Soc. Netw. Anal. Min. (2016) 6:56
DOI 10.1007/s13278-016-0365-6

ORIGINAL ARTICLE

A computational approach for the experimental study of EU case law: analysis and implementation

Nicola Lettieri[1] · Antonio Altamura[2] · Armando Faggiano[2] · Delfina Malandrino[2]

Received: 7 December 2015 / Revised: 24 June 2016 / Accepted: 20 July 2016 / Published online: 2 August 2016

Abstract In recent years, the encounter between network analysis (NA) and Law has issued new challenges both on a scientific and application level. If, on the one hand, it is fostering new computational-inspired approaches to visualize, retrieve, manipulate and analyze legal information, on the other hand, it is inspiring the creation of innovative tools allowing legal scholars without technical skills to start dealing with NA and visual analytics on their own. This paper presents an ongoing research project aiming to explore how approaches and techniques at the boundaries between Network analysis, Legal informatics and Visualization can help shedding new light into legal matters. The attention is focused, on EuCaseNet, an online toolkit allowing legal scholars to apply NA and visual analytics techniques to the entire corpus of EU case law.

The authorship of this work can be attributed as follows: Nicola Lettieri: concept and functional design of EUCaseNet; Legal, Computational Social Science and Legal Informatics profile. Delfina Malandrino, Antonio Altamura and Armando Faggiano: system design, Computer Science profiles. The case study is the result of a joint effort of the authors.

✉ Nicola Lettieri
n.lettieri@isfol.it

Antonio Altamura
antonio.altamura7@gmail.com

Armando Faggiano
armando.faggiano@gmail.com

Delfina Malandrino
dmalandrino@unisa.it

1 ISFOL, Rome, Italy

2 Computer Science Department, University of Salerno, Salerno, Italy

Keywords Computational legal studies · Legal informatics · Network analysis · Visual analytics

1 Introduction

In the last years, Network Science, the "study of network representations of physical, biological, and social phenomena", has gradually spread in extremely heterogeneous research areas ranging from Physics to Biology and Genetic, from Mathematics and Computer Science to Sociology and Economics. Nowadays, Network Analysis techniques (NA) lie at the heart of many scientific and practical applications including visual analytics (VA), data aggregation and mining, network propagation modeling, user behavior analysis, social sharing, filtering and recommendation, marketing intelligence and even law enforcement activities (Golbeck 2013).

In the legal field, new research challenges are being raised by the intersection between NA and Law. If, from one hand, scientifically it is fostering new computational approaches to retrieve, manipulate and analyze legal information, on the other hand, it is inspiring the creation of innovative visual tools supporting both domain experts and laymen in legal documents analysis, allowing them to derive useful insights in a growingly easy and effective way. This work presents an ongoing research project exploring the potentialities of NA and VA in supporting the study of European Court of Justice (hereinafter ECJ) case law. Our effort is driven by three main goals: The first one is to devise and propose new analysis techniques to address the above mentioned scientific challenges. The second one is the design and the implementation of *EUCaseNet*, an online and freely exploitable laboratory allowing the interactive application of VA and NA techniques to the corpus of EU case law. The third one, finally, is to start using our tool trying to answer specific research questions also in the light of the current legal literature.

The rest of the paper is organized as follows. In Sect. 2, we present the background, making reference to research experiences at the boundaries between Law, NA and VA. Section 3 is devoted to the discussion of the main motivations of our work. In Sect. 4, we introduce *EUCaseNet*, its architecture and its main functionalities. Two case studies that highlight the potentialities of *EUCaseNet* are presented in Sect. 5. Finally, in Sect. 6 we conclude with some ongoing works and future directions.

2 Background

As witnessed by a series of works made mainly (but not only) in the Legal Informatics research area (for an overview see, among the others Winkels et al. 2014), law has developed a growing interest toward the insights offered by NA and, more in general, by the intersection between graph theory, visualization techniques and legal issues. Thanks to its high level of abstraction and to the ability to support the understanding of structural and functional features of real networks, NA is spreading in the legal domain through both scientific and commercial applications exploited purposes that span from visual analytics to criminal network analysis. The most relevant and inspiring experiences made in this direction are briefly described in the following three sections.

2.1 Visualize, analyze and retrieve legal information

Network analysis is strictly connected with Visual Analytics, a fledgling research field aiming to devise innovative ways to turn large dataset into knowledge while also enabling users to act upon their findings in real-time. The term "visual analytics" appeared around 2000 within the computer science research area (Wong and Thomas 2004) and then gradually moved in other contexts giving birth to a multidisciplinary field that combines "visualization, human-computer interaction, data analysis, data management, geo-spatial and temporal data processing, spatial decision support and statistics" (Keim et al. 2010). In the last decade, VA techniques have become relevant in a growing number of areas spanning from Physics to Business intelligence in which large information datasets need to be analyzed. The study of financial markets, just to give an example, requires the analysis of huge amounts of daily generated data and a crucial challenge, therefore, is to analyze these data to monitor the market, understand historical situations, forecast trends or identify patterns of recurring events.

In this scenario, law is not an exception: The legal regulation of social life is the result of many heterogeneous sources belonging to different institutional levels (local, national, and international), an intricate universe of documents encompassing not only laws, but also case law, legal literature, administrative practices that form together a unitary and complex whole. Picking the way in this universe is a tough task that is crucial both for lawyering and scholarly investigation. When combined with visualizations techniques, NA can represent an interesting innovation in this regard: It does not simply make easier and more intuitive information retrieval (a feature that is particularly useful when managing large quantities of documents), but it also offers new insights in the legal world as it is already happening in many other research fields.

There already are interesting applications of VA and NA in the legal field. Lawiz Lettieri et al. (2014), is a recent experimental web application designed for the visualization, the exploration and the analysis of documents coming from heterogeneous sources of the Italian law (legislation, parliamentary works, Supreme Court case law, Constitutional case law, legal doctrine). Starting from a legislative measure given as input by the user, the application implements two visual analytics functionalities aiming to offer new insights on the legal corpus under investigation. The first one is an interactive node graph depicting relations and properties of the document of interest toward the whole document base. The second one is a zoomable treemap showing the topics, the evolution and the dimension of the legal literature settled over the years around the norm of interest.

Ravel Law[1] is a visual commercial search engine that exploits both natural language processing, machine learning and graph visualization to help lawyers in sorting through legal information. Whereas traditional legal databases present results in a column, often hiding important cases pages back in search results, Ravel Law visually represents the most important cases on a particular topic as the node of a network, with numerous edges pointing to subsequent cases that have cited it. The size of the hub reflects the relative number of cases that cite it. The frequency with which courts cite a particular case often signals the influence of the case over a given area of law or its relevance to a particular legal concept.

Another interesting (yet simpler) experiment is Lexmex,[2] an on line system developed in France at the Compiegne University of Technology that allows to visually display the relations between texts of law. The attention of the experiment is focused on the French Civil Code and related legislation; Lexmex "translates" the legislation in

[1] http://www.ravellaw.com.

[2] http://www.lexmex.fr/.

nodes and links generating a graph. The semantics of the visualization is simple: A node is bigger depending on the number of connections it has with other nodes of the graph. The colors correspond to the cluster detected by means of an algorithm allowing to detect of clusters or "communities"). The tool implements essential functionality for dynamic navigation such as zooming, node selection to show contextual information, search by keywords.

In more general terms, the three mentioned projects can be brought back to the efforts made by Information Visualization and Visual Analytics research (Cook and Thomas 2005; Keim et al. 2010; Wong and Thomas 2004) to allow people to turn data into knowledge facilitating analytical reasoning by means of interactive visual interfaces. The issue is connected with "exploratory search", a topic emerged in the information retrieval area (Duarte et al. 2015; Marchionini 2006) that has devoted a great deal of research to design systems and interaction techniques overcoming the limits of classical lookup search where the results are presented in list-based widgets. The idea underlying the exploratory paradigm is to more actively involve users providing them with a more direct control over the search process combining the traditional search activities with higher level goals (e.g., comparison, analysis, synthesis and evaluation).

2.2 Investigate the societal dimension of crime

Recent years has witnessed a growing interest toward the use of Social Network Analysis (SNA) techniques in the study of criminal organizations both for scientific and investigation goals. Sociality has a tremendous influence on crime: A large part of criminal phenomena from pornography trafficking to hacking and other cybercrimes is strongly conditioned (inhibited or facilitated) by relational dynamics. Criminals are often highly social actors: They communicate among them, collaborate and form groups in which it is possible to distinguish leaders, subcommunities and actors with very different roles. Therefore, SNA techniques seem to be perfectly suited to the needs of organized crime studies. This task is facilitated by the growing amount of digital information spanning from e-mails, to credit-card operations today available. The data deluge and the availability of increasingly advanced data mining techniques are offering both researchers and law enforcement agencies new tools and methodologies to unveil and understand structures and dynamics of criminal networks.

Criminal Network Analysis (CNA) is today a well-established interdisciplinary research area (Calderoni 2010) in which network analysis techniques are employed to analyze large volume of relational data and gain deeper insights about the criminal network under investigation. There are many examples of application of the principles of SNA in the analysis of criminal organizations. Among the most interesting works, we can cite (Stewart 2001; Xu and Chen 2005a). Specifically, in Xu and Chen (2005a) authors state that effective use of SNA techniques to mine criminal network data can have important implication for crime investigations, mainly because they can aid enforcement agencies fighting crime proactively, for example, by intervening when a crime takes place and also with the required police efforts. They developed *CrimeNet Explorer* (Xu and Chen 2005b), a system that, in addition to the visualization functionality, detects subgroups in a network, discover interaction patterns between groups and identify central members in a network.

