

# An Exploratory Analysis of Precedent Relevance in the Brazilian Supreme Court Rulings

Fernando A. Correia  
PUC-Rio  
Rio de Janeiro, RJ, Brazil  
fjunior@inf.puc-rio.com

José Luiz Nunes  
PUC-Rio  
Rio de Janeiro, RJ, Brazil  
jnunes@inf.puc-rio.com

Guilherme F. C. F. de Almeida  
Fundação Getulio Vargas  
Rio de Janeiro, RJ, Brazil  
guilherme.almeida@fgv.br

Alexandre A. A. Almeida  
PUC-Rio  
Rio de Janeiro, RJ, Brazil  
alex.aaa@aluno.puc-rio.com

Hélio Lopes  
PUC-Rio  
Rio de Janeiro, RJ, Brazil  
lopes@inf.puc-rio.com

## ABSTRACT

The new Brazilian Code of Civil Procedure (CPC) has elevated the importance of precedents in the legal decision-making process. This increased the need to find relevant precedents for a given issue or dispute. Precedents play a central role in judicial thinking by providing information to judges about the legal relevance of particular facts and by establishing legal rules. Precedents are also an important argumentative tool, enabling lawyers to present arguments based on previous decisions. The automated search for relevant precedents is an unattended issue in the Brazilian scenario, partly due to the court's massive production of decisions — only in 2018 the Brazilian Supreme Court (STF) produced more than 121.000 new rulings — and partly due to the technical challenges arising from the unstructured nature of the court's practices. In this paper, we present a study of precedent relevance, taking into account the uniqueness of the Brazilian legal system and of STF. To do so, we conducted an exploratory investigation over the precedent network extracted from 1.152.963 decisions published by the STF between 2008 and 2018. This exploratory analysis, although interesting in itself, reveals important challenges that need to be overcome by future research in order for the technology to have the kind of impact it can have on legal practice and academia. In our conclusion, we set out possible paths forward, briefly considering some of the most promising ways to sort out the signal from the noise.

## CCS CONCEPTS

• **General and reference** → **Evaluation**; • **Applied computing** → **Law**; Document searching; • **Information systems** → **Link and co-citation analysis**; • **Computing methodologies** → **Information extraction**; • **Mathematics of computing** → **Exploratory data analysis**.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

Berlin '19, September 23–26, 2019, Berlin, BE

© 2019 Association for Computing Machinery.

ACM ISBN 978-1-4503-6887-2/19/09...\$15.00

<https://doi.org/10.1145/3342558.3345416>

## KEYWORDS

Legal Rulings, Legal Documents, Precedents, Network Analysis, Legal Arguments, Brazilian Law

### ACM Reference Format:

Fernando A. Correia, José Luiz Nunes, Guilherme F. C. F. de Almeida, Alexandre A. A. Almeida, and Hélio Lopes. 2019. An Exploratory Analysis of Precedent Relevance in the Brazilian Supreme Court Rulings. In *ACM Symposium on Document Engineering 2019 (DocEng '19)*, September 23–26, 2019, Berlin, Germany. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3342558.3345416>

## 1 INTRODUCTION

The new Brazilian Code of Civil Procedure (CPC) [2], was created to increase legal certainty, predictability, isonomy, and rationality, bring more efficiency to the Brazilian legal system, and to thereby decrease the judiciary's workload [4]. To reach those goals it is necessary to respect court's precedents. Precedents stems from the *stare decisis* doctrine (to stand by things decided) according to which courts are bound by its previous rulings (*horizontal* precedent, see [12]). Courts are also bound by decisions of hierarchically superior courts (*vertical* precedent, [12]). The idea of precedent, specially horizontal precedent, is one of the defining traits of the common law system, under which courts are supposed to decide cases by looking to past precedents in a given area of the law [7] and to incrementally develop ever-improving rules based on concrete cases [1]. Distancing itself more from the idea that judges are bound only by statutes and formal rules — the prototypical civil law trait —, the new CPC takes the Brazilian legal system one step closer to common law.

In jurisdictions like the U. S., Canada, and India — all of which are common law systems — precedent relevance detection is an issue that has attracted considerable attention. Studies like [7], [14], [10] and [8] have proposed measures of the relevance of precedents based on common techniques used in Network Science, where decisions and its relationships are represented in network structures, also dubbed citation networks, where nodes stand for decisions and links depict citations between them. Two decisions  $x$  and  $y$  are directly connected if  $x$  cites  $y$ , that is, if the decision  $x$  refers to  $y$  in its text. Unfortunately, the techniques deployed in those studies cannot be directly applied to the context of the Brazilian legal system. As we will see, there are important differences between

the way Brazilian courts and common law courts deal with precedents, creating the need to develop novel methods to quantitatively measure the relevance of precedents.

