Analyzing the Supreme Court Citation Network

December 23, 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 (Knight and Epstein, 1996; Gillman, 2001; Richards and Kritzer, 2002; Hansford and Spriggs, 2006; Bailey and Maltzman, 2008, 2011). 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., Fowler et al., 2007; Fowler and Jeon, 2008; Bommarito II, Katz and Zelner, 2009; Lupu and Voeten, 2012; Pelc, 2014).

We are not the first to ask what predicts the citations in US Supreme Court Opinions. Indeed, a voluminous body of work has sought to explain how many times an opinion is cited (e.g., Cross, 2010; Benjamin and Desmarais, 2012), when in a lifecycle an opinion is cited (e.g., Black and Spriggs, 2013; Spriggs and Hansford, 2001), and how many cases are cited by an opinion (e.g., Lupu and Fowler, 2013)—all focused on the US Supreme Court. One common feature of the research design in all of these studies is that the observations are at the case or case-time level. The outcome variables in these analyses are defined as measures of the number of citations to a case over a period of time, the number of citations to a case at a particular time, or a measurement on the cases cited by a case. None of these studies treats citation in its micro-level form—as a relationship between two opinions, the citing opinion and the cited opinion. We are aware of one prior study, Clark and Lauderdale (2010), in which a statistical model is used to analyze dyadic citations between cases. However, citetclark2010locating uses a dyadic latent variable model in order to estimate ideal points for Supreme Court Opinions, but does not use any explanatory variables to predict the formation of citation ties between opinions. We build upon this literature both methodologically and substantively. Methodologically, we develop a novel statistical framework for modeling directed dyadic citations. Second, we apply this methodology to a half-century of directed dyadic citations between US Supreme Court opinions.

There exist two broad reasons why empirical analyses of citations are best defined on the directed dyad level, not the case level. The first is that directed dyadic analyses can test both dyadic and case-level hypotheses. For example, case-level analyses can model whether opinions supported by a liberal majority coalition are more likely than those supported by a conservative majority coalition to be cited heavily in the future, but they cannot precisely model whether liberal cases will be cited more by liberal cases than by conservative cases. Thus, the first reason for analyzing citations at the dyadic level it to expand the set of hypotheses that can be represented in the model. The second reason for studying citations at the directed dyadic level is that, as articulated in the growing literature on legal citation networks, citations form complex networks in which a citation at one point in time may influence future citations. This phenomenon of complex dependence is very common in networks of many types, but processes specific to Supreme Court citations create interdependence in citations. For example, if opinion i relies heavily on opinion j as precedent, opinion i is likely to discuss the legal basis for opinion j, and as a consequence cite some of the opinions cited by opinion j. Suppose opinion k is cited by opinion j. Opinion k is more likely to be cited by opinion i because opinion i relies heavily on j, and opinion j cites k. This is a special case of a very common process on networks referred to as "triad closure". Complex dependence is theoretically interesting on its own merits, but the effects of covariates cannot be reliably identified—either in terms of coefficient values or standard errors—without accounting for the interdependence inherent in networks (Desmarais and Cranmer, 2017).

In what follows we develop a theoretical case that citations on the US Supreme Court are characterized by forms of complex dependence that are common in networks. We then develop an extension of a model—the exponential random graph model—that can incorporate both exogenous covariates and complex forms of interdependence into a directed dyadic analysis of citations. Finally, we develop and estimate a specification of this model in an analysis of US Supreme Court citations between 1937 and 2001. We find robust support for the inherent complexity underpinning the formation of citation ties, and show that incorporating complex dependence into the model of citation formation significantly improves the model's predictive performance.

2 Complex Interdependence in Supreme Court Citations

When it comes to the development and testing of theory, the defining feature of networks is that the fundamental element under study—the relationship between two units (i.e., the citation from one opinion to another) is a piece of a complex web of relations. The formation (or lack thereof) of that relationship cannot be fully understood without considering how the relationship fits into the complex web. Analytical designs that account only for covariates in explaining tie formation

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

Sociality Citations sent beget citations sent. For each case discussed, often related cases to discuss, even those that do not cite each other. For each case that applies, there is likely a case that needs to be discussed in terms of why it does not apply. Justices often clarify not only those legal principles that apply, but often those that do not. When multiple lines of legal reasoning are drawn upon to justify a decision, often more are discussed in terms of their lack of applicability.

