

Analyzing the Supreme Court Citation Network

December 2, 2017

Abstract

1 Introduction

Majority opinions written by the United States Supreme Court exercise their authority and influence, in part, through their roles as precedents in future Supreme Court decisions and opinions. The findings regarding the extent and exact nature of the influences of precedent have been mixed, but the balance of the literature finds that past decisions exert some form of influence on the justices' decision making (?????). Despite a considerable body of research that focuses on the way in which relevant precedents shape decision making on the Court, relatively little work has focused on understanding which past opinions are cited by an opinion. Our focus in this paper is to provide what is, to our knowledge, the first comprehensive analysis of exactly which cases are cited by a case. We follow an emerging body of work on legal citations, and treat the system of citations as a network (e.g., ?????).

2 Prior work on court citations

Measures of citations have been used in several recent studies of courts...? study the citations between US state supreme courts.

3 Network Approaches to Studying Citations

TED, would you write a section that covers the ways in which network scholars have studied citations in the past, focusing on scientific papers, patents, and court cases?

4 Complex Dependence in Court Citations

Transitivity: Discussing past edges in the citation network leads to transitivity. If i cites j , it is also likely to discuss the cases on which j is based.

The statute thus addresses the risk of a morally unjustifiable death sentence, not by minimizing it as precedent unmistakably requires, but by guaranteeing that in equipoise cases the risk will be realized, by “placing a thumb [on] death’s side of the scale,” *Sochor v. Florida*, 504 U. S. 527, 532 (1992) (quoting *Stringer v. Black*, 503 U. S. 222, 232 (1992); alteration in original).

<https://supreme.justia.com/cases/federal/us/548/163/dissent2.html>

“Never before the decision in *Union Gas* had we suggested that the bounds of Article III could be expanded by Congress operating pursuant to any constitutional provision other than the Fourteenth Amendment. Indeed, it had seemed fundamental that Congress could not expand the jurisdiction of the federal courts beyond the bounds of Article III. *Marbury v. Madison*, 1 Cranch 137 (1803). The plurality’s citation of prior decisions for support was based upon what we believe to be a misreading of precedent. See *Union Gas*, 491 U. S., at 40-41 (SCALIA, J., dissenting). The plurality claimed support for its decision from a case holding the unremarkable, and completely unrelated, proposition that the States may waive their sovereign immunity, see *id.*, at 14-15 (citing *Parden v. Terminal Railway of Ala. Docks Dept.*, 377 U. S. 184 (1964)), and cited as precedent propositions that had been merely assumed for the sake of argument in earlier cases, see 491 U. S., at 15 (citing *Welch v. Texas Dept. of Highways and Public Transp.*, 483 U. S., at 475-476, and *n. 5*, and *County of Oneida v. Oneida Indian Nation of N. Y.*, 470 U. S., at 252).”

<https://supreme.justia.com/cases/federal/us/517/44/case.html>

Mutuality: When deciding a set of related cases, the courts opinions regularly reference each other in order to reinforce their arguments...

On a more specific level, *Western* argues that flight engineers must meet the same stringent qualifications as pilots, and that it was therefore quite logical to extend to flight engineers the FAA’s age 60 retirement rule for pilots. Although the FAA’s rule for pilots, adopted for safety reasons, is relevant evidence in the airline’s BFOQ defense, it is not to be accorded conclusive weight. *Johnson v. Mayor and City Council of Baltimore*, ante at 472 U. S. 370-371. The extent to which the rule is probative varies with the weight of the evidence supporting its safety rationale and “the congruity between the . . . occupations at issue.” Ante at 472 U. S. 371. In this case, the evidence clearly established that the FAA, *Western*, and other airlines all recognized that the qualifications for a flight engineer were less rigorous than those

required for a pilot. [Footnote 28]

<https://supreme.justia.com/cases/federal/us/472/400/case.html>

The city, supported by several amici, argues for affirmance nonetheless. It asserts first that the federal civil service statute is not just a federal retirement provision unrelated to the ADEA, but in fact establishes age as a BFOQ for federal firefighters based on factors that properly go into that determination under the ADEA, see *Western Air Lines, Inc. v. Criswell*, post p. 472 U. S. 400. Second, the city asserts, a congressional finding that age is a BFOQ for a certain occupation is dispositive of that determination with respect to nonfederal employees in that occupation. We consider each of these contentions in turn.

