Mapping landmark cases in the U.S. legal system

Rilder S. Pires[1], Erneson A. Oliveira[1,2,3], Carlos G. O. Fernandes[3,4], João A. Monteiro Neto[5] and Vasco Furtado[3,6]

[1]*Laboratório de Ciência de Dados e Inteligência Artificial, Universidade de Fortaleza, 60811-905 Fortaleza, Ceará, Brasil*
[2]*Mestrado Profissional em Ciências da Cidade, Universidade de Fortaleza, 60811-905 Fortaleza, Ceará, Brasil*
[3]*Programa de Pós Graduação em Informática Aplicada, Universidade de Fortaleza, 60811-905 Fortaleza, Ceará, Brasil.*
[4]*Banco do Nordeste do Brasil S.A., 60743-902 Fortaleza, Ceará, Brasil.*
[5]*Centro de Ciências Jurídicas, Universidade de Fortaleza, 60811-905 Fortaleza, Ceará, Brasil.*
[6]*Empresa de Tecnologia da Informação do Ceará, Governo do Estado do Ceará, 60130-240 Fortaleza, Ceará, Brasil.*

Abstract

Court decisions and emblematic legal cases are central elements of Law. They influence practitioners, scholars and public officers at the same time they define and shape the legal reality and its boundaries. Despite being an overlooked area, understanding how legal cases establishes connections and relationships can provide important insights not only about how influence and impact are built, but also to identify influential cases that are not listed as landmarks. Here, we explore data from the U.S. legal system by modeling it as a citation network fed with 360 years of legal cases. We characterize the probability distributions for degree-centrality measures of the network and find a power-law behavior for the in-degree probability distribution with an exponent $\alpha \approx 2.66$. We also obtain the probability distribution for landmarks according to their in-degree and out-degree in order to find the region in the in-degree$\times$out-degree space where landmarks are more likely to be found. Finally, we highlight some extreme special cases and make some considerations about the ratio between the number of landmarks and the total number of legal cases in a given spot of the in-degree$\times$out-degree space.

Keywords

U.S. legal system, Legal cases, Landmark cases, Complex networks, Citation networks,

1. Introduction

Precedents are fundamental to understanding the United States (U.S.) legal system since they are the pillars of *Common Law*, the judicial system in Anglo-Saxon countries. Unlike *Civil Law* (where statutes are the foundations of the judicial apparatus), Common Law is based on the previous judgments of the courts, which establish the rules to be followed, persuading or binding judges to that series of decisions [1, 2]. By legal point of view, there are some special cases, namely "landmark cases", that become relevant by setting key legal concepts or interpretations and in doing so they influence a great number of other cases along the years. In recent years, an increasing number of studies have been proposed to characterize legal networks through mathematical models [3, 4, 5, 6]. Despite the efforts of such studies, properly defining the

RELATED - Relations in the Legal Domain Workshop, in conjunction with ICAIL 2021, June 25, 2021, São Paulo, Brazil
rilder@unifor.br (R. S. Pires); erneson@unifor.br (E. A. Oliveira); carlosgustavo@edu.unifor.br (C. G. O. Fernandes); joaoneto@unifor.br (J. A. M. Neto); vasco@unifor.br (V. Furtado)
0000-0003-4873-5308 (R. S. Pires); 0000-0001-6413-8445 (E. A. Oliveira)

properties of a landmark case through quantitative approaches remain an open problem in law research areas.

Here, we use a model based on concepts of complex networks fed with 360 years of data in order to characterize landmark cases in the U.S. legal system. We know that vertices with high degree-centrality measures play an important role in the information dynamics of the network, especially in citation networks [7]. Such fact suggests that "landmarks" could be easily identified as "outliers" with the highest in-degrees and that out-degree may have some importance for landmarks since their degree of "innovation" could be measured through it. For these reasons, the main contribution of our study is to perform a topological map suggesting the location of landmark cases. We observe that the most cited cases in U.S. legal system are not landmarks. In addition, we find a well-located region with all landmark cases and, even within this region, the top cited landmark cases are exceptions. This allows us to shed some light on understanding the structure and evolution of U.S. law.

