Legislation as a complex network: Modelling and analysis of European Union legal sources

Marios KONIARIS, Ioannis ANAGNOSTOPOULOS and Yannis V ASSILIOU

*KDBSL Lab, School of ECE, Nat. Tech. Univ. of Athens, Greece

Dep of CS & Biomedical Inf., Univ. of Thessaly, Greece

Abstract. Legislators, designers of legal information systems, as well as citizens face often problems due to the interdependence of the laws and the growing number of references needed to interpret them. Quantifying this complexity is not an easy task. In this paper, we introduce the “Legislation Network” as a novel approach to address related problems. We have collected an extensive data set of a more than 60-year old legislation corpus, as published in the Official Journal of the European Union, and we further analysed it as a complex network, thus gaining insight into its topological structure. Results are quite promising, showing that our approach can lead towards an enhanced explanation in respect to the structure and evolution of legislation properties.

Keywords. Legal Digital Libraries, Network Analysis, Complex networks, Legislation, Knowledge extraction

Introduction

Legislation is a large collection of different normative documents, which keeps growing and changing with time. As legislation increases in size and complexity, finding a relevant norm may be a challenging task even for experts.

Furthermore, the process of drawing up a consistent and coherent legislation framework becomes a more and more challenging task. Drafting of new and amending existing legislation are very complicated processes. As a result authorities at European, national and local level, often consider proposed regulations for months or years before they finally become effective. Thus, it is critical to firstly quantify the legal complexity and then work towards the provision of a model that will assist us to reveal the emergent dependencies among the legislation corpus.

Typically, legal documents refer to authoritative documents and sources (e.g. most commonly regulations, treaties, court decisions, and statutes). Computer scientists and legal experts have used citation analysis methods, in order to construct case law citation networks, as well as to further model and quantify the complexity of the legislation corpus [5,13,11,17].

1This is an author’s copy version. The final publication is available at http://ebooks.iospress.nl/volumearticle/38450
However, the relations between legal documents on the studied networks are simple references. Thus the hierarchical structure of the normative system is vaguely underestimated and absent from the adopted model. In the present analysis, we propose an alternative approach to model and quantify legal complexity. Our model can be applied to civil law collections, such as the laws of the European Union.

Our approach differs from previous works handling the specific problem, as we do not define the legislation corpus in terms of a citation network. Instead we employ a multi relationship model in which two or more legal documents, belonging to the same or different types, may be linked to others by more than one relationships. Unlike previous studies of legal citation networks, our model provides a more realistic view of legislation. It encompasses many aspects such as hierarchy between the sources of law and the different types of relations between legal documents. Our modelling approach transforms legislation corpus into a multi-relational network: a network with a heterogeneous set of edge labels that can represent relationships of various categories in a single data structure.

We investigate the topological structure of the Legislation Network to discover properties and behaviours that transcend the abstraction. The results are quite promising, showing that the Legislation Network is a power law, small-world network. This can be reflected as an evolutionary advantage since power law, small-world networks are more robust to disturbance than other network architectures [12].

The rest of this paper is organized as follows. Section 1 briefly reviews related work and approaches. In Section 2 we introduce the examined datasets and our network construction model. In Section 3 we utilize our model to analyse the structure of the legislation corpus. Finally, Section 4 outlines the conclusions as well as future work issues.

1. Related work

Citation analysis has been used in the field of law to construct case law citation networks. The American legal system has been the one that has undergone the widest series of studies in this direction. Fowler et al. [14] at first experimented with methods to identify the most central decisions of the US Supreme Court and afterwards [5] they studied how the norm of stare decisis had changed over time in the jurisprudence of the US Supreme Court in order to identify the doctrine’s most important related precedents.

In contrast van Opijnen [16] concluded that network algorithms, which have been used in previous research, especially in-degree, HITS and PageRank [6], might not be the most appropriate to measure legal authority. The same researcher proposed a Model for Automated Rating of Case law which incorporates data from the publication and the citation of legal cases to estimate the legal importance of judgments [15].

Smith observed that the network of US Supreme Court decisions followed a power-law distribution [11]. The authors of [17] described a visualization-based interactive legal research tool that allows users to easily navigate in the legal semantic citation networks and study how citations are interrelated.

However, these studies focus on Common law: a law developed by judges through decisions of courts, which is fundamentally different with the Civil law that is used across

---

2 A legal norm inherited from English common law that encourages judges to follow precedent by letting the past decision stand.
the European Union. For quantifying the complexity of the legislation corpus through network analysis in the Civil law domain, Winkels et al have used a sample of 15,053 cases taken from the Dutch Supreme Court [4]. The authors verified that Fowler results also hold for the citation network of the sampled Dutch legal system. Similarly, the complexity of the French legal code was analysed in [13]. In this work the authors identified structural properties of the French legal code network, with a sample of fifty two legal codes.