<https://supreme.justia.com/cases/federal/us/472/353/case.html#370>

Also popularity If j served as precedent for i , j is authoritative regarding the rule of i , and should also be cited. Citations beget citations. Citation confers authority. Cases are even discussed in opinions in terms of the extent of influence they have had on other cases,

"The first case in which this Court struck down a statute under the Equal Protection Clause of the Fourteenth Amendment was *Strauder v. West Virginia*, 100 U. S. 303, decided in the 1879 Term. [Footnote 2/1] In the 1961 Term, we squarely held that the manner of apportionment of members of a state legislature raised a justiciable question under the Equal Protection Clause, *Baker v. Carr*, 369 U. S. 186. That case was followed by numerous others, e.g.: that one person could not be given twice or 10 times the voting power of another person in a state-wide election merely because he lived in a rural area..."

<https://supreme.justia.com/cases/federal/us/400/112/case.html>

5 The Exponential Random Configuration Model

Let $c(t) \in \{0, 1\}^N$ be a vector indicating which Supreme Court case has been cited at time t , where $c_i(t) = 1, i \in \{1, \dots, N\}$ indicates that the i th case has been cited at time t and $c_i(t) = 0$ indicates that the i th case has not been cited at time t . Furthermore, let

$$\mathcal{C}_t(N) = \{c(t) \in \{0, 1\}^N : c_i(t) \in \{0, 1\}\}$$

be the set of all possible citation combinations at time t . Note that the cardinality of $\mathcal{C}_t(N)$ increases exponentially for every newly added case, which results in 2^N elements.

The probability function of the ERCM is defined as

$$P_{\theta}(c(t) \mid c(t-)) = \frac{\exp(\theta^T \cdot h(c(t) \mid c(t-)))}{\sum_{c(t)^* \in \mathcal{C}} \exp(\theta^T \cdot h(c(t) \mid c(t-)))} \quad (1)$$

where $c(t-) \in \{0, 1\}^{N \times (t-1)}$ is a matrix that indicates which cases have been citing in each other before time t , $\theta \in \mathbb{R}^q$ is a q -dimensional vector of parameters, $h : \mathcal{C}_t(N) \rightarrow \mathbb{R}^q$, $(t) \rightarrow (h_1(c(t)), \dots, h_q(c(t)))^T$ is a q -dimensional vector of different statistics and $\kappa(\theta) := \sum_{c(t)^* \in \mathcal{C}} \exp(\theta^T \cdot h(c(t) \mid c(t-)))$ is a normalization constant that ensures that $(?)$ defines a probability function on \mathcal{C}_t .

The generative process of a model are informed by the decision regarding which network statistics $h(\cdot)$ are incorporated. We include the following statistics for the Supreme Court citation network:

$$h_{edges} : \mathcal{C}(N) \rightarrow \mathbb{R} \quad , \quad c(t) \rightarrow \sum_{i=1}^N c_i(t)$$

the number of citations made at time t .

$$h_{outstar} : \mathcal{C}(N) \rightarrow \mathbb{R} \quad , \quad c(t) \rightarrow \sum_{j < i}^N c_i(t) \cdot c_j(t) \cdot \sqrt{\frac{(t-a)(t-a-b)}{t^2}}$$

the number of weighted outstars occuring at time t . We argue that it should be more likely to cite more recent cases than cases that have been decided further in the past. For the weight

$$w(a, b) := \sqrt{\frac{(t-a)(t-a-b)}{t^2}}$$

we define a and b as the elapsed time since case i and j have been introduced to the network.

$$h_{triangle} : \mathcal{C}(N) \rightarrow \mathbb{R} \quad , \quad c(t) \rightarrow \sum_{j < i}^N c_i(t) \cdot c_j(t) \cdot c_j(t_{-i}) \cdot w(a, b)$$

where $c_j(t_{-i})$ indicates whether case j was cited at the time case i was introduced into the network. Just as for the outstar statistic, we include a weighting factor to favor more recent cases.