2. Datasets

We use the *Caselaw Access Project* (CAP) and the *Historic Supreme Court Decisions* (HSCD) open datasets [8, 9], which are compiled and made available by the Harvard and the Cornell Law Schools, respectively. The former is composed of citation records and other retrieved information for each digitalized court decision (*legal case*) in the U.S. legal system. In this work, we only use the citation records from the CAP dataset, which is a CSV file with ≈ 5 million records (≈ 380 MB), where each row has a legal case ID and all other legal cases IDs that cite it. The later is a list of referral court decisions (*landmark cases*) from the Supreme Court. The HSCD dataset is also a CVS file with 538 records (≈ 4 KB).

3. The Model

In order to characterize the U.S. legal system, we use the *citation network model*, where the vertices v_i are legal cases and the directed edges $e_k = (v_i, v_j)$ are citations in the CAP dataset, from a newer legal case v_i to an older legal case v_j. In fact, our model is a network similar to those used to describe paper citations [7]. Furthermore, we remove a few duplicated edges in order to have a perfect acyclic network. The numbers of vertices and edges of the citation network are $n \approx 5$ million and $m = 43$ million, respectively. We emphasize that such amount of citations unveils an additional computational challenge in our modeling. Mathematically, a network is characterized by its *adjacency matrix* $\mathbf{A}$. This matrix is defined in such way that each of its elements $A_{i,j}$ are either “1” if there is an edge between v_i and v_j or “0” otherwise [10, 11]. Some measures, namely *vertex-centrality measures*, can be derived from $\mathbf{A}$ in order to define some kind of global importance of a given vertex in the network. Two basic examples of these measures are the *degree-centrality measures* [7], defined as

$$K_{\text{in}}(i) = \sum_{j}^{n} A_{j,i} \quad \text{and} \quad K_{\text{out}}(i) = \sum_{j}^{n} A_{i,j}, \tag{1}$$

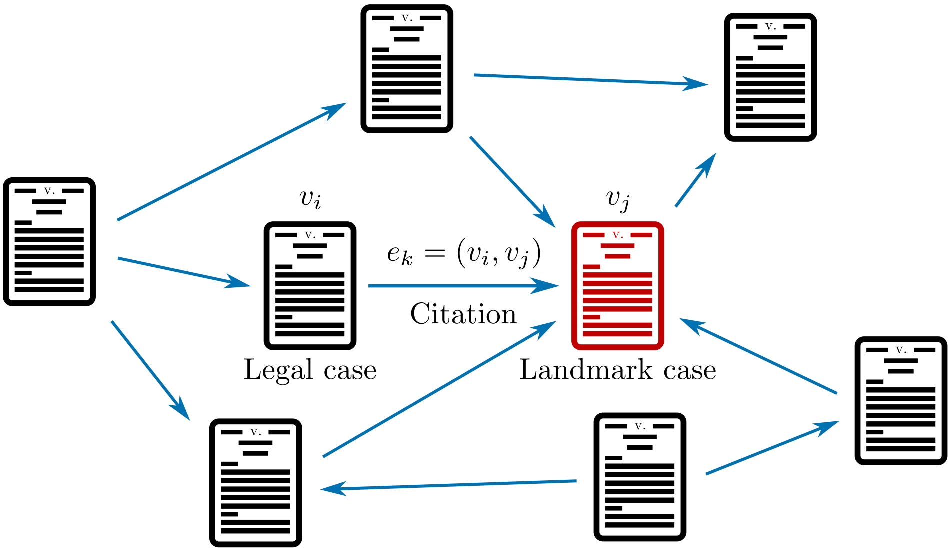

Figure 1: Schematic representation of the citation network. The documents and arrows represent the vertices (legal cases) and edges (citations), respectively. The citations, $e_k = (v_i, v_j)$, are assigned from vertices where arrows start, v_i, to the vertices where arrows end, v_j. Therefore, we say that v_i is citing v_j in this formalism. Furthermore, some legal cases are identified as landmark cases (red document).