In all of the above studies, the law graph is treated as a citation network, thus showing the effectiveness of network analysis in the legal domain. In one hand it was proven that case law citation networks contain valuable information, capable of measuring legal authority, identifying authoritative precedent, evaluating the relevance of court decisions, or even predicting the cases that will receive more citations in the future. Yet, on the other hand, citation network analysis over the legislation corpus, provides us information over a single dimension view. Edges on the graph are of the same type and just simple references between documents.

However, in the real-life paradigm of legal domain, there are multiple and heterogeneous networks, each representing a particular kind of relationship, and each kind of relationship plays a distinct role in a particular legal norm. Thus, in order to construct a network model that simulates legislation in a quite robust way, we have to take into account the multi-scale structure of law. Distinct features of the law as the hierarchy between the sources of law, or different types of relations between legal documents should be properly carved and incorporated into a model, as we analyse next.

2. Legislation Modelling

The way laws are correlated leads to a natural representation of the Legislation, the graph. In fact, if we consider the set of laws as nodes, and the identified references as edges, we get a directed graph. However, in order to fully model the legislation corpus we have to properly analyse it and carefully identify its unique features. In the following sections we describe the dataset used for the legislation analysis and the way the respective network is constructed.

2.1. Dataset used for legislation analysis

European Union law consists of founding treaties and legislation, such as Regulations and Directives, which have direct or indirect effect on the laws of European Union member states. There are three sources of European Union (EU) law:

a primary, the Treaties establishing the EU
b secondary, regulations and directives which are based on the Treaties
c supplementary law, the case law of the Court of Justice, international law and the general principles of law.

The official legal portal of the European Communities is offered by the EUR-Lex[3], a free public service for the dissemination of EU law. EUR-Lex contains all documents printed in the Official Journal of the EU dating back to 1951. For the purposes of our

Table 1. Explanation of the EUR-Lex sector classification mechanism (#docs corresponds to the numbers of documents within each sector, as of July 2013)

<table>
<thead>
<tr>
<th>EUR-Lex Classification</th>
<th>Sec.</th>
<th>Title</th>
<th>Explanation</th>
<th># docs</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Treaties</td>
<td>Treaties establishing the EU / supplementing Treaties</td>
<td>8,652</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>International agreements</td>
<td>Agreements between the EU and other sovereign countries</td>
<td>8,564</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Legislation</td>
<td>Secondary legislation to implement EU policy</td>
<td>120,550</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Complementary legislation</td>
<td>Agreements between Member States</td>
<td>1,231</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Preparatory acts</td>
<td>Proposals for future legislation/ opinions</td>
<td>73,123</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Jurisprudence</td>
<td>Case law (judgments, orders, interpretations and other acts)</td>
<td>37,570</td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td></td>
<td></td>
<td></td>
<td><strong>249,690</strong></td>
</tr>
</tbody>
</table>

work we have downloaded all documents within and extracted unnecessary html formatting option in order to obtain a text copy of the European Communities legal database.

Within this database, documents are organized into sectors. Table 1 summarises the sectors of the EUR-Lex database with their corresponding number of documents, as of July 2013. We have extracted all legislation concerning Sectors 1 to 6 from the database, in accordance with the three sources of EU law, accounting for a total number of 249,690 documents.

EUR-Lex database offers analytical metadata for each document. The bibliographic notes of the documents contain information such as dates of effect and validity, the legal form of the document, authors, the subject matter, the legal document from which the document draws its authority, as well as various relationships to other documents and classifications.

We considered that fields, which provide links to other documents in the database, are of particular significance and importance for our study. In Figure 1 we provide a visual representation example of a sequence of modifications imposed to a legal document in the form of amendments. The council directive 370L0220, dated 20 March 1970, was amended by directive 383L0351 in 16 June 1983 and then further amended by directive 389L0491 in 17 July 1989. Note that this is a bidirectional relationship. This is because, since directive 383L0351 modifies/amends directive 370L0220 then directive 370L0220 is amended by 383L0351.