The individual entries $c_i(t)$ can be taken as a manifestation of single Bernoulli variables $C_i(t)$. This interpretation allows the following calculation regarding the conditional distribution of $C_i(t)$:

$$\begin{aligned}
 \frac{P_\theta(C_i(t) = 1 \mid C_i(t)^c = c_i(t)^c)}{P_\theta(C_i(t) = 0 \mid C_i(t)^c = c_i(t)^c)} &= \frac{P_\theta(C_i(t) = 1, C_i(t)^c = c_i(t)^c)}{P_\theta(C_i(t) = 0, C_i(t)^c = c_i(t)^c)} \\
 &= \frac{P_\theta(C(t) = c_i^+(t))}{P_\theta(C(t) = c_i^-(t))} \\
 &= \frac{\exp(\theta^T \cdot h(c_i^+(t) \mid c(t-)))}{\exp(\theta^T \cdot h(c_i^-(t) \mid c(t-)))} \\
 &= \exp(\theta^T \cdot (h(c_i^+(t) \mid c(t-)) - h(c_i^-(t) \mid c(t-))))
 \end{aligned}$$

This implies the following equation:

$$\text{logit}(P_\theta(C_i(t) = 1 \mid C_i(t)^c = c_i(t)^c)) = \theta^T \cdot (h(c_i^+(t) \mid c(t-)) - h(c_i^-(t) \mid c(t-))) \quad (2)$$

In the equation above the following notations were used:

- $c_i^+(t)$ emerges from $c(t)$, while assuming $c_i(t) = 1$
- $c_i^-(t)$ emerges from $c(t)$, while assuming $c_i(t) = 0$
- The condition $C_i(t)^c = c_i(t)^c$ is short for: $C_j(t) = c_j(t)$ for all $j \in \{1, \dots, N\}$ with $i \neq j$
- The expression $(\Delta c_i)(t) := h(c_i^+(t) \mid c(t-)) - h(c_i^-(t) \mid c(t-))$ is called the *change statistic*. The k th component of $(\Delta c_i)(t)$ captures the difference between citation networks $c_i^+(t)$ and $c_i^-(t)$ on the k th integrated statistic in the model

6 Estimation

Maximum Pseudo-Likelihood Estimator

One can assume that the dyads are independent of each other, which means that the random variables $C_i(t)$ inside the random vector $C(t)$ are independent of each other. In this case, the equation (??) reduces to

$$\text{logit}(P_\theta(C_i(t) = 1)) = \theta^T \cdot (\Delta c_i)(t)$$

This corresponds with the logistic regression approach, where the observations of the dependent variables are simply edge values of the observed citation vector, and the observations of the covariate values are given as the scores of every single change statistic. Therefore, the resulting likelihood function is of the following form:

$$\text{lik}(\theta) = P_\theta(C(t) = c(t)) = \prod_i \frac{\exp(\theta^T \Delta(c_i)(t))}{1 + \exp(\theta^T \Delta(c_i)(t))} \quad (3)$$

Maximum Likelihood Estimator

The more rigorous technique is to estimate the parameters directly with the log-likelihood function derived from (??), which has the following form:

$$\text{loglik}(\theta) = \theta^T \cdot h(c(t)|c(t-)) - \log(\kappa(\theta)) \quad (4)$$

The problem resulting from estimating the parameters with (??) is that the term

$$\kappa(\theta) := \sum_{c(t)^* \in \mathcal{C}(N)} \exp(\theta^T \cdot h(c(t)^*|c(t-)))$$

which sums up the weighted statistics of all possible binar vectors of length N , has to be evaluated. However, the cardinality of $\mathcal{C}(N)$ ($\#(\mathcal{C}) = 2^N$) is incredibly large and a direkt calculation of this sum is for already small N not feasible.

An solution for this limitation is based on the following consideration: Fix a vector of parameters $\theta_0 \in \Theta$ from the underlying parameter range Θ and compute for $\theta \in \Theta$ the expected value

$$\begin{aligned} \mathbb{E}_{\theta_0} \left[\exp \left((\theta - \theta_0)^T \cdot \Gamma(C(t)) \right) \right] &= \sum_{c(t) \in \mathcal{C}(N)} \exp \left((\theta - \theta_0)^T \cdot \Gamma(c(t)) \right) \cdot \mathbb{P}_{\theta_0}(C(t) = c(t)) \\ &= \sum_{c(t) \in \mathcal{C}(N)} \exp \left((\theta - \theta_0)^T \cdot \Gamma(c(t)) \right) \cdot \frac{\exp(\theta_0^T \cdot \Gamma(c(t)))}{\kappa(\theta_0)} \\ &= \frac{1}{\kappa(\theta_0)} \sum_{c(t) \in \mathcal{C}(N)} \exp \left(\theta^T \cdot \Gamma(c(t)) \right) \\ &= \frac{\kappa(\theta)}{\kappa(\theta_0)} \end{aligned}$$

