Generative Dynamics of Supreme Court Citations: Analysis with a New Statistical Model

Christian S. Schmid*   Ted Hsuan Yun Chen †   Bruce A. Desmarais ‡

April 29, 2020

Abstract

The significance and influence of US Supreme Court majority opinions derive in large part from opinions' roles as precedents for future opinions. A growing body of literature seeks to understand what drives the use of opinions as precedents through the study of Supreme Court case citation patterns. We raise two limitations of existing work on Supreme Court citations. First, dyadic citations are typically aggregated to the case level before they are analyzed. Second, citations are treated as if they arise independent of one another. We make the case that studying Supreme Court citations as a dynamic citation network overcomes both of these limitations. We present a methodology for studying citations between US Supreme Court opinions at the dyadic level, as a network. This methodology—the citation temporal exponential random graph model—enables researchers to account for the effects of case characteristics and complex forms of network dependence in citation formation. We apply this methodology to a network that includes all Supreme Court cases decided between 1950 and 2015. We find evidence for dependence processes, including reciprocity, transitivity, popularity, and activity. The dependence effects that we identify are as substantively and statistically significant as the effects of the exogenous covariates we include in the model. The summary result from this analysis is that theoretical models of Supreme Court citation formation should consider both the effects of case characteristics and the structure of past citations.[1]

*Department of Statistics, The Pennsylvania State University, schmid@psu.edu

†Department of Political Science, The Pennsylvania State University, thc126@psu.edu

‡Department of Political Science, The Pennsylvania State University, bdesmarais@psu.edu

1 Introduction

United States Supreme Court opinions exercise authority and influence, in part, through their roles as precedents affecting future jurisprudence in the US. The findings regarding the nature of the influences of precedent on the Supreme Court 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 in an opinion. We follow an emerging body of work on legal citations, and treat the system of citations as a network (e.g., Harris, 1982; Caldeira, 1988; 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 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-year 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. These are case-level studies in that, based on the unit of analysis, it is impossible to determine both the origin and target case of a citation that contributes to the dependent variable.

An alternative approach to case-level analysis of citations would be to model them in the directed dyadic form through which they arise. A case decided at time t can cite (or not cite)

each case decided previously, and in the US Supreme Court, each other case decided at time t. We are aware of one prior study, Clark and Lauderdale (2010), in which a statistical model is used to analyze directed dyadic citations between cases. However, Clark and Lauderdale (2010) use a dyadic latent variable model in order to estimate ideal points for Supreme Court Opinions, but do not use this model to understand the relationships between explanatory variables and the formation of citation ties between opinions. We build upon the literature on citation analysis both methodologically and substantively. Methodologically, we develop a novel extension of a statistical model for networks, which we adapt to the network structural constraints of court citations. Second, we apply this methodology to a half-century of directed dyadic citations between U.S. Supreme Court opinions.

There exist two broad benefits of conducting empirical analyses of citations at the directed dyad 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 is 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 this paper 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 (ERGM)—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 1950 and 2015. 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 fit.

2 Network Processes in Supreme Court Citations

When it comes to the development and testing of theory, the defining feature of networks is that the micro-level unit of analysis—the relationship between two entities (i.e., the citation from one opinion to another) is a component of a complex system of relations. The formation (or lack thereof) of that relationship cannot be fully understood without considering how the relationship fits into the system. Analytical designs that account only for covariates in explaining tie formation are incomplete theoretically, and, as a consequence, are subject to a form of omitted variable bias (Cranmer and Desmarais, 2016). Citations in legal opinions are unique in terms of the windows into network dependencies offered by the texts of the opinions. A number of common structural dependencies that are found in networks are likely to apply to citations in Supreme Court opinions. In this section we present these dependence forms, and document the mechanisms by which they arise through archetypal passages in example opinions.

We should note that we do not distinguish between positive and negative citations in this theoretical framework. The dynamics we outline are not specific to a particular type of citation. The only distinction we draw (in the empirical analysis) is the instance in which a case has been overruled. In the extreme instance of overruled precedents, we assume that (and test whether) a case is much less likely to be cited after it has been overruled.

The first network property that we theorize in the context of Supreme Court citations is transitivity. In a network of directed relations (e.g., A cites B, but B doesn't cite A) transitivity refers to the tendency for A to send a tie to C if A sends a tie to B and B sends a tie to C (Holland and Leinhardt, 1971). In undirected networks, transitivity is simply the process by which friends of friends become friends (i.e., a friend of a friend is a friend). The term, "transitive closure" refers to a tie forming from A to C in response to extant ties from A to B and B to C. When writing opinions, Supreme Court justices present the legal bases for their rulings, which often involves discussing the most primary/relevant precedents underpinning these legal bases, but also the precedents and legal rules on which the primary precedents were based. This process of presenting several layers/levels of precedent in an opinion follows the structure of transitive closure exactly—opinion A cites opinion B as a primary precedent, and then cites opinion C because opinion B cites opinion C. The two examples presented below illustrate this process.

In the first example, a passage from Kansas v. Marsh (548 U.S. 163, 2006)—a case considering the constitutionality of a death sentence statute in Kansas. In this example, the case Stringer v. Black is cited by Kansas v. Marsh as a case that is quoted by Sochor v. Florida. The primary precedent under discussion in this passage of the opinion is Sochor v. Florida, but Stringer v. Black is cited as a result of its role in the Sochor v. Florida opinion.

> 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).

The second example, which we illustrate visually in Figure 1 is a passage from Seminole Tribe of Fla. v. Florida, (517 U.S. 44 1996)—a case addressing the rights of groups and citizens to sue states in federal court. In this example, Pennsylvania v. Union Gas Co (491 U.S. 1, 1989) is the primary precedent being critiqued, and several cases are cited and discussed in terms of their roles as precedents in the Union Gas opinion. We highlight one—Parden v. Terminal R.
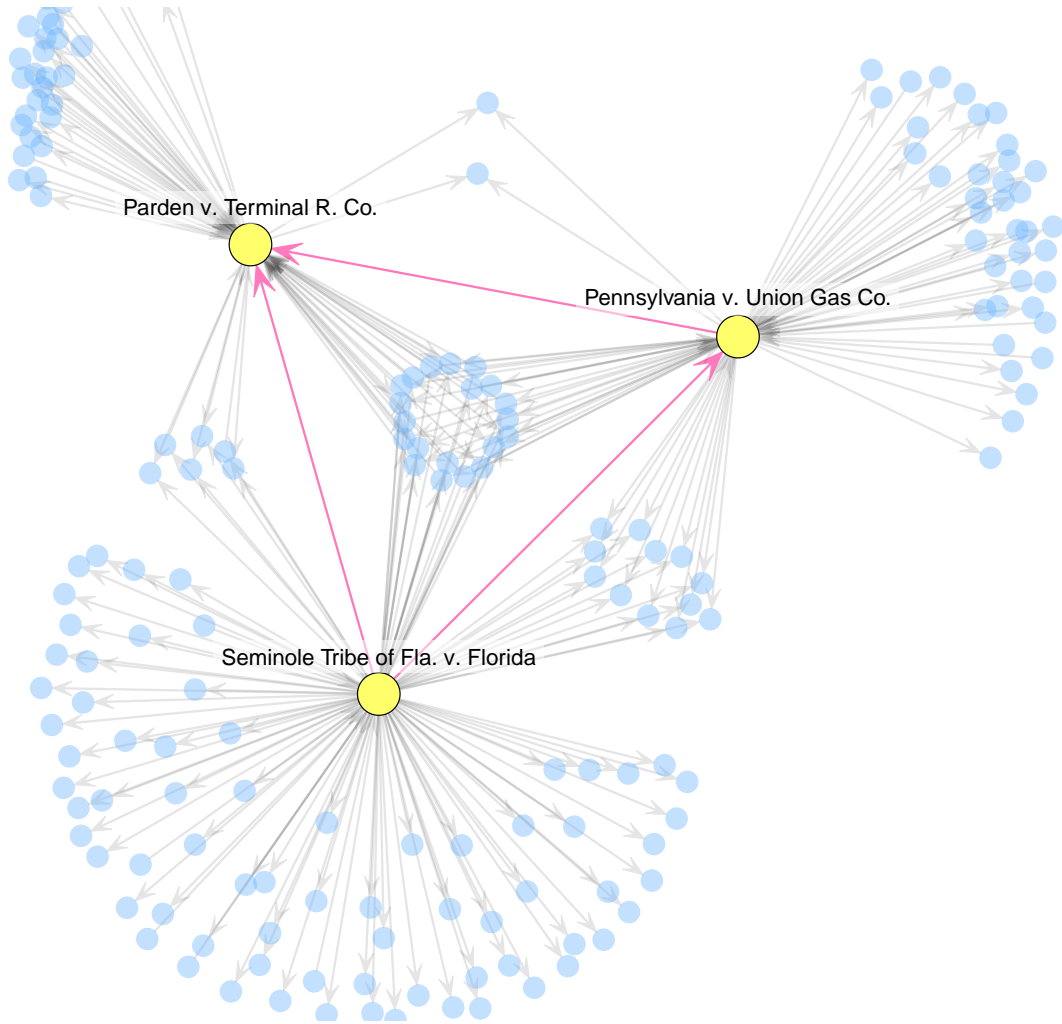
Figure 1: Illustration of transitive triangle connecting US Supreme Court opinions through citations.

Co., which is cited and discussed in both the Seminole and Union Gas opinions.

> 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).'

The second network property we consider in the context of Supreme Court citations is reciprocity. Reciprocity (also referred to as mutuality) is the tendency for node B to send a tie to node A in response to or coordination with A sending a tie to B (Garlaschelli and Loffredo, 2004). It is typically not possible for reciprocated ties to form in legal opinions. Most citations reference past opinions that were issued before the citing opinion's case was even argued before the Court. However, opinions written within the same Supreme Court term are often drafted in tandem, and can cite each other reciprocally. Opinion A citing opinion B within the same term represents a signal that opinion B is relevant to the legal reasoning underpinning opinion A. Unlike the citations themselves, the applicability of legal rules or lines of reasoning across cases is not directed—if A is relevant to B, B is highly likely to be relevant to A. We expect opinions written within the same term to exhibit a high degree of reciprocity. Below we provide two example passages from opinions that illustrate the phenomenon of within-term reciprocity.
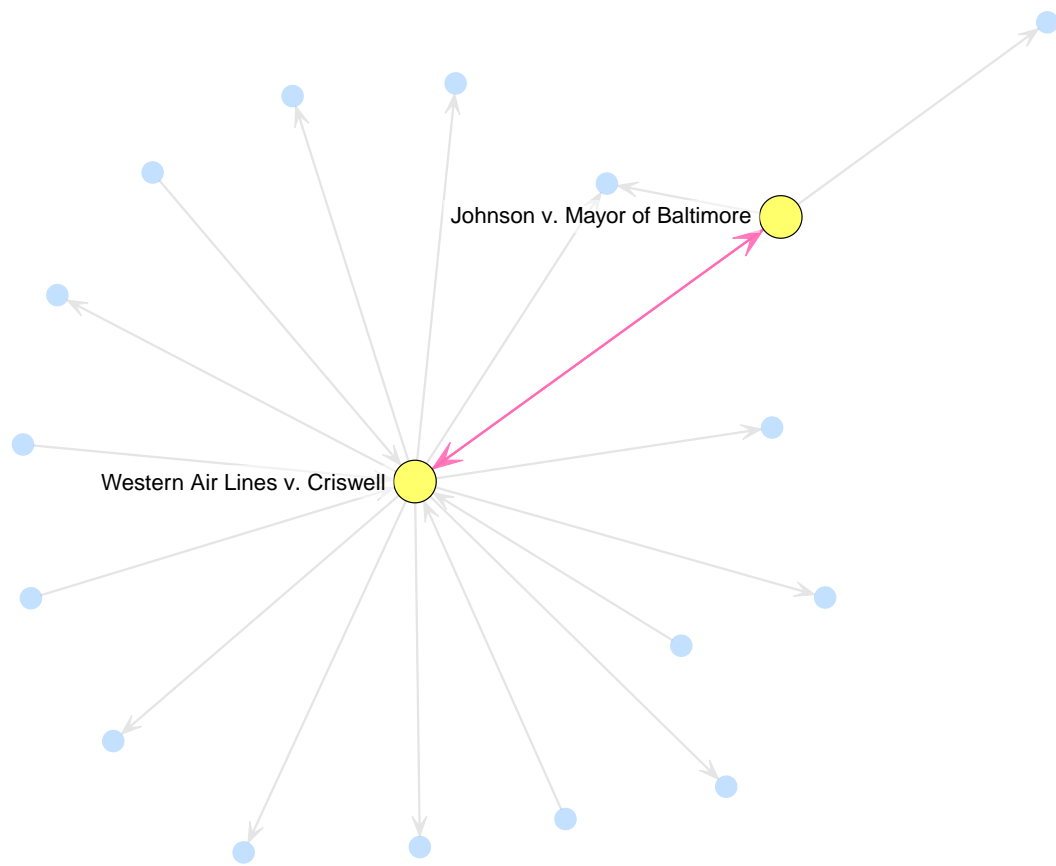
Figure 2: Illustration of a reciprocal tie between two US Supreme Court opinions.

