paper, we propose to exploit the information contained in the citation network by studying the two structures at the same time. As such, we follow here a previous work proposed in \([11]\) in the context of the International Criminal Court.

Formally, one defines a hybrid bipartite graph as a 4-upplet \( H = (V_J, V_A, E_J, E_A) \) where \( V_J \) stands for the set of judgments, \( V_A \) the set of articles, \( E_J \subseteq V_J \times V_J \) the set of
citations a judgment makes towards former judgments (directed graph) and \( E_A \subseteq V_A \times V_A \) the set of references a judgment makes to articles (bipartite graphs). An example of such structure is depicted in Fig 1.

2.2.2. Inferring the content of a case

Once the information has been represented as a network, one can turn to different statistics to describe the general structure of the network as well as the specific position of certain nodes. For instance, it is quite natural in our context to assess the importance of a judgment by relying on the in-degree of a node \( v \in V_J \), that is the number of other judgments that cite \( v \). Formally, one defines first the set of (in-)neighbours \( N_J(v) \) by:

\[
N_J(v) = \{ u \in V_J \mid (u,v) \in E_J \}
\]  

Then one defines the in-degree of a node \( v \in V_J \) as \( d_J(v) = |N_J(v)| \). Similarly, one can also estimate the importance of an article by computing the number of times it is referred to in the network. Formally, one computes for every article \( a \in A \) the value \( d_A(a) = |N_A(a)| \) where \( N_A(a) \) is defined as:

\[
N_A(a) = \{ v \in V_J \mid (v,a) \in E_A \}
\]  

The two properties above are an application of standard metrics defined for graphs (either directed or bipartite). But having the hybrid bipartite network in hand, one can exploit the structure to elaborate more intricate properties revealing the nature of a node. In particular, since we are interested in inferring the content of a judgment, we propose to rely on the articles referred to by the neighbourhood of the nodes. Formally, given a judgment \( v \in V_J \), let us define the set of in-articles related to \( v \) by:

\[
\text{Art}_{in}(v) = \{ a \in A \mid \exists u \in N_J(v), (u,a) \in E_A \}
\]  

This set contains thus all the articles referred to by judgments citing \( v \). For instance, going back to the example of Fig. 1, although the judgment \( J_1 \) only refer to Art. 1, its set of in-articles are Art.1, 2, 3 and 4 because the judgments that cite \( J_1 \) refer to those articles. Doing so, one can compute the number of times in-articles are referred to by computing the following value:

\[
\text{Freq}_{in}(v,a) = |\{ u \in N_J(v) \mid (u,a) \in E_A \}|
\]  

In our example \( \text{Freq}_{in}(J_1, \text{Art.2}) = 2 \) because \( J_2 \), and \( J_3 \) both refer to Art.2. We claim now that the content of a judgment \( J \) will be well approximated by the content of the articles that have a high value. In order to simplify the analysis, we propose in this paper to study only the main article, defined as the article of the set \( \text{Art}_{in}(v) \) which has the maximum number of citations. Formally, one defines the value \( \text{Main}_{in}(v) \) as follow:

\[
\text{Main}_{in}(v) = \arg \max_{a \in \text{Art}_{in}(v)} \text{Freq}_{in}(v,a)
\]  

Finally, although the computation proposed above might seem natural, it is well know that it would give a strong bias towards what one might call popular articles, that is
articles often referred to but, as such, not very indicative of the real topic. This is why we propose the following alternative which gives less weight to popular articles that have been cited in abnormal amounts:

$$\text{Main}_{in}^w(v) = \arg\max_{a \in \text{Art}_in(v)} \frac{\text{Freq}_{in}(v,a)}{\log(d_A(a))}$$ (6)

The article computed by this quantity is referred in the following as the main in-article (or MAININ in short) of judgment \(v\). In our former example, one can check for instance that the main in-article of \(J_1\) is \(\text{Main}_{in}^w(J_1) = \text{Art. 2}\).

Obviously, we can define the dual notion of main out-article (MAINOUT) by computing a similar value but based on judgments that are cited by \(v\) instead of judgments that cite \(v\). In our former example, the MAINOUT of \(J_6\) would be Art. 3.

Note that the two notions defined above are complementary. Given a judgment \(J\), MAINOUT looks at what the judgments cited by \(J\) are referring to – thus inferring what is the main article invoked by the judgments on which \(J\) relies – while MAININ looks at what the judgments citing \(J\) refer to – thus indicating how \(J\) is perceived by future judgments. Thus, those elements are two sides revealing the real content of the judgment \(J\).

