

Performance in the Courtroom: Automated Processing and Visualization of Appeal Court Decisions in France

Paul Boniol
paul.boniol@polytechnique.edu
LIX, École Polytechnique
Paris, France

George Panagopoulos
george.panagopoulos@polytechnique.edu
LIX, École Polytechnique
Paris, France

Christos Xypolopoulos
christos.xypolopoulos@polytechnique.edu
LIX, École Polytechnique
Paris, France

Rajaa El Hamdani
el-hamdani@hec.fr
HEC Paris
Paris, France

David Restrepo Amariles
restrepo-amariles@hec.fr
HEC Paris
Paris, France

Michalis Vazirgiannis
mvazirg@lix.polytechnique.fr
LIX, École Polytechnique
Paris, France

ABSTRACT

Artificial Intelligence techniques are already popular and important in the legal domain. We extract legal indicators from judicial judgment to decrease the asymmetry of information of the legal system and the access-to-justice gap. We use NLP methods to extract interesting entities/data from judgments to construct networks of lawyers and judgments. We propose metrics to rank lawyers based on their experience, wins/loss ratio and their importance in the network of lawyers. We also perform community detection in the network of judgments and propose metrics to represent the difficulty of cases capitalising on communities features.

CCS CONCEPTS

• Applied computing → Law; • Computing methodologies → Information extraction.

KEYWORDS

Natural Language Processing, Named Entity Recognition, Graph Mining, Network Analysis, Case-Law Analysis, Legal Text

1 INTRODUCTION

Recent advances in Artificial Intelligence (AI) and Natural Language Processing (NLP) allow the analysis of large numbers of legal documents in aggregate in contrast to traditional methods. A long-standing application of NLP to legal documents is information extraction and retrieval from judicial decisions. The interest in mining data from judgments can be explained by the critical role they play in the administration of justice in both common and civil law systems. The objective of our work is to analyze judgments by French courts to gain insights about the operation of the French judicial system, which could in turn help developing an interface for laypersons. As explained in [22], a legal user interface could shield the user of the legal system from the complexity of the underlying legal system. Ordinary people perceive the legal system as too complex [2], which results in part from the asymmetry of information in the market of legal services, where ordinary people are disadvantaged in comparison with providers of legal services [1]. The asymmetry of information adds to the access-to-justice gap, such that a layperson lacks the right information and tools to choose the right lawyer at an affordable cost, and might prefer to self represent herself or refrain from filing a lawsuit. According to

[10] "one of six Americans is a self-represented litigant in a newly filed case each year," however, the resolutions are in favor of litigants represented with a lawyer. Both [1, 10] suggests "the ease of access to information" is a solution to address the gap in accessing justice. Access to free basic legal information could help the user to navigate the justice system easily, understand better the legal area his problem falls into, and choose a lawyer with experience on the subject matter of the dispute. In our work, we extract and represent information from past judgments to increase the transparency of judicial procedures and make them more accessible to laypersons. First, we pre-process judgments by extracting relevant legal entities, such as the lawyers of each party, by using Named Entity Recognition (NER) models. Second, we analyze the win/loss rate of lawyers by building two lawyers' networks: an opposing network of lawyers and a collaborative network of lawyers. Third, we use network analysis of judgments to suggest a measure of case difficulty based on case types/communities with distinct win/lose rates.

2 RELATED WORK

Numerous research have been carried out on case-law corpora focusing on specific objectives. One of the long-time objectives is the prediction of case outcomes. One of the first approaches in this field was [14] to manually convert a case factual elements into numerical values, compute their sum and predict a decision in favor of the petitioner if the sum is above a manually selected threshold. Recent efforts [6, 13, 15, 24] have used machine learning techniques to build outcome prediction models. Judicial judgments are rich in data, which could be used to analyze the operations of the legal system. The authors of [8, 19] used empirical methods to understand and describe judicial decision-making. Other researchers extract information to empower legal decision-makers and legal practitioners [11, 16, 17, 26]. Judicial decisions lend themselves to the use of network analysis techniques. Networks of case law have been used several times [7, 9] to measure the importance of a case. The theory of graphs provides tools well adapted to analyze the complexity of case law networks; for example, [25] employs a hybrid version for bipartite graphs to clarify procedural aspects of the International Criminal Court. Judgments are expressed in natural language, therefore to scale their automatic processing, several researchers have been developing natural language processing techniques for the legal domain. Some adapt NLP techniques built for the general

language to the legal language. [23] build their model of sentence boundary detection (SBD) for legal documents. Researchers from the Lynx project [20] developed a set of NLP services to extract a variety of information from legal documents: term extraction, text structure recognition, and NER. NER techniques have several applications in the legal domain. [5] improved existing NER models and used the resulting models to extract, from French judgments, entities that should be anonymized before the public release of the judgments.