The first case in our example reciprocal dyad is a passage from Western Air Lines v. Criswell (472 U.S. 400, 1985)—a case considering mandatory retirement in the context of age discrimination laws. The second case in the dyad, Johnson v. Mayor of Baltimore (472 U.S. 353, 1985) is another case considering whether mandatory retirement violates the Age Discrimination in Employment Act. The mutual edge connecting these two cases is visualized in Figure 2. These cases addressed very similar legal questions, which increased the likelihood that they would inform each other, and the opinions were written within the same term, which made it possible for them to cite each other.

> *From Western Air Lines:* 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.

> *From Johnson:* 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.

The third, and final, network property we consider in the context of Supreme Court citations is popularity. Popularity, also termed "preferential attachment" is the tendency for ties to be sent to nodes to which many ties have already been sent (Barabási and Albert, 1999). Citations to an opinion signal both the Court's awareness of the legal reasoning of the case and the Court's evaluation that the opinion is an authoritative precedent. The more citations, the stronger this signal. Landmark cases, or those that establish new legal rules, are particularly authoritative and accrue citations from most future opinions that follow the respective line of reasoning. The passage below, from Oregon v. Mitchell (400 U.S. 112, 1970)—a case on the legality of state age restrictions on voting in federal elections—illustrates this popularity dynamic. In this opinion passage Baker v. Carr is cited in reference to its role as a landmark precedent, and noted for the number of other cases by which it has been followed. and for which an authoritative opinion is referenced, and even discussed in terms of the number of other cases by which it was followed. The language in this passage suggests that the attention to Baker v. Carr in previous Court opinions is in part responsible for its authority in Oregon v. Mitchell. The citations to Baker v. Carr are visualized in Figrue 3.

> 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..."

These three properties—transitivity, reciprocity, and popularity form the core of our network theory of U.S. Supreme Court citations. We also seek to model the effects of covariates (i.e., case features) on citation formation. In the next section, we introduce a modeling framework that can incorporate all of these effects into a single model, and adapt to the peculiar structure
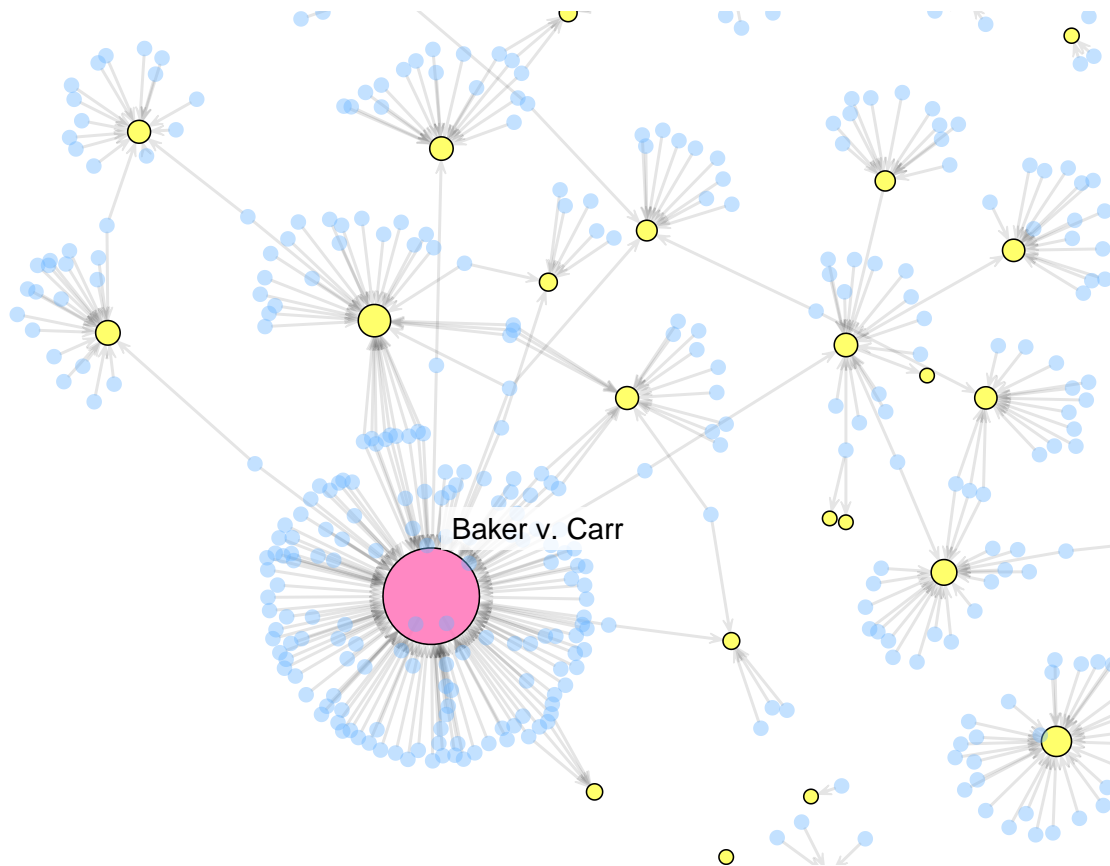
Figure 3: Illustration of ties sent to a landmark Supreme Court opinion via citations.

of citation matrices.

3 The Citation Exponential Random Graph Model

We develop a methodology that can be used to jointly test for the effects of covariates on citations—as have been studied in prior research, and test for dependence effects, as we hypothesize above. To accomplish this, we extend a model that has been developed to jointly represent covariate and dependence effects in network data—the exponential random graph model (ERGM) (e.g., **??**Graif, Lungeanu and Yetter, 2017; **?**) The network structures for which ERGMs are currently designed are insufficient to account for the structure of citation networks.

3.1 c-ERGM definition

We develop the citation ERGM (c-ERGM), to account for the structural constraints that apply to the network of Supreme Court citations. These structural constraints amount to three departures from the structure of networks for which ERGMs are currently designed. First, the citation network is partially acyclic. If two cases are decided during the same term, they can cite each other, forming a mutual edge (or two-cycle). However, if case i is decided before case j, case j can cite case i, but case i cannot cite case j. Second, new edges can be created over time, but cannot be eliminated. Unlike in, e.g., an alliance network, in which two countries can dissolve an alliance, once a citation exists in a citation network it cannot be dissolved. Third, the set of nodes in the network must increase for new edges to be created. In conventional ERGMs, the number of nodes in the network can increase or decrease in each time period, and is typically stable over time. In a citation network, new edges (and non-edges) are introduced over time via the introduction of new nodes. The structure of the citation network is depicted in Figure 4—a hypothetical citation network established over three time periods, with three cases decided in each time period. We denote C_t to be the set of citation and non-citations added to the network at time t (i.e., via the addition of three cases), $C_{<t}$ to be the citations among cases decided before time t, and $C_{\leq t}$ to denote the entire set of citations and non-citations on which C_t can

depend through the c-ERGM specification.

The likelihood function of the c-ERGM is given by

$$l(\boldsymbol{\theta}, C_{\leq t}) = \frac{\exp\left[\boldsymbol{\theta}'\boldsymbol{h}(C_t, C_{<t})\right]}{\sum_{C_t^* \in \mathcal{C}_t} \exp\left[\boldsymbol{\theta}'\boldsymbol{h}(C_t^*, C_{<t})\right]}, \tag{1}$$

where $\boldsymbol{\theta}$ is a vector of real-valued parameters, and $\boldsymbol{h}(C_t, C_{<t})$ is a vector of scalar-valued functions that each quantify a feature of the citation network (e.g., the relationship between citation ties and a case attribute, the number of mutual edges in the citation network). $\exp\left[\boldsymbol{\theta}'\boldsymbol{h}(C_t, C_{<t})\right]$ is a positive weight that is proportional to the probability of observing any particular form of the citations and non-citations added to the network at time t. The denominator of Equation 1 represents a normalizing constant, in which the positive weight is summed over all possible configurations of C_t, from the network in which the cases added to the network at time t send no citations at all, to the network in which the cases added at time t cite every possible case, and everything in between.