### 2.3. Experimental setting and validation

As indicated before, for the purpose of assessing the pertinence of the proposed approach we selected only a sample of the total cases available in HUDOC (the Courts database) and chose to focus on Art. 9 cases. To do so, we first extracted all cases that cite at least once Art. 9, which led to a set of 148 judgments. Then we expanded the network by also considering all cases that cite or are cited by at least one judgment in the former set. This step expanded the network to a total of 730 decisions. From these 730 decisions, we extracted the references to all articles in order to generate our hybrid bipartite graph as defined in Section 2.2.1. Once the hybrid bipartite graph has been defined, we turned to the computation of the main out-article (MAINOUT) and main in-article (MAININ) for every node, as explained in Section 2.2.2.

In order to test the relevance of the proposed approach to identify the content of a case, we selected 9 cases among the cases with the highest in-degree and 1 important case with a low degree. Then we checked whether the MAININ and/or MAINOUT articles correctly identify the main content of the case. We did this by manually reading the selected cases and identifying what we (and other authors of legal textbooks on the European Human Rights Convention) see as the main legal issue(s) dealt with in the case.

### 3. Results

Although we applied the method described above on the 730 judgments selected as explained in Section 2.3, we show the concrete results on a small sample of those judgments, selected among the most important cases referring to article 9. The selected cases...
Judgments | Date | MAINOUT | MAININ | In-degree  
--- | --- | --- | --- | ---  
5095/71 – Case of Kjeldsen, Busk Madsen and Pedersen v. Denmark | 07-12-1976 | Article 10 | Protocol 1 - Article 2 | 40  
14307/88 – Case of Kokkinakis v. Greece | 25-05-1993 | Article 10 | Article 9 | 97  
17419/90 – Case of Wingrove v. The United Kingdom | 25-11-1996 | Article 10 | Article 10 | 56  
24645/94 – Case of Buscarini and Others v. San Marino | 18-02-1999 | Article 9+14 | Article 9 | 35  
34369/97 – Case of Thlimmenos v. Greece | 06-04-2000 | Article 6-1 | Article 14 | 69  
30985/96 – Case of Hasan and Chaush v. Bulgaria | 26-10-2000 | Article 9 | Article 9 | 75  
41340/98 – Case of Refah Partisi (The Welfare Party) and Others v. Turkey | 31-07-2001 | Article 10 | Article 11 | 57  
45701/99 – Case of Metropolitan Church of Bessarabia and Others v. Moldova | 13-12-2001 | Article 9 | Article 9 | 49  
44774/98 – Case of Leyla Şahin v. Turkey | 29-06-2004 | Article 9 | Article 9 | 56  
18136/02 – Case of Siebenhaar v. Germany | 03-02-2011 | Article 10 | Article 11 | 2  
Correct inference | | | | 50 %  
| | | | 90 %  

Table 1. A selection of important cases in the network and their corresponding MAINOUT and MAININ articles. Articles highlighted in bold text are articles that are correctly inferred by the method. The last row gives the ratio of good inference over all considered cases.

With their MAININ and MAINOUT article are listed in Table 1. In the following, we detail our findings after investigating manually (through the reading of the judgments as well as textbooks) what is the content of the cases and comparing it to the MAINOUT (Section 3.1) and MAININ (Section 3.2) articles inferred by the proposed approach.

### 3.1. MAINOUT and case identity of article 9 cases

In this section, we investigate to which extent the MAINOUT article of a case could be indicative of its subject matter. Looking at Table 1, one immediately sees that, for a number of most prominent article 9 judgments\(^4\), this is the case. For example *Hasan and Chaush v. Bulgaria, Metropolitan Church of Bessarabia and Others v. Moldova* and *Leyla Şahin v. Turkey* all have Art. 9 as their MAINOUT, correlating with their subject matter (Interference in administration of religious communities, Recognition of a Church and Right to wear religious headscarfs respectively\(^5\)).