where $K_{\text{in}}(i)$ and $K_{\text{out}}(i)$ are usually called in-degree and out-degree of vertex i, respectively. In practical terms, $K_{\text{in}}(i)$ represents the number of legal cases that cite the legal case i and $K_{\text{out}}(i)$ corresponds to the number of legal cases that is cited by the legal case i.

4. Results

In order to understand the structure of citations in the U.S. legal system, we perform topological measures on the citation network. Precisely, we focus our analysis in the K_{in} and K_{out} centrality measures due the large size of the network. These measures have the advantage of being simple, *i.e.*, they are easily computed even for large networks. Despite the simplicity of these measures, they allows us to understand basic aspects of this network.

We show both K_{in} and K_{out} probability distributions for the U.S. legal system (Fig. 2a). We find that the probability distribution for K_{in} is described by a power law with exponent $\alpha \approx 2.66$ characteristic of other scale-free networks [12, 13, 14]. The α exponent was obtained using *Maximum Likelihood Estimation* [15] and assuming that a power-law distribution describe the data. For the K_{out} probability distribution, we are able to identify a long-tailed behavior, however we have no strong evidence to support a power-law hypothesis. In Figure 2b, we show both K_{in} and K_{out} probability distributions for the landmark cases. Such distributions are drastically different from those shown in Fig. 2a. Numerically, landmark cases have in average $\overline{K}_{\text{in}} \approx 1,252.5$ and $\overline{K}_{\text{out}} \approx 44.6$. In contrast with the general case, where $\overline{K}_{\text{in}} \approx \overline{K}_{\text{out}} \approx 8.96$, landmark cases have $\overline{K}_{\text{in}}$ and $\overline{K}_{\text{out}}$, respectively, ≈ 140 times and ≈ 5 times higher than usual legal cases.

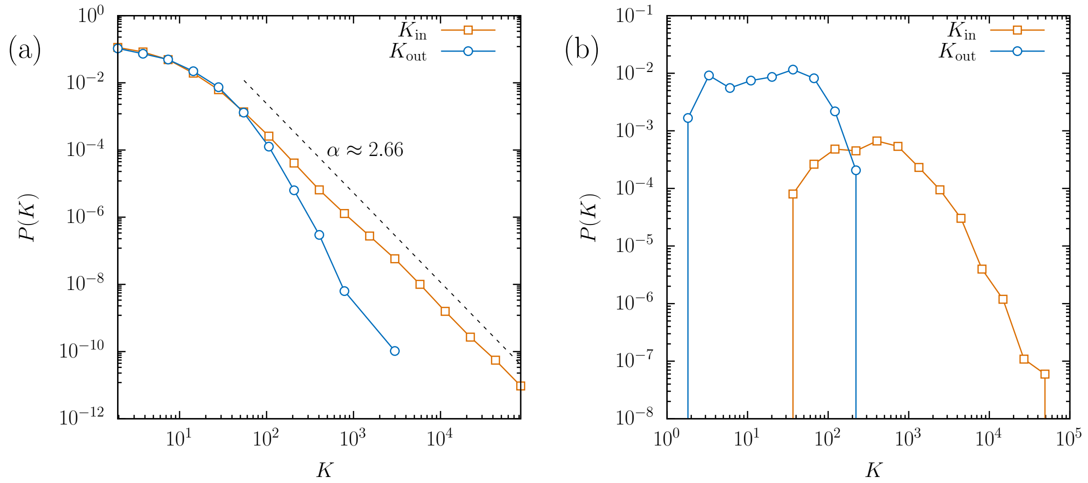

Figure 2: Probability distributions of K_{in} and K_{out}. (a) Probability distributions of K_{in} and K_{out} for legal cases in the citation network. As we can see, the distributions show a long-tailed behavior and $P(K_{\text{in}})$ show a pronounced power-law behavior with at least three orders of magnitude. The dashed line is a power-law $P(x) \propto x^{-\alpha}$ with $\alpha \approx 2.66$. (b) Probability distributions of K_{in} and K_{out} for the landmark cases.