As also discussed in Hutchins and Benham-Hutchins (2010), the most common challenge to address when designing and implementing a new software system is the ability to use information and document from many sources. Specifically, data can be structured and unstructured, whereas the latter representing the most important barrier for researchers that want to develop a new system for the domain experts. Additionally, it must be emphasized that data can be available from many domain sources as well as from many papers and electronic sources. Authors in Hutchins and Benham-Hutchins (2010) address this challenges by proposing a system software that supports both qualitative (visual) and quantitative (mathematical) analysis of criminal network data. Similar to other analyzed works the proposed system only applies SNA metrics to analyze the structure of the criminal network in order to derive useful insights about roles' entities and communication patterns.

After 9/11, social network experts began to look explicitly at the use of network methodology in understanding and fighting terrorism. Some important works include the analysis performed by Krebs (2002) that collected publicly available data on the Al-Qaeda hijackers and applied the social network metrics to derive useful insights about the connections among the terrorists and their roles in the network (Muhammad Atta was identified as the key leader of the network). A similar analysis was performed by taking into account newspaper articles and radio commentary (Rothenberg 2001). Additional works studied the potential use of SNA to destabilize terrorist networks (Memon and Larsen 2006). In an interesting work in the field of criminal network analysis (Ferrara et al. 2014), the authors presented *LogAnalysis*, an expert system specifically designed to allow statistical network analysis, community detection and visual exploration of mobile phone network data.

Finally, in a very recent work (Lettieri et al. 2016), authors defined a holistic methodology that combines document-enhancement, social network analysis and visualization techniques to support public prosecutors and criminal investigation departments in exploring the societal dimension of criminal groups. This approach has been deployed in a computational framework, named CrimeMiner and validated with a case study based on data coming from real criminal investigations.

2.3 Analyze legal documents

As highlighted by a growing literature, NA can play a role also in addressing legal analysis issues. The structural and functional analyses enabled by the NA approach appear to be a viable way to shed new light into "traditional" legal research questions. It is for this reason that NA is having diffusion in the legal analysis domain thanks to works focusing on four different research priorities. The first one is to determine the relevance of legal documents and sources, according to the different meanings acquired by the concept of "relevance" itself in different legal contexts. The second priority is to analyze the structural and functional features of more or less wide areas of legal systems (e.g., the level of complexity of legislation, case law, legal literature). Another relevant research goal is to study the relations between different expressions of the legal phenomenon (relationship between legislation and case law, or between supranational case law and domestic case law, etc.). Finally, an emerging research topic is the study of the evolution of a legal order also in a predictive fashion.

The determination of the relevance of legal sources and documents with NA techniques is one of the most discussed issues in recent times. The study of networks has been frequently used to analyze case law especially in common law countries where precedents have a prominent role (Fowler et al. 2007; Post and Eisen 2000; Smith 2007). An interesting and seminal work is the one presented in Fowler et al. (2007) that exploits the HITS algorithm (Kleinberg 1999) and the patterns in citations networks within and across cases to create importance scores identifying the most legally relevant precedents in the network of US Supreme Court law. By comparing different centrality measures, the authors show how the HITS algorithm allows to identify, with an higher level of precision than the other centrality measures, those cases in the network that are "influential" (inwardly relevant) or "well founded in law" (outwardly relevant). Another relevant research experience in this area is the one recently described in a work from Malmgren (2011) in which the PageRank algorithm is combined with In-degree and the HITS algorithm to investigate the relevance of the decisions of the ECJ.

As to the uses of NA to measure structural characteristics of legal systems, such as the complexity of large bodies of law, a paradigmatic example is the research proposed by Katz and Bommarito (2014) in a work focusing on the legal complexity of the United States Code that borrows concepts and tools from a range of academic disciplines, including computer science, linguistics, physics and psychology. By developing a knowledge acquisition protocol that considers structure, language and interdependence of Code's norms as elements contributing to its complexity, authors employ NA methods to test the interdependence level of the norms of the Code, concluding that the more norms are connected, the higher is the amount of energy an individual should expend in order to acquire knowledge regarding the content of the United States Code. Dealing with a somehow similar challenge is the work presented in Winkels and Boer (2013) where the authors try to build the "legal neighborhoods" of Dutch norms aiming to verify if, given the network of an article in focus, other relevant documents could be automatically identified just using "objective" meta-information deriving from the norm graph.

Quantifying the complexity of a given legal system is also close to the goal of a recent research project, by Koniaris et al. (2015), which aims to explore the interdependence between different kinds of legal documents, such as jurisprudence and doctrine or cases law and norms, presenting a multi-relational network modeling the hierarchy and showing the relationships between different legal sources of the European Union law (Treaties, International agreements, Legislation, Complementary legislation, Preparatory acts, Jurisprudence) in order to shed light on properties and behaviors of its different elements. The different subgraphs built by authors have been also employed to realize a temporal analysis of the evolution of the European legislation network, as well as to test the "resilience" of the entire system, measuring its vulnerability under specific cases that may lead to possible breakdowns. In the same vein, i.e., predicting the future developments of a given legal system, in Fowler et al. (2007), the network centrality measures are used to forecast which cases will be most cited in the future.

All the mentioned experiences gave a significant contribution in building a bridge between Law, NA and VA creating the basis for further scientific and application developments. In this scenario, another factor of paramount relevance is represented by the involvement of legal scholars in NA experiments in order to support both the design of research activities and the validation of the obtained results. Legal experts' knowledge is essential in identifying mistakes or ambiguities emerging from the legal network analysis performed by non-legal experts compared to the real features of the law.

As a matter of fact, there is not necessarily an overlap between the results that can be simply achieved through a naive application of NA measures to legal data and what could be considered relevant for the law. It could happen that some typical network properties, usually measured by computer scientists interested to the computational features of a specific graph (e.g., the resilience), were not usefully applicable to networks in legal context, especially without an expert explaining the legal consequences of the raw outcomes. From this perspective, it is essential to create tools allowing legal scholars without technical skills to exploit computational approaches to Law with two goals: i) make experiments with NA and VA, thus pushing new ideas both in legal and NA science; (ii) use NA and visualization in their daily activities (e.g., legal analysis and information retrieval).

3 EUCaseNet: motivation and functional requirements

The reflection about the use of information technologies in the legal field has already highlighted (Brownsword 2015) how the design of new research methodologies and applications cannot be delegated only to people outside the legal world. The understanding of the way in which new computationally grounded methodologies can be exploited by legal scientists and practitioners necessarily needs that the domain experts are gradually put in condition to experience in person new approaches and tools in order also to generate new research questions. On the basis of this consideration, the *EUCaseNet* research aims to experiment new approaches for the analysis of legal corpora, creating an online laboratory freely exploitable by legal scholars to start conducting experiments by themselves.

The project so far has moved in a twofold direction. The first one consisted in the design of an online toolkit allowing the interactive application of VA and NA techniques to the EU case law corpus. The idea is to enable the exploration of new methodologies for the analysis of the legal matter by means of an inductive process: In a first stage, having even basic computational/NA skills and making reference to its professional experience the domain expert chooses the data to consider and a set of algorithms to exploit. In a second step, the expert tries to compare the results obtained through the computational analysis with those achieved through traditional legal analysis methods in order to establish which combination of data and measures better maps the interpretations and the opinions offered by legal studies. The other research direction is centered on the identification of new metrics and algorithms to be validated according to the state of the art of the legal literature. Inspired by an in-depth debate with EU law scholars, *EuCaseNet* currently allows users to make basically two kinds of activities:

- *Study the features of the ECJ case law network.* Users can interact with a network built starting from the analysis of the citations between ECJ judgments (a set of 9706 judgments adopted in a period that goes from 1954 to the to the first semester of 2013) verifying if it is possible to extract meaningful (in a legal perspective) insights from the computational analysis of its features.
- *Visually analyze the ECJ judgments' metadata*. Users can (also visually) dig into the metadata (country, classification, date and other data, see Sect. 4) associated with every single ECJ judgment in order to discover relevant properties (correlations, dynamics etc) of the case law system.

Functionalities so far made available to the domain experts can be grouped in two fundamental classes, and we will describe them in the following sections.

3.1 Analysis of the features of the ECJ judgments corpus

A first group of functionalities aims to extract meaningful insights from the computational analysis of the ECJ case law corpus. From a scientific point of view, the goal is to shed a new light on the features belonging, at macro level, to the entire corpus of EU case law (or groups of judgments) or, at the micro level, to single pronouncements. In more in details, this set of features encompasses the following functionalities:

- Analyze the EU case law system represented as a network. This feature allow the exploration of "network properties" of the ECJ case law corpus (properties deriving from the application of NA measures like Modularity or Betweenness) so that to discover legal properties of single judgments or group of them. The attention is focused on issues like (e.g., content-related between judgments like presence of clusters suggesting (interpretable, for instance, in semantic terms as the sign of the influence of one judgment on a groups of pronouncements).
- Study the evolution of judgments' citations. This functionality exploits the computational analysis of the citation networks of specific judgments to derive potentially relevant insights about the evolution of their impact on the EU legal system (or about the rise and fall of the relevance of the topics discussed in the judgment taken into account).
- Visual Semantic Analysis of ECJ judgments. This functionality is an attempt to give a "visual answer" to questions like "what are the most *hot* topics in EU case law?". What are the topics more recently treated by the EU Court of Justice? Is it possible to trace the evolution of their relevance over the time?

3.2 Analyze the features of some of the most relevant ECJ judgments

A second set of functionalities is focused on the analysis of a specific group of judgments universally considered by legal doctrine as an example of the most relevant ECJ case law. The goal is twofold: visually analyze the properties of what can be considered a gold; standard implemented a sort of supervised "features extraction mechanism" aiming to discover the computational landmarks of relevant precedents (e.g., by stating that relevant judgments are always characterized by high Page Rank values). The functionalities so far implemented are the following:

- Analyze the impact of a judgment on EU case law over time. This functionality consists in the possibility to measure and plot the trends in the citations of every judgment belonging to the gold standard. Even if very simple and preliminary, this feature is a first attempt to quantitatively maps the impact of a judgment on the legal system
- Semantically, compare the topics of the gold standard with ECJ corpus' ones This functionality allows domain experts to analyze the relation existing between the topics discussed in the judgment of the gold standard and the most "hot" topic of EU case law: The idea is to provide legal experts with a quantitative base to establish if there an overlap between the topic of what are considered the pillars of EU legal system and the topics most frequented by the ECJ corpus as a whole.