The MAINOUT identifier is however not always accurate. Several cases have a MAINOUT which is different from how textbook authors see the case, *Kokkinakis v. Greece*, for example, which is the most cited article 9 judgment has Art. 10 (Freedom of

---

\(^4\)By prominent cases we here mean cases which have a high in-degree in the network, i.e. cases which have been cited a lot by other cases, and which also appear in leading textbooks as important cases for Art. 9.  
\(^5\)Our identification of subject matter relies on what textbook authors say about these cases. We have used [6]
expression) as its MAINOUT. Manual investigation revealed that it is due to the fact that Kokkinakis cites five cases, one mostly about Art. 9, one mostly about Art. 6 and three that are mostly about Art. 10. For this reason the MAINOUT of Kokkinakis is Art. 10 – even though the court in its deliberation (i.e. in its legal reasoning) seem to focus almost entirely on article 9.

The earliest case in the set of cases citing article 9 is the Danish case Kjeldsen, Busk Madsen and Pedersen, concerning an objection to compulsory sex education in the Danish school system. There was a religious motivation behind parts of the objection and article 9 in conjunction with Protocol 1, article 2 was invoked. While the primary subject of the case concerns P1-2 and the Right to Education (which includes a duty of the state to ensure such education and teaching in conformity with the religious and philosophical convictions of parents), the case has been cited by a number of article 9 judgments, including Kokkinakis. Similar to Kokkinakis, Kjeldsen has article 10 as its MAINOUT, although the case does not focus on freedom of speech. This is however clearly explained by the fact that it is an early precedent on the right of education. As such, it has obviously no prior cases on P1-2 to cite and the MAINOUT article is therefore not suited to infer the content of the case. It is worth noting however that the subject matter of the case is well inferred by the MAIN property which succeeds in identifying P1-2 as the real content of the case.

Finally, we examined Thlimmenos v. Greece which is also a highly cited case. It is interesting as it deals, not only with article 9, but also Art. 14. The case is about a positive obligation for member states to differentiate between conscientious objectors to compulsory military service and other previous felons. Thlimmenos however has Art. 6 as its MAINOUT. The reason is that it cites three Art. 6 cases in a relatively short section on the procedural issues involved in the case (the court cites cases in support of the standard for reasonable length that they use in the case). The case then once again shows that MAINOUT is more indicative of what citation resources in terms of prior cases were available at the time than it shows the identity of the case in terms of the main substantive legal issues decided in the case.

Overall the results on MAINOUT are the reflect that a judgment can only cite existing prior judgments. This means that the possibilities for citing is limited. When the Court has to decide a case on an issue which it has not been decided before, there is no prior cases to cite. Still, the court do cite prior cases on issues where it is possible and this explains why MAINOUT is not always indicative of the main legal content of the case.

### 3.2. MAININ and case identity of article 9 cases

Can MAININ show anything about the judgment which MAINOUT does not? As indicated and explained above there are some important article 9 cases that are not indicated as such by their MAINOUT property. We now investigate whether MAININ better identifies the content of the case. MAININ, as we explained in Section 2.2.2, is a measure that indicates the most cited article, by those cases that cite a given case. The working hypothesis behind this is that the content of a case C can be revealed by looking at the identity of the cases that cite C. Rather than focusing on who the C cites, we now focus on who cites C. The advantage of this is that we can rely on a broader perception of the case as a way to identify what its most important content is.

The MAININ article shows a more condensed selection of the cases that form the center of the article 9 case law. Investigating all the 148 judgments that cite at least one
Art. 9, 37 of them have Art. 9 as their MAIN and 8 of them have Art. 14+9 as their MAIN, i.e. a total of 45 cases have some variation of article 9 as their MAIN. All the most important Art. 9 cases are captured by this measure. The complete distribution of the MAIN article over all the 148 cases is depicted in Figure 2 below.

Going back to Table 1, one can see that MAIN gives positive results in the article 9 case law, and help filter out cases that may have cited article 9 but that do not deal substantively with the right. Refah Partisi (The Welfare Party) and Others v. Turkey, for instance, is an important case on the subject of Freedom of Association, dealing with the dissolution and ban on the Turkish Welfare Party. The political party was banned because it infringed on the Turkish separation of religion and state, and there was therefore also an alleged violation of article 9. This was not, however, deemed necessary to examine by the Court. The MAIN of the case, article 11, is then completely in line with the content of the decision.

Another example is Wingrove v. The United Kingdom which dealt with the subject of blasphemy in regards to freedom of speech. The case concerns an artistic short film on the life of St. Teresa, dealing with gratuitously erotic scenes. The case was decided under Article 10 as a Freedom of Speech case, and article 9 was not invoked. The MAIN of the case, article 10, again totally supports this.