"1. The "overbreadth" doctrine is not applicable here. There is nothing in the record to indicate that §28.04 will have any different impact on any third parties' interests in free speech than it has on appellees' interests, and appellees have failed to identify any significant difference between their claim that §28.04 is invalid on overbreadth grounds and their claim that it is unconstitutional when applied to their signs during a political campaign. Thus, it is inappropriate to entertain an overbreadth challenge to §28.04. Pp. 466 U. S. 796-803."

https://supreme.justia.com/cases/federal/us/466/789/

"No allegations of racial or class-based invidiously discriminatory animus are required to establish a cause of action under the first part of 1985(2). The statutory provisions now codified at §1985 were originally enacted as §2 of the Civil Rights Act of 1871, and the substantive meaning of the 1871 Act has not been changed. The provisions relating to institutions and processes of the Federal Government (including the first part of §1985(2)) – unlike those encompassing activity that is usually of primary state concern (including the second part of §1985(2) and the part of ¶1985(3) involved in Griffin, supra – contain no language requiring that the conspirators act with intent to deprive their victims of the equal protection of the laws. Thus, the reasoning of Griffin is not applicable here, and, given the structure of §2 of the 1871 Act, it is clear"

https://supreme.justia.com/cases/federal/us/460/719/

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 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 ith case has been cited at time t and $c_i(t) = 0$ indicates that the ith 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

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

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, $h : \mathcal{C}_t(N) \to \mathbb{R}^q$, $\c(t) \to (h_1(c(t)), \dots, h_q(c(t)))^T$ is a q-dimensional vector of different statistics and $\kappa(\theta) := \sum_{c(t)^* \in \mathcal{C}} \exp(\theta^T \cdot h(c(t)|c(t-)))$ is a normalization constant that ensures that (1) 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:

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

the number of citations made at time t.

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

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.

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

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:

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

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 kth component of $(\Delta c_i)(t)$ captures the difference between citation networks $c_i^+(t)$ and $c_i^-(t)$ on the kth integrated statistic in the model

5 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 (2) reduces to

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

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:

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

Maximum Likelihood Estimator

The more rigorous technique is to estimate the parameters directly with the log-likelihood function derived from (1), which has the following form:

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

The problem resulting from estimating the parameters with (4) 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 $c^{(1)}(t), \ldots, c^{(L)}(t)$ out of a distribution $\mathbb{P}_{\theta_0}$ appropriately, one gets with the law of big numbers and a big enough sample L the following relation:

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

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

Next, we will discuss how a sufficient number of suitable drawings $c^{(1)}(t), \ldots, c^{(L)}(t)$ can be sampled from the distribution $\mathbb{P}_{\theta_0}$.
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 $\mathbb{P}_{\theta_0}$. 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 $\pi := \dfrac{\exp(\theta^T \cdot \Delta(c_i)(t))}{1 + \exp(\theta^T \cdot \Delta(c_i)(t))}$
Draw a random number Z from Bin(1,π)
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 $\mathbb{P}_\theta$ using Gibbs sampling

Using the depicted algorithm, a sequence of random vectors $c^{(0)}(t), ..., c^{(L)}(t)$ can be simulated. Since the orignial 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 $k \in \{0, ..., L-1\}$ recursively proceed as follows:

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

6 Prior work on court citations

Measures of citations have been used in several recent studies of courts...Hinkle and Nelson (2016) study the citations between US state supreme courts.

