

The Italian Civil Code Network Analysis

Lucio La Cava¹, Andrea Simeri¹ and Andrea Tagarelli¹

¹Dept. Computer Engineering, Modeling, Electronics, and Systems Engineering (DIMES),
University of Calabria, 87036 Rende (CS), Italy

Abstract

The regulation of private law is a focal element in the metamorphosis of society, and by gaining a broader vision of the legal domain we might grasp novel or potentially hidden nuances and characteristics. In this work, we explore the Italian Civil Code (ICC) from an unprecedented perspective based on network analysis. We develop a text processing method to identify and extract article references from the ICC and, upon these, we define network models capturing their relation structures either at a book and corpus scale. The exploitation of the main structural features of these networks leads us to unveil meaningful patterns, holding within and across the books composing the ICC. Furthermore, by leveraging a community detection task, we investigate whether the formation of a community is related to the topic coherence of its assigned articles over the portions of books involved. Our findings reveal useful indicators that may help legal experts and practitioners enhance their knowledge from a novel perspective provided by the network of article references through the ICC.

Keywords

civil law, law article citation networks, structural properties of networks, community detection

1. Introduction

The Italian Civil Code, hereinafter referred to as ICC, is the legislation source containing norms that regulate private law in Italy. Enacted by Royal decree no. 262 of March 16, 1942, the ICC has been involved in a perpetual process of refinements and enhancements, and subjected to numerous reorganizations to stay updated with respect to legislative needs and social development.

The ICC is compiled as an organic corpus relating to the fundamental and constitutional civil laws. In addition to the organization into various books and their sections, the corpus and its constituent articles have an additional structure that is described by *cross-article references*. These are citations that may occur in the content of an article to refer to one or more articles, from either the same or different books, and hence they are exploited by legislators to clarify the scope and the semantics of specific articles. The ICC article references can naturally be modeled as a network, so to enable the discovery and analysis of relation patterns among the article contents beyond the original, logical organization of the corpus.

Research in artificial intelligence and law has traditionally focused on a number of problems that are typically addressed by natural language processing and machine learning models and methods. Despite a relative gap with respect to the network science discipline, a few works

RELATED - Relations in the Legal Domain Workshop, in conjunction with ICAIL 2021, June 25, 2021, São Paulo, Brazil

✉ lucio.lacava@dimes.unical.it (L. La Cava); andrea.simeri@dimes.unical.it (A. Simeri);

andrea.tagarelli@unical.it (A. Tagarelli)



© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

have been proposed to investigate the complexity in a legal corpus domain by leveraging network analysis methods. For instance, Fowler et al. [1] use a network-based representation to characterize the most important precedents at the U.S. Supreme Court. Moser et al. [2] study the structural properties of the citation networks inferred from Austrian Supreme Court decisions. Koniaris et al. [3] model the legislation network of the European Union law, and explore its topological structure and evolution over time, also evaluating its resilience. Mazzega et al. [4] study the network of French Legal codes. Moreover, Zhang et al. [5] propose a software tool to visually explore semantic-based citation networks to ease the analysis of the relations and the evolution of legal issues.

To the best of our knowledge, the Italian Civil Code has not been studied through its article-reference-based structure so far. Therefore, our main goal in this work is to contribute with a study of the networks that can be inferred from the ICC corpus based on article references. This allows us to extend our scope beyond the canonical exploration boundaries of the legal domain, by providing new insights on the interpretation and evaluation of the Italian civil law.

Our main contributions are summarized as follows:

- The ICC articles are not commonly available as hypertexts. Therefore, we develop a text processing method to identify and extract article references that are valid w.r.t. the ICC, hence discarding references to legal sources that are external to the ICC.
- Based on the article-reference relation, we define two network models by differentiating at level of single book or entire ICC corpus.
- We perform an analysis of the book-induced and corpus networks in order to unveil main structural features of such networks and interesting patterns underlying the article references within and across the books of the ICC. Our analysis of the networks is twofold, as we consider macro-scale and meso-scale properties of the networks.
- Upon the outcomes of our performed task of community detection, we also investigate on relations between the formation of communities and the topic coherence within and across books of the ICC. Our findings reveal useful indicators that may help legal experts and practitioners integrate their knowledge from a novel perspective that lays on a unifying, mesoscopic organization of the network of article relations spanning over the whole ICC.

The remainder of the paper is organized as follows. Section 2 introduces the ICC corpus, describes our process of extraction of valid article references from the ICC, and presents our defined network models. Section 3 contains our analysis of the structural characteristics of the ICC networks. Finally, Section 4 concludes the paper.