3 DATA

3.1 Data collection

Our dataset consists of 40.000 rtf files that were crawled through L'Aligifrance¹, a French legal publisher providing access to law codes and legal decisions. To navigate and crawl through L'Aligifrance we used Selenium², a python framework that simulates a real web browser. For our experiments, we used a sample of cases from the court of appeal consisting of 17.215 cases. We limit our first analysis to cases from the court of appeal due to the specificity of cases from trial courts and the Court of Cassation. For future works and analysis, a sample of cases from the Court of Cassation could also be used (more than 400.000 documents available on L'Aligifrance). We decided to focus first on cases decided by civil courts and to exclude both criminal, administrative, and specialized courts. We also remove procedural judgments, such as court orders. Judgments analyzed here are solely final decisions called "arrêt de Cour d'appel."

3.2 Data preprocessing

Data preprocessing was the most challenging part of the project. The structure and wording of the legal documents, which vary between different courts and dockets, as well as the use of legal formal language, were challenging obstacles to conduct the text mining tasks. Below we analyze in detail how we approached each part, from segmenting the documents to extracting the persons taking part in each court case and their roles.

3.2.1 Segmentation.

Analysis of the macrostructure of cases. The decisions of courts of appeals in France follow an overall similar structure. First, the documents state practical information about the litigation such as dates, jurisdiction, and the different entities involved in the trial, listed in the following order:

- **Appellant (appelant in French):** The name of the party is always anonymized, for example: "Monsieur Jean X."
- **Appellant's counsel:** can be anonymized but always start with the keyword "Me" or "Maître," for example, "Me Jean Dupont."
- **Appellee (intimé in French):** this entity has the same format as the appellant.
- **Appellee's counsel:** Same format as the appellant's counsel.
- **Court Entities** (non-fixed order):

- **Judge (magistrats, conseillers in French):** could be anonymized but always start with the keyword "Président."
- **Clerk (Greffier in French):** Can be anonymized but is always expressed near the word "Greffier."

After listing the entities, decisions from French appeal courts continue with the debate. The debate describes all the facts and procedure leading to the appeal. It also states the different arguments brought forth by the parties, and follows with the reasoning of the court. Finally, it closes with the conclusion which states the final decision. The keywords separating the different parts vary significantly, and are sometimes absent, which makes the segmentation task complex. Keywords may vary from one appeal court to another. We use graphs to compare the structure of judgments of appeal courts in several territorial jurisdictions. Figure 1 represents the flow of cases in three different jurisdictions. Each graph is built by parsing judgments from the same jurisdiction into sentences and then linking consecutive sentences by an edge. The name of a node is the text of the sentence. When the sentence is more than five words, the name of the node is "Long_Text_i_j" where i is the index of the case in the whole dataset, and j is the index of the sentence within the case. The size of the node is the occurrence of its text in all the judgment of the considered jurisdiction. To account for keywords that have small variations across documents, we use the Jaro distance³ to identify these variations. The Jaro distance is used to contract nodes of similar sentences, such that if two sentences have a Jaro distance larger than 0.8, then they are considered belonging to the same node. Therefore the big nodes are common parts from all documents, and therefore they represent the structure of these documents.

Figure 1 points out the difference in structure and flows that documents from different jurisdictions can have. For instance, Agen will use "ENTRE" to announce the appellant, and "ET" to announce the appellee, whereas Douai will use respectively "APPELANT" and "INTIMEE."

Nevertheless, we empirically observed that all the decisions, whatever the jurisdiction was, shared the same keyword "PAR CES MOTIFS" to announce the final decision of the court (last big node in the three flows of figure 1).