4 EUCaseNet: system overview

EUCaseNet is a toolkit[3] composed by different modules designed to allow the experimentation of new ways of interaction with the European case law for both professional and scientific purposes. It implements a set of visualization and provides different analysis features that can be used by domain expert to explore on the fly the EU case law corpus as well as single judgments within it. As discussed in Sect. 3, *EUCaseNet* currently hosts two different sections devoted to support, respectively, the analysis of the EU case law network and the study of a set of judgments that, as we will see in more detail in Sect. 5, can be considered as a gold standard for the relevance of legal precedents.

The reference document base used in this work is represented by a set of ECJ judgments freely available from the EUR-Lex portal (see Fig. 1) for a total and 3.5 GB of textual data. Starting from the judgments' identifier (the so called CELEX number), we crawled the EUR-Lex portal through HTTP requests and we saved the data in XML format. EUR-Lex provides, for each judgment, a large number of information. We are currently taking into account a subset of them: (1) title and reference (Celex number) of the judgment; (2) name of the Advocate-General who delivered their conclusion; (3) Classification according to the EUR-Lex classification schemas; (4) document date; (5) presentation date; (6) involved EU country; (7) relationships (expressed in terms of citations) among judgments; and (8) list of references to published comments on judgments (expressed as a number).

4.1 Functionalities

In this section, we describe the two different types of analysis that can be performed through: (1) network analysis and visualization starting from the citation network and (2) quantitative analysis and visualization starting from the judgments' metadata.

4.1.1 Network analysis and visualization of a citation network

This first category enables the application of NA metrics and quantitative analysis on the citation network generated from the judgments of ECJ. Specifically, we envisioned the following functionalities.

Graph generation, visualization and analysis. This feature has been implemented taking into account both the whole corpus of ECJ, shown in Fig. 2 and the gold standard citation network dataset, shown in Fig. 3. The render of the graph was performed by using Sigma.js,[4] a JavaScript library developed for graph drawing in the browser. Starting from an initial view, the graph allows users to manipulate and explore the visualization by applying forced-directed graph layout algorithms, such as ForceAtlas2 (Jacomy et al. 2014; Fruchterman and Reingold 1991). Users can perform several actions such as, zoom, drill-down and roll-up, drag and drop, and filtering by node attributes. The visualization also allows redefinitions of the size and color of the nodes according to the application of NA metrics. By clicking on a node, the focus switches from the general overview to the active node view, by showing contextual information about the judgment.

Citations line graph. This functionality exploits Javascript libraries to generate line graphs (see Fig. 4), starting from the quantitative analysis of the citations received by each judgment of the gold standard over the years.

[3] *EUCaseNet* prototype: http://goo.gl/HcQv5z.

[4] http://sigmajs.org/.

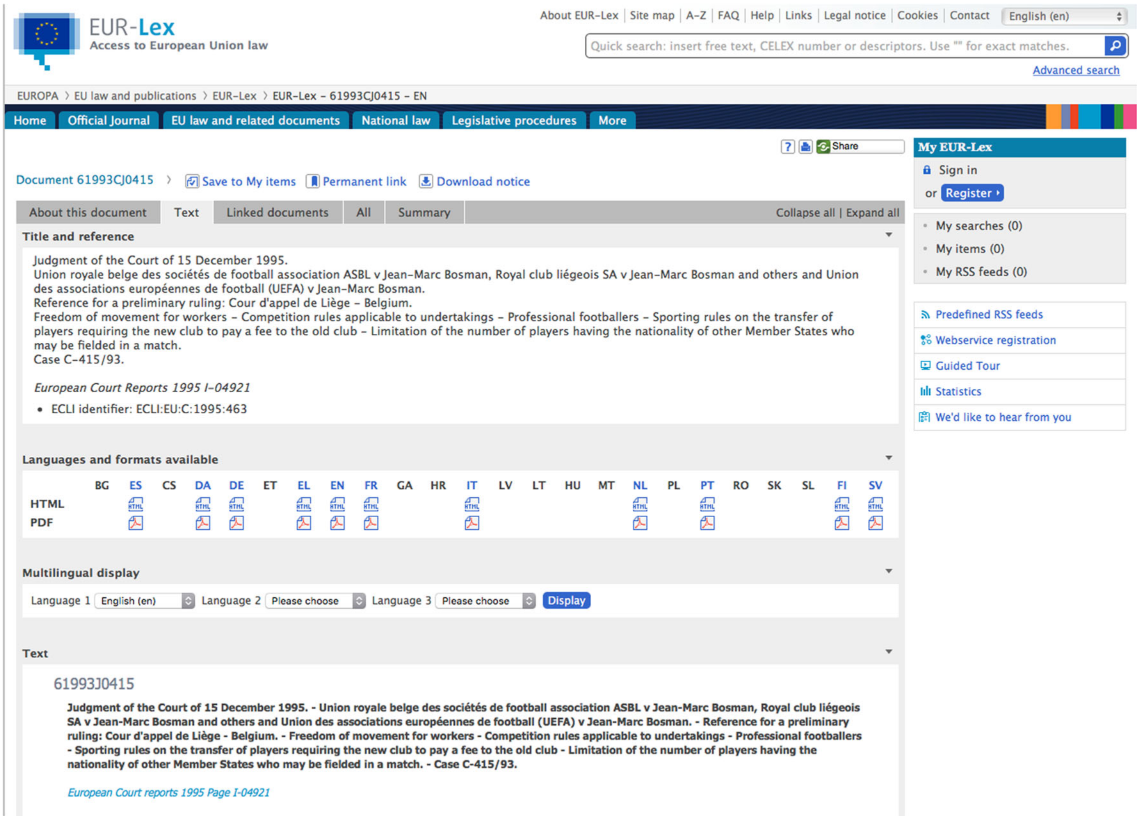


Fig. 1 The Eur-Lex portal

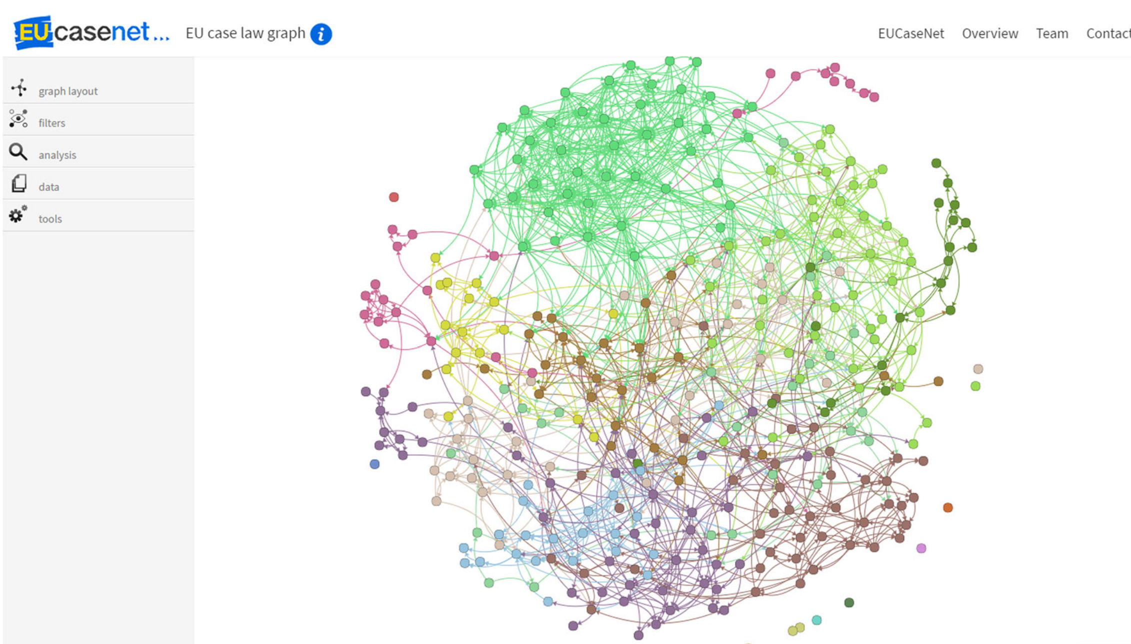


Fig. 2 Citations network graph of the first 350 judgments in terms of in-degree and their citation network