Though the likelihood function of the c-ERGM may appear quite different from that of statistical models conventionally used in political science, analysis of the conditional probability of a single citation from case i to case j reveals that we can interpret the parameters similar to logistic regression coefficients.

$$P(C_{ij,t} = 1 | C_{-ij,t}, C_{<t}) = \frac{\exp\left[\boldsymbol{\theta}'\boldsymbol{h}(C_t, C_{<t}|C_{ij,t}=1)\right]}{\exp\left[\boldsymbol{\theta}'\boldsymbol{h}(C_t, C_{<t}|C_{ij,t}=1)\right] + \exp\left[\boldsymbol{\theta}'\boldsymbol{h}(C_t, C_{<t}|C_{ij,t}=0)\right]},$$

$$= \frac{1}{1 + \exp\left[-\boldsymbol{\theta}'\left(\boldsymbol{h}(C_t, C_{<t}|C_{ij,t}=1) - \boldsymbol{h}(C_t, C_{<t}|C_{ij,t}=0)\right)\right]},$$

where $C_{ij,t} = 1$ indicates that case i cites case j, $C_{ij,t} = 0$ indicates that case i does not cite case j, $C_{-ij,t}$ is the observed elements of C_t except $C_{ij,t}$, and $\left(\boldsymbol{h}(C_t, C_{<t}|C_{ij,t}=1) - \boldsymbol{h}(C_t, C_{<t}|C_{ij,t}=0)\right)$ is the change in $\boldsymbol{h}(C_t, C_{<t})$ that results from toggling $C_{ij,t} = 0$ to $C_{ij,t} = 1$. This re-arrangement illustrates that the parameters can be interpreted in terms of the change in the log odds of a citation from i to j given a one-unit increase in the corresponding element of $\boldsymbol{h}$, conditional on the other citations observed in the network. For example, if the value of θ corresponding to an
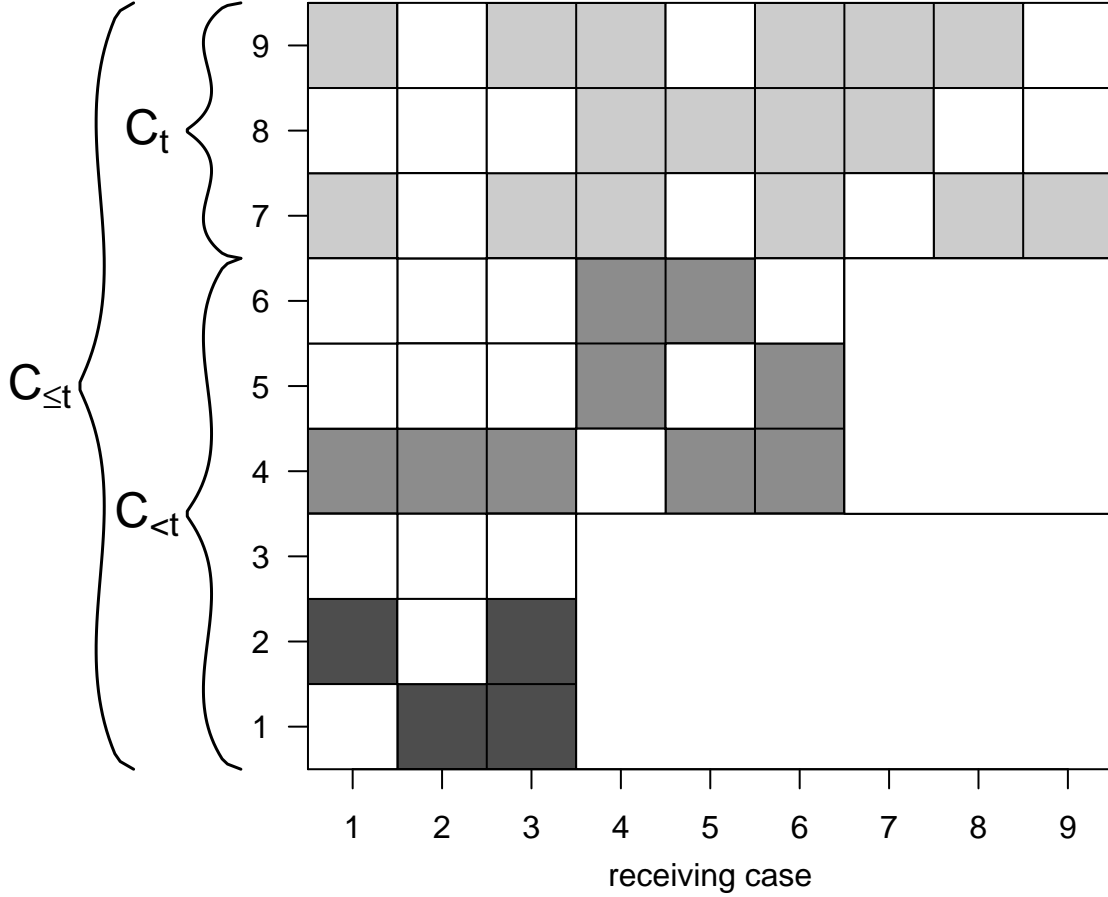
Figure 4: Illustration of temporal structure of the Supreme Court Citation Network. $C_{\leq t}$ is the entire set of citations (and non-citations) on which citations and non-citations at time t (i.e., C_t) depend. C_t are conditioned on the citations and non-citations established before time t (i.e., $C_{<t}$). The shaded small squares are hypothetical observed citations, and the white small squares are citations that could have been observed but were not. The regions of the citation matrix that are represented by large white rectangles are citations that could not have been observed since the citing case would have been decided in a term that preceded the term of the cited case. The citing case ID is given in the row and the prospective cited case is given in the column.

element of h that counts the number of mutual edges in the network is 0.5, then the log odds of observing $C_{ij,t} = 1$ increases by 0.5 if case i is cited by case j (as compared to the configuration in which case j does not cite case i). The logit form conditional probability is well known for the ERGM family (Goodreau, Kitts and Morris, 2009).

We should note that there is nothing about the c-ERGM (or any ERGM family model, for that matter) that permits the identification of causal effects of one tie on the formation of another. For example, if we find a positive popularity effect, the term for that dynamic would indicate that there is an increase in the log odds that case i cites case j with each other case that has cited case j. That finding could be driven by one of two (or some combination of both) dynamics. First, it could be the case that a large volume of citations to a case in past opinions draws the attention of the justices to the "popular" case—as in the Baker v. Carr passage cited above. Second, it could be that we have omitted a case-level variable that is strongly associated with the number of citations accumulated by a case, and the number of citations in past cases is operating as a proxy for this unmeasured case characteristic. We cannot statistically differentiate between these two dynamics. However, in both of these conditions, and analogous conditions related to other dependence properties, citations (or non-citations) to a case are not independent conditional on the other variables, and the c-ERGM can be specified to account for the dependence in the data.

3.2 c-ERGM Estimation

The normalizing constant in Equation 1 is intractable. For example, in the simple case of adding three cases to a network in which six cases already exists—like that depicted in Figure 4—there are 16,777,216 unique configurations of C_t that could be observed. The typical Supreme Court term involves adding hundreds of cases to a network that already includes thousands of previous cases. This means that straightforward methods of maximum likelihood estimation (MLE) are infeasible with the c-ERGM.

The common alternative relies on Monte Carlo methods to approximate the normalizing constant by simulating a large set of networks (Hunter and Handcock, 2006; Hummel, Hunter and

Handcock, 2012). The resulting estimator, the Monte Carlo MLE (MCMLE), is approximately consistent, meaning that it converges to the MLE as the sample size, i.e., the number of simulated networks, increases. However, one drawback is that with the number of nodes in the Supreme Court citation network being in the order of $10,000$, obtaining the MCMLE is computationally expensive (Schmid and Desmarais, 2017) and the success of the algorithm heavily relies on the starting parameter vector θ_0, which is ideally chosen in the proximity of the unknown MLE (Hummel, Hunter and Handcock, 2012). The prevailing choice for θ_0 is the maximum pseudo-likelihood estimation (MPLE) (Strauss and Ikeda, 1990), a fast estimation method that is defined as maximizing the log product of the conditional probability of each citation (and non-citation), conditional on the other elements of the observed citation network. The joint probability of all citations is replaced by the product over conditional probabilities, which, as we demonstrated above, assume a logit form. The MPLE is simple to obtain, but does not guarantee a starting value close to the MLE (Schmid and Hunter, 2020).

We estimate the MCMLE for every term of the c-ERGM between 1950 and 2015. While the network for the 1950 term consists of 1962 nodes, the size of the network monotonically increases to $10,020$ nodes for the 2015 term, making estimation more challenging as the network grows. The growth in the network also serves as our reason for estimating a separate model for each term. First, the network is large, with each term introducing the potential for hundreds-of-thousands of new citations. This means we have the statistical power necessary to estimate a separate set of parameter values for each term. Second, based on the findings of **?**, we know that the set of parameter values that best fits a network should change as more nodes are added to the network. This means it would be inappropriate to assume a constant set of parameter values as more cases are added each term.

The MCMLE of networks up until the 90s was obtainable in a reasonable time frame starting at the MPLE and sampling $10,000$ networks to approximate the normalizing constant. However, the estimation of most networks in the 90s with the MPLE as starting values was not feasible in an reasonable time frame anymore. Instead, we improved the choice of starting value θ_0 by fixing it at the MCMLE of the previous term $t - 1$ and successfully obtained the MCMLE of

the network at term t. But even this approach started to fail for networks around the turn of the millennium. Neither the MPLE nor the MCMLE of previous terms as starting values led to successful estimation, and neither did the Stepping algorithm (Hummel, Hunter and Handcock, 2012). The MCMLE for these large citation networks was obtained by setting the starting value according to an approach introduced by Schmid and Hunter (2020). This method is based on the fact that the MLE of exponential family distributions is solely a function of the vector of sufficient statistics $h(C_t, C_{<t})$, meaning that the MLE of two networks A and B is equal if $h(A) = h(B)$. However, the MPLE of networks with the same sufficient statistics is not necessarily the same. Instead of starting the MCMLE algorithm at the network's MPLE, Schmid and Hunter (2020) propose searching for a network C_t^* using simulated annealing (Kirkpatrick, Gelatt and Vecchi, 1983), with $h(C_t, C_{<t}) = h(C_t^*, C_{<t})$ and a weak dependence structure among unfixed ties. The MPLE of such a network is similar to the MLE, making it an effective starting value for the MCMLE algorithm.

4 Empirical Analysis

Our three data sources for this study include the Supreme Court Database (SCDB) (Spaeth et al., 2014), Martin-Quinn scores (Martin and Quinn, 2002), and Supreme Court citation data provided by the CourtListener Free Law Project (Lissner and Carver, 2010). In the next section we explain the variables we construct using these data sources. We limit the Supreme Court terms included in our analysis to those that are covered by all three of these data source (1937–2015)[2]. The Supreme Court citation network from $1937 - 2015$ consists of $9,945$ cases. The breakdown of the data by the Court's Chief Justice is presented in Table 1. The network has a total of $111,986$ citation ties.[3] The in- and outdegree distributions (i.e., the distributions of the

---

[2]There were 145 cases that were listed in the SCDB but could not be matched to a case in the CourtListener data. We decided to exclude these 145 cases from our analysis. Additionally, since the most commonly used data on Supreme Court citations in political science come from Fowler et al. (2007), we compared the Fowler data to the CourtListener data in the time interval that they overlap (1937–2001). We found considerable agreement—over 95% of the citations recorded in the CourtListener dataset were also in the Fowler dataset, and over 96% of the citations recorded in the Fowler data were also in the CourtListener dataset.

[3]In order to focus on citation actions that were not intended to totally invalidate an opinion, we exclude 315 citations that caused the cited case to be overruled. The data on overruling citation came from US. Congress.

number of citations sent and received by cases, respectively), are visualized in Figure 5. The maximum indegree (i.e., number of cases citing to a case) is 230 and the maximum outdegree (i.e., number of cases cited by a case) is 162. The majority of cases cite to and/or are cited by twenty or fewer other cases.

	Terms	Total Number Cases	Cases/Term
CE Hughes*	1937 - 1941	628	125.6
HF Stone	1942 - 1946	756	151.2
FM Vinson	1946 - 1953	789	98.63
E Warren	1954 - 1969	2149	126.41
WE Burger	1970 - 1986	2805	155.83
W Rehnquist	1987 - 2005	2022	106.42
J Roberts **	2006 - 2015	796	79.6

Table 1: For the time range of interest (1937 - 2015) 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.
** J Roberts still serves as chief justice (retrieved 4/2018).