Figure 1. Cross-reference links between legal documents in the EUR-lex. Amended by and Amendment to are bidirectional relationships

References in the legislation can be divided into two different categories: (a) read-only references that do not modify the target document and (b) edit references that mod-

4Internal reference is a reference that points to an article in the same regulation and thus, by intuition, is excluded from the scope of our study
Table 2. Type of references found in the EUR-Lex

<table>
<thead>
<tr>
<th>Type</th>
<th>Explanation</th>
<th>% of Ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amended by</td>
<td>The document is amended by another</td>
<td>9.50%</td>
</tr>
<tr>
<td>Amendment to</td>
<td>The document amends another document</td>
<td>9.50%</td>
</tr>
<tr>
<td>Legal basis</td>
<td>The document is authorized by the mentioning document</td>
<td>23.50%</td>
</tr>
<tr>
<td>Instruments cited</td>
<td>The document cites other docs</td>
<td>54.93%</td>
</tr>
<tr>
<td>Affected by case</td>
<td>The document was altered as of a case result</td>
<td>2.00%</td>
</tr>
<tr>
<td>Other</td>
<td>Various types of references.</td>
<td>0.57%</td>
</tr>
</tbody>
</table>

ify either the text or the lifecycle of the target document. Instruments cited is an example of the former, while amended by is an example of the latter.

Table 2 provides an overview of the major category types for the references found in the EU law database. It also identifies that the Instruments cited reference type consists of more than the half of the Legislation Network (close to 55%). This means that if we consider the respective corpus as simple instances of citation networks, like previous studies, then we would have nearly 45% of the total relations neglected. This also indicates that previous studies that focus solely on citation analysis over legal corpora, ignore a significant amount of the networks properties.

2.2. Legislation Network Construction

Generally, legislation consists of a number of normative documents that are cross-referred to each other. Thus, a directed network can be formed if a legal document refers to another (outgoing link), or is refereed by another document (incoming link).

Figure 2 displays the formation of EU law network from the legal document database. Nodes of the network represent the legal documents. Every document in the legal collection is analysed for cross references. If a cross reference is found between two documents, then a suitable edge connects those two nodes.

![Figure 2. A fraction of the EU Law Network](image)

Node types vary according to the corresponding sector of the legal document, as already explained in Table 1. Edges of the graph are multi type also. Their type follows the types of references found in the EUR-Lex database, as depicted in Table 2. In total the graph consists of 249,690 nodes and 998,902 edges connecting the nodes.

Nodes and edges on the legislation network have temporal attributes also. Each node is marked with a date of effect, the date that the legislation became effective and a date of
expiry, the date that the legislation will cease to effect. Quite often legislation is adopted without an explicitly stated expiry date, also called as sunset close. For those nodes, without a sunset close, we have set an expiration date for the year 9999. Edges follow the temporal distribution of the corresponding nodes. That is, an edge is considered valid only for the time periods between the effective dates and sunset close of the nodes they connect.

The EU Legislation Network, as many real-world networks, exhibits both temporal evolution and multi-scale structure. It is a multi-relational network, as it is a network with a heterogeneous set of edge labels, which represent references of various types (Legal basis, Instruments cited, etc.).

Due to the limited space, we will not present the whole analysis of this network, but we will confine the analysis on various sub networks that are of special importance for the understanding of the legislative process. Nevertheless, our approach is of general usage and any particular combination of node and edge filtering technique can be applied within our framework.

We have identified the following four sub networks, which we examine in detail through the rest of the paper:

1. **The current Legislation Network (LN)**. It is a sub graph of the original graph using all legislation that it was in effect on year 2013 (As of July 2013). Nodes and edges from the Legislation Network that have become invalidated, e.g. sunset close past date, were removed from the Legislation Network.

2. **The (current) network of Regulations (RN)**. In this network we keep track only of legal documents that belong to the sector Legislation of EUR-Lex. This means that this network accounts for the corpus of secondary legislation to implement EU policy. Invalidated nodes and edges have been removed.

3. **The (current) network of Instruments cited (ICN)**. The network of Instruments cited contains all documents of the Legislation Network and only those edges connecting nodes of type Instruments cited. It resembles a citation network as it is studied in previous works. Again invalidated nodes and edges have been removed.

4. **The network of Legal basis (LBN)**. Within this network we keep only the edges of type Legal basis. An edge is added the network from node legal document A to node legal document B if A is authorized by B. This network is of great importance for everyone trying to identify the internal hierarchy of the legislation corpus. Again, invalidated nodes and edges have been removed.

In order to construct the sub-networks we divided the Legislation Network into sub-graphs based on the following criteria: sector type, reference type, time period or even a combination of them. Algorithm 1 described below, helps us to divide the legislation graph in a sub-graph of specific sector of legislation. The corresponding EU legislation Sectors are presented in Table 1.