2. Data Preparation and Models

The Italian Civil Code (ICC) is divided into six, logically coherent books, each in charge of providing rules for a particular civil law theme:

- *Book-1*, on Persons and the Family (articles 1-455) — contains the discipline of the juridical capacity of persons, of the rights of the personality, of collective organizations, of the family;

- *Book-2*, on Successions (articles 456-809) — contains the discipline of succession due to death and the donation contract;
- *Book-3*, on Property (articles 810-1172) — contains the discipline of ownership and other real rights;
- *Book-4*, on Obligations (articles 1173-2059) — contains the discipline of obligations and their sources, that is mainly of contracts and illicit facts (the so-called civil liability);
- *Book-5*, on Labor (articles 2060-2642) — contains the discipline of the company in general, of subordinate and self-employed work, of profit-making companies and of competition;
- *Book-6*, on the Protection of Rights (articles 2643-2969) — contains the discipline of the transcription, of the proofs, of the debtor’s financial liability and of the causes of preemption, of the prescription.

The articles of each book are internally organized into a hierarchical structure based on four levels of division, namely (from top to bottom in the hierarchy): “titoli” (i.e., chapters), “capi” (i.e., subchapters), “sezioni” (i.e., sections), and “paragrafi” (i.e., paragraphs). It should however be emphasized that this hierarchical classification was not meant as a crisp, ground-truth organization of the articles’ contents: indeed, the topical boundaries of contiguous chapters and subchapters are often quite smooth, as articles in the same group often not only vary in length but can also provide dispositions that are more related to articles in other groups.

The ICC currently in force consists of 2 969 article numbers, which actually corresponds to 3 225 articles considering all variants and subsequent insertions. However, during its history, the ICC was revised several times and subjected to repealings, i.e., per-article partial or total insertions, modifications and removals; to date, 2 294 articles have been repealed.

2.1. Extraction of article references

An article in a specific book of the ICC may contain citations of one or more articles, which are from the same or a different book. We will refer to these citations as *article references*.

Unfortunately, identifying and extracting article references from ICC articles is not straightforward, because of two main reasons:

1. The ICC and its books are not designed as hypertexts, neither any index structure containing article references is originally available.
2. An article may also contain references to laws or legal items that do not correspond to ICC articles.

Therefore, since our goal is to infer citation networks from the lists of article references that are contained in each ICC book, we developed an approach to identify and extract *valid* article references, i.e., citations of articles within the ICC only. In the following, we elaborate on the text pre-processing steps that were carried out for the task at hand.

The ICC is obviously publicly available, in various digital formats. From one of such sources, we extracted the contents of each article from all books. Note that we normalized all variants and abbreviations of frequent keywords such as “articolo” (i.e., article), “decreto legislativo” (i.e., legislative decree), “Gazzetta Ufficiale” (i.e., Official Gazette), and finally we lowercased all letters.

An article reference is comprised of three parts:

- *Prefix*, which corresponds to a common root for all lexical variants of article; typically, the abbreviation “*art*” is used.
- *Article id*, which follows the numerical intervals specific to each book (as described earlier in this section).
- *Variant suffix*, which is optional and used to designate a variant of a given article, which is however not alternative, and hence must be treated as a separate article in the book; specifically, an article variant suffix corresponds to a Latin adverbial numeral, to express the multiplicity of occurrence, and hence subsequent versions of a given article id, i.e., “*bis*” (stands for “twice”), “*ter*” (“three times”), “*quater*” (“four times”), and so on.

As mentioned before, in addition to references to other articles in the ICC, an article may contain references to legal items that are external to the ICC corpus. Since they appear in similar textual format, it is not trivial to distinguish between the two types of references. Nonetheless, in the article contents, we can recognize *cue-words* to exploit as hints for the presence or not of references of either type:

- *Relevant* cue-words are used to characterize the presence of valid article references; for instance, “*codice civile*” (“civil code”) and its lexical variants.
- *Irrelevant* cue-words are instead used to locate references to external sources. These include “*legge*” (“law”), “*decreto legislativo*” (“legislative decree”), “*sentenza*” (“judgment”), and are usually followed by a date in some format. Moreover, they also include other words, such as “*comma*” (“paragraph”), which may follow a reference inside a portion of the context delimited by round brackets.

It should be noted that relevant cue-words are much fewer and less frequently used than irrelevant cue-words. Moreover, there are cases in which both types of cue-words are found within an article content, also with multiple occurrences of one or both types, while in other cases they are both missing. For example, the following is an excerpt from article 42-bis that contain multiple valid references:

”[...] si applicano inoltre art2499, art2500, art2500-bis, art2500-ter, secondo comma art2500-quinquies, art2500-nonies, in quanto compatibili. [...]”

By contrast, in the next example from article 86, references are to external sources only, and hence must be discarded from our analysis:

”[...] la legge 20 maggio 2016, 76 ha disposto (con art1, comma 35) che la presente modifica acquista efficacia dal 5 giugno 2016. [...]”

To process the article contents for the task of article reference extraction, we first segmented an article’s text into shorter passages, dubbed contexts, that are delimited by “.” or “;” and preceded by an alphanumerical sequence of at least four characters; the latter constraint is needed to avoid selecting false contexts, e.g., corresponding to one of the many abbreviations that are found in the ICC articles such as “d.p.r.”, “c.p.c.”, “c.p.p.”, “etc.”.