The degree distributions indicate that there is a long tail to both the number of citations sent and received. These long-tailed (i.e., high kurtosis) distributions provide preliminary evidence of substantial heterogeneity in the features that drive citations to and from cases (Strogatz, 2001). Figure 6 displays the citation network network from 1937–2001. We see here that the densest rates of tie formation tend to be between consecutive courts (e.g., the Stone Court is much more tightly tied to the Hughes Court than the Rehnquist Court is to the Hughes Court). This pattern lends preliminary/descriptive support to the hypothesis that the rate of citations to a case decrease over time.

4.1 c-TERGM specification

The c-TERGM we estimate includes two classes of terms—one that captures the effects of covariates on tie formation, and another that captures the complex dependence processes that we expect to observe in the Supreme Court Citation Network. We describe our model specification

---

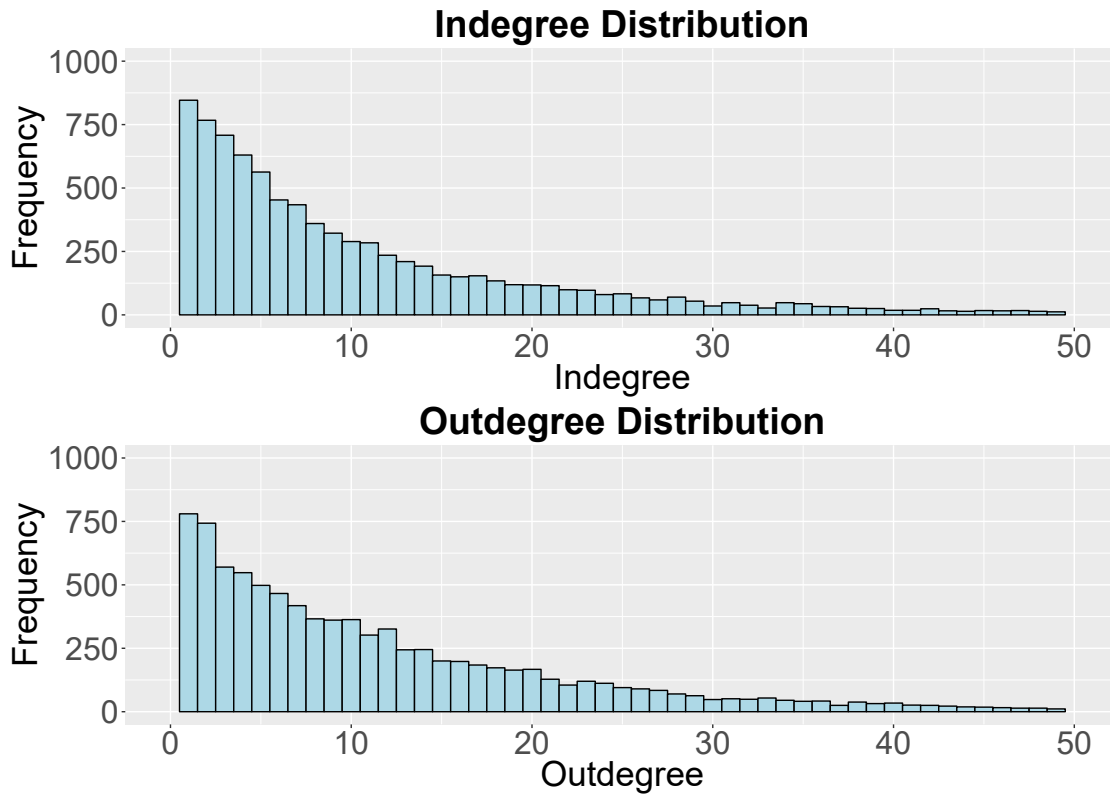Senate (2016). Our results are virtually unchanged if we include the overruling citations.

Figure 5: 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.
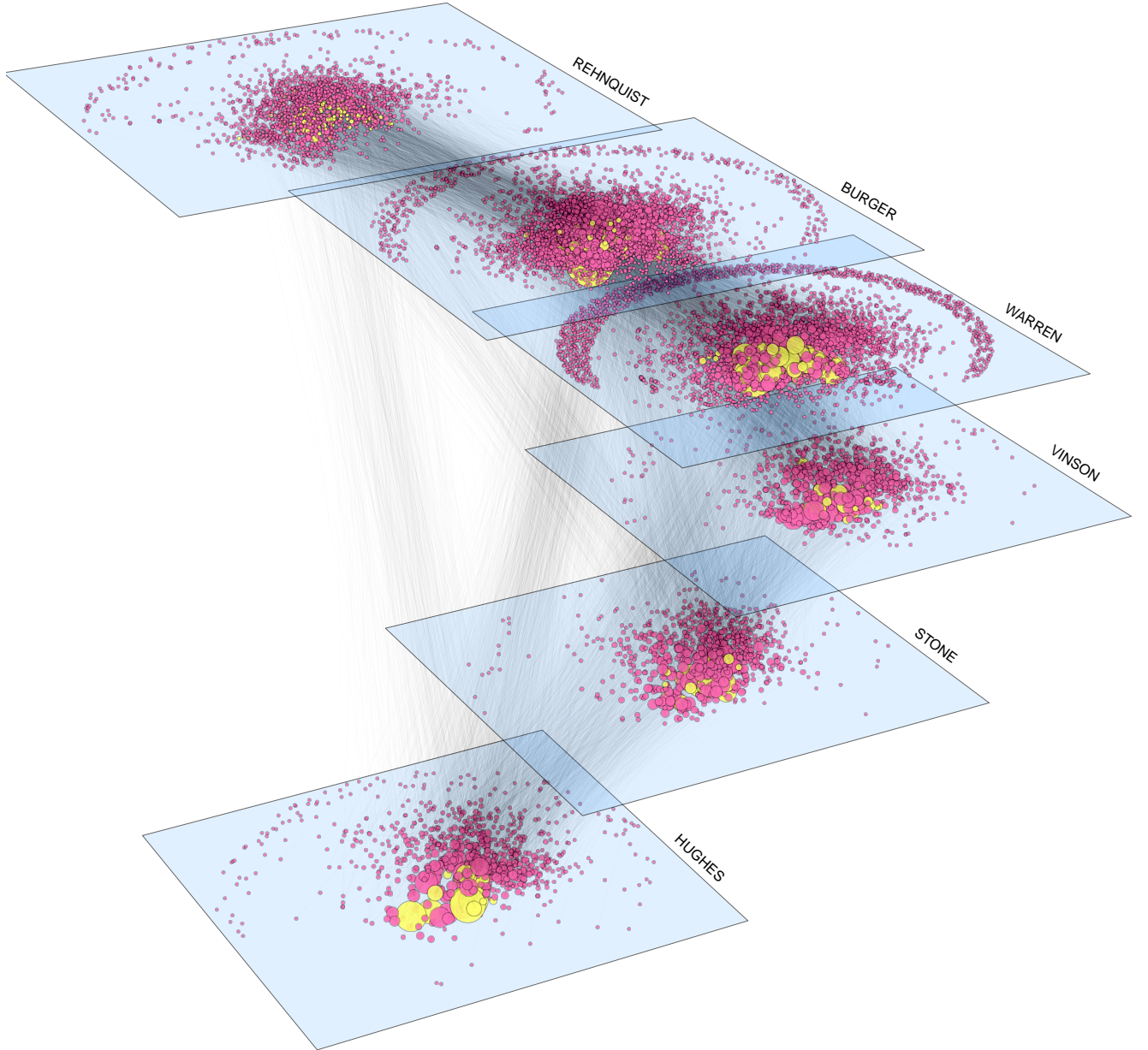
Figure 6: Supreme Court Citation Network, 1937-2001. Nodes are Supreme Court cases, with size based on incoming citations. Salient cases (*Oxford Guide*) are in yellow. Each network layer contains cases decided under a different chief justice. Because our data is temporally bound, cases for the Hughes and Rehnquist courts are not complete.

by defining the terms within these two classes.

4.1.1 Covariate terms

Covariate effects are accounted for in the c-TERGM via the term that is used to specify the effect of covariates in other ERGM family models, as

$$h_{covariate}(C_t, C_{<t}, X) = \sum_{ij} C_{ij,t} X_{ij}.$$

Since $C_{ij,t}$ is a binary indicator of whether case i cites case j, $h_{covariate}$ amounts to the sum of covariate values among directed dyads for which we observed a citation. The dyadic interpretation of the coefficient attributed to this term is the change in the log odds of a tie from i to j given a unit increase in X_{ij} (i.e., exactly the interpretation of the effect of a covariate in logistic regression). We include several exogenous covariates based on this standard statistic formulation.

The first covariate (i.e., X) we incorporate into the model accounts for the degree to which cases cite those that are similar in terms of the ideological positions of the justices who supported the decision. We account for this effect following Spriggs and Hansford (2001), who 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. Clark and Lauderdale (2010) estimates a latent coordinate model of Supreme Court opinions based on the network of case-to-case citations. They find that the majority opinion falls at the ideal point of the median member of the majority coalition in the case. We include a covariate term in which X_{ij} is the absolute difference between the Martin-Quinn scores (Martin and Quinn, 2002) of the median justices in the majority coalitions for cases i and j. We expect this variable to have a negative effect, which would correspond to cases citing those to which they are ideologically similar.

We include one covariate that accounts for the rate at which cases that fall under the same issue area cite each other. The variable X^{ij} indicates whether cases i and j share the same issue area, e.g., $X^{ij} = 1$ if cases i and j are classified into the same issue area, and 0 otherwise.

1	Criminal Procedure	8	Economic Activity
2	Civil Rights	9	Judicial Power
3	First Amendment	10	Federalism
4	Due Process	11	Interstate Relations
5	Privacy	12	Federal Taxation
6	Attorneys	13	Miscellaneous
7	Unions	14	Private Action

Table 2: Assigned numbers for the variable *Issue Area*. This information originates from the Supreme Court Database code book.

The Issue area data comes from the Supreme Court Database (SCDB) (Spaeth et al., 2014). We include these variables because Cross (2010) finds that the number of citations to Supreme Court opinions depends heavily on the issue area of the case. Table 2 reports the issue areas covered in our data.

We also include an indicator variable that models the rate at which justices cite themselves. Similar to the same issue area variable, this effect is modeled with a single indicator variable in which $X_{ij} = 1$ if the majority opinions in cases i and j were written by the same justice. The data is taken from the SCDB (Spaeth et al., 2014). Justification for including this variable?

We model the way in which citations to a case depend upon the age of a case. For this we use a second-order polynomial in which one covariate X^{ij} is defined as the age of case j at the time that case i is decided, and another term in which X^{ij} is the squared age of case j at the time that case i is decided. We include these covariates because Black and Spriggs (2013) find that the number of citations to a Supreme Court case over time depends significantly on the age of the case, characterized by a sharp drop off and leveling out with age.

Benjamin and Desmarais (2012) study the propensity for cases to be overruled and cited in other negative ways. They find that cases with majority coalitions that are large and ideologically broad are less likely to be cited negatively. In our data we do not differentiate between negative and positive citations, but since the overwhelming majority of citations are positive, we hypothesize that the effects they found will be reversed in our analysis. We include one covariate in which X^{ij} is the number of justices in the majority coalition for case j. We also include another covariate in which X^{ij} is the absolute difference between the maximum and

minimum ideal points of the justices in the majority coalition for case j. We expect both of these covariates to have positive effects.

