Network Analysis to Support Navigation and Use of Ever-Changing Legislation

GUIDO BOELLA [a] LUIGI DI CARO [a] LIVIO ROBALDO [a] and ANDREA VIOLATO [b]

[a] *Department of Computer Science, University of Turin*
[b] *Nomotika s.r.l.*

1. Introduction

In this abstract we introduce the working context related to the understanding of an heterogeneous network of references contained in the Italian regulatory framework. We then present an extended analysis of a large network of laws, providing several types of analytical evaluation that can be used within a legal management system for understanding the data through summarization, visualization, and browsing.

The importance of Information Technology tools lies on the ability of limiting the well-known complexity of the regulatory framework. Starting from the language used to express concepts and rules that often create serious problems of interpretation, *scientia iuris* is full of innumerable aspects that make it difficult to manage by domain experts. The use of computer systems can reduce this complexity for those who have to do with a legal reality that results to be more and more interdisciplinary, international and multifunctional. All this is even more true by taking into account one of the most complex aspects of any modern legal text, which is the extensive use of *references*. It is therefore the intention of this paper to first investigate the need for the development of tools capable of assisting the employee to manage the particular regulatory complexity characterized by cross-references, and then to simulate possible technological solutions.

The presence of thousands of stratifications in the Italian legal sources over the years has enabled the strengthening of the use of cross-references between legal texts in order to complete the content. This scenario is even more complex when considering a second and unique aspect of this discussion: the temporal dynamism that characterizes the use of normative references in any legal context. In fact, it is not a rarity that, over the years, the legislature put hand to the subjects adapting the contents to the stimuli given by socio-economic factors and the surrounding cultural context. This natural evolution is often mediated through a surgical use of references which permit to individuate the precise points under modification. In the Italian scenario, this assumption is far from being uncommon. Many, in fact, are the examples of legal decrees subsequently converted (even with modifications) into laws. Finally, it is worth noting how the use of normative references can also take place outside the purely legislative context. In fact, it is quite common to find a citation of a law inside a legal text that is not used with explicit references, but also by case law of the courts, which, in exercising a peculiar hermeneutical

activity of this rule, it implicitly complements the content, clarifying the more conceptual and obscure aspects and thus bringing important suggestions to the reader.

In [2] we introduced the Eunomos software, which is being developed in the context of the ICT4LAW project[1]. Eunomos is an advanced legal document management system based on legislative XML representation of laws which are retrieved automatically from institutional legislative portals, and incorporates a tool for building legal ontologies called Legal Taxonomy Syllabus [1].

2. Techniques and Applications

In legislation, all the data can be also viewed as complex networks where nodes are laws and links represent kinds of relationship like "*modification*", "*implementation*", "*substitution*", and so on. These information are often complex to treat, organize, and use since they are many and continuosly changing over time. Manual intervention, indeed, is often required in classical tasks because of the difficulty to get the "big picture", that is, to have an at-a-glance overview over the data under evaluation.

Social Network Analysis is quite a new field that inherits methods from Phisics (i.e., Complex Networks) and Mathematics (i.e., Theory of Graphs) to face problems related to the huge amount of data coming from social networks like Twitter[2], Facebook[3], Flickr[4], and so forth. These algorithms are useful to capture statistics about the type of the networks along their properties. More in detail, it is possible to analyze a network in terms of its evolution over time, important nodes, implicit relationships, etc. All these evaluations can be helpful also in a network of laws. [3] presents a large overview over the mathematical properties of graphs. In this section we present a set of analyses that may help the jurist to deal with these networks.

The first statistical analysis of a graph is given by the distribution of the number of neighbours of a node, also called *degree*. In real networks, the degree distribution has a tail that often follows a power law, that means that it contains many nodes with low degree and some node with large degree. This characteristic of real networks is called *community structure*. Such communities are also commonly named clusters, and they represent nodes that play similar roles within the graph. A common and often useful analysis of a network looks at such latent community structures, and it is based on some clustering approach. In our test we made use of the concept of *edge betweenness* of an edge, i.e., the extent to which it lies along shortest paths between all pairs of nodes in the network. This algorithm works by iteratively following two steps: computing the edge betweenness for all edges in the current graph and then removing the edge with the highest betweenness value. This analysis helps finding communities, since it iteratively removes central nodes to separate the graph in distinct subnetworks.

Relationships between nodes of a network may have a precise direction, that needs to be taken into account to understand the system as a whole. PageRank [4] is an algorithm

[1] ICT4LAW: ICT Converging on Law: Next Generation Services for Citizens, Enterprises, Public Administration and Policymakers funded by Regione Piemonte 2008-2013, call Converging Technologies 2007, website: http://www.ict4law.org

[2] http://www.twitter.com/

[3] http://www.facebook.com/

[4] http://www.flickr.com/

that assigns a weight to each node of a network in the World Wide Web domain, with the purpose of quantifying its relative importance within it. In spite of its original use on hyperlinks, it can be useful for several other domains dealing with directed graphs. Generally speaking, in the legal domain, if one law modifies a law B, one usually does not find on B a link back to A. In some case, there can be few relationships that may be reciprocal (i.e., citations). The PageRank algorithm measures the importance of a node by considering that of the nodes that link to it. Thus, this indicates a finer node evaluation with respect to the edge betweenness. An accurate analysis of the most important nodes within a network of laws can be helpful to understand and use the data.

The *diameter* of a graph can give useful insights. In detail, it is the largest number of nodes which must be traversed in order to travel from one node to another. In a network of laws it can be used to estimate the maximum complexity of modification/citation paths. Since a network of law is often disconnected, it is constituted by many subnetworks. In this case, the diameter considers the maximum distance found in all subnetworks.

3. Data and Results

In this section we present an analysis of the data coming from Eunomos, i.e., a network of around 10K laws interconnected by 7K links. The dataset contains different types of laws, as for instance "*stato:legge*", that indicates the principal type of law of Italy. The resulting distribution of the degree levels demonstrated that most of the laws have few connections with other laws, whereas some other laws have a large connectivity within the graph. This is in line with the majority of the networks automatic systems have to deal with.

The diameter of the network (actually, the maximum diameter of all the disconnected subnetworks in it) is 8. This means that, in the worst case, if a jurist has to navigate and understand a path between two laws in the network, it could go through other eight laws. Still, links can have different meanings, so the process may include multiple cateogories of operations. Having a tool that can support this navigation may represent a dramatically help in such process.

Finally, the distribution of the PageRank scores over the nodes/laws of the network gives an interesting result since it seems to perfectly split the important nodes (also called hubs) from the others. A system that makes use of these scores can filter out unimportant information rather than letting emerge crucial points within the regulatory framework.

References

[1] Gianmaria Ajani, Guido Boella, Leonardo Lesmo, Alessandro Mazzei, Daniele P Radicioni, and Piercarlo Rossi. Legal taxonomy syllabus: Handling multilevel legal ontologies. *Proceedings of Langtech 2008*, 2008.

[2] G. Boella, L. Humphreys, M. Martin, P. Rossi, and L. van der Torre. Eunomos, a legal document management system based on legislative XML and ontologies (position paper). *Applying Human Language Technology to the Law*, page 41, 2011.

[3] Béla Bollobás. *Modern graph theory*, volume 184. Springer, 1998.

[4] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The pagerank citation ranking: bringing order to the web. 1999.