- Clark and Lauderdale (2010) Estimates a latent coordinate model of supreme court opinions based on the network of supreme court citations. Find that the majority opinion falls at the ideal point of the median member of the majority coalition. Account for negative citations. WE COULD USE THIS DATASET AS A SECOND APPLICATION

- Cross (2010) Look at the total number of citations to Supreme Court opinions over a ten year period of the Rehnquist Court. (case level). Found case issue area, opinion ideology, opinion author, vote margin was not significant. Negative citations decreased future positive citations.

- Black and Spriggs (2013) Study the number of citations to a case over time, in order to understand the lifecycle of citation rates. Show that the rate of citations to a case goes down very quickly. THE EFFECT OF AN INDEPENDENT VARIABLE ON A PRECEDENT IS DEPENDENT UPON THE AGE OF THE CASE.

- Benjamin and Desmarais (2012) Study the propensity for cases to be overruled and cited in other negative ways. They find that cases with large and broad majorities rare less likely to be cited negatively.

- Spriggs and Hansford (2001) Find that cases are more likely to be overruled when the Court is ideologically distant from the median justice in the majority coalition that decided the case, when the legal issues under consideration are complex, whn the cases ha been negatively treated previously, when there were many concurring opinions, and less likely to be overruled if it was a unanimous coalition.

- Lupu and Fowler (2013) The hub score, which is a measure of the degree to which a case cites other important cases increases with salience [GET THE AMICUS BRIEFS, NYT COUNTS]. Lower if stautory cases. Increases with ideological variance of the majority coalition. Increases with the number of concurring opinions.

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.

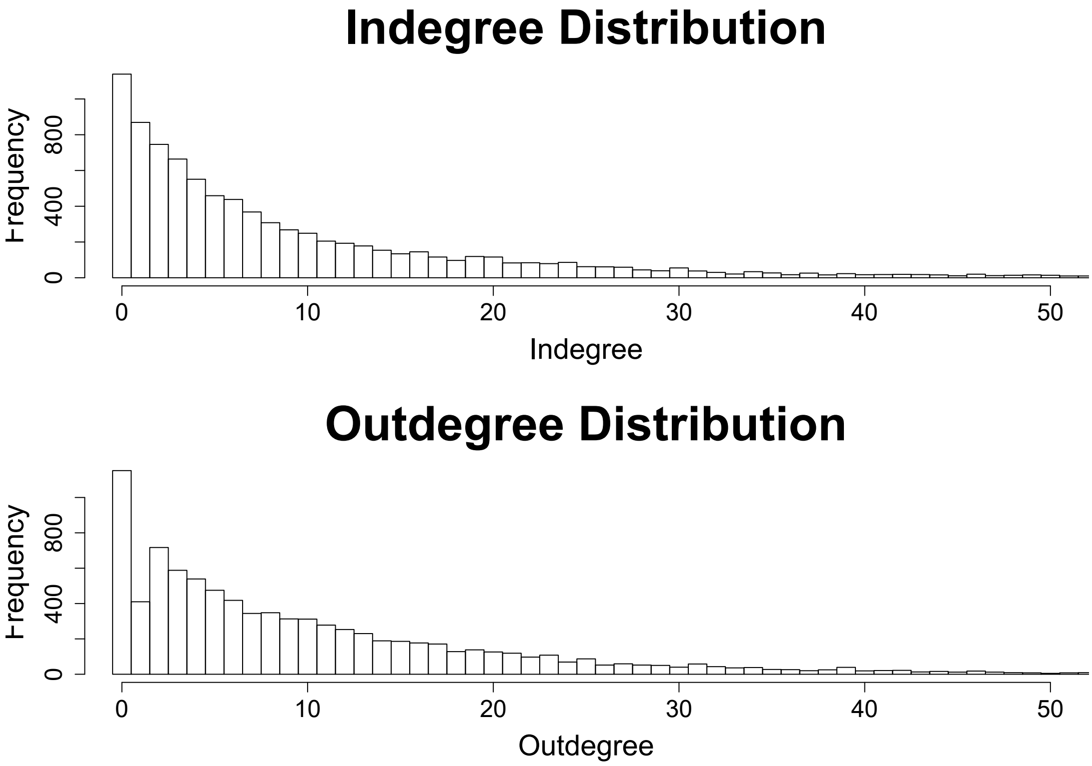

Figure 1: The in- and outdegree distribution of the Supreme Court Citation Network from 1937 - 2001. There are cases with an indegree >50, but they are not captured in this figure.