The final control variable we include in the model is one in which $X_{ij} = 1$ if case j was overruled prior to the term in which case i was decided. This variable, quite simply, models the effect of being overruled on the rate of citation to a case after the overruling citation. Fowler and Jeon (2008) find that citations to a case drop off quickly after the case has been overruled.

4.1.2 Dependence terms

We include dependence terms to account for each of the dynamics that we hypothesize will characterize the Supreme Court Citation Network—transitivity, reciprocity, popularity, and activity.

The common dependence effects for these four dynamics are the number of triangles (case i cites cases j and k, and case j cites case k), the number of mutual dyads (case i cites j and vice versa), the number of out-2-stars (case i cites j and k), and the number of in-2-stars (case i is being cited by cases j and k) in $C_{\leq t}$. Unfortunately, adding the triangle, out-2-star, or in-2-star statistic causes model degeneracy (Handcock, 2003). Model degeneracy is a common obstacle when fitting ERGMs and occurs when the model places most of its probability mass on just a few networks, usually the empty and the full network. In this scenario, the simulated networks are too different from the observed network, making the underlying distribution defined by the model extremely unrealistic.

The statistic included in the model for reciprocity counts the number of mutual dyads (i.e., dyads in which cases i cites case j, and case j cites case i) in $C_{\leq t}$. In practice, this is the number of mutual dyads in $C_{\leq t}$, since mutual dyads must emerge among opinions that were written in the same term. The log odds interpretation of the reciprocity statistic is that if opinions in case i cite case j, the log odds that case j also cites i change by the theta relative to the configuration in which case j does not cite case i.

Accounting for the popularity effect through the in-2-star statistic is prone to cause model degeneracy. In an attempt to stabilize the model against model degeneracy as well as to still

capture the popularity effect, Snijders et al. (2006) introduces the *alternating-in-k-star* statistic, which was shown to be equivalent to the *geometrically weighted indegree distribution* (gwidegree) statistic introduced by Hunter and Handcock (2006). Define $D_r(C_t, C_{<t})$ as the number of nodes with indegree $r, r \in \{0, \ldots, N_t - 1\}$, where N_t is the number of cases in the network at time t, then gwidegree represents a network's indegree distribution in a single statistic by geometrically weighting the degree distribution

$$h_{gwidegree}(C_t, C_{<t}, \lambda) = \lambda \sum_{r=1}^{N_t-1} \left(1 - (1 - \frac{1}{\lambda})^r \right) D_r(C_t, C_{<t}). \tag{2}$$

The decay parameter $\lambda \in [0, \infty]$ controls the decline of the weight put on $D_r(C_t, C_{<t})$ as r increases which means that the higher λ the more the statistic responds to changes for cases with a high indegree. We fixed $\lambda = 1$. Adding one single edge to the network changes the indegree distribution of the network so that a case with an indegree of s turns into a case with an indegree of $s + 1$. With our notation this means that D_r and D_{r+1} get replaced by $D_r - 1$ and $D_{r+1} + 1$, while no changes for $D_\ell, \ell \in \{0, \ldots, N_t - 1\} \setminus \{r, r - 1\}$ are made. This changes the log odds in the following way

$$\text{logit} P(C_{ij,t} = 1 | C_{-ij,t}, C_{<t}) = \exp\left(\theta \cdot (1 - \frac{1}{\lambda})^r \right).$$

As stated by Levy (2016) this means for the interpretation that a negative parameter value indicates the centralization of edges while a positive parameter indicates the dispersion of edges. Since we expect highly cited cases to be more likely to be cited again, we expect the gwidegree parameter to be negative.

In order to account for the activity effect, the straightforward thing to do is to add the *geometrically weighted outdegree* (gwodegree) statistic, which is defined as gwidegree with the difference that $D_r(C_t, C_{<t})$ represents the number of nodes with outdegree r. However, adding the gwodegree statistic causes the model to degenerate again. We circumvent degeneracy by adding the outdegree of every node that entered the network prior time t as a receiver effect. In other words, this statistic adds up the outdegree of cases that are being cited at time t and

examines whether cases that show a high degree of activity are being more likely to be cited in the future.

We include two statistics to account for different types of transitivity. The first transitivity statistic calculates the number of transitive ties, $i \to j$, where there is a directed path of length two from i to j through at least one case k, and j and k were written during a different term than i. We will refer to this statistic as the *different term transitivity* statistic. As illustrated in chapter 2 we theorize Supreme Court citations to pattern a high degree of transitivity. As a result, we expect positive parameter estimates for the different term statistics.

The second transitivity statistic is the *geometrically weighted edgewise shared partners* (gwesp) statistic introduced by Hunter and Handcock (2006). For directed networks the edgewise shared r-partners statistic (ESP_r) can take multiple forms, but we integrate the definition where cases i and j that are connected by a citation, also cite r common other cases. We follow the notation of Butts (2008) and call this ESP specification *outgoing shared partners* (OSP). Similar to the gwidegree-statistic the gwesp-statistic incorporates a network's edgewise shared partners distribution by replacing D_r by ESP_r. Note that unlike the different term transitivity statistic the gwesp statistic does not require the cited cases to be written in a previous term.

4.2 Predictive Performance

Our case for studying legal citations at the directed dyadic level hinges upon the contribution to modeling offered by incorporating network dependence. To quantify this contribution, we use AIC and BIC out-of-sample prediction. 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 (Jensen and Cohen, 2000; Ward, Greenhill and Bakke, 2010). 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. Out-of-sample prediction is a common way to evaluate methods for modeling ties in networks, and has recently been applied to TERGMs in particular (Desmarais and Cranmer, 2013; Cranmer and Desmarais, 2017).
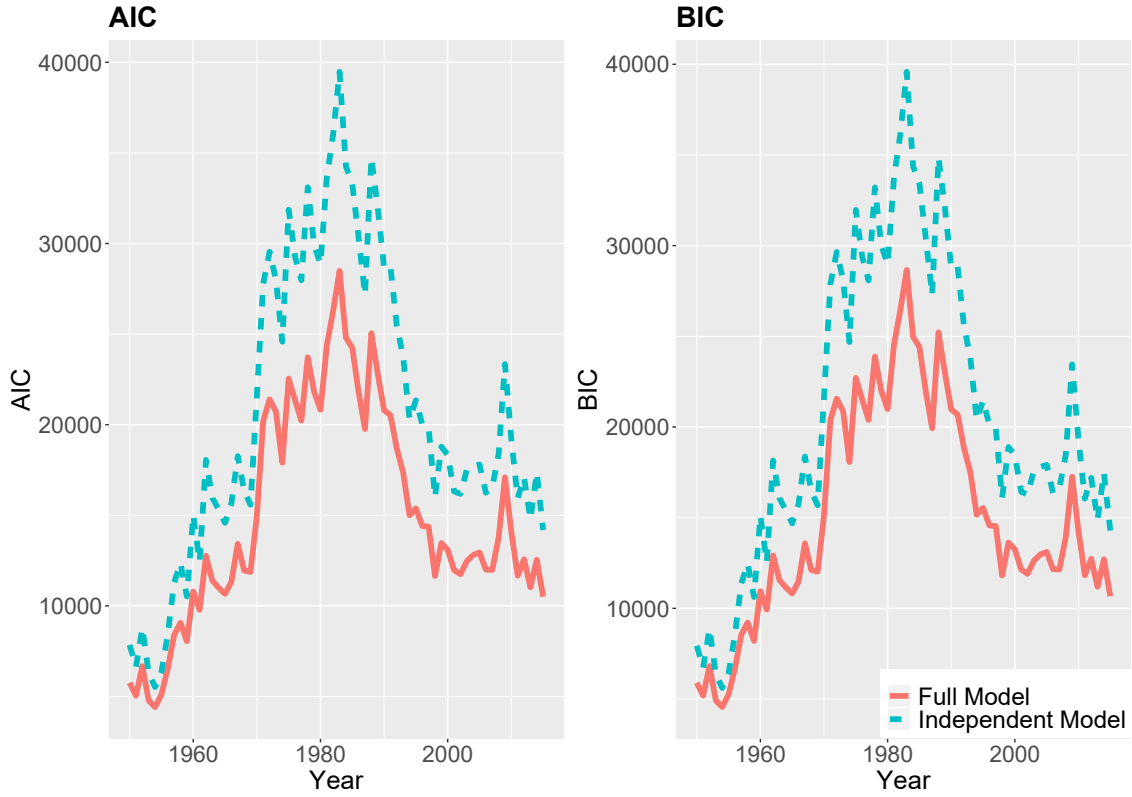
Figure 7: AIC and BIC for the full and the independent model for the time frame 1950-2015

Figure 7 depicts the AIC and BIC comparison for two different models for each term from 1950 to 2015. The full model is defined as the model that includes all covariate terms and dependence terms described in 4.1.1 and 4.1.2. The independent model only consists of covariate terms, not accounting for any structural dependencies in the networks. We see that the AIC and BIC is considerably smaller for the full model throughout the period of consideration. On average the independent model yields a 5525 point higher AIC and an average 5582 point higher BIC. The maximum difference is reached in 1983 with a difference of 10,990 points for the AIC and 10,930 points for the BIC.

4.3 c-TERGM Results

The results from the c-TERGM are presented Figures 8 and 9. In these figures we depict coefficient estimates and 95% confidence intervals for all of the effects for all the variables

we discussed in sections 4.1.1 and 4.1.2. The coefficients in an ERGM family model can be interpreted in the same way as logistic regression coefficients—they give the change in the log odds of a tie from i to j given a one unit increase in the respective variable. For this reason, the number of edges in a network corresponds to the intercept in an ERGM. We first discuss the dependence effect estimates in Figure 8. A circle indicates that the estimate was significant at an $\alpha = 0.05$ level, while and square translates to a p-value between 0.05 and 0.1. A triangle signals that the estimate for a given year was not significant at an $\alpha = 0.1$ level.