In Figure 3a, we show the probability distribution of cases of the U.S. legal system as function of K_{in} and K_{out}. We observe that, in general, the legal cases are more likely to be in a region with small values of K_{in} and K_{out}. The landmark cases, however, have a higher probability of being in regions of large K_{in} and K_{out} as shown in Fig. 3b. The black line, shown in Fig. 3a, highlights the areas where the probability distribution of landmark cases, shown in Fig. 3b, is greater than zero. In Figure 3a, we also observe that there are regions, where the probability distribution is higher than zero, beyond the limiting line defined by landmark cases.

Table 1
Special legal cases labeled in Fig. 3.

Label	ID	K_{in}	K_{out}	Legal case
A	6206897	66554	33	Anderson v. Liberty Lobby, Inc. (1986)
B	4147232	45	2944	Henry v. New Jersey Department of Human Services (2010)
C	6219632	1253	37	Idaho v. Wright (1990)
D	368014	1244	32	Skinner v. Oklahoma ex rel. Williamson (1942)

In Table 1, we show the special legal cases highlighted in Fig. 3. These cases were chosen according to the following criteria: case “A” is the case with highest K_{in}, case “B” is the case with highest K_{out}, and cases “C” and “D” are the legal and landmark cases, respectively, closer to the $(1, 252.5, 44.6)$ point. We choose the point $(1, 252.5, 44.6)$ as a reference, since it corresponds to coordinates of $\overline{K}_{\text{in}}$ and $\overline{K}_{\text{out}}$ for the landmark cases.

The case Henry v. New Jersey Department of Human Services, 204 N.J. 320, 9 A.3d 882 (2010),