Segmentation with keywords. We also sought to extract entities corresponding to the lawyers defending each party. As described above, legal entities are mentioned after the practical information in a fixed order. Moreover, domain experts confirmed these legal entities are mentioned in separate segments. These segments are often preceded by known keywords, as shown in figure 2. Once we have identified the beginning and end of each segment, we use them to extract lawyers' names as described in the following subsection.

3.3 Extraction of lawyers' entities

To detect lawyers' names throughout the document, we discard, first, all segments except the appellant and the appellee segments. Second, we segment them further into sentences using the sentence

¹www.legifrance.gouv.fr

²https://selenium-python.readthedocs.io/

³The Jaro distance is a similarity measure between two strings [12]

Table 1: Examples of Jaro distance between pairs of similar sentences contracted to a same node

Sentence 1	Sentence 2	Jaro distance
faits et procedure	faits procedure	0.86
procedure et pretentions des parties	procedure et moyens des parties	0.83
moyens et pretentions des parties	pretentions et moyens des parties	0.92

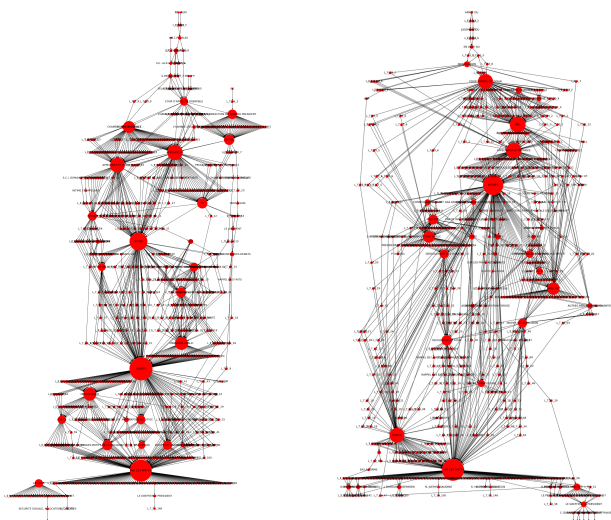


Figure 1: Most recurrent Keywords in decisions by courts in Grenoble, and Douai respectively. The nodes are part of the document. When the sentence is more than five words, the name of the node is attributed to Long_Text_i with i unique. Therefore, the big nodes are common parts from all documents, and therefore the structure.

tokenizer by Polyglot ⁴, which is a Python package providing multilingual natural language processing tools. Third, we only keep sentences containing honorifics used for lawyers such as "Me," "MaÃtre," or lawyers' keywords like: "representÃr par." We then use a well-established Named Entity Recognition model by Polyglot [3] to recognize person entities from the remaining sentences. The model uses pretrained word embeddings from Wikipedia [4] to classify whether a word is an entity or not based on its sentence. Last, we consider the extracted named entities appearing in the appellant segment as lawyers of the appellant, and the names appearing in the appellee segment as lawyers of the appellee. It should be noted that decisions without any reference to a lawyer on both sides were overlooked.

3.4 Extraction of the judge decision

From the initial segmentation, the final decision of the court is to be found in the conclusion segment of the judgement. Concerning

⁴<https://polyglot.readthedocs.io/>

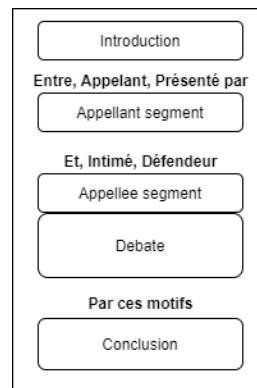


Figure 2: Segmentation of a legal case based on keywords, to facilitate entity recognition

judgments from appeal courts, the court will either confirm the first lower court decision (Tribunal judiciaire) or reverse it. However, the court can also partially confirm the judgment. In other words, the court can decide to accept one of the appellant's requests, and therefore change the first decision partially. Empirically, we noticed that certain words are present in certain types of decision, and after validation from the domain experts, we resorted to a keyword-based solution:

- "Confirme", "Rejete", "Irrecevable": keep the first decision (Appellee "wins")
- "Infirme", "Rectifier", "RÃlforme": change the first decision (Appellant "wins")

Out of a sample of 5832 cases, 570 conclusions (~10%) include at least a keyword representing both outcomes, in which case we keep the outcome that has most keywords. This is a temporary solution that requires refinement in the future.

