The Italian Civil Code Network Analysis

<u>Relations in the Legal Domain @ICAIL2021</u>

The 18th International Conference on Artificial Intelligence and Law

University of São Paulo, Monday June 21 – Friday June 25.

Lucio La Cava, Andrea Simeri, Andrea Tagarelli University of Calabria, Italy



DIMES





Introduction





Natural Language Processing **Network Science**

The Italian Civil Code

2

The Italian Civil Code (ICC)

- Divided into six logically coherent books
- Each book provides rules for a particular civil law theme
- The articles of each book are internally organized into a hierarchical structure
 - "titoli" (i.e., chapters)
 - "capi" (i.e., subchapters)
 - "sezioni" (i.e., sections)
 - "paragrafi" (i.e., paragraphs)
- The ICC currently in force consists of 2 969 article numbers, which actually corresponds to 3 225 articles considering all variants and subsequent insertions



Challenges









Article references

- ICC articles may contain citations of one or more articles, from the same or a different book, dubbed as *article references*
- Identifying and extracting article references from ICC articles is tough
 - No hypertext or indexes
 - References outside the ICC
- Need for an approach to identify and extract valid article references

Article reference composition

- *Prefix:* a common root for all lexical variants of an article (typically, the abbreviation "art" is used)
- *Article id:* follows the numerical intervals specific to each book
- *Variant suffix* (optional): used to designate a variant of a given article
 - Variants are not alternative, and hence they must be treated as separate articles
 - Suffixes correspond to Latin adverbial numerals and express the multiplicity of occurrence (e.g., "bis" stands for "twice", "ter" for "three times, and so on)

Cue words



Relevant: leveraged to find the presence of valid article references, e.g., *"codice civile"* ("civil code") and its lexical variants



Irrelevant: leveraged to find references to external sources, e.g., "*legge*" ("law"), "*decreto legislativo*" ("legislative decree"), "*sentenza*" ("judgment"). They are usually followed by a date in some format, and they also include other words, such as "*comma*" ("paragraph"), which may follow a reference inside a portion of the context delimited by round brackets



Relevant cue-words are much fewer and less frequently used than irrelevant cue-words

Extraction of article references

Algorithm 1: Extraction of article references

```
Data: Set of articles A in the ICC; list of relevant cue-words art-refs_words and list of
       irrelevant cue-words ext-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-refs words, c) and not search(art-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
```

Insights on the ICC books

	Book-1	Book-2	Book-3	Book-4	Book-5	Book-6
# articles	395	345	364	891	713	331
<pre># articles w/ references</pre>	123	71	45	77	258	88
<pre># article-references</pre>	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

Models notation

- ${\mathcal A}$ set of articles in the ICC
- \mathcal{B} 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 $\in \{1, ..., 6\}$ with $i \neq j$, where $\mathcal{A}_i \subset \mathcal{A}$ is the subset of articles assigned to book i
- r: A → 2^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

Network models





Book-induced networks $G_i = \langle V_i, E_i \rangle$

Global or corpus network $G_{\mathcal{B}} = \langle V, E \rangle$

$$V = \bigcup_{i=1..6} V_i$$
$$E = \bigcup_{i=1..6} E_i$$

 $V_{i} = \{a \in \mathcal{A}_{i} | r(a) \neq \emptyset\} \cup \{a \in \mathcal{A} \setminus \mathcal{A}_{i} | \exists a' \in \mathcal{A}_{i}, a \in r(a')\}$ $E_{i} = \{(a, a') | a, a' \in V_{i}, a \in r(a')\}$

The ICC Network

- Only a portion of the articles is actually involved in article reference relations, i.e., 1 147 articles (about 36% of the articles within the ICC)
- Expected self-explanatory trait that characterizes most of the articles
- Non-negligible fraction of ICC articles that leverages referencing to improve their readability
- The exploration of the latter fraction might reveal fascinating traits!

	GB	G_1	G_2	G_3	G_{4}	G_5	G_{6}
			144	~ <u>.</u>	1/1	400	171
#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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 1 2 8	218	142	93	161	427	166
<pre>#weakly connected components *</pre>	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

	$G_{\mathcal{B}}$	G_1	G_2	G_3	G_4	G_5	G_6
#nodes	1 1 4 7	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 2 8	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

	$G_{\mathcal{B}}$	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 2 8	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

	$G_{\mathcal{B}}$	G_1	G_2	G_3	G_4	G_5	G_6
#nodes	1 1 4 7	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 2 8	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

	$G_{\mathcal{B}}$	G_1	G_2	G_3	G_4	G_5	G_6
#nodes	1 1 4 7	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 2 8	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

	$G_{\mathcal{B}}$	G_1	G_2	G_3	G_4	G_5	G_6
#nodes	1 1 4 7	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 2 8	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

			~	~	~	~	~
	$G_{\mathcal{B}}$	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 2 8	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

The ICC Network at the macro-scale

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).



	$G_{\mathcal{B}}$	G_1	G_2	G_3	G_4	G_5	G_6
#nodes	1 1 4 7	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 1 2 8	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

	$G_{\mathcal{B}}$	G_1	G_2	G_3	G_4	G_5	G_6
#nodes	1 1 4 7	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 (full averaging)*	0.081	0.106	0.097	0.039	0.055	0.072	0.091
#strongly connected components	1 1 2 8	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

The ICC Network at the meso-scale

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).



- Both cohesive (i.e., mostly formed by articles from the same book) and mixed communities arise
- Hints at consistency with the themes of a (portion of) book or themes that are shared by (portions of) different books
- The exploration of the topics characterizing the articles in the top-10 communities revealed interesting patterns

- Community capturing most representative topics of a particular book
- Book memberships in the community are mostly or fully homogeneous
- For instance, in the top-1 community, the topic "administration of the capital of a company" emerges, 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)
- Ties formed through article references that cross the boundaries of two or more books
- For instance, in the top-8 community, the pattern "succession" norms (Book-2) in a context of "marital separation" (Book-1) emerges



- Community induced from reinforcement of topic(s) from a book with related topics that differently contribute to the contents of other books
- Mixed book memberships distributed over substructures built upon acrossbook article references
- For instance, in the top-6 community, the topic "transcription" (Book-6) is reinforced with the topic "properties" (Book-1 and Book-2), jointly with the topic "community (Book-1)

Conclusions

- Novel and diversified perspectives to gain further insights and understandings
- Potential enhancement of the legal domain
- First study of the citation networks that can be inferred from the ICC articles
- Unveiling of valuable hidden patterns
- Proposal of a methodology that can be easily generalized to other civil law code systems built upon a similar organization as the ICC

Thanks for your attention!