These efforts are, however, still in its infancy, as legal precedents – specially horizontal – have only recently become the focus of systematic interest in the Brazilian legal community. In this paper, we focus on decisions issued by the Brazilian Supreme Court, *Supremo Tribunal Federal* (STF). STF renders upwards of 100,000 decisions per year and most of those decisions cite precedents. The absence of effective filters for reaching STF, as well as the multiple natures of cases decided by the court [5] creates a highly heterogeneous body of citations, in addition to a much larger citation network than other Supreme Courts. Understanding the use of precedents in STF can teach us important lessons with regards to Brazilian law.

With this in mind, we have conducted an exploratory investigation over the precedent network extracted from decisions published by STF between 2008 and 2018. After analyzing measures already found in the literature, we describe challenges for possible future work on the subject.

We organize this paper in five sections. Section 2 introduces the description of the citation network building process and some relevant network properties. Section 3 explores the role of precedents in STF and its development throughout time. The problem of precedent relevance is discussed in Section 4, through both qualitative and quantitative points of view, revealing important challenges that need to be overcome in the future. Finally, our conclusion sets out possible paths forward, briefly considering some of the most promising ways to sort out the signal from the noise.

## 2 THE NETWORK REPRESENTATION

Our work draws from the results of the *Supremo em Números* Project (Supreme Court by the Numbers), a research group of the *FGV Direito Rio* law school, which, among other things, created and maintains the database used in this study. The citation network used is the result of the extraction of citations from the 1,152,963 decisions published by the STF between 2008 and 2018. Those decisions are classified by the Court in 57 different categories, known as procedural classes.

It must be stressed that the most logical path to create a citation network would be to use each decision as a node. However, this was not feasible. The citation extraction algorithm was limited by the writing patterns used by the Justices for precedent citation. Unlike the U. S., there is no formal standard procedure for precedent citation in the Brazilian judiciary. As such, we were not able to reliably identify which specific decision from a given procedure was being cited in the text. In order to circumvent this limitation, we have opted to group all opinions issued in a case, and use cases as nodes. Since each case can have multiple decisions, this method creates some distortions in our network. In average, 1.24 decisions are rendered on each STF case (with  $\sigma = 0.67$ ), where 65.32% of those leave the court after the first decision and 23.77% after the second decision. To minimize error in nodes' time-stamps, we used the date of the case's first decision as the node date.

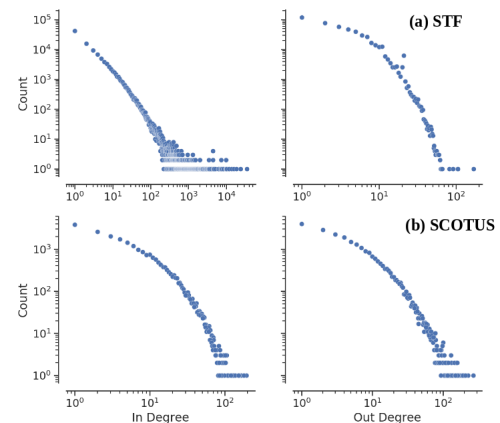
Another issue caused by this is that our citation network – which are typically acyclical – contains cycles. If the first decision of case  $a$  cites the first decision of the case  $b$  and the second decision in  $b$ ,

which came after  $a$ , cites the first decision in  $a$ , there will be a cycle between those two nodes.

### 2.1 Network Properties

The initial version of the network was composed of 906,843 nodes, 40.0% of which have no connection to any other node. Since no information regarding precedent relevance can be extracted from these nodes we shall not care for them in this analysis. Thus, the final network representation ended up with 543,759 nodes connected by 2,404,616 of edges, where 90.6% of the nodes cite at least another node and 21.8% are cited at least once. Despite the number of edges we have very sparse graph with a density of  $8.13 \cdot 10^{-6}$  – for comparison, according with [6], the Supreme Court of the U. S. (SCOTUS) has a density of  $2.4 \cdot 10^{-4}$ , 2 orders of magnitude higher.