The panel showing the results of the reciprocity statistic does not provide estimates for ever term. In these missing terms there were no two cases that would cite each other, making the estimation of the reciprocity statistic impossible. A similar scenario is visible for the *Overruled Cases* variable. In the time period from 1953-1956 no cases that have been overruled have been cited which results in a non-estimable effect.
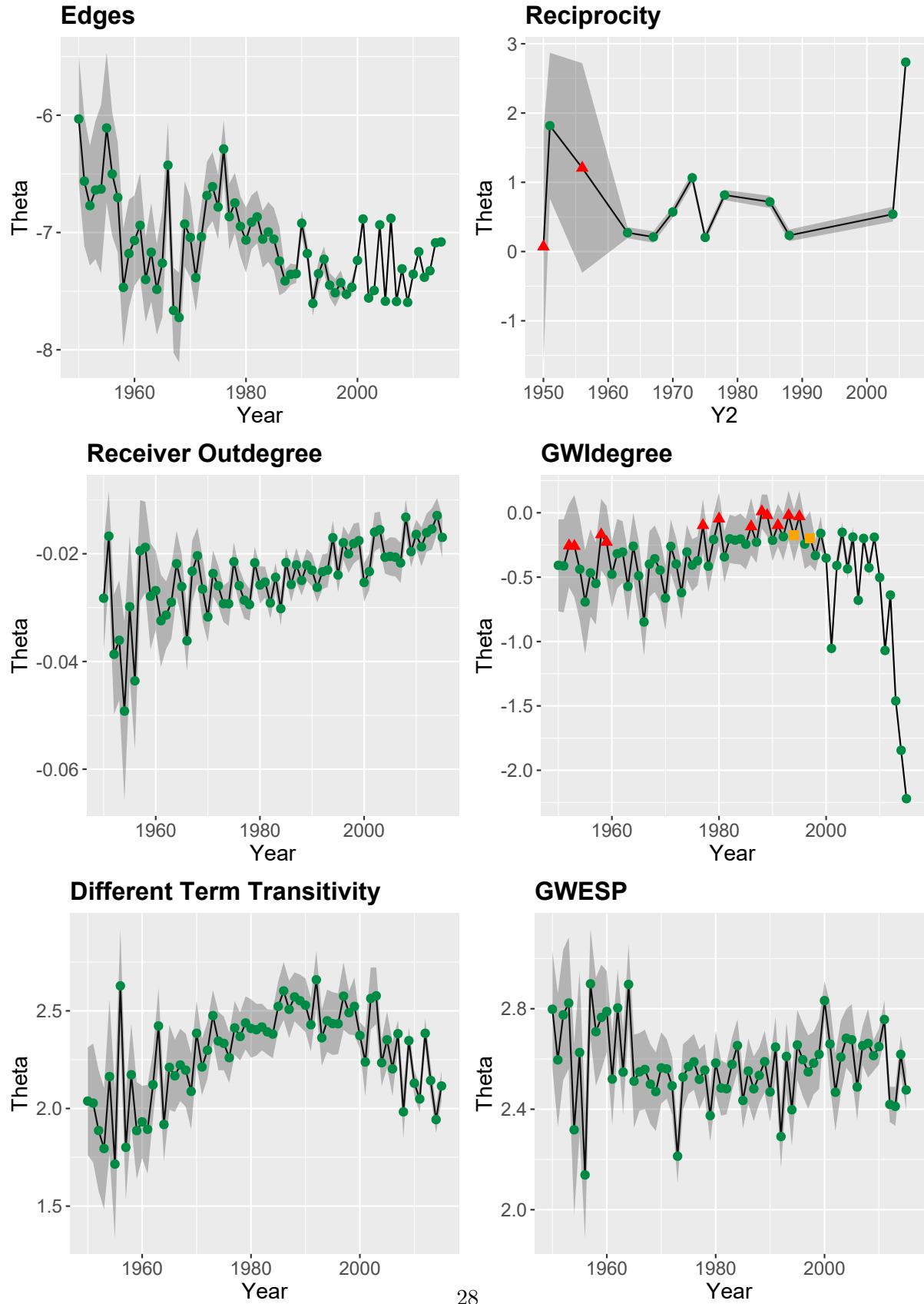
Figure 8: ERGM results for the dependence terms. Circles indicate a p-value smaller than 0.05, squares a p-value between 0.05 and 0.1 and triangles a p-value greater than 0.1.
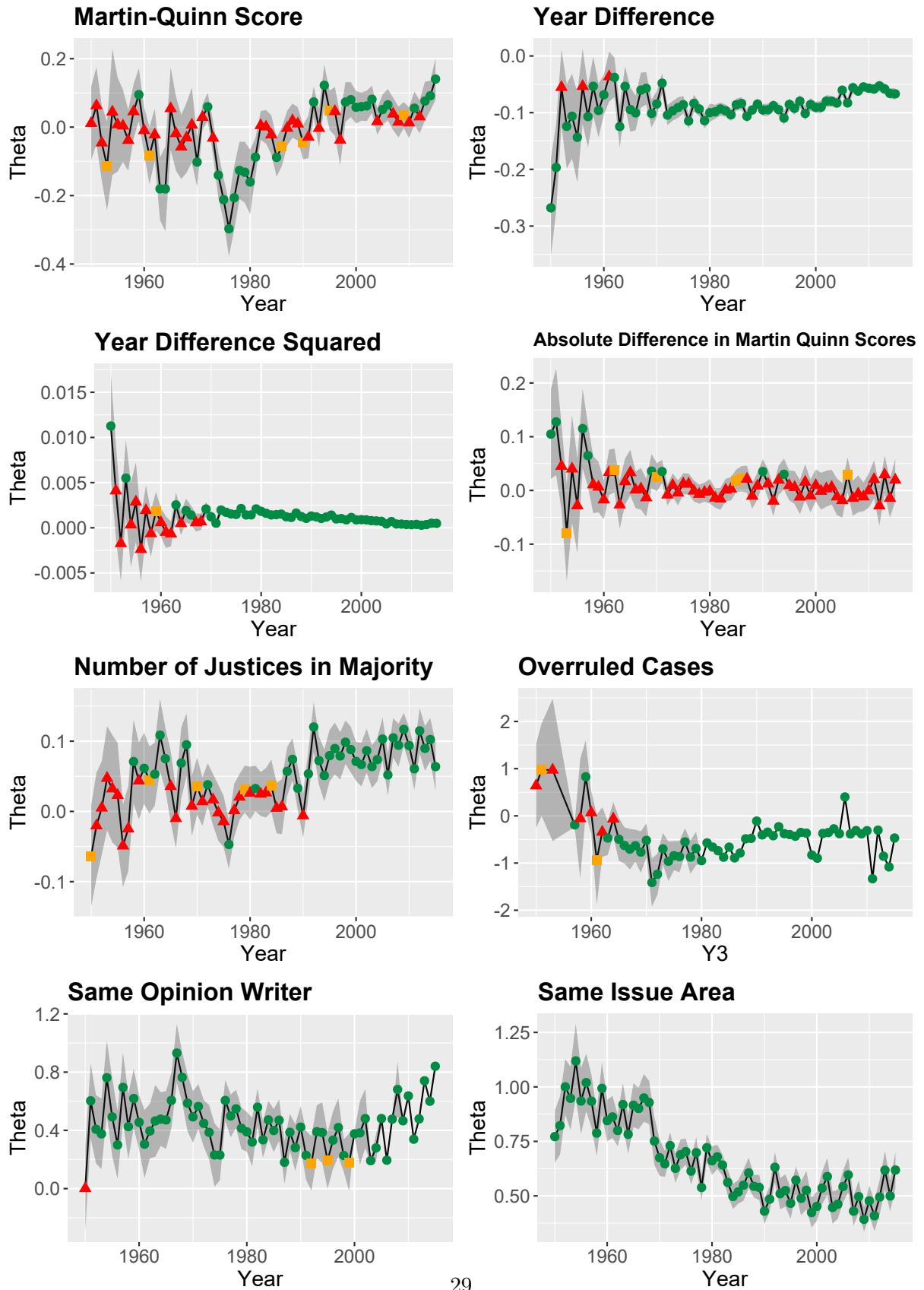
Figure 9: ERGM results for the covariate terms. Circles indicate a p-value smaller than 0.05, squares a p-value between 0.05 and 0.1 and triangles a p-value greater than 0.1.

5 Conclusion

We present a methodology for studying citations between US Supreme Court opinions at the dyadic level, as a network. This methodology—the citation-ERGM—enables researchers to include both exogenous covariates such as the ideological predisposition and age of a case, and dependence terms, such as transitivity and reciprocity, as explanations for citation formation. We apply this methodology to a network that includes all Supreme Court cases decided between 1937 and 2015. We find, somewhat counterintuitively, that Supreme Court citations are highly reciprocal. We also find that citations are driven by dependencies such as triad closure, popularity, and activity. The dependence effects that we identify are as substantively and statistically significant as the effects of the exogenous covariates we include in the model. The summary result from this analysis is that theoretical models of Supreme Court citation formation should consider both the effects of case characteristics and the structure of past citations.

We recognize several limitations of the current analysis, which should be addressed in future iterations of this work or in future studies. First, our theoretical claims and analyses treat all citations—positive and negative—as equivalent. Both data and methodological limitations will need to be overcome to consider these two citation types. Second, there is nothing in the c-ERGM that governs the generation of new cases (i.e., cases are assumed to arise exogenously, unrelated to the history of the citation network). As new cases are often taken up by the Supreme Court in order to test, clarify, and/or challenge existing precedents, it is likely inappropriate to assume that new cases arise exogenous to the existing citation network. Overcoming this limitation would require methodological innovation in the c-ERGM as well as theoretical development regarding exactly how citation structure might predict case emergence.

6 Appendix

6.1 Goodness-of-Fit

We evaluate the goodness-of-fit of the model following Hunter, Goodreau and Handcock (2008) by examining the distribution of four hyper statistics, e.g., the out- and indegree distribution and the distribution for two different edgewise shared partners statistics. OTP stands for *outgoing two-paths* and refers to the number of cases r that are cited by case i and that cite case j, while j is also directly cited by i. The second ESP statistic is the OSP specification that has been introduced in section 4.1.2. Figure 10 visualizes the goodness-of-fit results for the citation network for the 1950 (top) and 2015 (bottom) term. The solid black line indicates the statistic's distribution in the Supreme Court citation network of that given term and the boxplots depict the statistic's distribution of 1000 networks that have been simulated from the ERGM defined by the MCMLE. This means that in the ideal case the solid black line passes through ever single boxplot.

We see that our models do a good job capturing the out and indegree distribution of the citation network, since the black line falls almost exclusively within the ranges spanned by the boxplots. For the ESP distributions we can observe that the number of ties with $r = 0$ shared partners is captured well for both the OTP as well as for the OSP statistic. However, the model overestimates the number of $r = 1$ shared partners and then, especially in the 2015 term network, underestimates the number of ties with more than $r = 1$ shared partners.
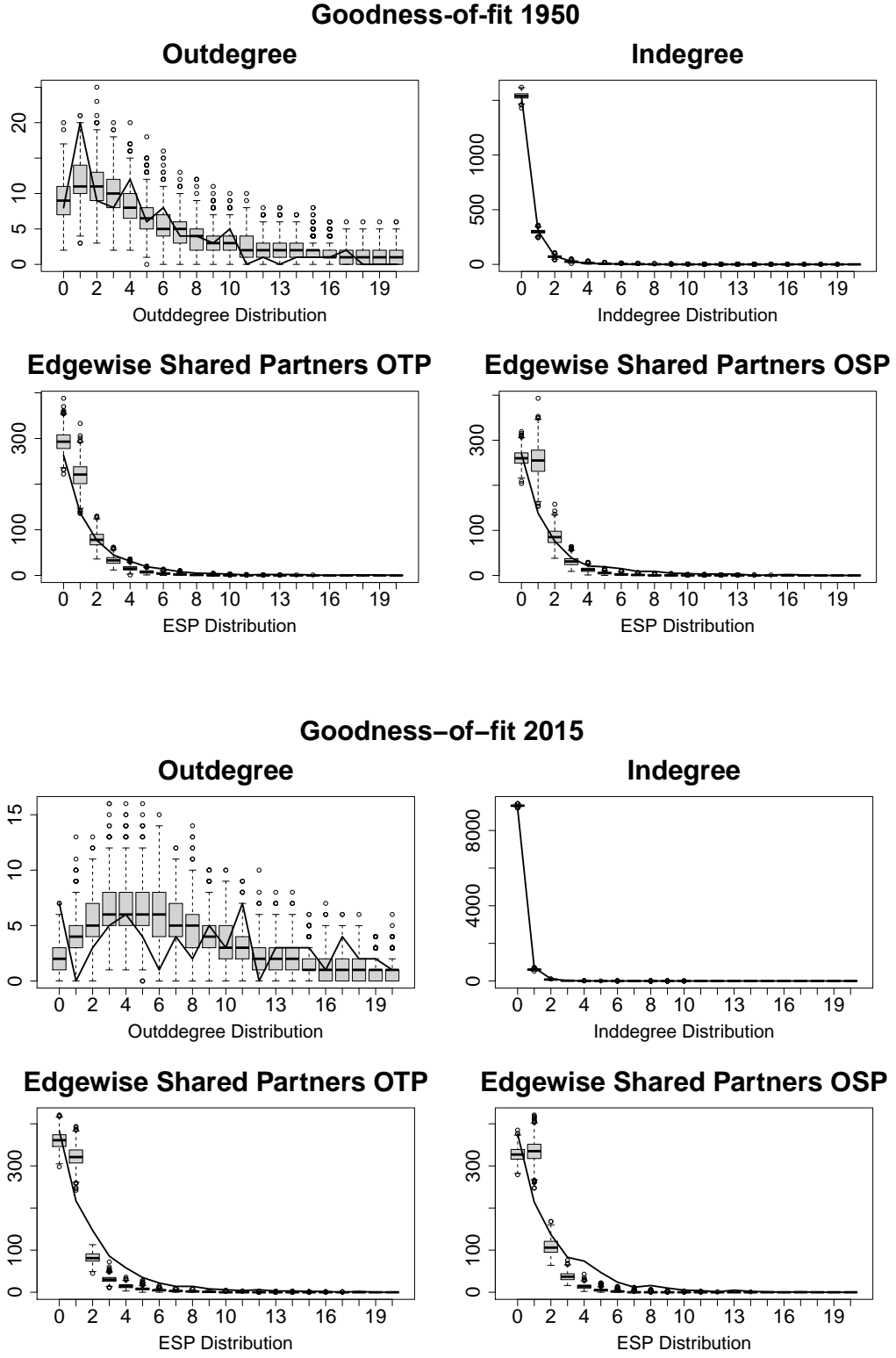
Figure 10: Goodness-of-fit diagnostic for the 1950 network (top) and the 2015 network (bottom).