4 NETWORK ANALYSIS OF LAWYERS

Once the entities' recognition is complete, we extract all the instances (and their function) in every document. Since courts tend to have a limited number of lawyers, judges and court clerks, cases share the same entities. Therefore, all the cases can be considered a big graph where entities interact with each other.

4.1 Opposing network of lawyers

We extracted the winning and losing lawyers in each decision. From this, we can define a directed weighted network. We draw an edge between lawyers if they have been opposed. The edge from lawyer

i to lawyer j is weighted by the wins $wins_{i,j}$ of lawyer i to lawyer j:

$$wins_{i \rightarrow j} = aw_{i \rightarrow j, acc} + bw_{i \rightarrow j, def}$$

Where $w_{i \rightarrow j, acc}$ is the number of wins of lawyer i as an appellant and $w_{i \rightarrow j, def}$ is the number of wins of lawyer i as an appellee. Parameters 'a' and 'b' are used to weigh more winning as an appellant than winning as an appellee since it is known by legal experts that the event of winning an appeal is less frequent than losing it. We also confirm this intuition by counting the rate of appeals' rejection from our dataset. We get a rejection rate of 0.9. We collapse both edges between two lawyers into one directed edge weighted by: $|wins_{i \rightarrow j} - wins_{j \rightarrow i}| \log(wins_{i \rightarrow j} + wins_{j \rightarrow i} + 1)$. In this case the edge direction is determined by the sign of $wins_{i \rightarrow j} - wins_{j \rightarrow i}$ such that the edge target is the lawyer with most wins. To visualize the most important nodes, we remove lawyers with only one case (899 out of 2146), which leaves us with a network with 1247 nodes and 2182 edges. The resulting network appears in figure 3. The

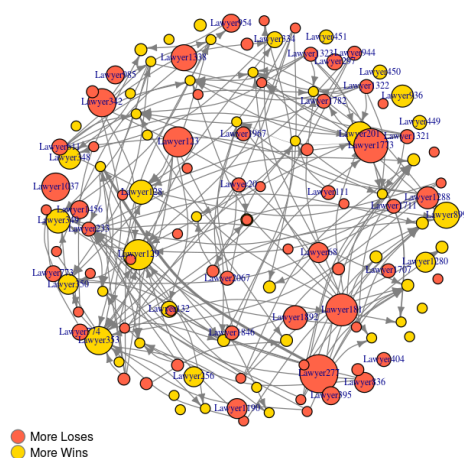


Figure 3: The network of lawyers that have opposed in appeal cases. The width of the edge is defined by the number of wins the source node has over the target. The node color defines whether the lawyer has more losses or more wins, and the size is analogous to total number of cases, large yellow nodes mean the lawyer has won much more cases than he lost and vice versa.

edge goes from a "losing" to a "winning" lawyer, and the width of the edge represents the difference in the number of wins. Node size is the number of appeal cases where the lawyer appears and the color captures the win-loss difference.

4.2 Collaboration network of lawyers

The collaboration network in figure 4 indicates lawyers that have been on the same side during an appeal case. The edges are weighted

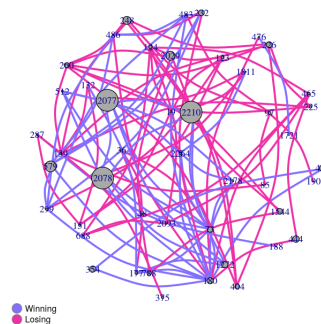


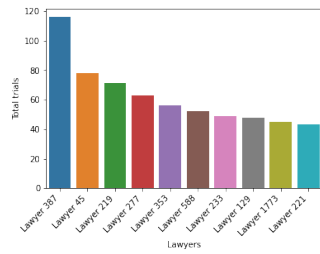
Figure 4: Network of lawyers that have collaborated in court cases. The edge color signifies the sign of the win-loss metric and the width of the absolute value.

based on the wins minus the losses, so the network can capture which collaborations are the most successful. We have removed nodes with number of collaborations below a fixed threshold to obtain a decluttered visualization of the network. We obtain a network of 47 nodes and 94 edges out of 2182 nodes and 2950 edges.