We also looked again at the cases examined in the MAINOUT section, to see whether their MAIN articles were qualitatively more indicative of their content. Kokkinakis, which has article 10 as MAINOUT, has article 9 as its MAIN, which is completely in line with its role. It is the most cited of the article 9 cases and it was the first case to find a state in violation of article 9. Similarly, Thlimmenos has article 14 as its MAIN which is also fitting its content, considering the much wider network of article 14 in which Thlimmenos has a prominent role (there are for instance 637 cases citing art. 14 [9]). While Thlimmenos sets a precedent with article 9 cases, it set a much larger precedent within all article 14 cases (concerning the obligation to treat differently persons whose situations are significantly different, see [6] p.790).

All in all, MAIN turned out to be particularly useful to detect the real content of a case (assimilated in this study as one main article). However, it must be noted here that, by definition (see Section 2.2.2), the computation of MAIN requires for a case to be
cited at least by another judgment, which is not always the case. Thus it makes it difficult for the techniques to be applied on recent cases, like in Siebenhaar for instance where the approach fails in finding the main relevant article.

4. Discussion

The main challenge for computational driven legal analysis is of course to translate quantitative thin data to meaningful qualitative information. Obviously case or article citation patterns do not necessarily reveal anything substantial about the legal reasoning in a case. An example can illustrate this.

Refah Partisi shows a potential issue with the MAINOUT article. The MAINOUT of the Refah Partisi case is article 10. When reading the case however, one immediately sees that the arguments in the case focus on the right to freedom of association (article 11 – the case is about a compulsory resolution of a political party). The case is also cited mostly by cases who mostly cite Art. 11 and it therefore has article 11 as its MAININ. We have shown above how some cases, such as early precedents, will have a MAINOUT that differs from the subject matter of the case, because first cases (i.e. first on some issue) do not have any prior case they can cite on the matter they are dealing with. This however is not the reason why Refah Partisi has article 10 as its MAINOUT. Refah Partisi cites several cases and these cases cite both article 9, 10 and 11. The reason is that the case very principled since it concerned a political party with a strong public support and a religious political agenda. The Court therefore naturally touched upon and made references to cases within all three fields of rights protection.

Although Refah Partisi is perhaps special in having such a close overlapping of different right provisions which are all relevant to the case, the Court generally pursues a judicial policy of trying to assure some overall coherence in its case law. This means that it will often draw argumentative support in area of its case law from some other area of its case law than what is immediately relevant.

We believe that part of the issues raised above are due to the simplification we made in this first step towards exploiting citations to identify case content, as we reduced the content to a unique article (MAININ or MAINOUT). We think that by extending the methodology to identify a profile which would cover a set of main articles instead of a single one, it would provide a better and more comprehensive picture of the real content of a case\(^6\). We claim that the evidence provided in this study show that the main principles proposed in this approach are worth being pursued.

5. Conclusion

Legal precedent plays an important role in almost all legal systems. A precedent is a former judgment that is cited as a legal source, and hence as part of the legal grounds (justification) for the decision in a present case. Lawyers who either argue or decide cases, therefore have an interest in identifying legal precedent. In this paper we aimed at devising a new method for computationally identifying relevant precedents for lawyers who are involved in litigating European human rights cases. By relying on existing case-

---

\(^6\)it could be extended to a complete list of related articles provided with a weight indicating each importance.
to-case and case-to-article citation networks, we have proposed a computation approach that results in the assignment of one convention article to each case. We have exploited two different information (MAINOUT and MAININ) and we have found that we get the best overall results by using MAININ. The MAININ of a given case A shows what the most cited article is, in those cases that cite A as precedent. We tested the results by selecting a number of cases from the case law of the European Court of Human Rights. By manually reading through a number of complex cases with a high in-degree (i.e. cases that has been cited many times by subsequent cases) we found that the article number assigned to a case through the MAININ computation corresponds well with our qualitative findings of legal content in most of the cases (90%). As a result of our research, we can now combine existing methods and findings which already makes it possible to assign both in-degree and hub score to cases in a given network, with MAININ, thereby computationally generating lists of what are the most important precedents in relation to a given article in the courts total network. In a follow up study we intend to investigate how best to implement this.

Another interesting perspective would be to combine the approach proposed in the present paper to the techniques developed in [13] that propose an alternative way to handle the problem of recent cases that have not attracted a lot of citations yet.

References