Most decisions cite relatively few other decisions, while a minority of decisions cite many precedents, a fact that can be observed in Figure 1.a. Similarly, the vast majority of the decisions are cited by relatively few other, while a small number of decisions are cited very often. A formal hypothesis test was not made as to whether it is plausible to see the curves as a power law distribution coming from a preferential attachment model, although the distribution does resemble these networks.



**Figure 1: Distribution of inward and outward citations in the STF network, 2008-2018, and in comparison with the SCOTUS, 1791-2005 (Adapted from [6]).**

For a better understanding about the connectivity of the graph, we also observed the number of weakly connected components. The network shows 1,911 weakly connected components, where we found one big component with 539,200 nodes, surrounded by components with 2.39 nodes in average ( $\sigma = 1.29$ ). As stated before, our graph is not a directed acyclic graph. We can verify this by the existence of strongly connected components, which by itself imply the existence of directed cycles. We found as much as 586 strongly connected components with more than one node, the biggest one being composed of as much as 1,454 nodes. Beside this one, the other components have on average 3.50 nodes ( $\sigma = 6.65$ ), and from those, 559 were part of the biggest weakly connected component.

### 3 PRECEDENT AND TIME

As suggested by [6] the number of relevant precedents grow over time and an increase in the number of citations is also expected. An interesting aspect to investigate is for how long decisions are cited. This would help understand the life cycle of cases, or how a given case “ages” and for how long it stays relevant. As a first insight into the time dimension of the network, the Figure 2 displays the distribution of the term the opinion was issued (year of the precedent creation) in respect to the years cited (when the precedent cited was issued). For illustration, the Figure shows that in 2016, the majority of precedents cited in that year were created in 2015 and 2013.

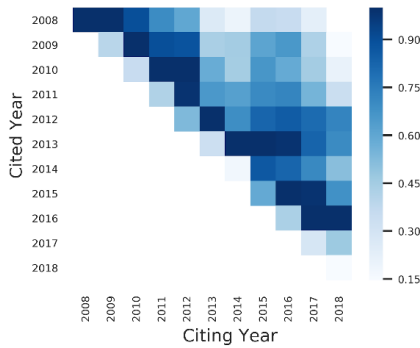


Figure 2: Number of citations to each year.

The Figure 2 states that, except for 2015 and 2018, the year following its creation is when the precedent has higher chance to be used. We can also assert that in any given year, justices are more likely to cite cases from recent years, an indication that decisions might have a short life cycle, with most citations being received right after it is issued.

As mentioned in the first section, the STF rules on a wide variety of cases due to its broad jurisdiction. To tease apart those jurisdictions, we have grouped cases and decision in five main groups:

- Constitutional: proceedings that deal with the constitutionality of laws in abstract;
- Appeal: appeals that seek to reverse rulings from lower courts on constitutional grounds, analogous to SCOTUS decisions;
- Criminal: when the Court acts as a regular fact finding case court for some federal public officials;
- Wrist: decisions on writs, including those judged directly and on appeal which do not require constitutional review arguments;
- Others: All remaining classes that did not fit the previous classifications.

Looking over the precedent life-cycle over this perspective – Figure 3 shows the distribution of the time between an opinion and its citation – the first conclusion we can draw is that indeed there are multiple citation profiles within the Court, and that different types of cases present distinct behaviour.

Appeal and Writ cases present a similar temporal behaviour, where most citations occur very early in the decision’s life. After

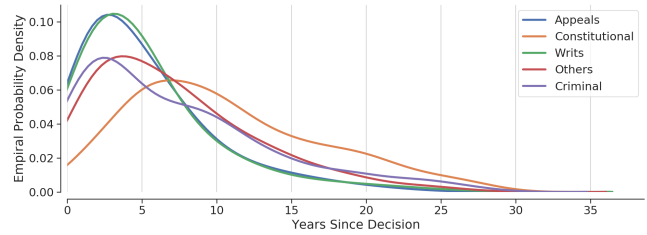


Figure 3: Precedents life-cycle per case’s type.

the first few years, there is a sharp decrease in citation to these decisions, which indicate that its content rapidly loses relevance on new cases’ decisions. On the other hand, the Constitutional persona of the Court presents a much larger tail, and it still receives a relevant amount of citations even 15 years later. In fact, most citations to those cases occur around 8 years after the decision. In comparison, the majority of citations to cases belonging to all other groups occur before that.