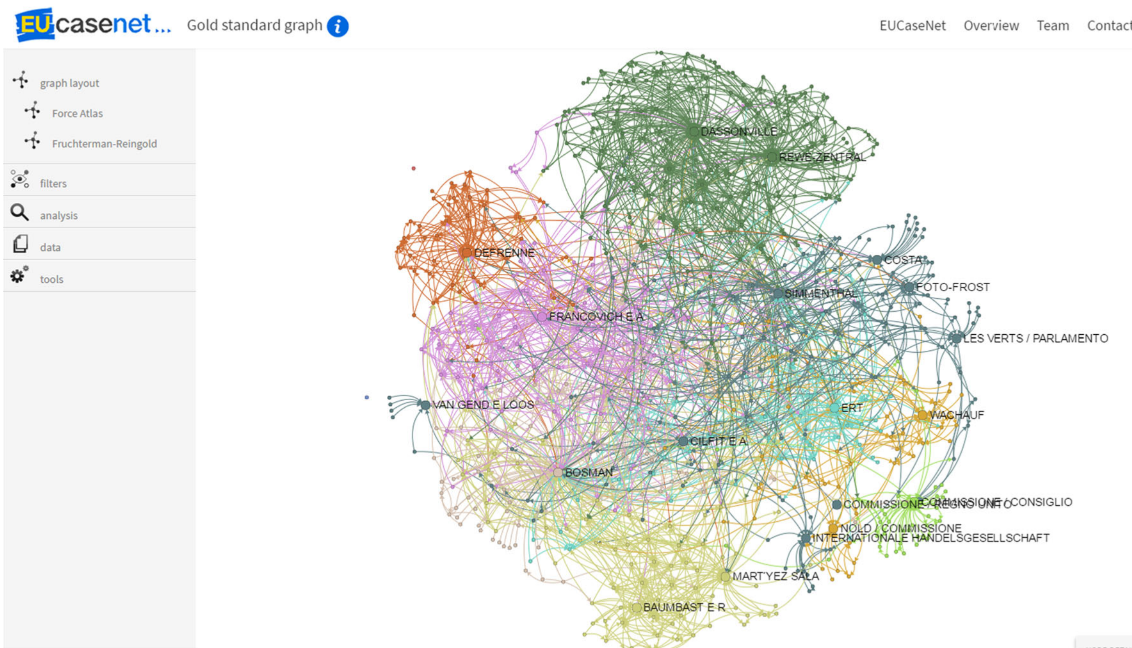


Fig. 3 Citations network of the judgments of the gold standard

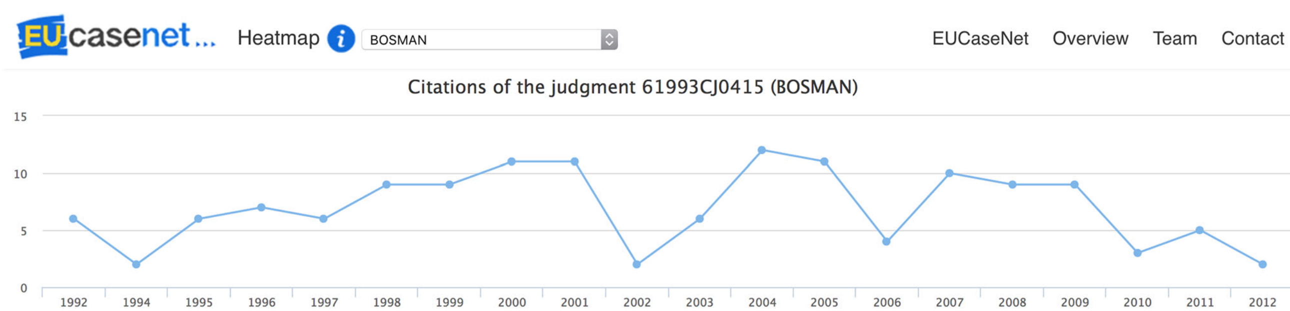


Fig. 4 Citation line graph of a judgment of the gold standard

4.1.2 *Quantitative analysis and visualization of judgments' metadata*

Based on the domain expert guidelines, we focused on the information extracted from EU case law classification for a semantic analysis of the judgments. This functionality allows the user to visualize the arguments (semantic classification officially provided by ECJ Web site) offering a general overview about the frequency (and therefore the relevance) and the evolution of the topics treated by the ECJ case law. This way we are able to somehow extract insightful informations from the metadata associated to every single judgment. To achieve that we implemented achieved the following features:

Treemap generation. The treemap (Brownsword 2015) offers a "semantic fingerprint" of the ECJ case law by simply visualizing colored areas generated by exploiting the metadata provided by the European Court of Justice. The treemap is a well-known method for displaying structured data. Every aspect of the treemap is defined according to a specific metric and conveys information: The size of the sector is proportional to the argument frequency and sorted from the bigger to the smaller. The color is defined on a color scheme (from dark green to light green): each sector takes the color associated with its statistical mode computed on the document year (the newer the lighter green, see Fig. 5). By clicking on a sector, a line graph of the temporal trend of the active argument is shown accordingly.

Heatmap generation. The module take into account the gold standard citations and the arguments of each judgment. The heatmap highlights the arguments of the documents of the gold standard. The aim is to give a "semantic fingerprint" of the gold standard. The sectors in this case are sorted and colored by argument frequency (see Fig. 6).

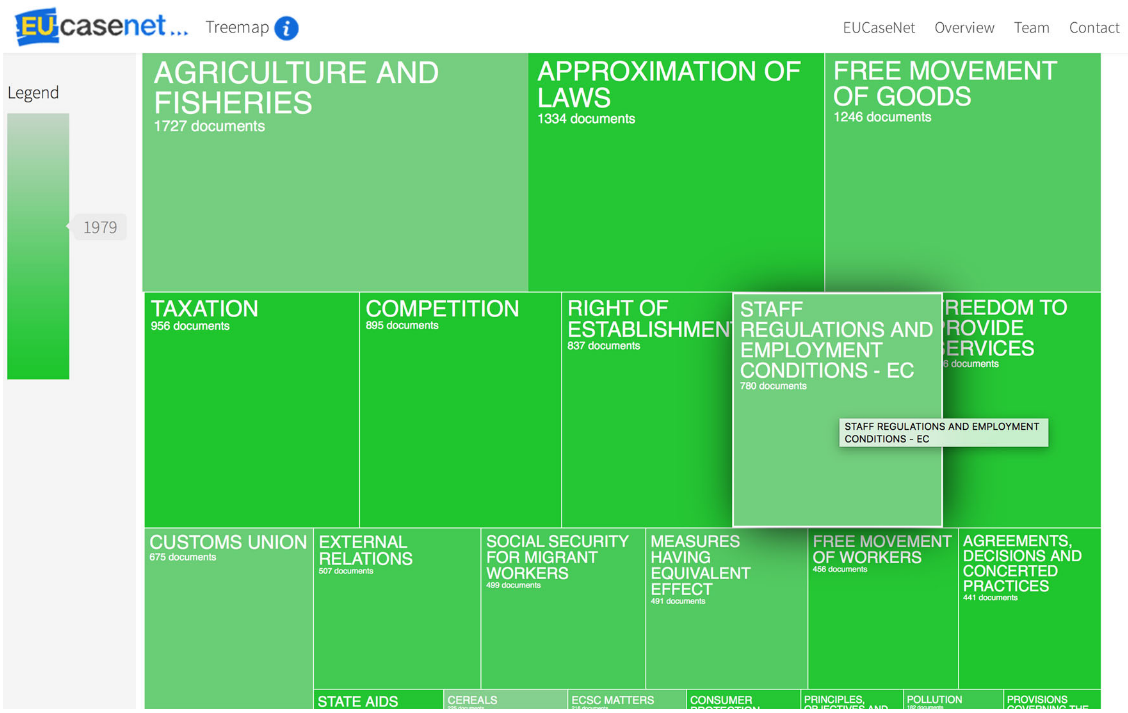


Fig. 5 Treemap user interface

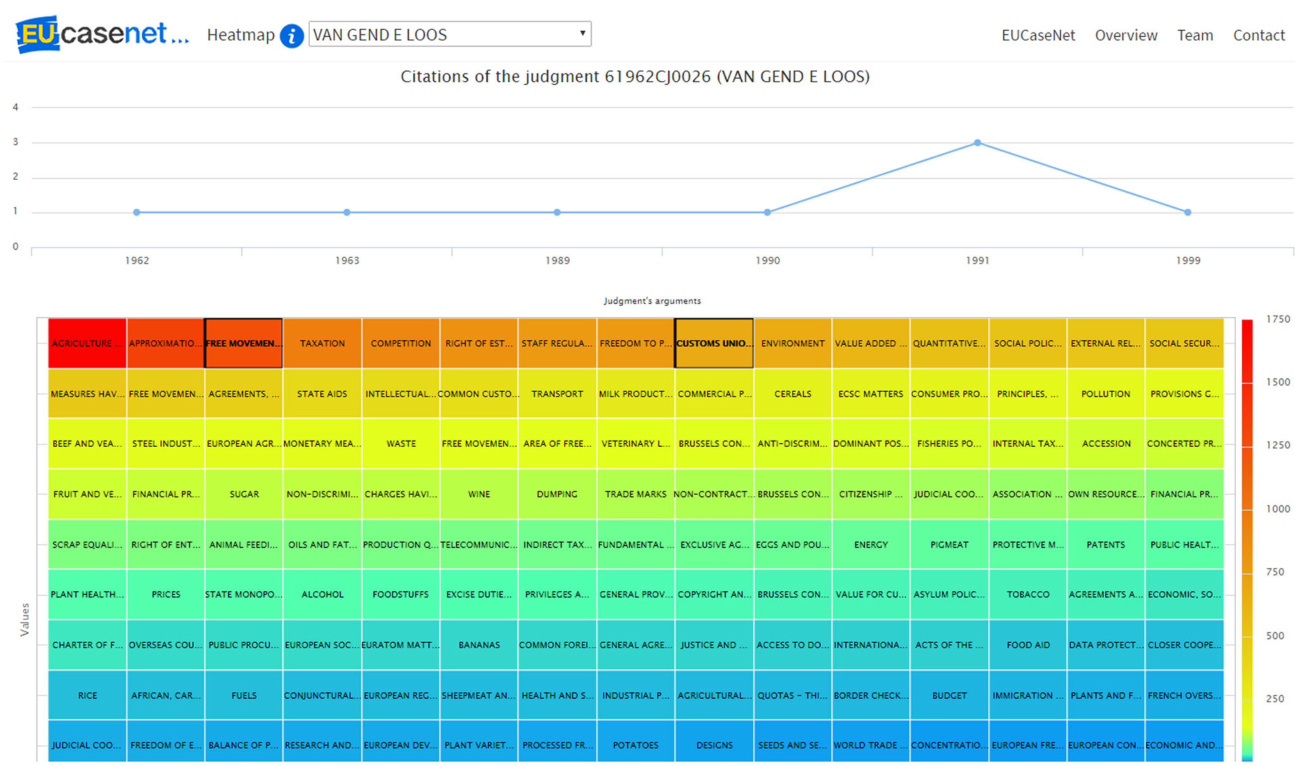


Fig. 6 Heatmap of arguments. The black border of the sectors highlights the specific arguments of the judgment under analysis

Line graph generation. This feature computes and plots the frequency distribution of each argument over the years. It allows the selection of specific arguments of interest, giving an useful insight by offering a method to visually compare their trends and evolution over the years (see Fig. 7).