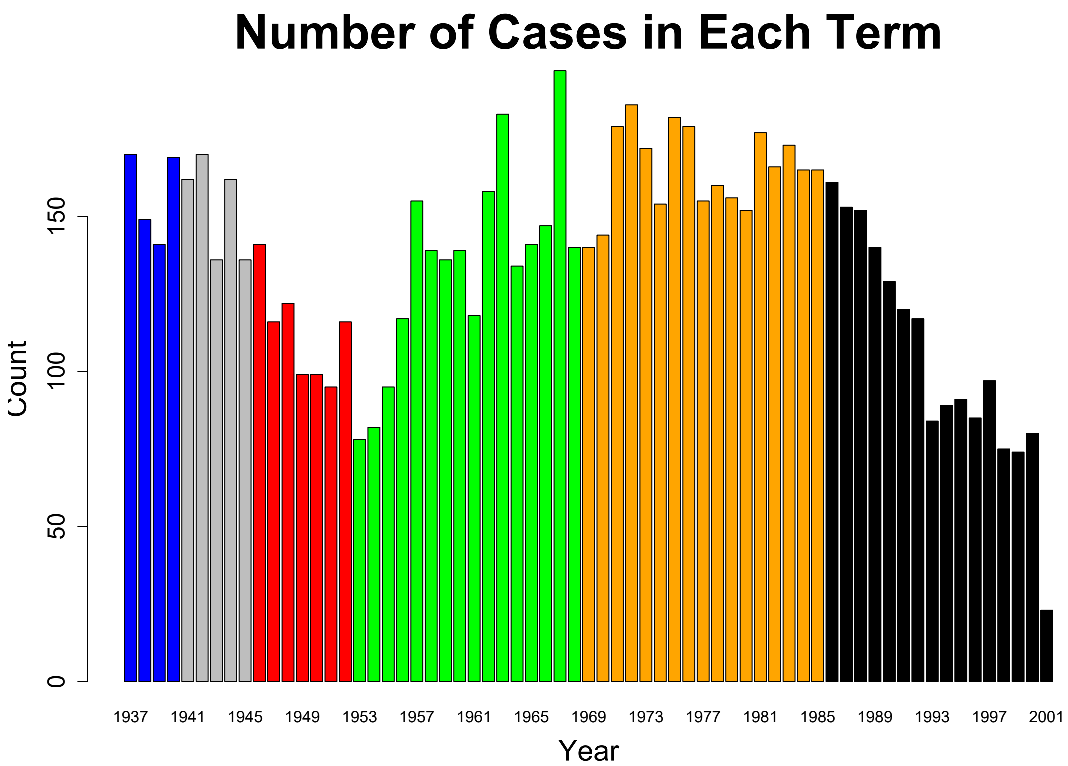

Figure 2: Number of cases in each term. Different colors indicate different chief justices.