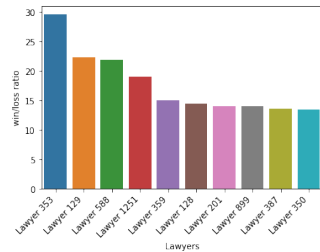
4.3 Lawyers Ranking

In this section, we suggest three metrics to rank and compare between lawyers. First, we measure the experience of a lawyer by the number of judgments mentioning him as the appellant's or appellee's lawyer. Second, we compute the win-loss rate of lawyers. Third, we calculate the centrality of a lawyer in the opposing network.

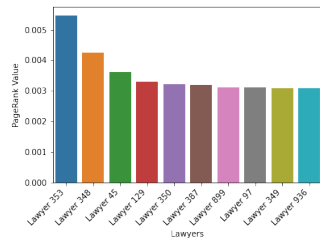
In figure 5a, lawyers are ranked by their experience in going in front of the court of appeal. However, this measure alone does not indicate the performance of the lawyer. Thus we need to refer to the win/loss ratio to evaluate the performance. Lawyer 353 ranks first in terms of the win/loss ratio instead of fifth in terms of the total number of cases. In this case, lawyer 353 performs better than lawyer 387, who is ranked first in figure 5a but ranks 9 in terms of win/loss ratio. A weakness of the win/loss ratio ranking is that it does not consider the experience of the opposing lawyer; while the opponent's worth can be a measure of the win's value. To this end, we compute the weighted directed PageRank of the opposing network 5c. As explained in section 4.1, the weights are the number of wins such that wins as an appellant's lawyer counts more than wins as an appellee's lawyer. So edges directed towards lawyers who win more representing an appellant have higher weights than edges directed towards a lawyer who wins more representing an appellee. Therefore top lawyers in figure 5c are lawyers who won against experienced lawyers and who win most as an appellant's lawyer. Lawyer 387 is ranked best than lawyer 350 in terms of win/loss ratio, but worst in terms of PageRank measure. We could explain this difference in the ranking by the fact that the majority of wins of lawyer 350 wins as an appellant's lawyer, while the majority of wins of lawyer 387 wins as an appellee's lawyer. Thus



(a) Lawyers ordered by the total number of cases



(b) Lawyers ordered by the win/loss ratio



(c) Lawyers ordered by their importance using PageRank algorithm

Figure 5: Ranking of lawyers using three different measures

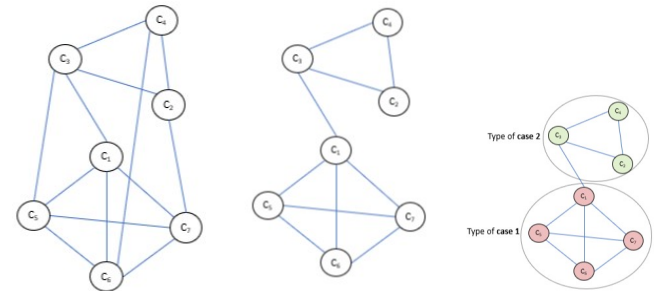
it is recommended for an appellant to choose lawyer 350 rather than lawyer 387.

4.4 Network analysis of judgments

In this section, we develop a method to assess cases' difficulty from the perspective of the appellant. More precisely, the aim is to compute the difficulty withing a group of cases dealing with the same legal issues. First, we built a network of cases to discover communities of cases about similar legal issues. Second, we use the win/loss rate of the appeal as a proxy to its *difficulty*.

Graphs encode knowledge and patterns more efficiently [18, 21]. The crucial element is the edges representing some kind of similarity/affiliation among the nodes. In order to build a graph of cases, we needed to connect them with some property that represented similarity. The profound way we came up with was that the similarity of two cases is defined by the number of common law articles mentioned in the text of the cases reflecting apparently the thematic similarity among them. Thus we build a network of judgments to discover the communities' structure and natural divisions among

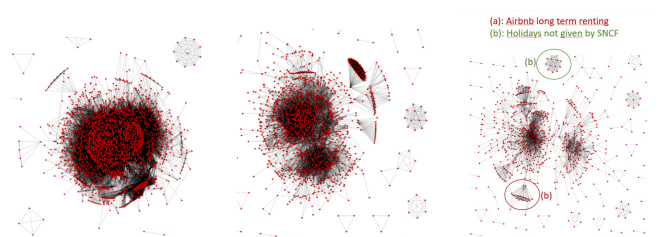
the set of studied cases. First, we prepare cases by extracting cited articles of law. We extract articles by using regular expressions. Then we create an edge between two cases if they cite at least k same articles. Figure 6 shows graphs of judgments for different values of k. It is evident that as we increase k the graph becomes smaller with the cases having higher similarity due to the higher number of common articles.