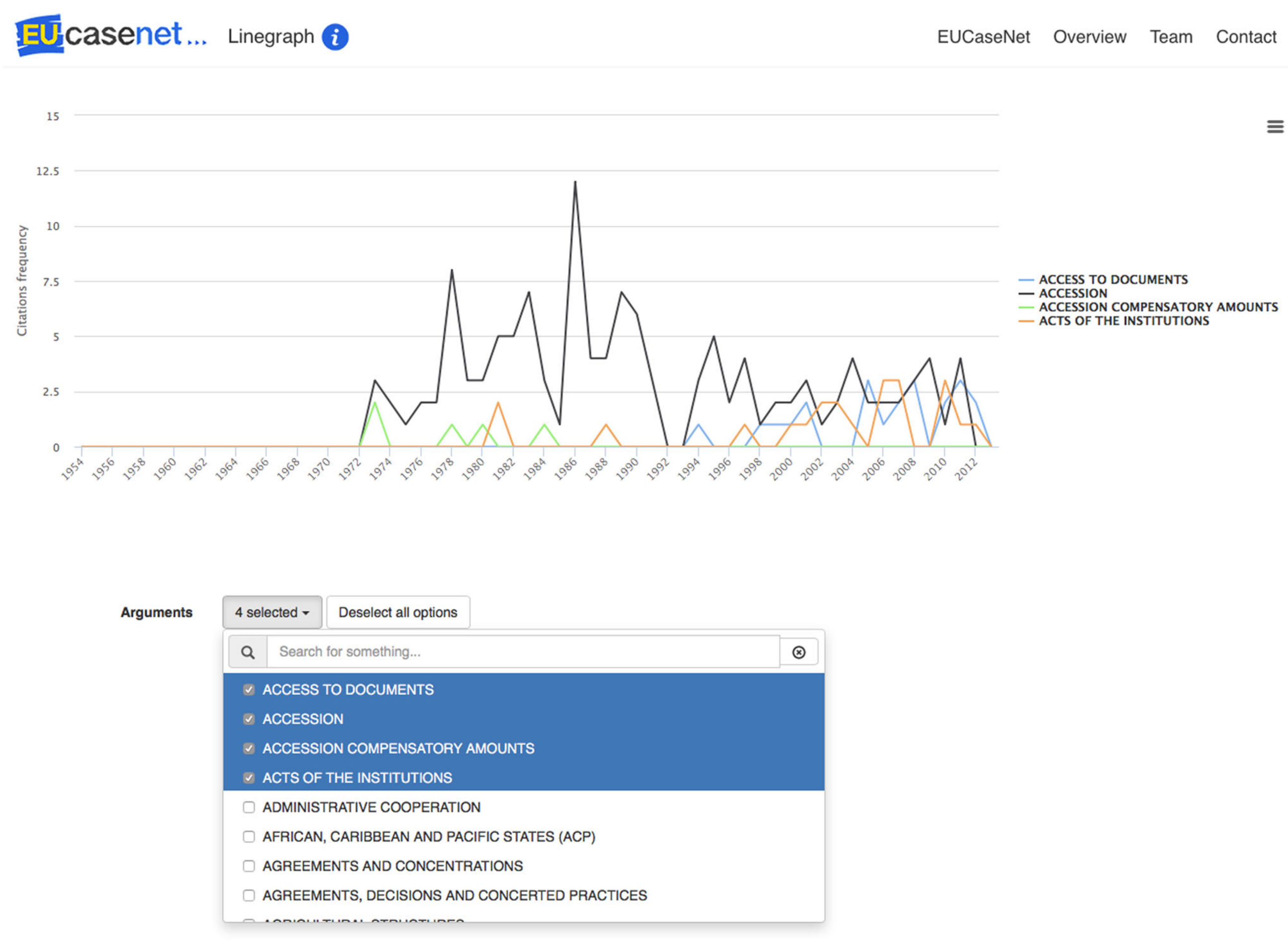


Fig. 7 The evolution of the arguments over the years

A functionality that somehow lies in the middle between the above listed functionalities is the Data analysis feature. Users can access the EU case law judgments in form of table, query it and export the results (see Fig. 8).

4.2 Architecture

In this section, we briefly describe the software architecture of *EUCaseNet* and how components interact each other. As you can see in Fig. 9, *EUCaseNet* is a three-tier architecture, implemented by following a typical Model-View-Controller layer architecture, very common for Web applications. Specifically, the tool is composed of three components: (1) data persistence layer, (2) business layer and (3) user interface layer.

The data persistence and business layers are implemented server-side, through Java Servlet components, within Apache Tomcat.[5] The user interface layer is implemented with commonly used JavaScript libraries. We will discuss these components and all provided functionalities in the following.

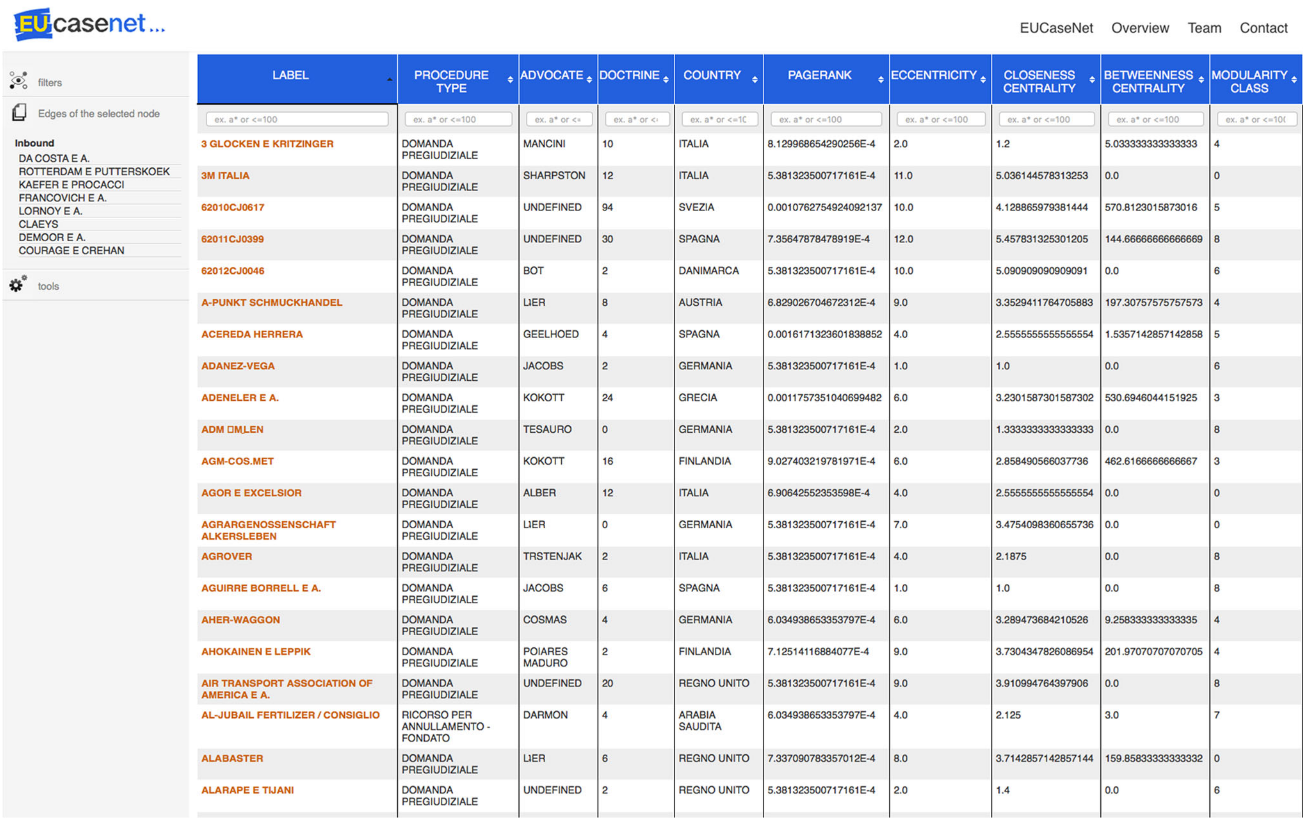
Data persistence layer. It contains all the information about Court of Justice sentences. This data is organized in a MySQL database and in a XML file (gexf) to be computed by Gephi library.[6]

Business layer. It is responsible for the creation, management and manipulation of data, by implementing the functionalities provided by Gephi Library that allows the generation and analysis of the graph. SNA techniques have been implemented in server-side algorithms exploiting Gephi libraries. To enhance the user experience, SNA algorithms have been precomputed and stored in the data persistent layer.

User interface layer. It allows the user to interact with *EUCaseNet* features. The computation of the layout graph algorithms has been implemented here to improve the user experience. As an example, when a user decides to apply a specific graph layout algorithm (i.e., Fruchterman-

5 http://tomcat.apache.org/.

6 http://gephi.org.