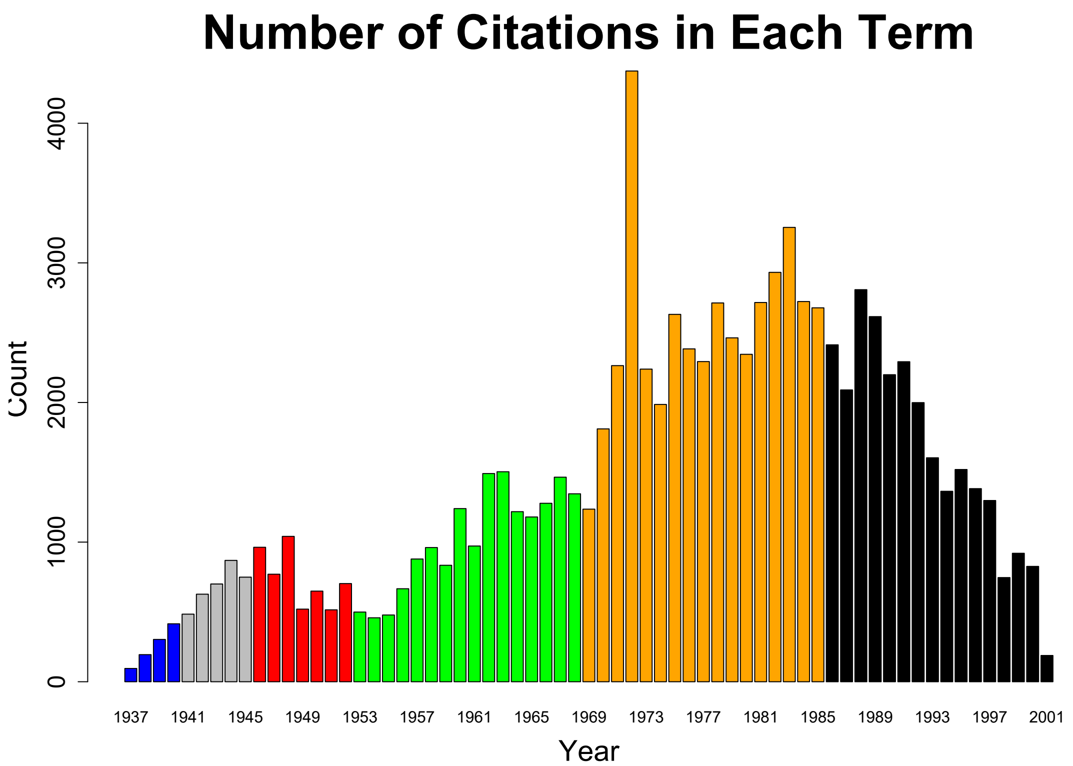

Figure 3: Number of citations for the 1937-2001 time period. Citations for cases prior 1937 are not considered in this figure. Different colors indicate different chief justices.