This equation offers the following possibility: If one draws L random vectors $c^{(1)}(t), \dots, c^{(L)}(t)$ out of a distribution \mathbb{P}_{θ_0} appropriately, one gets with the law of big numbers and a

big enough sample L the following relation:

$$\frac{1}{L} \cdot \sum_{i=1}^L \exp \left((\theta - \theta_0)^T \cdot \Gamma(c^{(i)}(t)) \right) \approx \mathbb{E}_{\theta_0} \left[\exp \left((\theta - \theta_0)^T \cdot \Gamma(C(t)) \right) \right] = \frac{\kappa(\theta)}{\kappa(\theta_0)} \quad (5)$$

This approximate can then be used to approximate the log likelihood function.

Next, we will discuss how a sufficient number of suitable drawings $c^{(1)}(t), \dots, c^{(L)}(t)$ can be sampled from the distribution \mathbb{P}_{θ_0} .

For this purpose, the Markov Chain Monte Carlo (MCMC) methods can be used.

Gibbs sampling for the ERCM

To be able to compute the approximate likelihood function one needs a sufficiently large number of random vectors from the distribution \mathbb{P}_{θ_0} . Snijders ? introduces an approach to sample random networks for the ERGM framework by using *MCMC methods*. We adapt this approach for sampling appropriate binary vectors for the ERCM.

Gibbs sampling

Choose any vector $c^{(0)}(t) \in \mathcal{C}(N)$ (e.g. observed vector)

for i in $1:N$ **do**

Compute $\pi := \frac{\exp(\theta^T \cdot \Delta(c_i)(t))}{1 + \exp(\theta^T \cdot \Delta(c_i)(t))}$
Draw a random number Z from $\text{Bin}(1, \pi)$
if $Z=1$ then
set $c_i^{(k+1)} = 1$ and $c_j^{(k+1)} = c_j^{(k)}$, if $i \neq j$
else
set $c_i^{(k+1)} = 0$ and $c_j^{(k+1)} = c_j^{(k)}$, if $i \neq j$
end

end

Start all over using $c^{(k+1)}$

Algorithm 1: Simulation of vectors of \mathbb{P}_{θ} using Gibbs sampling

Using the depicted algorithm, a sequence of random vectors $c^{(0)}(t), \dots, c^{(L)}(t)$ can be simulated. Since the original vector was chosen randomly and the first simulated vectors are very dependent on the chosen mvector (only one entry is changed per iteration!), usually the first B vectors, where $N \ll B \ll L$, are discarded as the so

called *Burn-In*.

Metropolis Hastings for the ERCM

Choose any vector $c^{(0)}(t) \in \mathcal{C}(N)$ to start with (e.g., the observed vector). For $k \in \{0, \dots, L-1\}$ recursively proceed as follows:

1. Randomly choose a number $i \in \{1, \dots, N\}$
2. Compute, using the equation (??) the value

$$\pi := \frac{\mathbb{P}_\theta(C_i(t) \neq c_i^{(k)}(t) \mid C_i(t)^c = c_i(t)^c)}{\mathbb{P}_\theta(C_i(t) = c_i^{(k)}(t) \mid C_i(t)^c = c_i(t)^c)}$$

3. Define $\delta := \min\{1, \pi\}$ and draw a random number Z from $\text{Bin}(1, \delta)$. If
 - $Z = 0$, let $c^{(k+1)}(t) := c^{(k)}(t)$
 - $Z = 1$, define $c^{(k+1)}(t)$ as

$$c_p^{(k+1)}(t) = \begin{cases} 1 - c_p^{(k)}(t) & \text{if } p = i \\ c_p^{(k)}(t) & \text{if } p \neq i \end{cases}$$

4. Start at step 1 with $c^{(k+1)}(t)$.

The first $B \ll L$ vectors are discarded as Burn-In.