LABEL	PROCEDURE TYPE	ADVOCATE	DOCTRINE	COUNTRY	PAGERANK	ECCENTRICITY	CLOSENESS CENTRALITY	BETWEENNESS CENTRALITY	MODULARITY CLASS
ex. a* or <=100	ex. a* or <=100	ex. a* or <=	ex. a* or <=	ex. a* or <=10	ex. a* or <=100	ex. a* or <=100	ex. a* or <=100	ex. a* or <=100	ex. a* or <=10
3 GLOCKEN E KRITZINGER	DOMANDA PREGIUDIZIALE	MANCINI	10	ITALIA	8.129968654290256E-4	2.0	1.2	5.033333333333333	4
3M ITALIA	DOMANDA PREGIUDIZIALE	SHARPSTON	12	ITALIA	5.381323500717161E-4	11.0	5.036144578313253	0.0	0
62010CJ0617	DOMANDA PREGIUDIZIALE	UNDEFINED	94	SVEZIA	0.0010762754924092137	10.0	4.128865979381444	570.8123015873016	5
62011CJ0399	DOMANDA PREGIUDIZIALE	UNDEFINED	30	SPAGNA	7.35647878478919E-4	12.0	5.457831325301205	144.66666666666669	8
62012CJ0046	DOMANDA PREGIUDIZIALE	BOT	2	DANIMARCA	5.381323500717161E-4	10.0	5.090909090909091	0.0	6
A-PUNKT SCHMUCKHANDEL	DOMANDA PREGIUDIZIALE	LIJER	8	AUSTRIA	6.829026704672312E-4	9.0	3.3529411764705883	197.30757575757573	4
ACEREDA HERRERA	DOMANDA PREGIUDIZIALE	GEELHOED	4	SPAGNA	0.0016171323601838852	4.0	2.5555555555555554	1.5357142857142858	5
ADANEZ-VEGA	DOMANDA PREGIUDIZIALE	JACOBS	2	GERMANIA	5.381323500717161E-4	1.0	1.0	0.0	6
ADENELER E A.	DOMANDA PREGIUDIZIALE	KOKOTT	24	GRECIA	0.0011757351040699482	6.0	3.2301587301587302	530.6946044151925	3
ADM □M⎵LEN	DOMANDA PREGIUDIZIALE	TESAURO	0	GERMANIA	5.381323500717161E-4	2.0	1.3333333333333333	0.0	8
AGM-COS.MET	DOMANDA PREGIUDIZIALE	KOKOTT	16	FINLANDIA	9.027403219781971E-4	6.0	2.858490566037736	462.6166666666667	3
AGOR E EXCELSIOR	DOMANDA PREGIUDIZIALE	ALBER	12	ITALIA	6.90642552353598E-4	4.0	2.5555555555555554	0.0	0
AGRARGENOSSENSCHAFT ALKERSLEBEN	DOMANDA PREGIUDIZIALE	LIJER	0	GERMANIA	5.381323500717161E-4	7.0	3.4754098360655736	0.0	0
AGROVER	DOMANDA PREGIUDIZIALE	TRSTENJAK	2	ITALIA	5.381323500717161E-4	4.0	2.1875	0.0	8
AGUIRRE BORRELL E A.	DOMANDA PREGIUDIZIALE	JACOBS	6	SPAGNA	5.381323500717161E-4	1.0	1.0	0.0	8
AHER-WAGGON	DOMANDA PREGIUDIZIALE	COSMAS	4	GERMANIA	6.034938653353797E-4	6.0	3.289473684210526	9.258333333333335	4
AHOKAINEN E LEPPIK	DOMANDA PREGIUDIZIALE	POIARES MADURO	2	FINLANDIA	7.12514116884077E-4	9.0	3.7304347826086954	201.97070707070705	4
AIR TRANSPORT ASSOCIATION OF AMERICA E A.	DOMANDA PREGIUDIZIALE	UNDEFINED	20	REGNO UNITO	5.381323500717161E-4	9.0	3.910994764397906	0.0	8
AL-JUBAIL FERTILIZER / CONSIGLIO	RICORSO PER ANNULLAMENTO - FONDATO	DARMON	4	ARABIA SAUDITA	6.034938653353797E-4	4.0	2.125	3.0	7
ALABASTER	DOMANDA PREGIUDIZIALE	LIJER	6	REGNO UNITO	7.337090783357012E-4	8.0	3.7142857142857144	159.85833333333332	0
ALARAPE E TIJANI	DOMANDA PREGIUDIZIALE	UNDEFINED	2	REGNO UNITO	5.381323500717161E-4	2.0	1.4	0.0	6

Fig. 8 Table view of the whole EU case law judgments

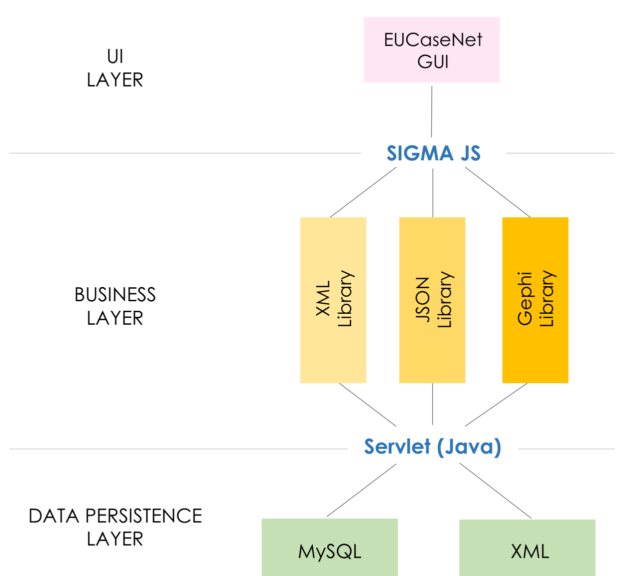


Fig. 9 EUCaseNet architecture overview

Reingold algorithm), he will be able to perceive the node movements and interact with nodes for specific analysis.

A typical interaction, shown in Fig. 10, can be summarized as follows:

- The client-side component (or Front-end), invokes a service through the HTTP protocol using asynchronous requests (AJAX);
- The server-side component (or Back-end) instances a Servlet that provides the requested service and response in JSON format;
- The client-side component receives the response in JSON format and visualizes them into the browser.

5 Experiments and analyses

In this section, we describe the two analyses have been done so far. We first analyze the properties of the whole EU case law corpus, then study the relevance of the judgement of the gold standard.

5.1 Properties of the EU case law corpus

A first part of our study was devoted to investigate the basic structural and functional features of EU legal system. More specifically, we focused on the evolution of the EU case law network over time and on more "structural" issues trying to ascertain the topological nature of the EU

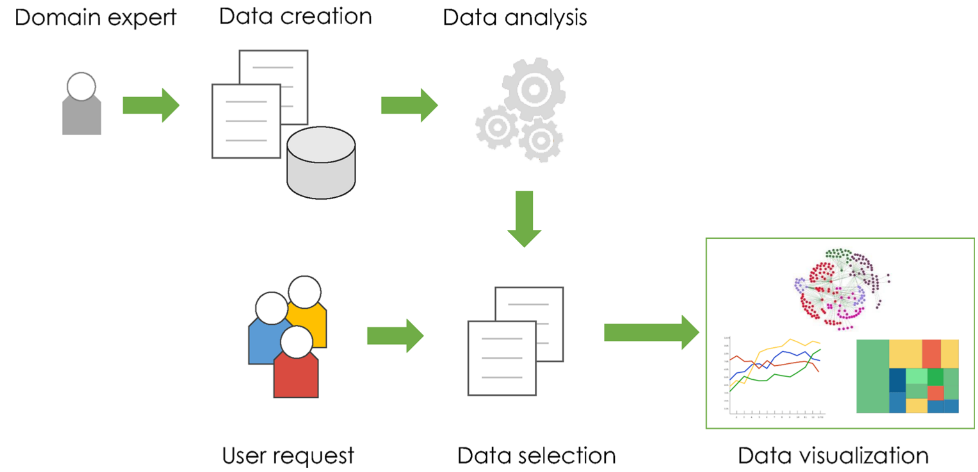


Fig. 10 Application workflow from the data definition to visualization

case law corpus (namely the scale free structure and the small-world topology). As pointed out in a well-established literature in this field (see, among the others Fowler et al. 2007; Koniaris et al. 2015) the mentioned analyses can offer a good starting point for more domain specific research questions studies 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).

Dynamic of the citations over time. Following a trend similar to the one that characterizes the US Supreme Court case law network (Krebs 2002), the citation of EU precedents clearly shows a time-dependent dynamic.

Figure 11 shows the number of outward citations as a function of the year in which citing cases were decided. That is, it portrays the number of citations to EU judgments contained within judgments released in a given year (i.e., outward citations). One obvious feature of the citation of precedent is that, over time, the EU Court has become more inclined to cite a greater number of prior decisions. A similar trend can be found in the number of inward citations mapped in Fig. 11b.

Structure of the case law network. Another relevant outcome achieved over the years by network analysis research is that networks in natural, social and technological systems are not random: The evolution of their structure follows a series of organizing principles that distinguish them from randomly linked networks (Smith 2007). As highlighted in Koniaris et al. (2015), a well-known 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. The feature is common to large scale communication, biological and social systems (Barabâsi et al. 2002) and to the network of US Supreme Court decisions (Fowler and Jeon 2008). A first result of our analysis was the confirmation that also the EU case law network, like the network of the