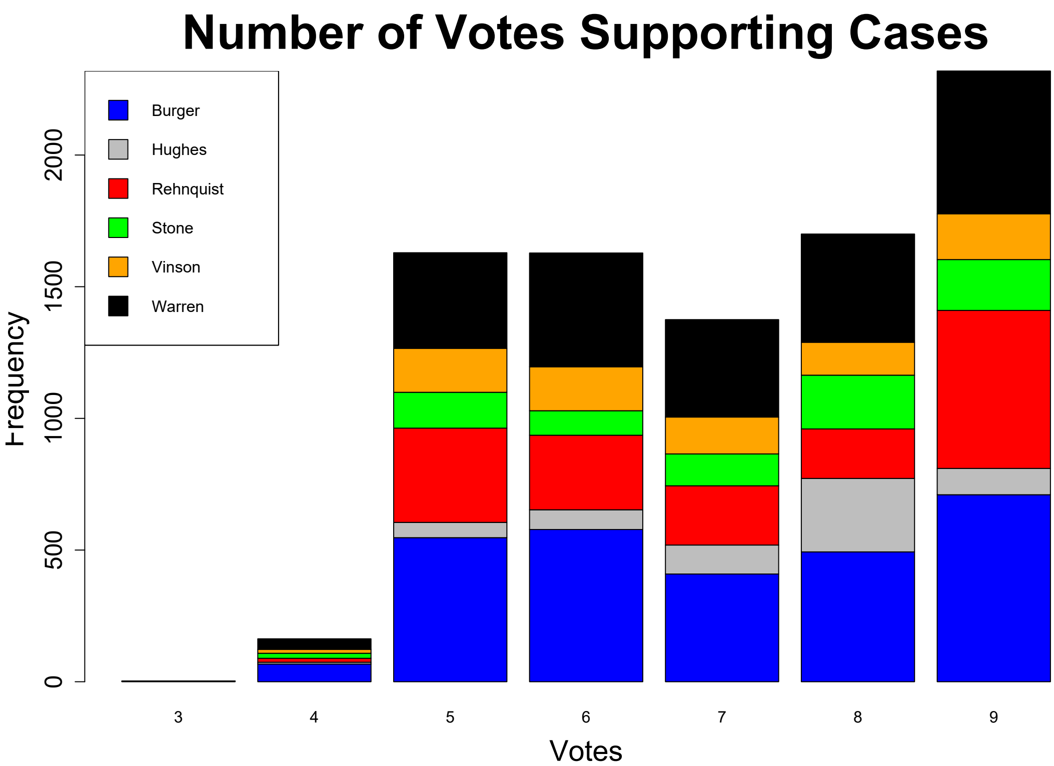

Figure 4: Number of Votes that were supporting cases between 1937-2001. Different colors indicate terms with different chief justices.

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	Significance
Edges	-5.533	-5.733	-5.388	*
Instar(2)	0.031	0.027	0.036	*
Outstar(2)	0.022	0.020	0.032	*
Mutual	3.316	2.622	3.983	*
Triangle	1.490	1.410	1.560	*
Martin Quinn Score	0.080	0.022	0.126	*
Same Issue Area Homophily	1.378	1.313	1.451	*
Year Difference	-0.077	-0.090	-0.064	*
(Year Differnce)^2	0.0029	0.0025	0.0032	*
Receiver Abs Diff MQ Score in Majority	-0.030	-0.048	-0.014	*
Receiver Number Justices in Majority	-0.115	-0.135	-0.089	*
Receiver Sender Year	0.0007	0.0006	0.0009	*
Sender Same Issue Area 2	0.162	0.105	0.217	*
Sender Same Issue Area 3	-0.317	-0.403	-0.236	*
Sender Same Issue Area 4	0.497	0.426	0.570	*
Sender Same Issue Area 5	0.332	0.174	0.495	*
Sender Same Issue Area 6	0.527	0.402	0.635	*
Sender Same Issue Area 7	0.327	0.251	0.394	*
Sender Same Issue Area 8	0.076	0.021	0.129	*
Sender Same Issue Area 9	0.261	0.209	0.315	*
Sender Same Issue Area 10	0.369	0.301	0.432	*
Sender Same Issue Area 11	-0.046	-0.233	0.151	
Sender Same Issue Area 12	0.0073	-0.075	0.089	
Sender Same Issue Area 13	0.381	0.162	0.564	*
Sender Same Issue Area 14	0.151	-0.019	0.300	
Receiver Same Issue Area 2	0.231	0.166	0.300	*
Receiver Same Issue Area 3	-0.101	-0.209	-0.006	*
Receiver Same Issue Area 4	0.516	0.425	0.612	*
Receiver Same Issue Area 5	0.351	0.183	0.534	*
Receiver Same Issue Area 6	0.461	0.284	0.628	*
Receiver Same Issue Area 7	0.361	0.281	0.437	*
Receiver Same Issue Area 8	0.153	0.097	0.221	*
Receiver Same Issue Area 9	0.279	0.221	0.346	*
Receiver Same Issue Area 10	0.492	0.413	0.570	*
Receiver Same Issue Area 11	0.247	0.056	0.411	*
Receiver Same Issue Area 12	0.339	0.246	0.430	*
Receiver Same Issue Area 13	1.060	0.845	1.239	*
Receiver Same Issue Area 14	0.557	0.255	0.793	*
Instar(2) $\times$ Sender Year	-0.00025	-0.00034	-0.00017	*
Outstar(2) $\times$ Sender Year	-0.00036	-0.00056	-0.00028	*
Mutual $\times$ Sender Year	-0.010	-0.043	0.029	
Triangle $\times$ Sender Year	-0.0004	-0.0021	0.0016	
Martin Quinn Score $\times$ Sender Year	-0.0025	-0.0035	-0.0013	*
Same Issue Area $\times$ Sender Year	-0.0058	-0.0076	-0.0043	*
Year Difference $\times$ Sender Year	0.0004	0.0002	0.0007	*
Year Difference$^2 \times$ Sender Year	-0.000037	-0.000044	-0.000030	*
MQ Score in Majority $\times$ Sender Year	0.0013	0.0009	0.0017	*
Justices in Majority $\times$ Sender Year	0.0029	0.0023	0.0035	*

Table 2: Bootstrapped MPLE Results for the time period 1937 – 2001. A '*' indicates that the 2.5th and 97.5th quantile of the variable does not include '0' and as a result is statistically significant.