6.2 Checking for Model Degeneracy

A common challenge when fitting ERGMs is model degeneracy. Model degeneracy occurs when the probability distribution defined by the parameter vector does not predominantly yield networks with similar statistics as the observed network. Generally, model degeneracy results in simulated networks with no ties or all possible ties. In a non-degenerate model the statistics of the networks that were simulated from the probability distribution defined by the MCMLE fall in the proximity of the observed network's statistics. Figures 11 and 11 depict trace and density plots for the dependence terms in the 1950 and 2015 term citation network. The histograms on the left visualize a statistic's density from 1000 simulated networks, while the right side shows the statistic's trace plot of the same 1000 networks. The solid black line indicates the statistic's value in the actual citation network. Both figures indicate that this model is non-degenerate and that the simulated network's statistics fall almost evenly around the observed statistic. The density and trace plots for the ERGM of the terms not depicted provide similar results.

Degeneracy Check 1950

Density of Edges

Edges

Density of Receiver Outdegree

Receiver Outdegree

Density of GWIDEGREE

GWIDEGREE

Density of Different Term Transitivity

Different Term Transitivity
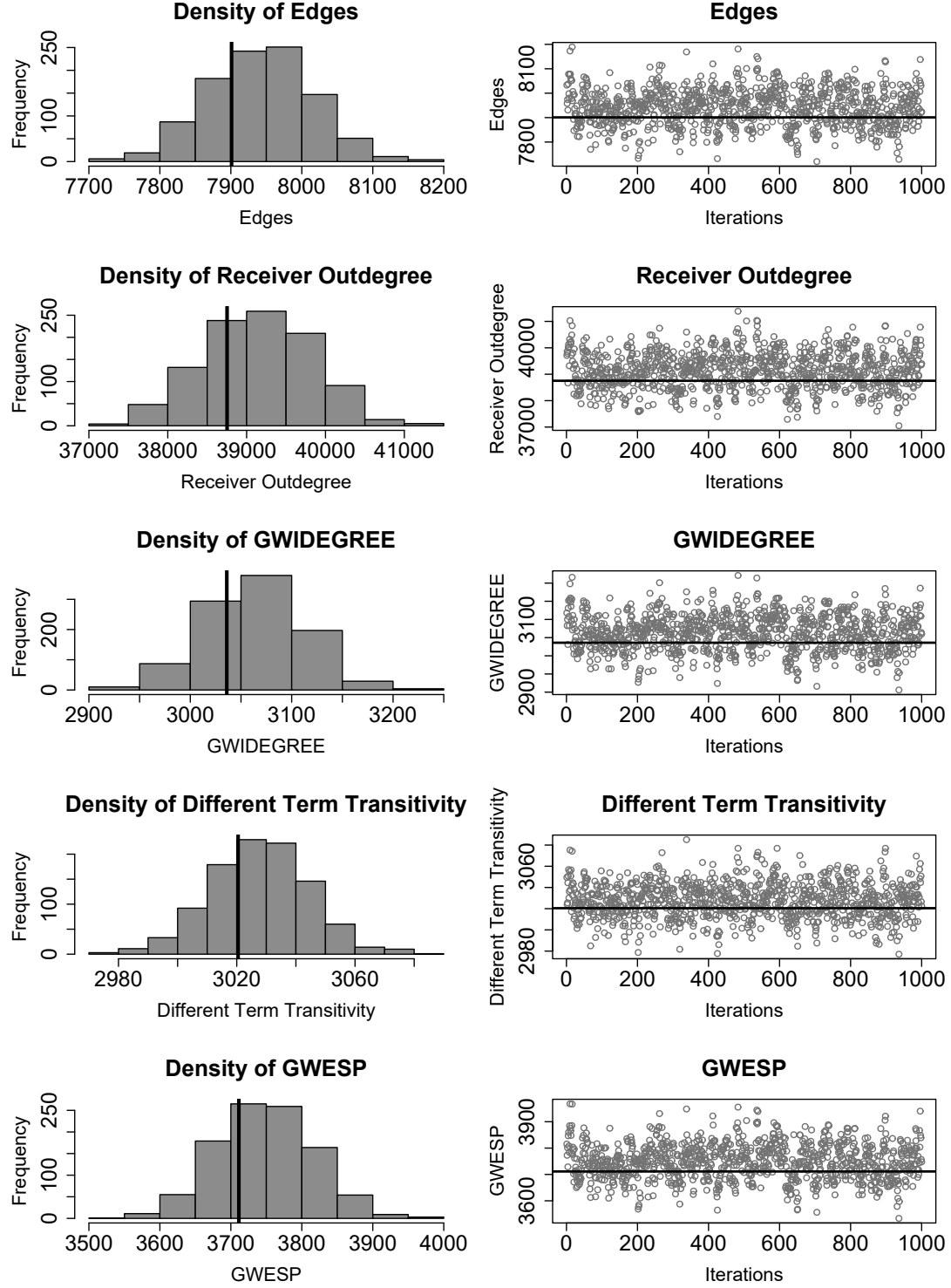
Density of GWESP

GWESP

Figure 11: Density and trace plots for the dependency terms of the 1950 term citation network.
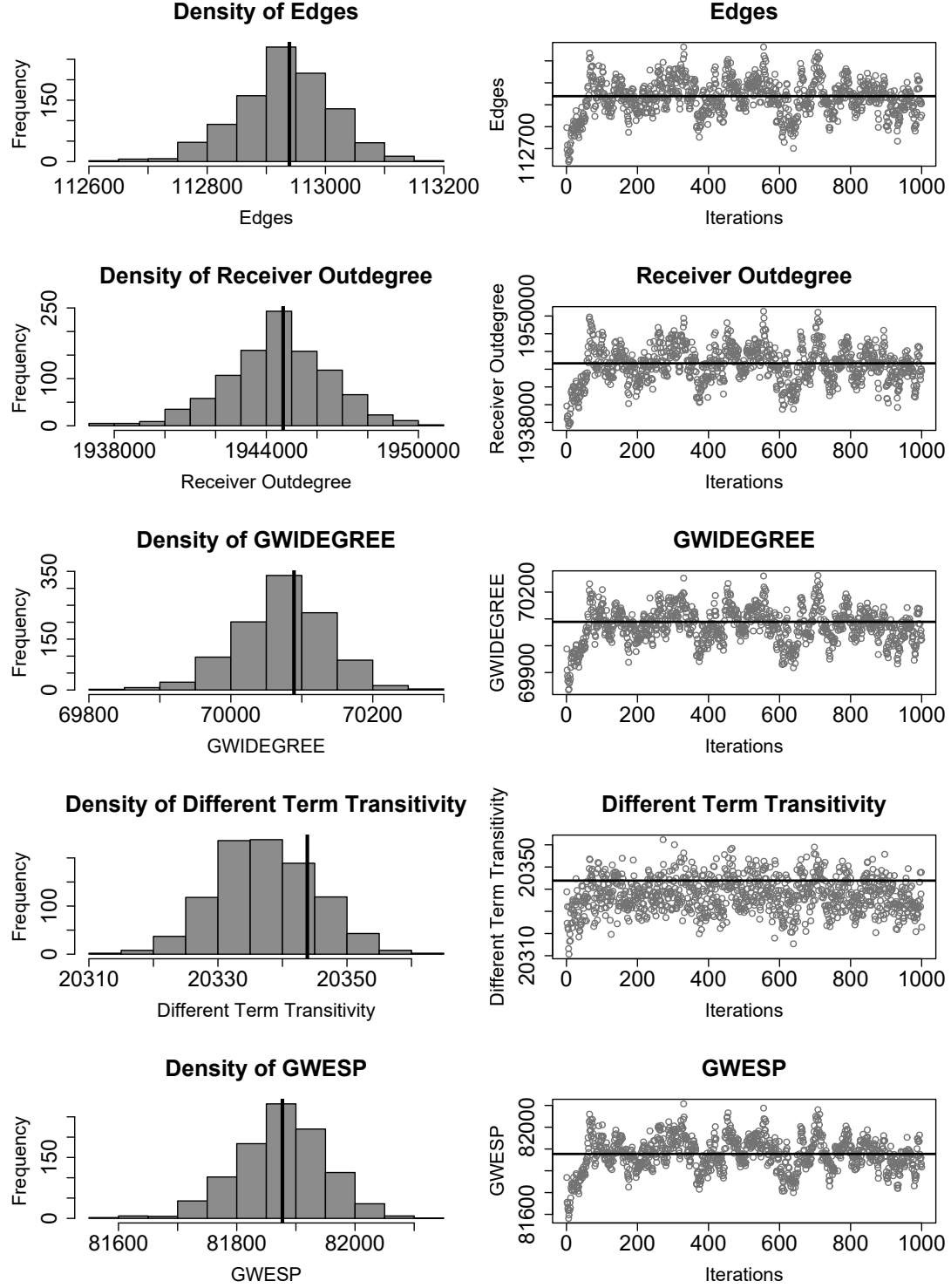
Degeneracy Check 2015

Figure 12: Density and trace plots for the dependency terms of the 2015 term citation network.

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.

Barabási, Albert-László and Réka Albert. 1999. "Emergence of scaling in random networks." *science* 286(5439):509–512.

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.

Butts, Carter T. 2008. "A Relational Event Framework for Social Action." *Sociological Methodology* 38(1).

Caldeira, Gregory A. 1988. "Legal precedent: Structures of communication between state supreme courts." *Social networks* 10(1):29–55.