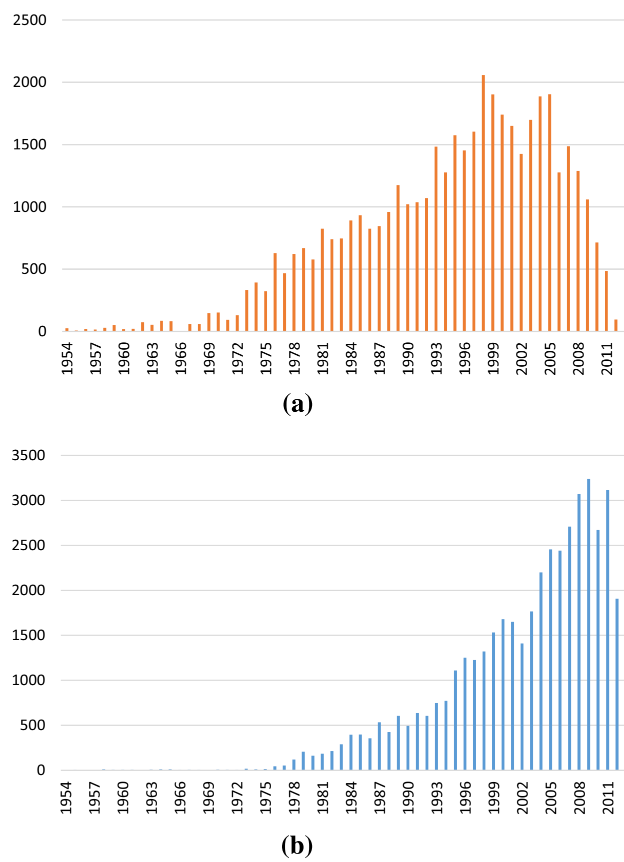
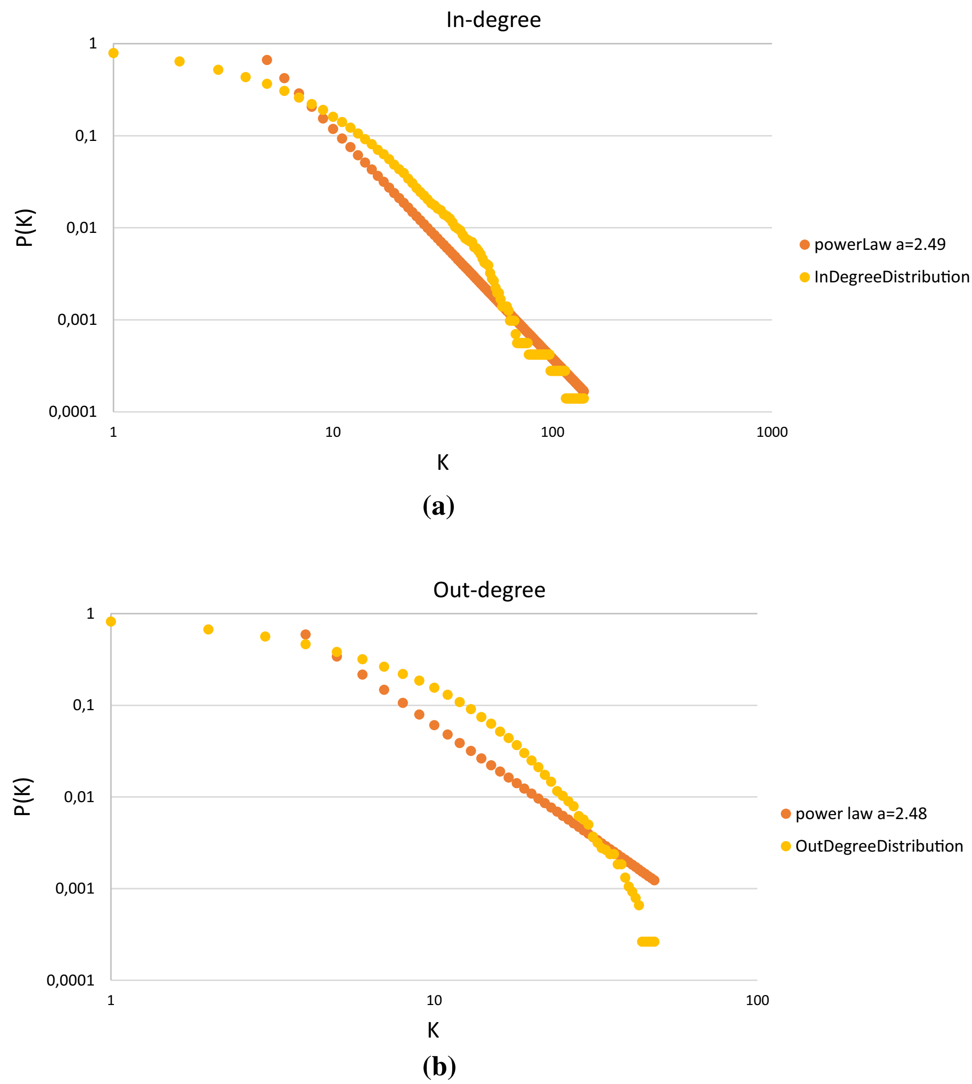


Fig. 11 Number of outward citations in the EU Court of Justice network, 1954–2013. **b** Number of inward citations in EU Court of Justice network, 19542013. **a** In-degree. **b** Out-degree

US Supreme Court decisions, presents a scale free topology.

As shown in Fig. 12, the approximate straight-line form of the distribution function allow us to infer that, in all the examined subgraphs, the Legislation Network follows a power-law distribution both the in-degree and the out-

Fig. 12 The in-degree and out-degree distribution of the judgments of the ECJ. Data are plotted on doubly logarithmic axes via the cumulative distribution. The fitted power-law distributions are also plotted. **a** In-degree. **b** Out-degree



degree distribution follow a power-law. A few judgments have a high number of citations (and, therefore, can be considered highly influential probably serving as reference point in the case law scenario), while a large number of judgment has a very low number of citations.

The small-world topology. We finally focused on the topology of the EU case law network trying to assess if it is a small-world network. To this aim, according to a well-established research practice, it is necessary to compare the values of the Average shortest path (L) and the Average clustering coefficient (C) of candidate network metrics with Erdos-Rényi random networks (Winkels et al. 2014) showing the same number of nodes and edges. As summarized by Humphries et al. (2006):

> When comparing a small-world network s to a random network r of the same T, k value, it is found that $Ls \approx Lr$ and that $Cs \gg r$. [..]. Typically, these relationships are calculated for datasets forming a single network, and so the presence of a small-world topology is decided by inspection of the generated L, C values. As we are generating large numbers of networks, we propose a simple quantitative definition of these relationships. This will allow us to explore the extent to which the small-world topology changes with parameter variation: In other words, it is a measurement of small-world-ness.

We considered both network as an undirected graph for accuracy purpose. Table 1 summarizes the results of our analysis: judgments of the ECJ network satisfy the small-world conditions.

Results show that the average shortest path of the ECJ network is comparable to average shortest path of random network and the average clustering coefficient is significantly larger than average cluster coefficient of the corresponding random model. We can trace back this finding to the nature of law and its hierarchical form. Legal

Table 1 EU case law small-world metrics

	Average shortest path	Average clustering coefficient
EUCaseNet	4.751	2.772
RandomNet	4.407	0.001

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.

5.2 Analyzing a gold standard

After the set of analyses described in the previous section, we decided to start facing the problem of the determination of the relevance of judgments in the specific context of EU case law. Being aware of the "multi-dimensionality" of the concept of relevance (and of the fact that its different profiles partly overlap and are often indicated by similar clues), we decided to focus on other analyses

After the set of analyses described in the previous section, we decided to start facing the problem of the determination of the relevance of judgments in the specific context of EU case law. Being aware of the "multi-dimensionality" of the concept of relevance (and of the fact that its different profiles partly overlap and are often indicated by similar clues), we decided to focus on other analyses

Relevance of precedents
Taking cue from this definition, the volume identifies a set of 19 judgments that we used as gold standard for our experiments. More in detail, keeping the cases listed in the volume as a reference, we used *EUCaseNet* to measure the values assumed, for each case, by attributes and measures typically considered in NA & Law field: *(1) Citation by legal literature* (as they result from EUR-Lex database); *(2) Centrality measures* (In-degree, Out-degree, Closeness, Betweenness, and Eigenvector); *(3) Eccentricity*; *(4) Page Rank*. Our goal was twofold: On the one hand, start testing the functionality of the *EUCaseNet* toolkit; on the other hand, like other authors do (see, e.g., Malmgren 2011) start verifying, even if in an initially "rough" way, the performance of each measure in mapping the relevance of a case as defined in our gold standard. Table 2 shows the judgments and the values assumed by the measures and the attributes taken into account for each of them. Table 3 provides an overview of the performance of each of the considered parameters by comparing the relevance assigned to the 19 judgments by the gold standard (in our analysis these judgments can be considered as equally relevant) and the position that the same 19 judgments occupy in a ranking based on NA measures values.

As our preliminary results clearly show, when applied to a "simple" citation network, traditional NA measures have poor performances. Incongruities arise, just to give an example, if we consider the rankings assigned by the

Table 2 Measures and attributes of the case study judgment

Judgment	Doctrine	In-degree	Out-degree	Eccentricity	Closeness	Betweenness	PageRank
VAN GEND E LOOS	46	8	0	0	0	0	0.00049
COSTA	14	23	0	0	0	0	0.00111
INTERNATIONALE HANDELSGESELL	2	19	0	0	0	0	0.00120
COMMISSIONE/CONSIGLIO	20	37	0	0	0	0	0.00121
NOLD/COMMISSIONE	3	13	0	0	0	0	0.00072
DASSONVILLE	9	117	0	0	0	0	0.0020
DEFRENNE	17	48	0	0	0	0	0.0018
SIMMENTHAL	35	49	1	2	1.666	3315.4763	0.00075
REWE-ZENTRAL	23	70	0	0	0	0	0.00194
CILFIT E A.	19	22	1	2	1.5	1506.7138	0.00031
LES VERTS/PARLAMENTO	11	14	2	1	1.0	2921.9004	0.00066
FOTO-FROST	13	20	2	2	1.777	9078.7165	0.00046
WACHAUF	3	20	1	2	1.666	2554.2996	0.00040
ERT	8	42	10	8	3.352	54365.1861	0.00069
FRANCOVICH E A.	74	55	14	7	3.4431	54442.0397	0.00093
BOSMAN	82	140	21	10	4.0945	272540.6123	0.00139
MARTiNEZ SALA	14	36	0	0	0	0	0.00026
COMMISSIONE/REGNO UNITO	16	1	12	9	4.8045	0	0.00006
BAUMBAST E R	10	22	11	12	5.7630	26859.8971	0.00018

Table 3 The cases in which the judgment stands in one of the top 20 positions of the ranking calculated using the measure considered are highlighted in italics