Algorithm 1: Extraction of article references

Data: Set of articles \mathcal{A} in the ICC; list of relevant cue-words $art\text{-}refs_words$ and list of irrelevant cue-words $ext\text{-}ref_words$

Output: Article reference sets AR

$AR \leftarrow \{\}$

foreach $a \in \mathcal{A}$ **do**

$AR_a \leftarrow \{\}$

$contexts \leftarrow getContexts(a)$

foreach $c \in contexts$ **do**

$isIrrelevant \leftarrow search(ext\text{-}refs_words, c)$ **and not** $search(art\text{-}refs_words, c)$

if not $isIrrelevant$ **then**

$\langle a, refs \rangle \leftarrow findAll("art[0-9]+[a-z]*", c)$

$AR_a \leftarrow AR_a \cup \{\langle a, refs \rangle\}$

end

end

$AR \leftarrow AR \cup AR_a$

end

Algorithm 1 sketches our designed procedure to identify and extract article references from the text of ICC articles. The procedure takes in input two lists of words, which are used to find indicators of presence of relevant and irrelevant *cue-words* for the task at hand.

In the extraction procedure, the *search* function takes in input a list of cue-words and a context to check whether any cue-word occurs within the context. If irrelevant cue-words are found in the context but relevant cue-words are not, then the next context is processed; otherwise, the context is scanned left-to-right through the *findAll* function, which returns all non-overlapping matches of the article-references pattern in the given context.

By applying Algorithm 1 to the aforementioned examples, the output for the first text will be a list of 6 references to associate with article 42-bis, while in second text the output is an empty list since the occurrence of word “art1” is correctly recognized as an external, hence irrelevant, reference, due to the presence of irrelevant cue-words (i.e., “la legge 20 maggio 2016” and “comma 35”).

2.2. Network models

Let us denote with \mathcal{A} the set of articles in the ICC, and with \mathcal{B} a partition of \mathcal{A} into six groups, each corresponding to a book in the ICC, i.e., $\mathcal{B} = \bigcup_{i=1..6} \mathcal{A}_i$, such that $\mathcal{A}_i \cap \mathcal{A}_j = \emptyset$, for all $i, j \in \{1, \dots, 6\}$ with $i \neq j$, where $\mathcal{A}_i \subset \mathcal{A}$ is the subset of articles assigned to book i . Let also $r : \mathcal{A} \mapsto 2^{\mathcal{A}}$ denote a function that associates each article a to a set of articles that are referred to by a , possibly including articles that are from the same book of a or a different one.

Based on the article-reference relation, and by differentiating at level of single book or entire corpus, we define the following network models of interest to the study of the ICC.

Book-induced networks. Our first defined network model focuses on representing the

Table 1

Statistics about the extraction of the article references from the ICC books.

	Book-1	Book-2	Book-3	Book-4	Book-5	Book-6
# articles	395	345	364	891	713	331
# articles w/ references	123	71	45	77	258	88
# article-references	243	132	70	118	551	180
avg. # article-references per-article	0.615	0.383	0.192	0.132	0.773	0.544
avg. # article-references per-article w/ references	1.976	1.859	1.556	1.532	2.136	2.045

relations between articles of each particular book and their article references, which may cross the boundary of the book itself. Given a book, the induced network is a directed graph built from the article-reference list of all articles of that book. Formally, for each book i , we define the book-induced network for i as the directed graph $G_i = \langle V_i, E_i \rangle$ such that $V_i = \{a \in \mathcal{A}_i \mid r(a) \neq \emptyset\} \cup \{a \in \mathcal{A} \setminus \mathcal{A}_i \mid \exists a' \in \mathcal{A}_i, a \in r(a')\}$ and $E_i = \{(a, a') \mid a, a' \in V_i, a \in r(a')\}$.

Global or corpus network. Our second network model encompasses all relations among articles observed in the entire ICC. Therefore, we define the ICC corpus network $G_{\mathcal{B}}$ as the directed graph obtained by merging all book-induced networks, i.e., $G_{\mathcal{B}} = \langle V, E \rangle$ such that $V = \bigcup_{i=1..6} V_i$ and $E = \bigcup_{i=1..6} E_i$.

3. Structural Analysis of the ICC Networks

In this section we present our analysis of the book-induced and corpus networks, which is aimed to characterize main structural features of such networks and to unveil interesting patterns underlying the article references within and across books of the ICC. We organize our presentation into two subsections: the first one (Section 3.1) is concerned with *macroscopic* structural properties of the networks, whereas the second (Section 3.2) is focused on *mesoscopic* structural properties based on the outcome of a community detection task.

3.1. Macroscopic properties

Table 2 reports main results of our macroscopic structural analysis of the ICC corpus and book-induced networks. As a first general remark, only a portion of the articles is actually involved in article reference relations; in particular, considering the global network, the number of nodes (1 147) corresponds to about 36% of the articles within the ICC. This is actually not surprising, since it is commonly expected that a law article should be self-contained or self-explanatory. Nonetheless, there are norms regulating specific sections of the law code which, to be completely specified, require one or more references to different articles, possibly crossing the boundaries of a book. Indeed, this is what happens for a fraction of the ICC, which is not negligible, and hence deserves to be investigated.