7 Results

Descriptive Results

The supreme court citation network from 1937 – 2005 consists of 8817 cases which got voted at 2116 different time points. The network has a total of 93,263 ties, of which 452 are mutual. The number of triangles in the network is 211,855. The in- and outdegree distribution is visualized in figures ?? and ?. The maximum indegree is 190 and the maximum outdegree is 159.

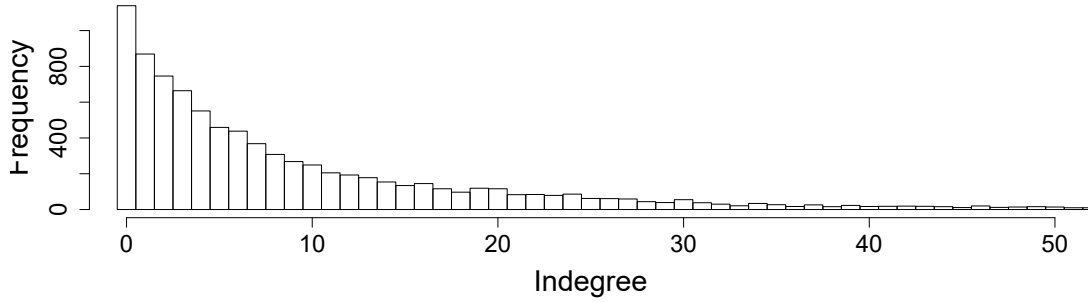
	Terms	Total Number Cases	Cases/Term
CE Hughes*	1937 - 1941	629	125.5
HF Stone	1942 - 1946	766	153.2
FM Vinson	1946 - 1953	788	98.5
E Warren	1954 - 1969	2159	127.0
WE Burger	1970 - 1986	2805	155.8
W Rehnquist**	1987 - 2001	1670	83.5

Table 1: For the time range of interest (1937 - 2001) this table displays the chief justices, the time range they served as chief justice, the number of cases in their time range as well as the average number of cases per year.

* CE Hughes served as chief justice from 1930 - 1941.

** W Rehnquist served as chief justice from 1987 - 2005.

Indegree Distribution



Outdegree Distribution

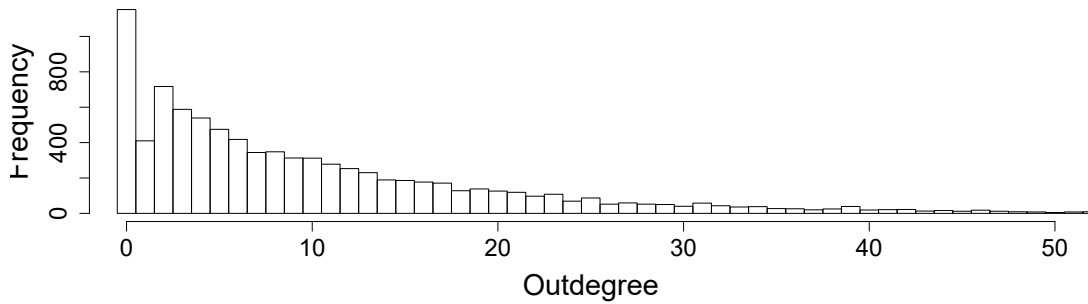


Figure 1: The in- and outdegree distribution of the Supreme Court Citation Network from 1937 - 2001. There are cases with an indegree >50 , but they are not captured in this figure.