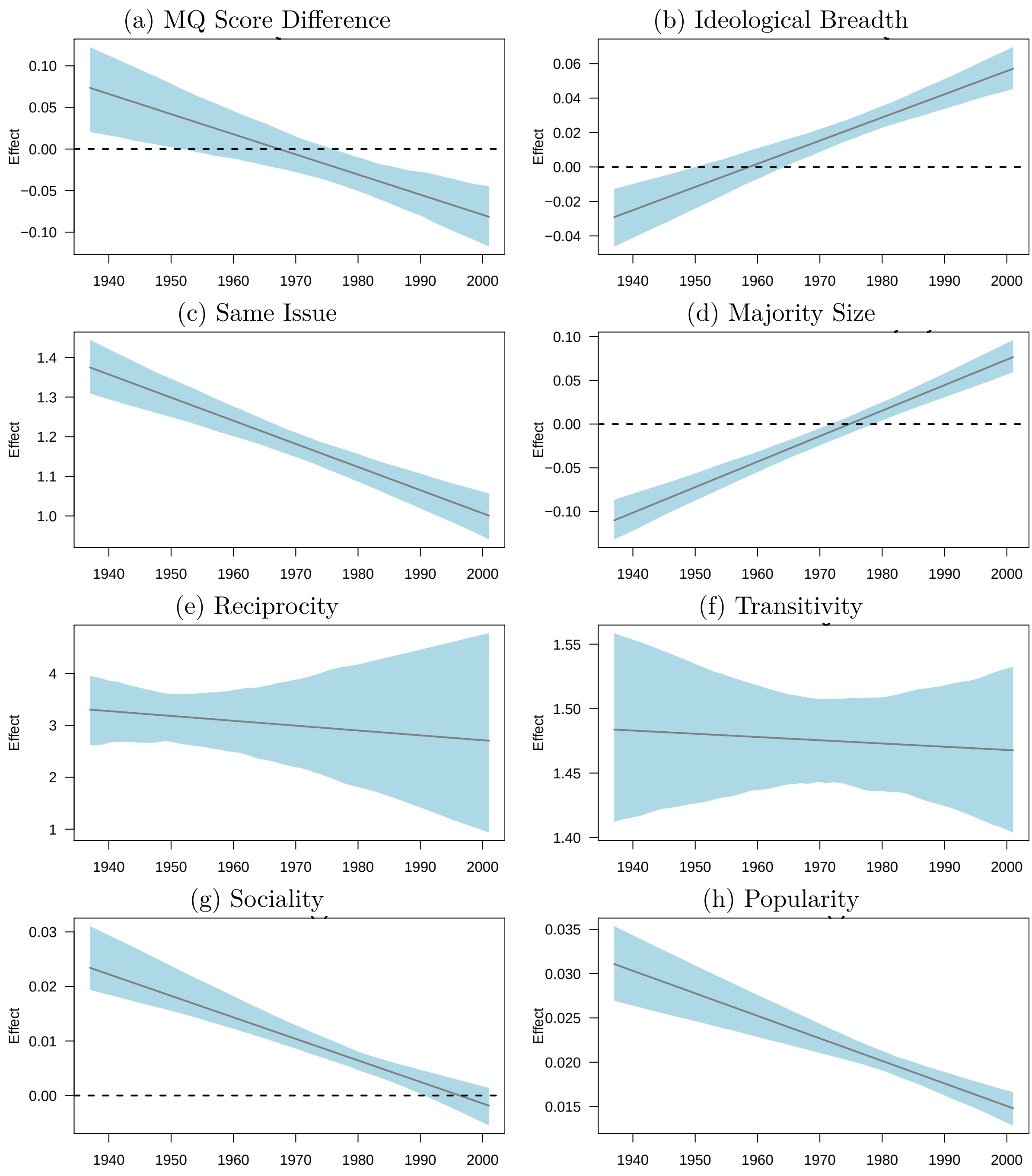

Figure 5: Effects

7.2 Predictive Performance

Our case for studying legal citaions at the directed dyadic level hinges upon the contribution to modeling offered by incorporating network dependence. To quantify this contribution, we use out-of-sample prediction [CITES]. Predicting out-of-sample offers an unbiased and general purpose way to evaluate the contribution, in terms of model fit, of one or more terms/parameters in a model. Unlike in-sample measures of model fit, out-of-sample methods are highly robust in avoiding overfitting, and work when we cannot accurately calculate the value of the likelihood function, as in the current case.

In our prediction experiment we randomly split the directed dyads into an 80% training set and a 20% test set. The parameters of the model are estimated using the directed dyads in the training set, and the parameters are used to form the conditional probability of a tie for all of the directed dyads in the test set. Directed dyads for which the conditional probability of a tie exceeds 0.5 are predicted to be citations. The experiment is run with the full model, and with a model that excludes all of the dependence terms (i.e., all terms involving reciprocity, in and out stars, and triangles)—the independent dyads model. We run this experiment for 10 iterations. Predictive performance is evaluated with three common and related measures—precision (i.e., the proportion of predicted citations that are actually citations), recall (the proportion of actual citations that are predicted to be citations), and the F1 score (the harmonic mean of precision and recall). All three measures are bounded between 0 and 1, with higher scores indicating better performance.

	Independent Model		Full Model	
	mean	range	mean	range
precision	0.5499	(0.5384, 0.5619)	0.8605	(0.8526, 0.8666)
recall	0.0827	(0.0811, 0.0843)	0.5858	(0.5817, 0.5891)
F1 score	0.1438	(0.141, 0.1463)	0.6971	(0.6941, 0.7008)

Table 3: The predictive performance of the directed dyadic models, over ten 80/20 train/test splits.