(a) Cases with at least 3 articles in common (b) Cases with at least 5 articles in common (c) A case graph displaying community structure: two groups of cases with dense internal connections and sparser connections between groups

Figure 6: Examples of cases graphs and communities for a number of 7 cases

We built networks for different values of k from cases of the last three months of 2018, as shown in figure 7. The network naturally groups similar cases in communities. For example, in figure 7c cases against the same appellee and about the same issue. We also notice, in figure 7c, that cases with the same win/loss rate are grouped in the same communities.



(a) Cases with at least 3 articles in common (b) Cases with at least 5 articles in common (c) A case graph displaying community structure: two groups of cases with dense internal connections and sparser connections between groups

Figure 7: Examples of cases graphs and communities for a number of 7 cases

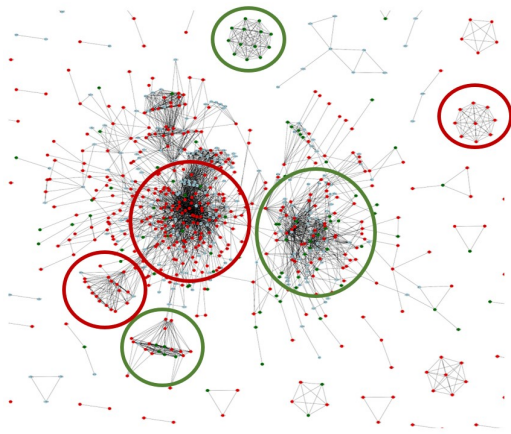


Figure 8: Examples of detected communities. Communities circled in red have a high losing rate. Communities circled in green have a high winning rate.

5 CONCLUSION

We used NLP methods to extract information from judgments of the French court of appeal. We constructed indicators about the difficulty of lawyers' performance and cases by using network analysis techniques on lawyers' networks and cases' networks. Our objective is to use these indicators to guide laypersons when confronted with the legal systems and contribute to the decrease of the access-to-justice gap by reducing the asymmetry of information characterizing the legal market. The lawyers' ranking could serve to build a system that guides an appellee in choosing a lawyer. However, the lawyers' ranking relies only on wins and losses of lawyers. In future work, we expect to produce a ranking that takes into account the legal area of the case and its difficulty, in such a way that the ranking could be more personalized to the needs of a layperson.