Similarly, Algorithm 2 separates the legislation graph in a sub-graph of specific relations. Applicable types of legislation references are presented in Table 2.

Table 3 summarizes various properties of the sub Legislation Networks that we present. For each network we indicate the number of nodes, the number of edges, the average degree, the diameter, the average path length, the size of the giant component (g.c.) and the number of isolated nodes.
Algorithm 1: Produce legislation graph of specific sector

**Input:** legislation graph \( G \), legislation sector \( s \)

**Output:** legislation graph \( G \) of specific sector

1. \( \textbf{Sectors} \leftarrow \text{list of legislation sectors} \)
2. \( \textbf{for each sector} \in \textbf{Sectors} \) do
3. \( \textbf{if} \ s \neq \text{sector} \) then
4. \( n \leftarrow \text{nodes in } G \text{ of sector type } s \)
5. \( e \leftarrow \text{edges}(n) \)
6. \( G \leftarrow G \subset (n, e) \)
7. \( \textbf{end if} \)
8. \( \textbf{end for} \)
9. return \( G \)

Algorithm 2: Produce legislation graph with specific relations

**Input:** legislation graph \( G \), relation type \( r \)

**Output:** legislation graph \( G \) of specific relation

1. \( \textbf{Relations} \leftarrow \text{list of legislation relations} \)
2. \( \textbf{for each relation} \in \textbf{Relations} \) do
3. \( \textbf{if} \ r \neq \text{relation} \) then
4. \( e \leftarrow \text{edges in } G \text{ of relation type } r \)
5. \( G \leftarrow G \subset (e) \)
6. \( \textbf{end if} \)
7. \( \textbf{end for} \)
8. return \( G \)

Table 3. E.U. Legislation network properties

<table>
<thead>
<tr>
<th>Metrics / Network</th>
<th>Legislation (LN)</th>
<th>Regulations (RN)</th>
<th>Instruments cited (ICN)</th>
<th>Legal basis (LBN)</th>
</tr>
</thead>
<tbody>
<tr>
<td># of nodes</td>
<td>164,932</td>
<td>62,451</td>
<td>164,932</td>
<td>164,932</td>
</tr>
<tr>
<td># of edges</td>
<td>524,724</td>
<td>72,659</td>
<td>387,990</td>
<td>74,489</td>
</tr>
<tr>
<td>Average degree</td>
<td>6:3629</td>
<td>2:3269</td>
<td>4:7049</td>
<td>0:9033</td>
</tr>
<tr>
<td>Network diameter</td>
<td>25</td>
<td>25</td>
<td>78</td>
<td>28</td>
</tr>
<tr>
<td>Average path length</td>
<td>7:5773</td>
<td>7:1997</td>
<td>6:8951</td>
<td>1:4790</td>
</tr>
<tr>
<td>Size of g.c.</td>
<td>116,790</td>
<td>29,583</td>
<td>78,140</td>
<td>49,038</td>
</tr>
<tr>
<td>Isolated nodes</td>
<td>42,821</td>
<td>26,102</td>
<td>79,504</td>
<td>113,014</td>
</tr>
</tbody>
</table>

3. Network Analysis

An important realization of network analysis is that networks in natural, technological and social systems are not random, but follow a series of basic organizing principles in their structure and evolution that distinguish them from randomly linked networks [12]. A much studied metric for networks is the degree distribution, \( P(k) \), giving the probability that a randomly selected node has \( k \) links. In many cases the probability \( P(k) \) decays as a power-law, following

\[
P(k) \propto k^\alpha
\]
Figure 3. The degree distribution of the four legislation sub networks. Each column accounts for one sub network with the top row for inward links while the bottom row represents outward links. Data has been binned logarithmically to reduce noise. The fitted power-law (red line) and log-normal (blue line) distributions are also plotted.

where $\alpha$ is a constant parameter of the distribution known as the exponent or scaling parameter, that typically lies in the range $2 < \alpha < 3$. This feature is common to large scale communication, biological and social systems [12,10,2] and to the network of US Supreme Court decisions [5,11]. In practice, few empirical phenomena obey power laws for all values of $x$. More often the power law applies only for values greater than some minimum $x_{min}$. In such cases we say that the tail of the distribution follows a power law.

Figure 3 (a,b) shows a cumulative distribution function of inward links (number of times each legal document was cross referenced) and outward links (the number of other documents each document cross-references) in the legislation network (LN), on a log-log scale. The majority of documents are cross referenced by only a few times, while there are a few documents that are widely linked.