We see from Table 3 that, based on all three measures, the predictive performance of the model improves dramatically from adding the network dependence terms. The recall of the full model is particularly impressive, indicating that it recovers over half of the actual citations in the test set. This provides clear evidence that the full model, which includes covariates and network dependence terms, represents a more accurate and complete model of the process of citation formation in US Supreme Court opinions.

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.

Benjamin, Stuart Minor and Bruce A Desmarais. 2012. “Standing the test of time: The breadth of majority coalitions and the fate of us supreme court precedents.” *Journal of Legal Analysis* 4(2):445–469.

Black, Ryan C and James F Spriggs. 2013. “The Citation and Depreciation of US Supreme Court Precedent.” *Journal of Empirical Legal Studies* 10(2):325–358.

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.

Clark, Tom S and Benjamin Lauderdale. 2010. “Locating Supreme Court opinions in doctrine space.” *American Journal of Political Science* 54(4):871–890.

Cross, Frank B. 2010. “Determinants of citations to Supreme Court opinions (and the remarkable influence of Justice Scalia).” *Supreme Court Economic Review* 18(1):177–202.

Desmarais, Bruce A and Skyler J Cranmer. 2017. Statistical Inference in Political Networks Research. In *The Oxford Handbook of Political Networks.*

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.

Lupu, Yonatan and James H Fowler. 2013. "Strategic citations to precedent on the US Supreme Court." *The Journal of Legal Studies* 42(1):151–186.

Pelc, Krzysztof J. 2014. "The politics of precedent in international law: A social network application." *American Political Science Review* 108(3):547–564.

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

Spriggs, James F and Thomas G Hansford. 2001. "Explaining the overruling of US Supreme Court precedent." *The Journal of Politics* 63(4):1091–1111.