REFERENCES

- [1] 2016. Access to Justice and Market Failure. *Slaw* (November 2016). <http://www.slaw.ca/2016/11/01/access-to-justice-and-market-failure/>
- [2] 2016. *Understanding Effective Access to Justice*. Retrieved April 4, 2020r from <http://www.oecd.org/gov/Understanding-effective-access-justice-workshop-paper-final.pdf>
- [3] Rami Al-Rfou, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2015. PolyglotNER: Massive multilingual named entity recognition. In *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 586–594.
- [4] Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2013. Polyglot: Distributed word representations for multilingual nlp. *arXiv preprint arXiv:1307.1662* (2013).
- [5] Valentin Barriere and Amaury Fourret. 2019. May I Check Again?—A simple but efficient way to generate and use contextual dictionaries for Named Entity Recognition. Application to French Legal Texts. *arXiv preprint arXiv:1909.03453* (2019).
- [6] Karl Branting, B Weiss, B Brown, C Pfeifer, A Chakraborty, L Ferro, M Pfaff, and A Yeh. 2019. Semi-Supervised Methods for Explainable Legal Prediction. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*. 22–31.
- [7] Mattias Derlén and Johan Lindholm. 2014. Goodbye van G end en L oos, Hello B osman? Using Network Analysis to Measure the Importance of Individual CJEU Judgments. *European Law Journal* 20, 5 (2014), 667–687.
- [8] Lee Epstein, William M Landes, and Richard A Posner. 2013. *The behavior of federal judges: a theoretical and empirical study of rational choice*. Harvard University Press.
- [9] James H Fowler, Timothy R Johnson, James F Spriggs, Sangick Jeon, and Paul J Wahlbeck. 2007. Network analysis and the law: Measuring the legal importance of precedents at the US Supreme Court. *Political Analysis* 15, 3 (2007), 324–346.
- [10] John M Greacen, Amy Dunn Johnson, and Vincent Morris. 2014. From market failure to 100% access: Toward a civil justice continuum. *UALR L. Rev.* 37 (2014), 551.
- [11] Jerrold Soh Tsin Howe, Lim How Khang, and Ian Ernst Chai. 2019. Legal Area Classification: A Comparative Study of Text Classifiers on Singapore Supreme Court Judgments. *arXiv preprint arXiv:1904.06470* (2019).
- [12] Matthew A Jaro. 1989. Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *J. Amer. Statist. Assoc.* 84, 406 (1989), 414–420.
- [13] Daniel Martin Katz, II Bommarito, J Michael, and Josh Blackman. 2014. Predicting the behavior of the supreme court of the united states: A general approach. *arXiv preprint arXiv:1407.6333* (2014).
- [14] Fred Kort. 1957. Predicting Supreme Court decisions mathematically: A quantitative analysis of the "right to counsel" cases. *American Political Science Review* 51, 1 (1957), 1–12.
- [15] Shangbang Long, Cunchao Tu, Zhiyuan Liu, and Maosong Sun. 2019. Automatic judgment prediction via legal reading comprehension. In *China National Conference on Chinese Computational Linguistics*. Springer, 558–572.
- [16] Dennis P Michalopoulos, Jessica Jacob, and Alfredo Coviello. 2019. AI-Enabled Litigation Evaluation: Data-Driven Empowerment for Legal Decision Makers. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*. 264–265.
- [17] Wai Yin Mok and Jonathan R Mok. 2019. Legal Machine-Learning Analysis: First Steps towards AI Assisted Legal Research. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law*. 266–267.
- [18] Giannis Nikolentzos, Antoine J-P Tixier, and Michalis Vazirgiannis. 2019. Message Passing Attention Networks for Document Understanding. *arXiv preprint arXiv:1908.06267* (2019).
- [19] Jeffrey J Rachlinski and Andrew J Wistrich. 2017. Judging the judiciary by the numbers: Empirical research on judges. *Annual Review of Law and Social Science* 13 (2017), 203–229.
- [20] Georg Rehm, Julian Moreno Schneider, Jorge Gracia, Artem Revenko, Victor Mireles, Maria Khvalchik, Ilan Kernerman, Andis Lagzdins, Mărcis Pinnis, Artus Vasilevskis, et al. 2019. Developing and orchestrating a portfolio of natural legal language processing and document curation services. In *Proceedings of the Natural Legal Language Processing Workshop 2019*. 55–66.
- [21] François Rousseau and Michalis Vazirgiannis. 2013. Graph-of-word and TW-IDF: new approach to ad hoc IR. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. 59–68.
- [22] JB Ruhl and Daniel Martin Katz. 2015. Measuring, monitoring, and managing legal complexity. *Iowa L. Rev.* 101 (2015), 223.
- [23] George Sanchez. 2019. Sentence boundary detection in legal text. In *Proceedings of the Natural Legal Language Processing Workshop 2019*. 31–38.
- [24] Octavia-Maria Sulea, Marcos Zampieri, Mihaela Vela, and Josef Van Genabith. 2017. Predicting the law area and decisions of french supreme court cases. *arXiv preprint arXiv:1708.01681* (2017).
- [25] Fabien Tarissan and Raphaëlle Nollez-Goldbach. 2016. Analysing the first case of the international criminal court from a network-science perspective. *Journal of Complex Networks* 4, 4 (2016), 616–634.
- [26] Thomas Vacek, Ronald Teo, Dezhao Song, Timothy Nugent, Conner Cowling, and Frank Schilder. 2019. Litigation Analytics: Case outcomes extracted from US federal court dockets. In *Proceedings of the Natural Legal Language Processing Workshop 2019*. 45–54.