In light of the above remark, one trait common to all the networks is the low *density*, which is the actual number of edges divided by the maximum possible number of links in the network,

Table 2

Summary of structural characteristics of the ICC corpus network (first column) and book-induced networks (subsequent columns).

	G_B	G_1	G_2	G_3	G_4	G_5	G_6
#nodes	1 147	223	144	95	161	432	171
#edges	1 294	243	132	70	118	551	180
reciprocity	3.4%	4.9%	3%	2.9%	0%	2.2%	7.8%
density	0.001	0.005	0.006	0.008	0.005	0.003	0.006
average degree*	2.218	2.126	1.806	1.453	1.466	2.523	2.023
average in-degree	1.128	1.090	0.917	0.737	0.733	1.275	1.053
% sources	37.2	35	35.4	41.1	43.5	35.4	31
% sinks	42.3	44.8	50.7	52.6	52.2	40.3	48.5
assortativity*	0.016	-0.184	-0.063	-0.058	-0.141	0.012	-0.173
assortativity	-0.035	-0.198	-0.051	-0.072	-0.158	-0.037	-0.196
average path length	2.241	1.639	1.393	1.104	1.064	2.384	1.568
diameter	7	6	3	3	3	7	5
transitivity*	0.099	0.119	0.135	0.160	0.107	0.098	0.109
clustering coefficient*	0.166	0.227	0.225	0.197	0.220	0.129	0.190
clustering coefficient (<i>full averaging</i>)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 128	218	142	93	161	427	166
#weakly connected components *	157	23	29	30	50	47	23
modularity*	0.891	0.866	0.882	0.914	0.946	0.807	0.876
#communities*	174	32	32	30	50	59	30
modularity	0.892	0.868	0.881	0.909	0.946	0.812	0.876
#communities	175	32	32	30	50	60	30

* Statistic calculated by discarding edge orientation

i.e., $|E|/(|V|(|V| - 1))$ for any directed graph $G = \langle V, E \rangle$. Partly related is also the low average *degree*, i.e., the average number of references involving an article, as well as low average *in-degree*, i.e., the average number of references to an article. Focusing on the latter, we observe that the ICC corpus network G_B , and the book-induced networks for Book-1, Book-5 and Book-6 have average in-degree above 1 (i.e., on average, each article is pointed by at least one other article), whereas the remaining networks exhibit a value lower than 1, indicating broader isolation. Indeed, by looking at an article's in-degree as a measure of its authoritativeness and usefulness to clarify and deepen the semantics of the articles that point to it, we find out that some articles indeed take a central role in the ICC. Figure 1 displays a visualization of the ICC corpus network, where articles of the same book are colored the same, and nodes with highest in-degree are made evident with associated label.

By contrast, the periphery of each of the networks represents a significant fraction of nodes, as it can be noted from the percentages of *source* and *sink* nodes reported in Table 2. We refer to source and sink nodes as those having only outgoing links and incoming links, respectively, i.e., source articles include others in their definition but are not referred to by other articles, whereas sink articles are used in the definition of one or more articles but they do not need others to complete themselves.