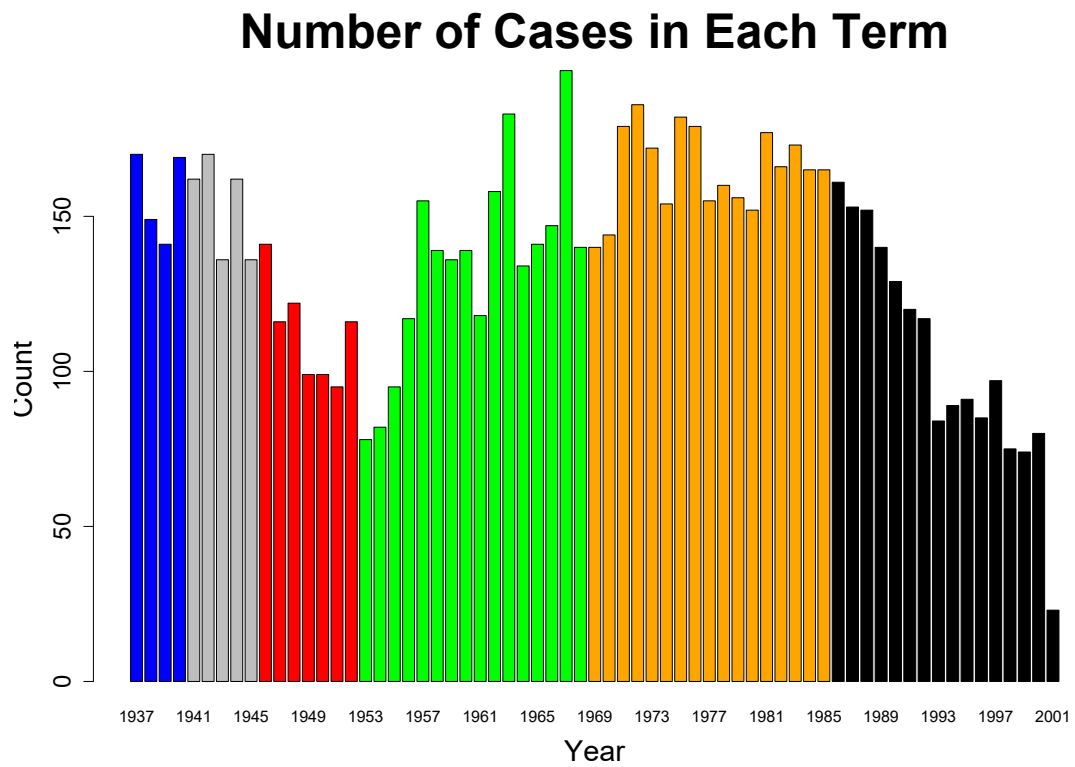


Figure 2: Number of cases in each term. Different colors indicate different chief justices.

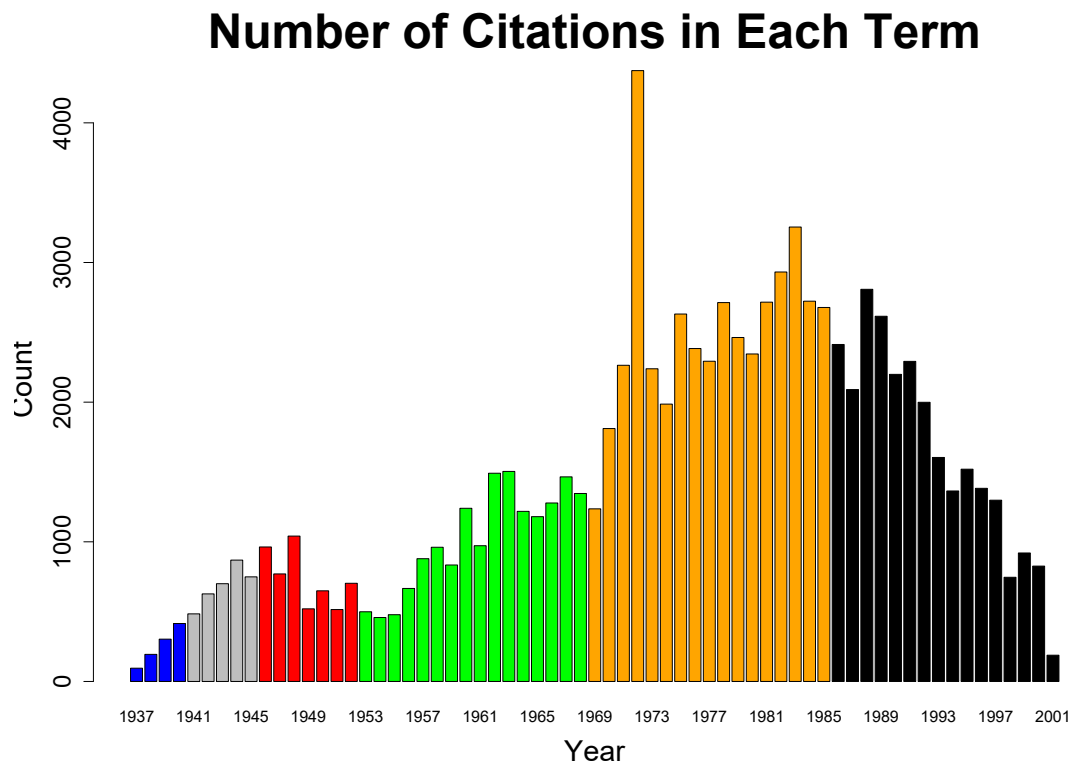


Figure 3: Number of citations for the 1937-2001 time period. Citations for cases prior 1937 are not considered in this figure. Different colors indicate different chief justices.