### 4 THE QUEST FOR RELEVANT PRECEDENTS

Before discussing the existing literature on quantitative measures for precedents relevance based on techniques used in Network Science, it is important to state clearly what counts towards the legal relevance of a precedent. Suppose that a lawyer meets with a new client, concerned with the legal treatment that the courts will likely give to his case. The client describes the facts that lead him to a law firm - the wrongful termination of a contract, for instance - and asks the lawyer: *What are my legal rights in this case?*. To provide a good answer, the lawyer must consult the sources of law. If one of these sources are legal decisions, he will need to go through the decisions of the competent court searching for precedents relevant to the case at hand.

We can describe this search in a semi-formal way. Given case  $C_1$ , with facts  $F_{c_1}, F_{c_2}, F_{c_3}$ , the most relevant precedent ( $P_1$ ) will be the one where the facts of the precedent ( $F_{p_1}, F_{p_2}, F_{p_3}$ ) are closest to those of the case ( $F_{c_1}, F_{c_2}, F_{c_3}$ ). There are, however, pragmatic constraints to this idea. Every legal case contains substantive (i.e., is a given speech act protected by free speech?) as well as procedural (was the appeal presented within the mandated deadline?) issues. Procedural issues cut across multiple legal areas. This can generate noise. Just for illustration, imagine a scenario where the lawyer wishes to answer a substantive question, such as the one posed by last paragraph’s client. If the lawyer’s search for precedents returns only procedural decisions, the search mechanism failed to return a relevant answer. As we’ll see, this is at play in STF.

On the quantitative dimension of relevance, we could use established techniques in Network Science, such as measures of network centrality such as PageRank or Authority and Hub scores. However, we do not know yet how well these metrics would be able to capture the legal relevance of cases. For a practical illustration of how these measures perform, we will use a simple experiment.

In order to do so, we apply a similar idea presented by Fowler [6], where we evaluated the correlation between each metric and future citations (after one year) of the case by the Court. The experiment result, presented in Table 1, shows no strong correlation between

any of those metrics. We have also included the degree as a comparison metric between the network statistics and straightforward number of citations made or received by the case.

**Table 1: Correlation between each metric and future citations.**

	Authority	Hub	PageRank	In-Deg	Out-Deg
<i>Pearson</i>	0.12	-0.00	0.33	0.33	-0.00
<i>Spearman</i>	0.43	-0.16	0.44	0.44	-0.08

## 5 CONCLUSIONS AND FUTURE WORKS

Traditional techniques to measure network centrality fail to properly capture legal relevance in the STF precedent network. Why is this so? We have a few hypotheses. STF, unlike most higher tier courts, does not have an effective entry filter [5]. As such, most of the court's workload consists of decisions stating that the court will not consider the case due to the absence of a constitutional issue underlying the dispute, or because the appeal makes reference to factual disagreements. These decisions are often called "defensive precedents"[3], as they are meant to protect the court from a overwhelming influx of cases. Defensive precedents are largely based on case law. As such, they are prominent in the STF's precedent network. A recent study [9] found that they dominate precedent citation rankings across all major branches of Brazilian law. The problem posed by this practice is that traditional centrality measures end up picking defensive precedents as the most relevant precedents for all kinds of different substantive issues, even when they are not at all legally relevant.

Another challenge arises from the way the court uses indirect citations. We suspect that, in certain cases, STF makes heavy use of indirect citation as a way to update its precedents without material change. Further research must quantify in what ways, if any, these idiosyncratic features of STF's practice impact Network metrics.

In the remainder of this conclusion, we briefly sketch out some avenues for research we envision as promising ways forward. We intend to pursue and evaluate each of these paths, looking for the best method to find legally relevant precedents.

Precedents may be modelled as infectious diseases. In this analogy, cases that settle important legal issues (as *Roe v. Wade* did for abortion law under the U.S. Constitution) are the patient zero of an outbreak that reaches — and controls — a multitude of posterior cases. Given that the field of epidemiology has used network analysis for decades in order to model disease spread [11], it is possible that the analogy will provide interesting insight into precedent networks. Precedents, as we've seen, are legal sources. Legal sources can be posited as major premises in canonically deductive legal arguments, see [13]. Identifying precedents, thus, may be the first step into an inquiry about the broader field of legal arguments. Understanding when precedents are indeed mentioned as part of an argument and when they're used in some other context may, in turn, prove helpful in reducing the noise in our data. By ignoring out-of-argument citations, we may end up with more sensible network data. We may also note how different precedents are used

concomitantly in support of the same conclusion. This captures another dimension of the relationship between precedents.