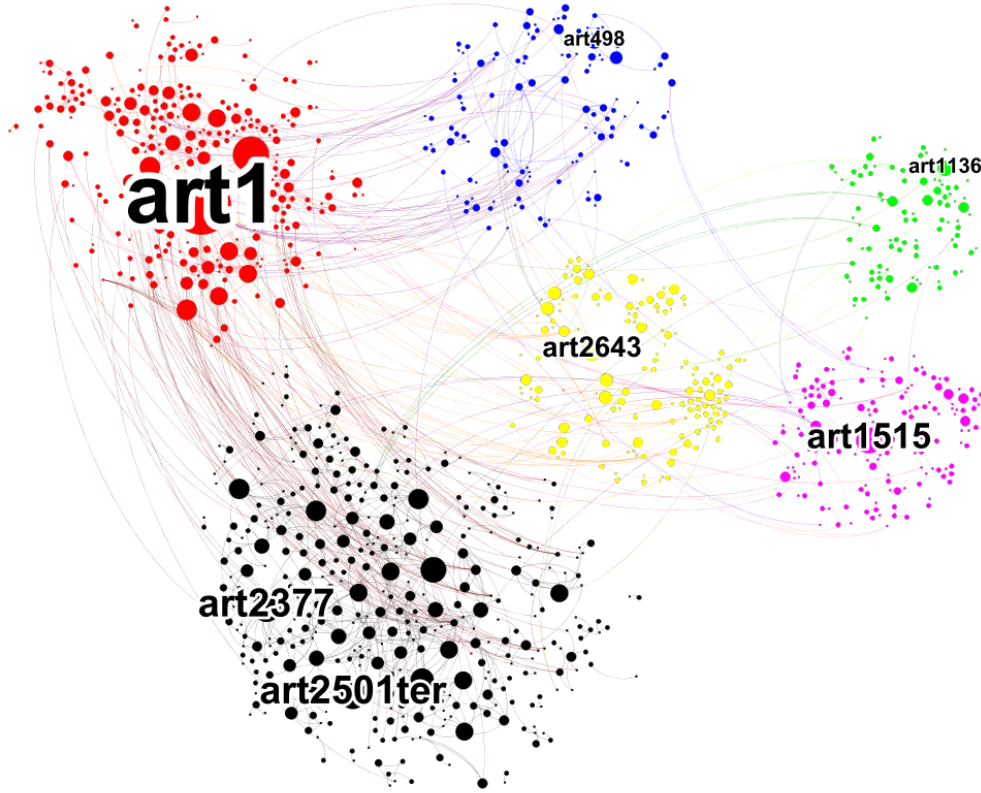


Figure 1: Articles in the ICC corpus network. Node sizes are proportional to the in-degree, and for each book a label representing the article id is associated to the node(s) with highest in-degree. Colors are used to distinguish the six books of the ICC as follows: red (Book-1), blue (Book-2), green (Book-3), magenta (Book-4), black (Book-5), yellow (Book-6).

The above duality also impacts on the degree correlation, also known as *assortativity*, which captures how the probability of links between two nodes is influenced by their degree [6, 7]. For all networks, we observe negative correlation, which means that articles with different degrees tend to link each other.

We also analyzed the tendency of articles to form strong ties with closure patterns, specifically dyadic closure or *reciprocity* and *triadic closure*. The former is computed as the fraction of reciprocal edges, and helps us understanding how articles might reinforce each other to enhance and refine their meaning. Reciprocity is found to be low for all networks, with 3.4% for the corpus network and upper bounded by 8% for the book-induced networks. For the latter, it is

interesting to observe two contrasting cases: the one corresponding to zero reciprocity, which characterizes the article references found in Book-4, and the other one corresponding to the maximum reciprocity found in Book-6, thus suggesting a more evident dyadic closure for the articles in that book than in others. As concerns the triadic closure, i.e., the probability of closing connected triplets of nodes, we resorted to both global and local measures [8]. In the former case, we evaluated the the probability that two incident edges are completed by a third one to form a triangle, namely *transitivity*. In the latter case, instead, we evaluated the transitivity at node level, i.e., how strongly connected are the neighbors of a node, namely *local clustering coefficient*. Although both measures show low values, it should be noted how the local clustering coefficient shows slightly greater values when ignoring source and sink articles.

We also investigated reachability aspects in the various networks. First, we examined the *average path length*, i.e., the average of the pairwise distances between nodes in a network, where the distance between two nodes corresponds to the length of a shortest path connecting them. As it can be noted from the table, the average path length is slightly above 2 for the corpus network and for the Book-5 network only. Also, the *diameter*, i.e., the shortest-path distance between the two most distant nodes in the network, is maximum in the Book-5 network (7) and determines the diameter for the corpus network as well; Book-1 and Book-6 networks have diameter equal to 6 and 5, respectively, while the remaining networks have diameter 3.

3.2. Mesoscopic properties

Reachability-based approaches to the identification of subnetworks with particular properties of connectivity focus on the existence of paths, regardless of the distance. In this respect, for each of the networks under study, we calculated the *strongly* and *weakly* connected components, which correspond to the maximal subgraphs where every node is reachable from every other node through a directed or undirected path, respectively. Given the outcome of the above discussed analysis steps, it does not come to our surprise that the observed number of strongly connected components is extremely high, which indicates the formation of small groups of articles that are involved in chains of references when accounting for edge orientation, and highlights the lack of *multi-hop references* (i.e., *art. x* refers to *art. y*, which in turn refers to *art. z*). By discarding edge orientation, instead, mutual connectivity clearly increases, and the detected number of weakly connected components is one order of magnitude lower than the node-set size. This trait confirms the existence of strategic articles, namely those referred by other ones and that support (undirected) connectivity between articles.

Connected components can be seen as a raw form of *communities*, based on minimal requirements of connectedness. However, according to a density hypothesis, communities should correspond to locally dense neighborhoods of a network. Therefore, we delved into each of the article reference networks by carrying out a mesoscale-level analysis to shed light on the underlying *community structure*, i.e., a division of a network into regions such that the nodes in each region should be highly linked with each other, whereas few links should exist between the regions. Note however that the amount of edges per se is not an optimal proxy to quantify the community structure: a good community structure is not merely one in which there are few edges between communities, rather it is one in which there are fewer than expected edges between communities. This is the key principle underlying the theory of *modularity* to discover