The same cumulative distribution function appears on the legal basis network (LBN) for the same year is depicted in Figure 3 (d). A few laws are highly influential, they serve as legal basis for several others and similarly a few laws have high support, a large number of laws form its legal basis. The approximate straight-line form of the distribution function allow us to infer that, in all the examined sub graphs, the Legislation Network follows a power law distribution.

Another important topological characteristic that many real graphs were found to exhibit is the so called small-world. According to [8] small-world networks are defined as having a small diameter and exhibiting high clustering. Many social, technological, biological and information networks have been studied and categorized as small-world networks [9]. Small-world networks can be seen as systems that are both globally and locally efficient, in terms of how efficiently information is exchanged over the network. [18]
Small world properties are measured by the average shortest path and clustering coefficient metrics. In order to classify a network as a small-world network, the candidate network metrics are compared with Erdős-Rényi random networks, with the same number of nodes and edges. If a network exposes the small world properties, then it is expected that average shortest path is slightly shorter than of a random network and the average clustering coefficient is of magnitude larger than that of a random network.

Similar to many studies on the small-world networks, our analysis is restricted to the giant components in the networks (i.e. the maximal connected sub-graph of the network). Table 4 summarizes the results of our analysis for the four legislation networks we present. The parameter network average shortest path and average clustering coefficient metrics are denoted by $L_{net}$ and $C_{net}$ and the corresponding random network ones are symbolized as $L_{rand}$ and $C_{rand}$ respectively.

<table>
<thead>
<tr>
<th></th>
<th>$L$</th>
<th>$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Legislation</strong></td>
<td>7.5773</td>
<td>0.0215</td>
</tr>
<tr>
<td>Random</td>
<td>7.9695</td>
<td>7.5291e-05</td>
</tr>
<tr>
<td><strong>Regulations</strong></td>
<td>7.1997</td>
<td>0.007067</td>
</tr>
<tr>
<td>Random</td>
<td>12.3370</td>
<td>0.0001235</td>
</tr>
<tr>
<td><strong>Instruments cited</strong></td>
<td>6.8951</td>
<td>0.03434</td>
</tr>
<tr>
<td>Random</td>
<td>7.2577</td>
<td>0.0001086</td>
</tr>
<tr>
<td><strong>Legal basis</strong></td>
<td>1.4790</td>
<td>0.0002776</td>
</tr>
<tr>
<td>Random</td>
<td>23.2799</td>
<td>4.2320e-05</td>
</tr>
</tbody>
</table>

Despite the variations in the metrics, all of the networks satisfy the small-world conditions. Comparing our results with other studies, as presented in [9], we see that the average shortest path lengths in the legislation sub graphs are distinctively smaller than the values of networks reported and of magnitude smaller than the theoretical average degree of the corresponding random model. We attribute this finding to the nature of law and its hierarchical form. Legal documents are made by the authority given by other legal documents, which reduces their total number of references well below the expected number from the random model.

4. Conclusions and future work

In this paper we introduce a novel approach to model the law: the Legislation Network. Our approach offers a model to create a systematic alternative structure to a naturally evolved normative system. The Legislation Network is a multi-relational network that accommodates the hierarchy between the sources of law and can represent relationships of various categories between legal documents, alongside their temporal evolution.

Characterizing the structural properties of a network is of fundamental importance to understand the complex dynamics of the modelled system. The Legislation Network is highly heterogeneous with respect to the number of edges incident on a node. The degree distribution of legal documents follows a power law and the connectivity of the

---

5 Average number of steps along the shortest paths for all possible pairs of network nodes.
6 Clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together.
Legislation Network relies on a small set of very important legal documents. We plan to perform a robust resilience test to assess its vulnerability under specific cases that may lead to possible breakdowns. We also plan to study the temporal evolution of the Legislation Network as to provide a richer model to better explain the structure and evolution of legislation.

Likewise our model can be exploited for visualizing the legal corpus. Graph visualizations are used to convey the content of a graph as they can highlight patterns, reveal clusters and related connections. We believe that a visualization system for the Legislation Network can be of great assistance to both citizens and legal experts, helping them to easily navigate the legislation corpus. Another great benefit of such an approach, lies in the fact that legislation can be exploited not only from the traditional point-of-view, but as a graph of hyper-textual information with temporal properties. As an example, it will be easier for lawmakers to monitor the effect of a possible change in the whole normative system, thus taking appropriate actions. In such a system, the use of domain-specific ontologies and linked data techniques would further enrich the added value of Legislation Network.

Finally, our modelling approach can be used to improve the effectiveness of legal information retrieval systems. Our hypothesis is that the Legislation Network can be exploited for text retrieval, in the same manner as hyperlink graphs on the Web.

References