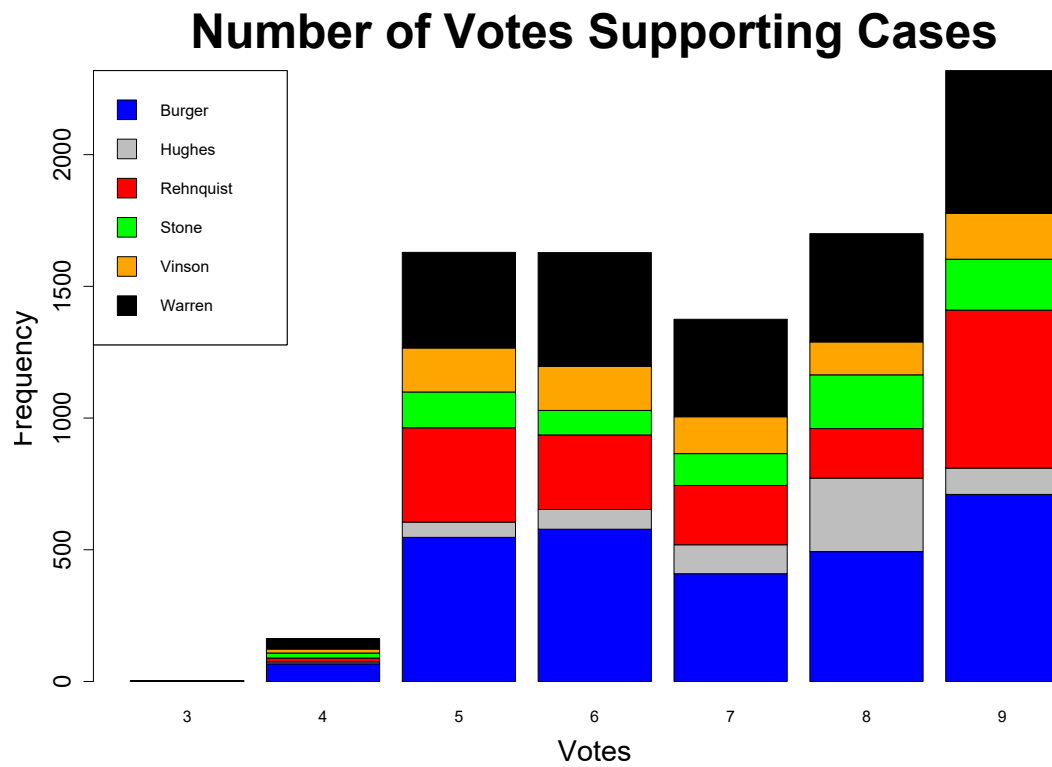


Figure 4: Number of Votes that were supporting cases between 1937-2001. Different colors indicate terms with different chief justices.

7.1 Inferential Results

Model updates

- Add sender issue area and receiver issue area node covariates (i.e., nodeifactor and nodeofactor applied to issue area).
- In the GLM, multiply each variable by the sender time covariate.
- Write the bootstrap coefficients to a file, send to Bruce so he can write up a code for summarizing the over-time trends in effects.
- Add a receiver variable (i.e., nodeicov) that equals the absolute difference between the maximum MQ score of a justice in the majority and a minimum MQ score of a justice in the majority.
- Add a receiver variable (nodeicov) equivalent to the number of justices in the majority

	Estimate	Lower Bound	Upper Bound
Edges	-5.319	-5.392	-5.243
Instar(2)	0.018	0.017	0.019
Outstar(2)	0.018	0.016	0.022
Mutual	3.179	2.694	3.565
Triangle	1.531	1.493	1.563
MQ Score	0.041	0.018	0.063
Same Issue Area	1.061	1.028	1.099
Year Difference	0.010	0.008	0.013