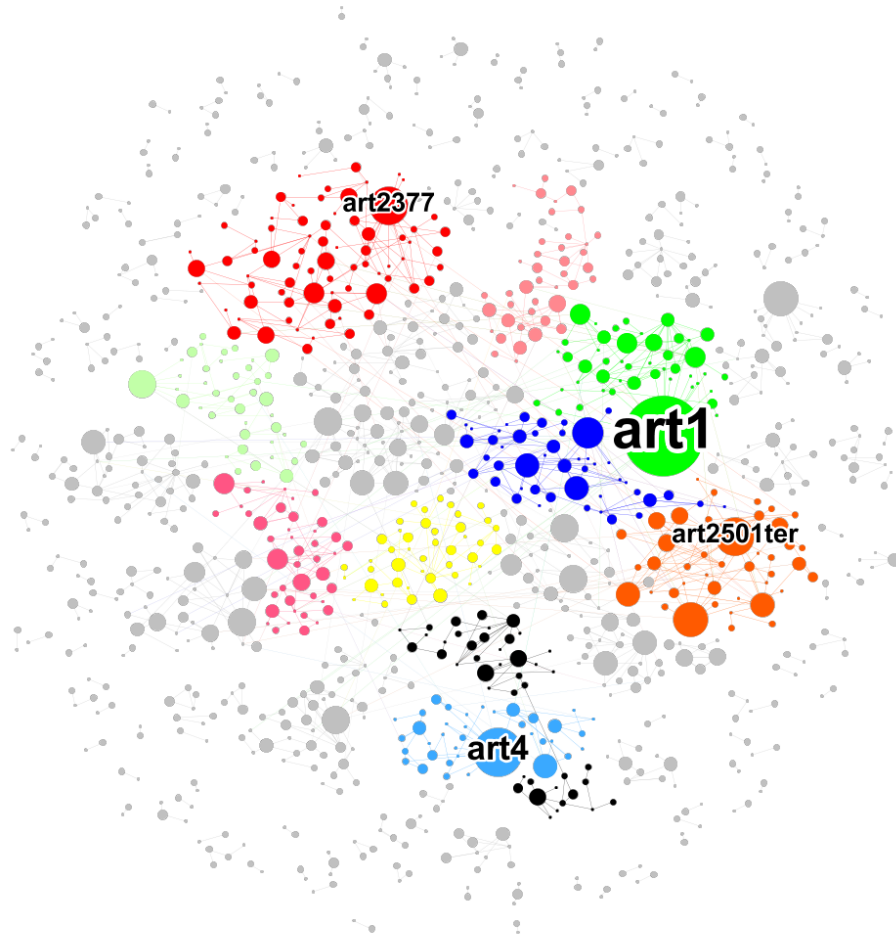


Figure 2: Community structure of the ICC corpus network. Node sizes are proportional to the in-degree, and nodes with in-degree greater than or equal to 10 are labeled with the associated article id. Colors are used to distinguish the top-10 largest communities (nodes assigned to the remaining communities are colored in gray).

a community structure in a network: intuitively, the modularity of a community is the total difference of the fraction of the edges within the community w.r.t. the expected such fraction if the edges were distributed at random, so that the higher this deviation, the better the community. In this work, we follow the widely-recognized line of research that resorts to a modularity maximization approach to community detection. In particular, we use the most popular method belonging to this category, namely the *Louvain* method [9], both in its original, undirected version and its variant that accounts for edge orientation while maximizing the modularity.¹

As reported in Table 2, both Louvain and its directed variant lead to the discovery of communities having high modularity (note that modularity is upper bounded by 1) in all networks.