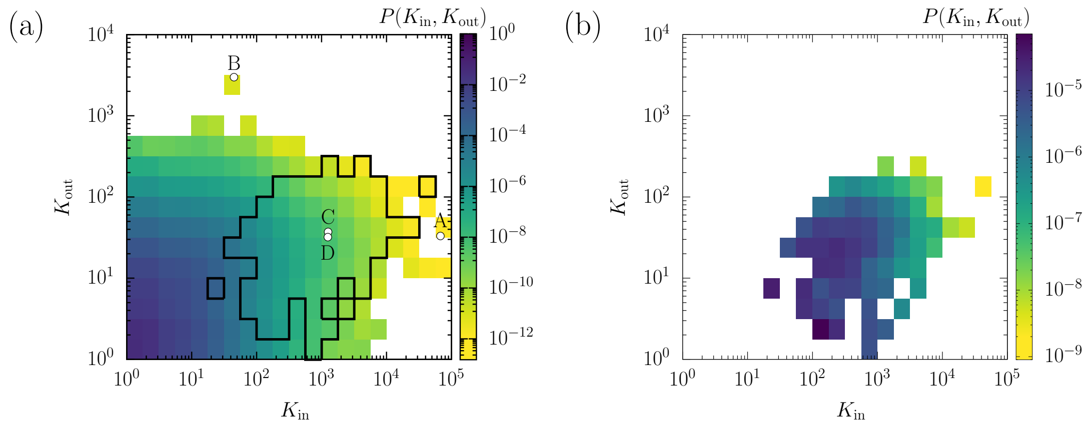

Figure 3: Probability distributions in the in-degree×out-degree space. (a) Probability distribution of legal cases as function of $K_{\rm in}$ and $K_{\rm out}$. (b) Probability distribution of landmark cases as function of $K_{\rm in}$ and $K_{\rm out}$. The black line in (a) represents the boundary of the area delimited by the distribution of landmark cases shown in (b). Due to the poor statistics caused by the limited number of landmark cases, this area shows some "islands" and even "holes" that makes its boundary disconnected. The points "A" and "B" in (a) are special legal cases with high values of $K_{\rm in}$ and $K_{\rm out}$ and the points "C" and "D" are special legal cases close to the $(1, 252.5, 44.6)$ point, which corresponds to the values of $\overline{K}_{\rm in}$ and $\overline{K}_{\rm out}$ for landmark cases.

cites more than $2,900$ cases. While it deals primarily with a claim related to the New Jersey Law Against Discrimination, the incredible high number of citations is connected to a secondary issue raised during the judgment. During the case, the Supreme Court of New Jersey debated an issue related to the composition of the Court as the Honorable in charge of delivering the opinion in this case was nominated to the court in a temporary assignment. This secondary issue debate led Justice Rivera-Soto to cite more than $2,000$ cases supporting his argument that the court was unconstitutionally constituted. The identification of anomalous cases can be an interesting staring point to qualitative approach investigating the relevant secondary issues that normally are shadowed by the initial discussion and do not receive the attention they need.

On the opposite side, Anderson v. Liberty Lobby, Inc. (1986) is a very important case where the U.S. Supreme Court set the standards guiding the acceptance or not of a summary judgment request. A summary judgment will happen when there are no factual issues to discuss and the trial court would analyze only matters of law. This case is highly influential, being cited more than $65,000$ times because it set the mandatory rules that every case pledging a motion of summary judgment needs to fulfill. Knowing better this network can reveal not only the level of influence of the case, but also indications of in which jurisdictions the case is more often cited and even if there is a rise of new cases questioning or reinforcing its standards.

The ratio $N_{\rm land}/N_{\rm all}$ in the in-degree×out-degree space is shown in Fig. 4. Here, $N_{\rm land}$ and $N_{\rm all}$ are the total number of legal and landmark cases, respectively, inside of a given bin defined by the same divisions used in Fig. 3. We observe that such ratio tends to be greater in regions of high K_{in} and K_{out}, *i.e.*, in regions of very low probability even for landmark cases.

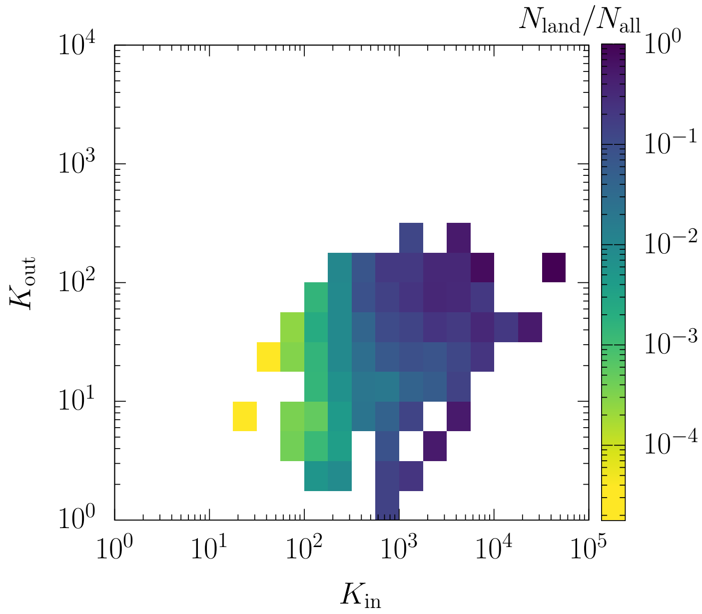

Figure 4: Ratio $N_{\mathrm{land}}/N_{\mathrm{all}}$ in the in-degree×out-degree space. We show that the ratio between the number of landmarks N_{land} and the total number of legal cases N_{all} tends to be greater in regions of high K_{in} and K_{out}.