Finally, we may also look at clustering techniques. Preliminary evidence shows that STF uses standardization as a strategy to deal with the overwhelming number of cases that reach the court. The court renders decisions with almost identical texts over many cases dealing with similar issues. These similarities may be captured by certain metrics and used to identify clusters of similar decisions. This, in turn, allows for a more succinct representation of the precedent network, one which hopefully may prove useful in identifying legally relevant case law. It may be the case that one of these strategies proves dominant over the others. Perhaps modelling precedents as diseases is sufficient to solve the problem of measuring legal relevance. It may also be that each kind of procedure is best described by its own methods. Maybe truly constitutional procedures can be described in a perfectly adequate way by the techniques employed by [6], while appellate court workload demands a solution mindful of indirect citations. Our goal is to provide answers to these questions by systematically exploring each possibility in order to identify the truly relevant precedents of STF.

## 6 ACKNOWLEDGEMENTS

This study was supported by the National Council for Scientific and Technological Development (CNPq), the Getulio Vargas Foundation (FGV) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

## REFERENCES

- [1] Larry Alexander and Emily Sherwin. 2008. *Demystifying Legal Reasoning*. Cambridge University Press, Cambridge.
- [2] Brasil. 2015. LEI Nº 13.105, DE 16 DE MARÇO DE 2015, Código de Processo Civil. [http://www.planalto.gov.br/ccivil\\_03/\\_ato2015-2018/2015/lei/13105.htm](http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2015/lei/13105.htm)
- [3] Monica Bonetti Couto. 2018. O NOVO CPC E A (ESPERANÇA DE) SUPERAÇÃO DA JURISPRUDÊNCIA DEFENSIVA. *Revista Eletrônica de Direito Processual* 19, 3 (Dec 2018).
- [4] Elpidio Donizetti. 2015. A FORÇA DOS PRECEDENTES NO NOVO CÓDIGO DE PROCESSO CIVIL. *Direito UNIFACS – Debate Virtual* 0, 175 (Jul 2015).
- [5] Joaquim Falcão, Pablo de Camargo Cerdeira, and Diego Werneck Arguelles. 2012. *1 Relatório Supremo em Números - o Múltiplo Supremo*. Escola de Direito do Rio de Janeiro da Fundação Getulio Vargas, Rio de Janeiro. <http://bibliotecadigital.fgv.br/dspace/handle/10438/10312>
- [6] James H. Fowler and Sangick Jeon. 2008. The authority of Supreme Court precedent. *Social Networks* 30, 1 (Jan 2008), 16–30.
- [7] Andrew James Green and Albert Yoon. 2016. *Triaging the Law: Developing the Common Law on the Indian Supreme Court*. Number ID 2816666. <https://papers.ssrn.com/abstract=2816666>
- [8] Matthew P. Hitt. 2016. Measuring Precedent in a Judicial Hierarchy. *Law & Society Review* 50, 1 (2016), 57–81.
- [9] Fernando Leal and Ana Paula Barlos. Forthcoming. *VII Relatório Supremo em Números: os precedentes no Supremo*. Escola de Direito do Rio de Janeiro da Fundação Getulio Vargas, Rio de Janeiro.
- [10] Greg Leibon, Michael Livermore, Reed Harder, Allen Riddell, and Dan Rockmore. 2018. Bending the law: geometric tools for quantifying influence in the multinet network of legal opinions. *Artificial Intelligence and Law* 26, 2 (Jun 2018), 145–167.
- [11] Douglas A. Luke and Jenine K. Harris. 2007. Network Analysis in Public Health: History, Methods, and Applications. *Annual Review of Public Health* 28, 1 (Apr 2007), 69–93.
- [12] Frederick Schauer. 2012. Precedent. In *The Routledge Companion to Philosophy of Law*, Andrei Marmor (Ed.). Routledge, New York, 123–136.
- [13] Fábio Perin Shecaira. 2018. The Value of Methodological Deductivism in Argument Construction. *Informal Logic* 38, 4 (Dec 2018), 471–501.
- [14] Fabien Tarissan, Yannis Panagis, and Urška Šadl. 2016. Selecting the Cases That Defined Europe: Complementary Metrics for a Network Analysis. In *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM '16)*. IEEE Press, 661–668. event-place: Davis, California.