Table 2: Bootstrapped MPLE Results for the time period 1937-2001

	Estimate	Lower Bound	Upper Bound
Edges	-5.391	-5.506	-5.244
Instar(2)	0.035	0.028	0.042
Outstar(2)	0.015	0.0003	0.017
Mutual	4.785	3.616	5.749
Triangle	1.377	1.271	1.513
MQ Score	-0.018	-0.095	0.061
Same Issue Area	1.229	1.126	1.327
Year Difference	-0.077	-0.096	-0.061

Table 3: Bootstrapped MPLE Results for the time period when Fred M. Vinson was chief justice.

	Estimate	Lower Bound	Upper Bound
Edges	-5.032	-5.140	-4.931
Instar(2)	0.028	0.025	0.032
Outstar(2)	0.018	0.015	0.022
Mutual	3.166	1.826	4.565
Triangle	1.520	1.449	1.599
MQ Score	-0.101	-0.149	-0.052
Same Issue Area	1.349	1.289	1.408
Year Difference	-0.031	-0.038	-0.026

Table 4: Bootstrapped MPLE Results for the time period when Earl Warren was chief justice.

	Estimate	Lower Bound	Upper Bound
Edges	-3.952	-4.017	-3.882
Instar(2)	0.015	0.013	0.016
Outstar(2)	0.004	0.002	0.006
Mutual	2.558	1.425	3.719
Triangle	1.383	1.333	1.440
MQ Score	-0.084	-0.101	-0.059
Same Issue Area	0.959	0.911	1.008
Year Difference	-0.017	-0.019	-0.015

Table 5: Bootstrapped MPLE Results for the time period when Warren E. Burger was chief justice.

	Estimate	Lower Bound	Upper Bound
Edges	-3.614	-3.701	-3.525
Instar(2)	0.015	0.013	0.016
Outstar(2)	0.002	-0.002	0.004
Mutual	9.635	6.426	10.834
Triangle	1.551	1.489	1.620
MQ Score	-0.007	-0.039	0.028
Same Issue Area	0.855	0.794	0.916
Year Difference	-0.016	-0.018	-0.014

Table 6: Bootstrapped MPLE Results for the time period when William Rehnquist was chief justice.

Bibliography

- Bailey, Michael A and Forrest Maltzman. 2008. "Does legal doctrine matter? Unpacking law and policy preferences on the US Supreme Court." *American Political Science Review* 102(3):369–384.
- Bailey, Michael A and Forrest Maltzman. 2011. *The constrained court: Law, politics, and the decisions justices make*. Princeton University Press.
- Bommarito II, Michael J, Daniel Katz and Jon Zelner. 2009. Law as a seamless web?: comparison of various network representations of the united states supreme court corpus (1791-2005). In *Proceedings of the 12th international conference on artificial intelligence and law*. ACM pp. 234–235.
- Fowler, James H and Sangick Jeon. 2008. "The authority of Supreme Court precedent." *Social networks* 30(1):16–30.
- Fowler, James H, 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):324–346.
- Gillman, Howard. 2001. "What's law got to do with it? Judicial behavioralists test the 'legal model' of judicial decision making." *Law & Social Inquiry* 26(2):465–504.
- Hansford, Thomas G and James F Spriggs. 2006. *The politics of precedent on the US Supreme Court*. Princeton University Press.
- Hinkle, Rachael K and Michael J Nelson. 2016. "The Transmission of Legal Precedent among State Supreme Courts in the Twenty-First Century." *State Politics & Policy Quarterly* 16(4):391–410.
- Knight, Jack and Lee Epstein. 1996. "The norm of stare decisis." *American Journal of Political Science* pp. 1018–1035.
- Lupu, Yonatan and Erik Voeten. 2012. "Precedent in international courts: A network analysis of case citations by the European court of human rights." *British Journal of Political Science* 42(2):413–439.
- Pelc, Krzysztof J. 2014. "The politics of precedent in international law: A social network application." *American Political Science Review* 108(3):547–564.

Bibliography

Richards, Mark J and Herbert M Kritzer. 2002. "Jurisprudential regimes in Supreme Court decision making." *American Political Science Review* 96(2):305–320.