5. Conclusions

In summary, we performed a numerical data analysis in order to establish a topological map characterizing the location of landmark cases in the U.S. legal system. Precisely, we modeled the U.S. legal system as a citation network fed with 360 years of digitalized documents and focused our analysis on the in-degree and out-degree centrality measures. By evaluating K_{in} and K_{out} for all legal cases, we obtained the probability distributions for these measures and found a power-law decay for the probability distribution of K_{in} with an exponent $\alpha \approx 2.66$. We also characterized the landmark cases and evaluated their K_{in} and K_{out}. Probability distributions of K_{in} and K_{out} for the landmark cases were found to be drastically different from those found for all legal cases, showing that landmark cases have $\overline{K}_{\mathrm{in}}$ and $\overline{K}_{\mathrm{out}} \approx 140$ times and ≈ 5 times higher than usual legal cases, respectively. Moreover, we showed that there is an area in the $K_{\mathrm{in}} \times K_{\mathrm{out}}$ space where landmarks are more likely to be found. Surprisingly, there are regions beyond of the limiting area of landmark cases where the probability distribution for all legal cases is higher than zero. This result could be confirmed by identifying some special legal cases where K_{in} or K_{out} were greater than the limiting values of the landmark region. Using a similar approach, we also observed that it is possible to find a landmark case and an usual legal case with very similar K_{in} and K_{out}. In order to understand the relative occurrence of landmark cases inside the region where they are more likely to be, we studied the $N_{\text{land}}/N_{\text{all}}$ ratio in this region and found that it is greater in regions of very low probability for both landmarks and usual legal cases. As perspective for future works, we will propose a temporal analysis of the citation network as well as the introduction of other centrality measures in order to better characterize legal cases.

Acknowledgments

We gratefully acknowledge CNPq, CAPES, FUNCAP, BNB, and the Edson Queiroz Foundation for financial support.

References

[1] J. V. Calvi, S. E. Coleman, S. Coleman, S. Coleman, American law and legal systems, Prentice Hall, 2000.

[2] F. Schauer, Is the common law law, HeinOnline, 1989.

[3] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, et al., Science of science, Science 359 (2018). URL: https://doi.org/10.1126/science.aao0185.

[4] R. Whalen, Legal networks: The promises and challenges of legal network analysis, Mich. St. L. Rev. (2016) 539.

[5] I. Carmichael, J. Wudel, M. Kim, J. Jushchuk, Examining the evolution of legal precedent through citation network analysis, NCL Rev. 96 (2017) 227.

[6] P. Zhang, L. Koppaka, Semantics-based legal citation network, in: Proceedings of the 11th international conference on Artificial intelligence and law, 2007, pp. 123–130.

[7] M. Newman, Networks, Oxford university press, 2010.

[8] H. L. School, Caselaw access project, 2021. URL: https://api.case.law/v1/citations/.

[9] C. L. School, Historic supreme court decisions, 2021. URL: https://www.law.cornell.edu/supct/cases/name.htm.

[10] G. Chen, X. Wang, X. Li, Fundamentals of complex networks: models, structures and dynamics, John Wiley & Sons, 2014.

[11] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, Introduction to algorithms, MIT press, 2009.

[12] S. Redner, How popular is your paper? an empirical study of the citation distribution, The European Physical Journal B-Condensed Matter and Complex Systems 4 (1998) 131–134.

[13] X. Li, H. Chen, Z. Huang, M. C. Roco, Patent citation network in nanotechnology (1976–2004), Journal of Nanoparticle Research 9 (2007) 337–352.

[14] Y.-H. Eom, S. Fortunato, Characterizing and modeling citation dynamics, PloS one 6 (2011) e24926.

[15] A. Clauset, C. R. Shalizi, M. E. Newman, Power-law distributions in empirical data, SIAM review 51 (2009) 661–703.