¹<https://github.com/nicolasdugue/DirectedLouvain>

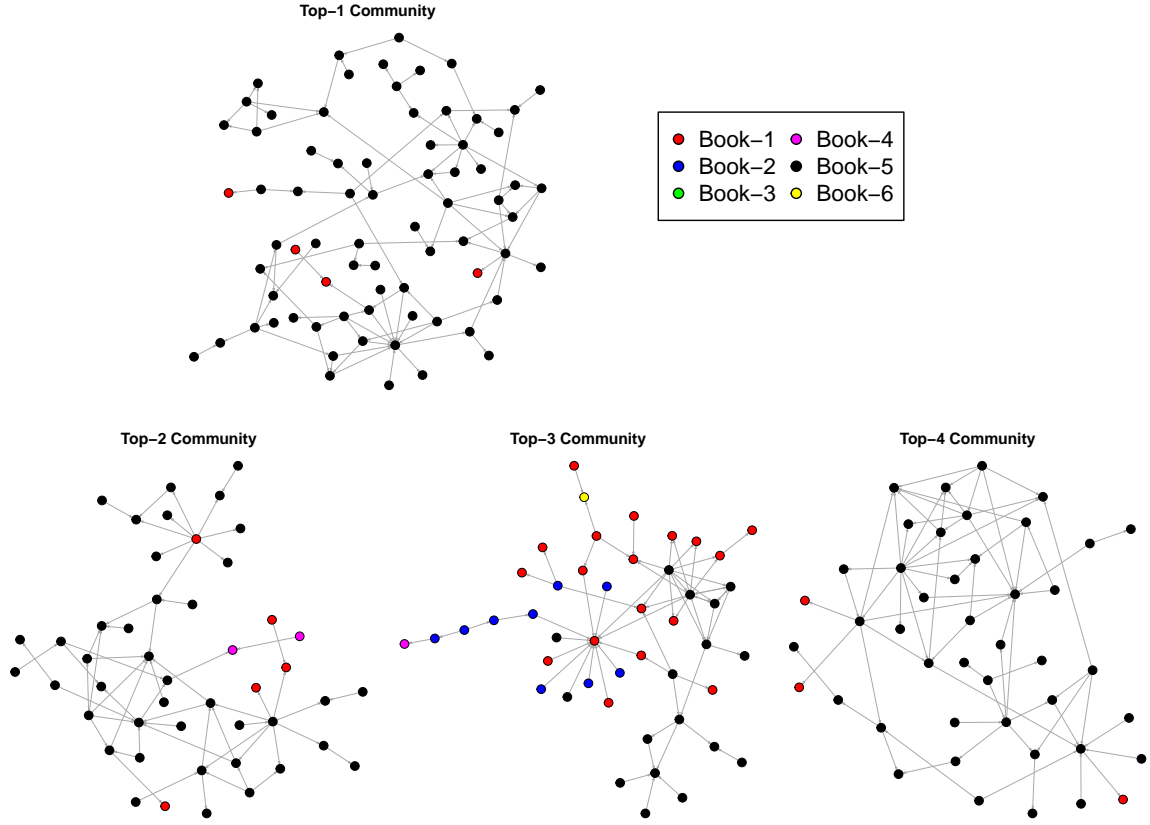


Figure 3: Top-10 largest communities discovered by the Louvain method in the ICC corpus network. Color codes correspond to the ICC books.

This is also consistent with the presence of several connected yet isolated regions within the networks, which are also comparably in size with the connected components previously discussed. Furthermore, it should be noted that accounting for edge orientation does not imply significant differences in modularity as well as number of communities with respect to the outcome of the undirected counterpart. Figure 2 illustrates the community structure identified in the ICC corpus network, where the top-10 largest communities are emphasized and distinguished by color. Note that these top-10 communities cover about 40% of the network.

One interesting question that arises is whether the different communities contain mixed book-memberships, i.e., whether a community may cross the book boundaries through the article reference relations. To this purpose, we explored the top-10 largest communities in the ICC corpus network, which are visualized in Figures 3–4. As it can be noted from the figures,

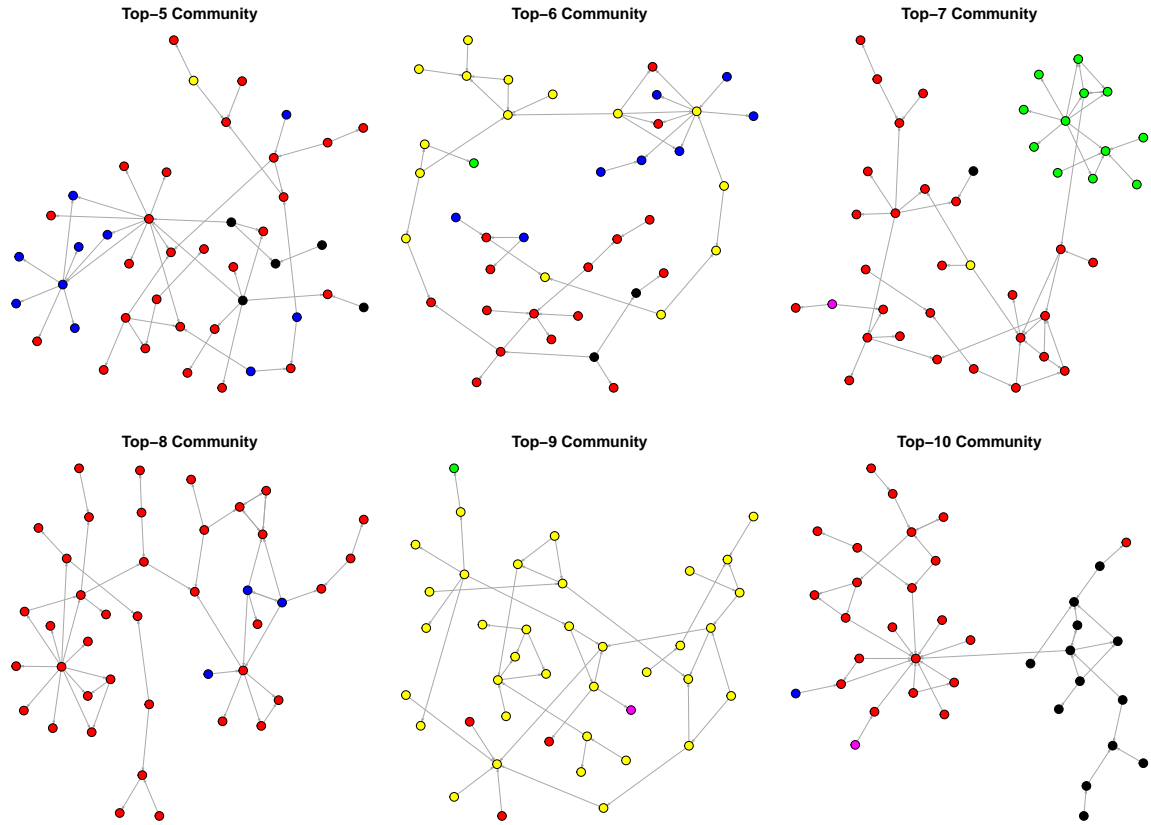


Figure 4: (Cont.) Top-10 largest communities discovered by the Louvain method in the ICC corpus network. Color codes correspond to the ICC books.

there are indeed cohesive communities, which are mostly formed by articles from the same book, as well as communities that contain articles from different books. This would suggest that the modular structure in the ICC corpus network can be consistent with either themes of a particular (portion of) book or themes that are shared by (portions of) different books. Clearly, the latter might originate from the opportunity of completing or enhancing the norms provided by a book’s article(s) with those provided by other books’ article(s).