	Judgment ranking							
	Doctrine (max: 52)	In-degree (max: 67)	Out-degree (max: 44)	Eccentricity (max: 22)	Closeness (max: 5301)	Betweenness	PageRank	Eigenvector centrality
VAN GEND E LOOS	*16*	59	44	22	5301	34	939	614
COSTA	39	44	44	22	5301	34	35	203
INTERNATIONALE HANDELSGESELLSCHAFT	50	48	44	22	5301	34	33	515
COMMISSIONE/CONSIGLIO	32	30	44	22	5301	34	31	256
NOLD/COMMISSIONE	49	54	44	22	5301	34	443	288
DASSONVILLE	45	*2*	44	22	5301	34	*6*	*6*
DEFRENNE	36	19	44	22	5301	34	*12*	33
SIMMENTHAL	*18*	*18*	43	20	5296	1557	386	44
REWE-ZENTRAL	32	6	44	22	5301	34	*10*	*13*
CILFIT E A.	34	45	43	20	33	2265	1360	209
LES VERTS/PARLAMENTO	41	53	42	21	20	1699	551	763
FOTO-FROST	39	47	42	20	*18*	81	1043	263
WACHAUF	49	47	43	20	5296	1832	1193	234
ERT	44	25	34	*14*	380	882	502	35
FRANCOVICH E A.	*2*	*12*	30	*15*	415	884	95	*15*
BOSMAN	*1*	*1*	23	*12*	655	*2*	23	*1*
MARTiNEZ SALA	39	31	44	*22*	5301	34	1444	174
COMMISSIONE/REGNO UNITO	36	66	32	*13*	1330	183	614	2679
BAUMBAST E R	42	45	33	*10*	3621	1791	1599	395

measures and the parameters listed in Table 3 to the Van Gend and Loos judgment: the most relevant precedent in EU case law—according to the Court itself [http://curia.europa.eu/jcms/jcms/P_95693/]—shows very low value of In-Degree centrality and 0 in all the other measures, although it collects a relevant number of citations from legal doctrine. A first explanation of this result could be related, among the other reasons, to the problem of "obliteration through incorporation" (Garfield and Merton 1979) occurring when an idea is considered so influential to make unnecessary to explicitly cite it. This kind of problem is frequent when citation analysis is the only method used for identifying the relevance over long run, since it allow to detect only the explicit links (i.e., an express reference), without any consideration for those just implicitly contained in a document (e.g., the case reinterpreted as general principles of law within an argumentative path). It is clear, in this vein, the importance to better reflect on the ontology of the phenomenon investigated (e.g., relevance) i.e., on its concrete expressions (data, relations) and on the algorithms more suitable to identify it.

6 Conclusion

The experiments and, more in general, the overall experience so far gained allow to draw some considerations both on a scientific and application level. From the first point of view, the feeling is that, even if very preliminary, the project suggests the opportunities potentially hidden in the adoption of computational approaches to the study of legal matters. The quantitative analysis of legal data promises not only to insightfully complement traditional legal studies, but also to provide points of departure for more advanced and effective tools to be used both for scientific and information retrieval purposes.

Our work, is clear, is just at the beginning: There is a huge series of scientific and technical challenges involving different expertise from computer science and linguistics etc., to be faced if we to make this kind of tools and approaches capable to give concrete answers to real world problems. The active contribution of legal experts, obviously, will be essential if we want to understand how NA, VA and, more on general, computational approaches can shed new light in our understanding of legal phenomena.

Moving a bit more on to the substance of the research, we can state that even if very preliminary, our results show that Network analysis measures offer disappointing performances when applied on a citation network based. Inconsistencies clearly arise, for example, if one considers the rankings assigned by the measures and the parameters listed in Table 2 to Van Gend and Loos, probably one of the most relevant judgments in ECJ case law. The ECJ pronouncement shows indeed a very low value of In-Degree centrality and 0 in all the other measures, although it collects a significant number of citations from legal doctrine. The reasons of this can be different. A first explanation of the result could be related, among the other things, to the so called "obliteration through incorporation" (Garfield and Merton 1979), a phenomenon occurring when an idea or a document is considered so influential to make unnecessary to explicitly cite it. This kind of problem is frequent when citation analysis is the only method used for identifying the relevance over long run, since it allow to detect only the explicit links (i.e., an express reference), without any consideration for those just implicitly contained in a document (e.g., the case reinterpreted as general principles of law within an argumentative path). In the light of this, it is clearly evident the need to better investigate the phenomenon (relevance of case law), its concrete expressions (data, relations) and the algorithms more suitable to identify it.

From an application standpoint, we might be said to have realized a tool that, even leaving large room for improvements, allows domain experts to make the first basic experiments with EU case law data.

As to the future developments, we are planning to invite domain experts to use the system involving them in the validation of both the tool and the methodology. To this aim, the features of the tool will be extended in order also to allow the analysis of different legal sources. In the same perspective, we are also planning to explore not only the use of more efficient graph database management systems (Neo4j,[7] OrientDB[8]), but also machine learning techniques to datasets generated from social network analysis.

Another crucial essential aspect will be the definition a more and more refined operational/computational definitions of legal research questions (starting from the concept of relevance) to be addressed by means of visualization and NA techniques identifying also the data sources more suitable to this aim. Going down this route, we will start investigating on other suitable measures, e.g., drawing inspiration by the techniques which are being developed in multimodal network analysis area.

[7] http://neo4j.com.

[8] http://orientdb.com.

Acknowledgments This paper stems from a talk given at the ARS'15 International Workshop "Large Networks And Big Data: New Methodological Challenges" held in Capri on April 29–30, 2015. Authors would like to thank Sebastiano Faro for the useful contribution, suggestions given about the topic discussed in this paper.

References

Barabâsi AL, Jeong H, Néda Z, Ravasz E, Schubert A, Vicsek T (2002) Evolution of the social network of scientific collaborations. Phys A Stat Mech Its Appl 311(3):590–614

Brownsword R (2015) In the year 2061: from law to technological management. Law Innov Technol 7(1):1–51

Calderoni F (2010) Morselli, Carlo: inside criminal networks. Eur J Crim Policy Res 16(1):69–70

Cook KA, Thomas JJ (2005) Illuminating the path: the research and development agenda for visual analytics. Tech. rep., Pacific Northwest National Laboratory (PNNL), Richland, WA, USA

Duarte EF, Oliveira Jr E, Côgo FR, Pereira R (2015) Dico: a conceptual model to support the design and evaluation of advanced search features for exploratory search. In: Human–computer interaction. Springer, Berlin, pp 87–104

Ferrara E, De Meo P, Catanese S, Fiumara G (2014) Detecting criminal organizations in mobile phone networks. Expert Syst Appl 41(13):5733–5750

Fowler JH, Jeon S (2008) The authority of supreme court precedent. Soc Netw 30(1):16–30

Fowler JH, Johnson TR, Spriggs JF, Jeon S, Wahlbeck PJ (2007) Network analysis and the law: measuring the legal importance of precedents at the us supreme court. Polit Anal 15(3):324–346

Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. Softw Pract Exp 21(11):1129–1164

Garfield E, Merton RK (1979) Citation indexing: its theory and application in science, technology, and humanities, vol 8. Wiley, New York

Golbeck J (2013) Analyzing the social web. Newnes

Humphries MD, Gurney K, Prescott TJ (2006) The brainstem reticular formation is a small-world, not scale-free, network. Proc R Soc Lond B Biol Sci 273(1585):503–511

Hutchins CE, Benham-Hutchins M (2010) Hiding in plain sight: criminal network analysis. Comput Math Organ Theory 16(1):89–111

Jacomy M, Venturini T, Heymann S, Bastian M (2014) ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PLoS One 9(6):1–12

Katz DM, Bommarito MJ II (2014) Measuring the complexity of the law: the United States code. Artif Intell Law 22(4):337–374

Keim DA, Kohlhammer J, Ellis G, Mansmann F (2010) Mastering the information age-solving problems with visual analytics. Florian Mansmann

Kleinberg JM (1999) Hubs, authorities, and communities. ACM Comput Surv (CSUR) 31(4es):5

Koniaris M, Anagnostopoulos I, Vassiliou Y (2015) Network analysis in the legal domain: a complex model for european union legal sources. arXiv preprint arXiv:1501.05237

Krebs VE (2002) Mapping networks of terrorist cells. Connections 24(3):43–52

Lettieri N, Vicidomini V, Altamura A, Faro S (2014) Nets of legal information connecting and displaying heterogeneous legal sources. In: International workshop on network analysis in law. In conjunction with Jurix'14, 27th international conference on legal knowledge and information systems

Lettieri N, Malandrino D, Vicidomini L (2016) By investigation, I mean computation. A framework to investigate the societal

dimension of crime. Trends Organ Crime. doi:10.1007/s12117-016-9284-1
Malmgren S (2011) Toward a theory of jurisprudential relevance ranking. Using link analysis on EU case law, Graduate thesis, Stockholm University 1
Marchionini G (2006) Exploratory search: from finding to understanding. Commun ACM 49(4):41–46
Memon N, Larsen HL (2006) Notice of violation of ieee publication principles practical approaches for analysis, visualization and destabilizing terrorist networks. In: First international conference on availability, reliability and security (ARES'06). IEEE, p 8
Post DG, Eisen MB (2000) How long is the coastline of law? Thoughts on the fractal nature of legal systems. J Legal Stud 29:545
Rothenberg R (2001) From whole cloth: making up the terrorist network. New York Times, New York
Smith TA (2007) Web of law, the. San Diego L Rev 44:309
Stewart T (2001) Six degrees of Mohamed Atta. Business 2(2):10
Winkels R, Boer A (2013) Finding and visualizing context in dutch legislation. In: Proceedings of NAiL, vol 2013
Winkels R, Lettieri N, Faro S (2014) Network analysis in law. Edizioni Scientifiche Italiane
Wong PC, Thomas J (2004) Visual analytics. IEEE Comput Graph Appl 24(5):20–21
Xu J, Chen H (2005a) Criminal network analysis and visualization. Commun ACM 48(6):100–107
Xu JJ, Chen H (2005b) Crimenet explorer: a framework for criminal network knowledge discovery. ACM Trans Inf Syst (TOIS) 23(2):201–226