Clark, Jennifer Hayes and Veronica Caro. 2013. "Multimember districts and the substantive representation of women: An analysis of legislative cosponsorship networks." *Politics & Gender* 9(1):1–30.

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

Cranmer, Skyler J and Bruce A Desmarais. 2016. "A critique of dyadic design." *International Studies Quarterly* 60(2):355–362.

Cranmer, Skyler J and Bruce A Desmarais. 2017. "What Can We Learn from Predictive Modeling?" *Political Analysis* pp. 1–22.

Cranmer, Skyler J, Bruce A Desmarais and Elizabeth J Menninga. 2012. "Complex dependencies in the alliance network." *Conflict Management and Peace Science* 29(3):279–313.

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. 2013. "Forecasting the locational dynamics of transnational terrorism: A network analytic approach." *Security Informatics* 2(1):8.

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.

Garlaschelli, Diego and Maria I Loffredo. 2004. "Patterns of link reciprocity in directed networks." *Physical Review Letters* 93(26):268701.

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.

Goodreau, Steven M, James A Kitts and Martina Morris. 2009. "Birds of a feather, or friend of a friend? Using exponential random graph models to investigate adolescent social networks." *Demography* 46(1):103–125.

Graif, Corina, Alina Lungeanu and Alyssa M Yetter. 2017. "Neighborhood isolation in Chicago: Violent crime effects on structural isolation and homophily in inter-neighborhood commuting networks." *Social Networks* .

Handcock, Mark S. 2003. *Assessing degeneracy in statistical models of social networks: https://www.csss.washington.edu/Papers/wp39.pdf.*

Hansford, Thomas G and James F Spriggs. 2006. *The politics of precedent on the US Supreme Court.* Princeton University Press.

Harris, Peter. 1982. "Structural change in the communication of precedent among state supreme courts, 1870–1970." *Social Networks* 4(3):201–212.

Holland, Paul W and Samuel Leinhardt. 1971. "Transitivity in Structural Models of Small Groups." *Small Group Research* 2(2):107–124.

Howard, Michael, Emily Cox Pahnke, Warren Boeker et al. 2016. "Understanding network formation in strategy research: Exponential random graph models." *Strategic management journal* 37(1):22–44.

Hummel, Ruth M, David R Hunter and Mark S Handcock. 2012. "Improving simulation-based algorithms for fitting ERGMs." *Journal of Computational and Graphical Statistics* 21(4):920–939.

Hunter, David R and Mark S Handcock. 2006. "Inference in curved exponential family models for networks." *Journal of Computational and Graphical Statistics* 15(3):565–583.

Hunter, David R., Steven M. Goodreau and Mark S. Handcock. 2008. "Goodness of Fit of Social Network Models." *Journal of the American Statistical Association* 103(481):248–258.

Jensen, David D and Paul R Cohen. 2000. "Multiple comparisons in induction algorithms." *Machine Learning* 38(3):309–338.

Kirkpatrick, S., C. D. Gelatt and M. P. Vecchi. 1983. "Optimization by Simulated Annealing." *Science* 220(4598):671–680.

Knight, Jack and Lee Epstein. 1996. "The norm of stare decisis." *American Journal of Political Science* pp. 1018–1035.

Levy, Michael. 2016. "gwdegree: Improving interpretation of geometrically-weighted degree estimates in exponential random graph models." *The Journal of Open Source Software* 1.

Lissner, Michael and Brian W. Carver. 2010. "CourtListener.com: A platform for researching and staying abreast of the latest law.".
URL: *http://www.courtlistener.com*

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.

Martin, Andrew D and Kevin M Quinn. 2002. "Dynamic ideal point estimation via Markov chain Monte Carlo for the US Supreme Court, 1953–1999." *Political Analysis* 10(2):134–153.

Masket, Seth and Boris Shor. 2015. "Polarization without parties: Term limits and legislative partisanship in Nebraska?s Unicameral Legislature." *State Politics & Policy Quarterly* 15(1):67–90.

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.

Schmid, Christian S and Bruce A Desmarais. 2017. "Exponential Random Graph Models with Big Networks: Maximum Pseudolikelihood Estimation and the Parametric Bootstrap." *IEEE International Conference on Big Data, http://ieeexplore.ieee.org/document/8257919/* pp. 116–121.

Schmid, Christian S. and David R. Hunter. 2020. "Some Results on MPLE." *Working Paper, The Pennsylvania State University, Department of Statistics* (1).

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins and Mark S. Handcock. 2006. "NEW SPECIFICATIONS FOR EXPONENTIAL RANDOM GRAPH MODELS." *Sociological Methodology* 36(1):99–153.

Spaeth, Harold, Lee Epstein, Ted Ruger, Keith Whittington, Jeffrey Segal and Andrew D Martin. 2014. "Supreme Court Database Code Book.".
URL: *http://scdb.wustl.edu*

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

Strauss, David and Michael Ikeda. 1990. "Pseudolikelihood estimation for social networks." *Journal of the American Statistical Association* 85(409):204–212.

Strogatz, Steven H. 2001. "Exploring complex networks." *nature* 410(6825):268.

US. Congress. Senate. 2016. *Constitution of the United State of America: Analysis, and Interpretation – Centennial Edition.* 112th Cong., 2d sess. S. Doc. 112-9. `https://www.govinfo.gov/app/details/GPO-CONAN-2017/`.

Ward, Michael D, Brian D Greenhill and Kristin M Bakke. 2010. "The perils of policy by p-value: Predicting civil conflicts." *Journal of Peace Research* 47(4):363–375.

Issue Area Activity
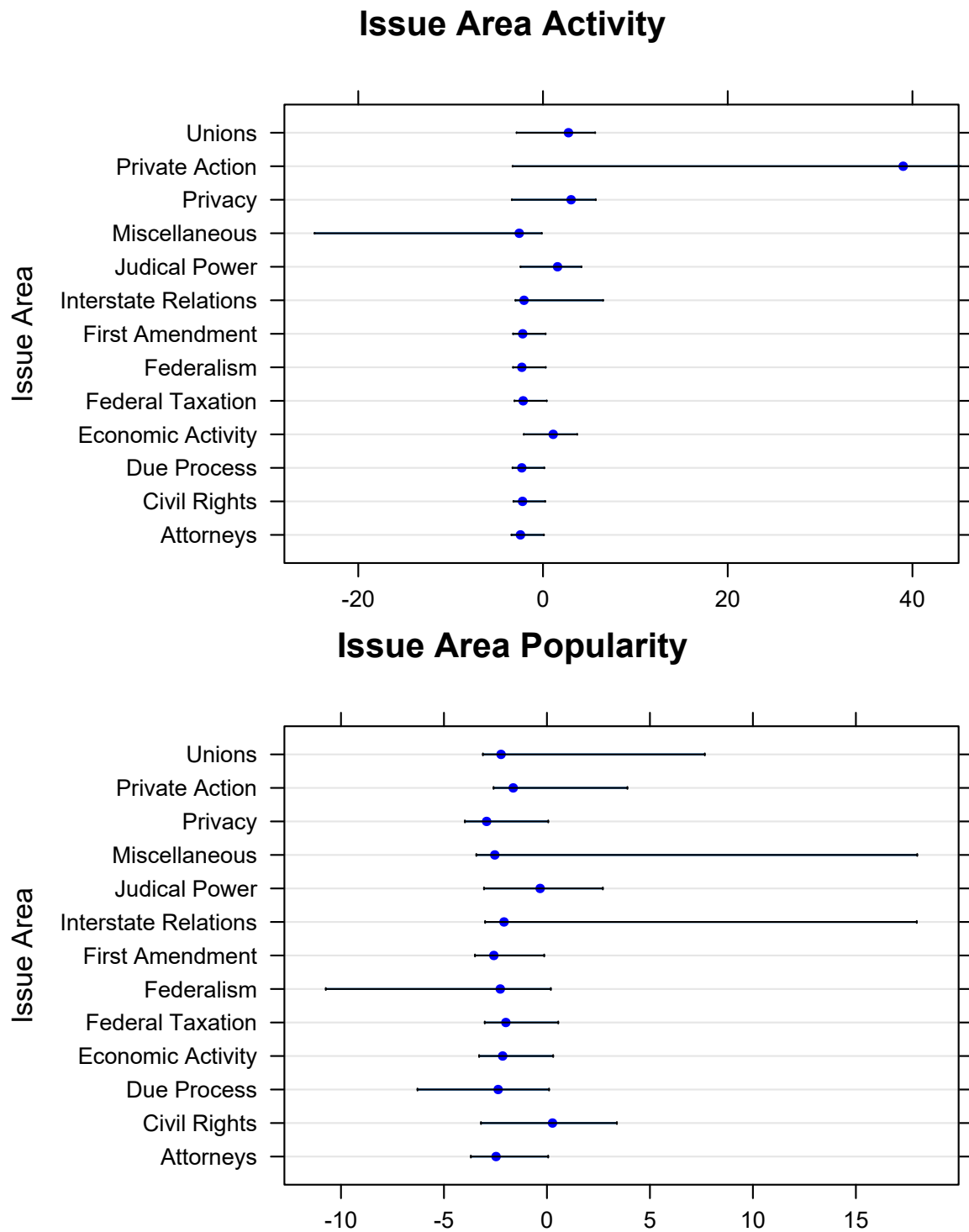

Issue Area Popularity

Figure 13: Issue Area Popularity and Activity (nodeifactor and nodeofactor). Baseline is Issue Area 'Criminal Procedure'.
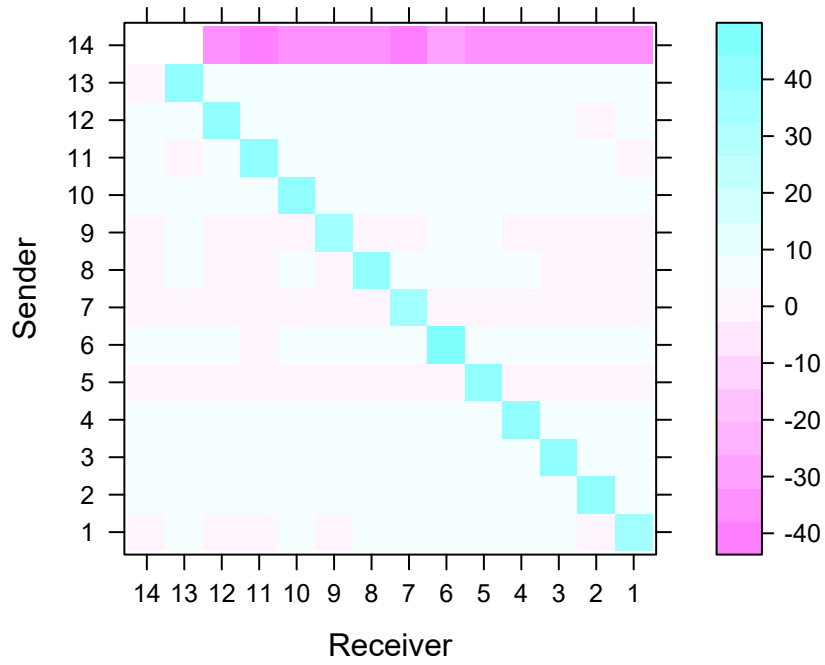
Figure 14: Nodemix Issue Area. Sender14Receiver14 and Sender14Receiver13 couldnt be estimated.