In this respect, we moved a step forward by exploring the contents of the articles involved in the top-10 communities, in order to gain insights into patterns underlying possible relations between the formation of communities and the topic coherence. From this analysis stage, several remarks stand out, which we try to summarize as follows:

- Community capturing most representative topics of a particular book. This is likely to happen when the book memberships in the community are mostly or fully homogeneous. For instance, the top-1 community contains articles focused on “administration of the capital of a company”, which is a representative topic of Book-5.
- Community unveiling fine-grain topical patterns that are mostly discussed in a book yet complemented with references to other book(s). This refers to sort of strong ties that

are formed through article references that cross the boundaries of two or more books. For instance, top-8 and top-10 communities are such a type of community. In the top-8 community, an interesting pattern is found out for “succession” norms (Book-2) in a context of “marital separation” (Book-1).

- Community induced from reinforcement of topic(s) from a book with related topics that differently contribute to the contents of other books. This type of community turns out to be characterized from mixed book-memberships that are distributed over substructures built upon across-book article references. For instance, the top-6 community contains node-articles that belong to five different books; moreover the topic “transcription” that is largely discussed in Book-6 (which is the mostly covered in the top-6 community) is reinforced with the topic “properties”, which is shared between Book-1 and Book-2, jointly with the topic “community”, which is discussed within Book-1.

In light of the above remarks, we can conclude that a mesoscopic view like that supplied by our discovered community structure can represent a valuable support to enhance the macroscopic (i.e., at book level) or microscopic (i.e., at article level) views that are primarily considered from legal experts and practitioners.

4. Conclusions

Artificial intelligence research in the legal domain is constantly growing and draws on various fields, ranging from natural language processing to machine learning and network science, thus achieving an interdisciplinary imprint. These diversified viewpoints can be valuable to enrich our understanding of the legal domain and to enhance the evolutionary process of law codes, also taking into account social development and needs.

In this work, by taking a network analysis and mining perspective, we presented the first study of citation networks that can be inferred from the ICC articles. Our analysis of the structural features of such networks has shed light on valuable hidden patterns, such as the linkage between community memberships of articles and their topical structure, paving the way for new interpretations and study of the ICC.

It is also worth noticing that our proposed methodology can easily be generalized to other civil law code systems presenting a similar organization as the ICC, i.e., developed upon a logical structure of the corpus into books, and their internal subdivisions.

References

- [1] J. H. Fowler, T. R. Johnson, J. F. Spriggs, S. Jeon, P. J. Wahlbeck, Network Analysis and the Law: Measuring the Legal Importance of Precedents at the U.S. Supreme Court, *Political Analysis* 15 (2007) 324–346. doi:10.1093/pan/mpm011.
- [2] M. Moser., M. Strembeck., An Analysis of Three Legal Citation Networks Derived from Austrian Supreme Court Decisions, in: *Proc. of the 4th International Conference on Complexity, Future Information Systems and Risk*, 2019, pp. 85–92. doi:10.5220/0007749900850092.

- [3] M. Koniaris, I. Anagnostopoulos, Y. Vassiliou, Network analysis in the legal domain: a complex model for European Union legal sources, *Journal of Complex Networks* 6 (2017) 243–268. doi:10.1093/comnet/cnx029.
- [4] P. Mazzega, D. Bourcier, R. Boulet, The Network of French Legal Codes, in: *Proc. of the 12th International Conference on Artificial Intelligence and Law, ICAIL '09*, 2009, p. 236–237. doi:10.1145/1568234.1568271.
- [5] P. Zhang, L. Koppaka, Semantics-based legal citation network, in: *Proc. of the 11th International Conference on Artificial Intelligence and Law, ICAIL '07*, 2007, pp. 123–130. doi:10.1145/1276318.1276342.
- [6] M. E. J. Newman, Assortative mixing in networks, *Physical Review Letters* 89 (2002). doi:10.1103/physrevlett.89.208701.
- [7] M. E. J. Newman, Mixing patterns in networks, *Physical Review E* 67 (2003). doi:10.1103/physreve.67.026126.
- [8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.-U. Hwang, Complex networks: Structure and dynamics, *Physics Report* 424 (2006) 175–308.
- [9] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *Journal of Statistical Mechanics: Theory and Experiment* 10 (2008) P10008. doi:10.1088/1742-5468/2008/10/p10008.