Generative Dynamics of Supreme Court Citations

Christian S. Schmid[*]   Ted Hsuan Yun Chen [†]   Bruce A. Desmarais [‡]
David R. Hunter [§]

May 2, 2018

Abstract

The significance and influence of US Supreme Court majority opinions derive in large part to 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 1937 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

[§]Department of Statistics, The Pennsylvania State University, dhunter@stata.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 US 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 1937 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 predictive performance.

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 fundamental element under study—the relationship between two units (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
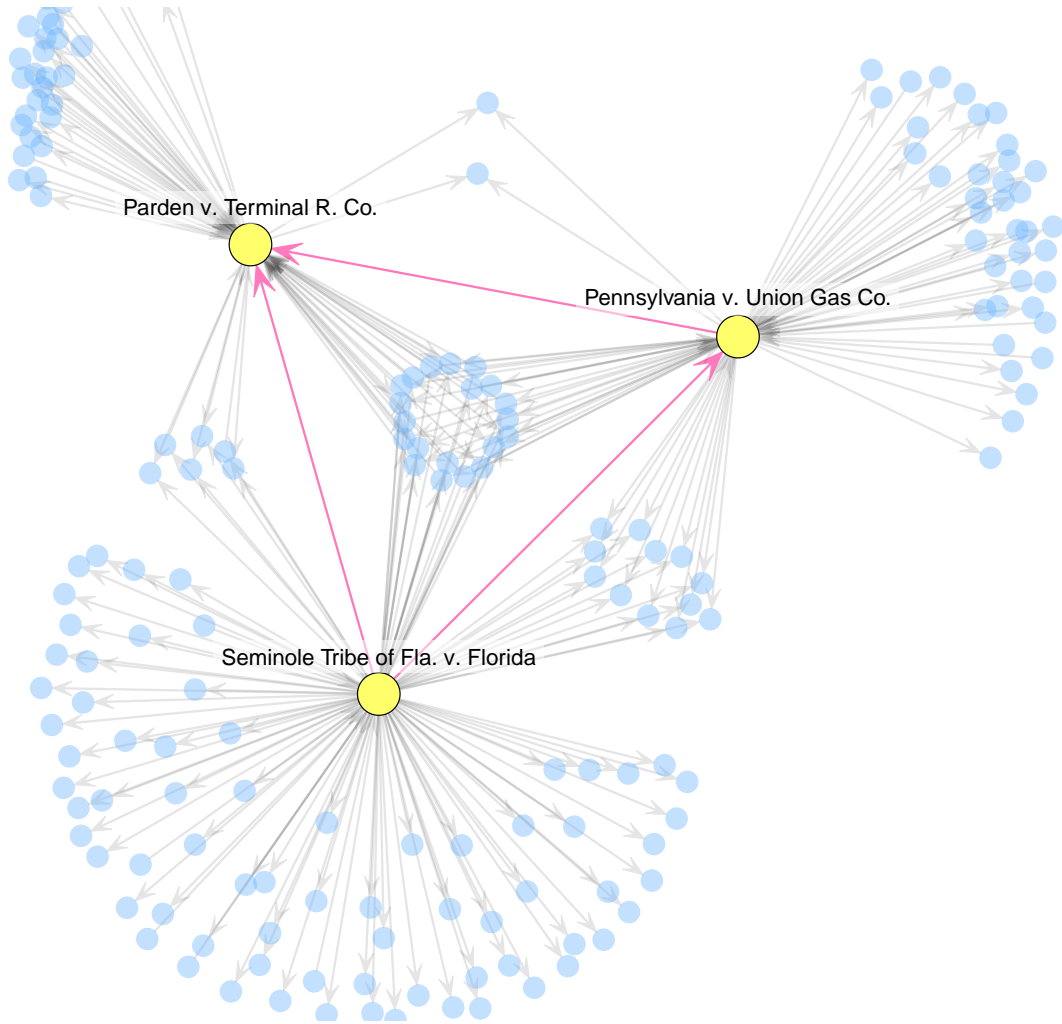
Figure 1: Illustration of transitive triangle connecting US Supreme Court opinions through citations.

their roles as precedents in the Union Gas opinion. We highlight one—Parden v. Terminal R. 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.
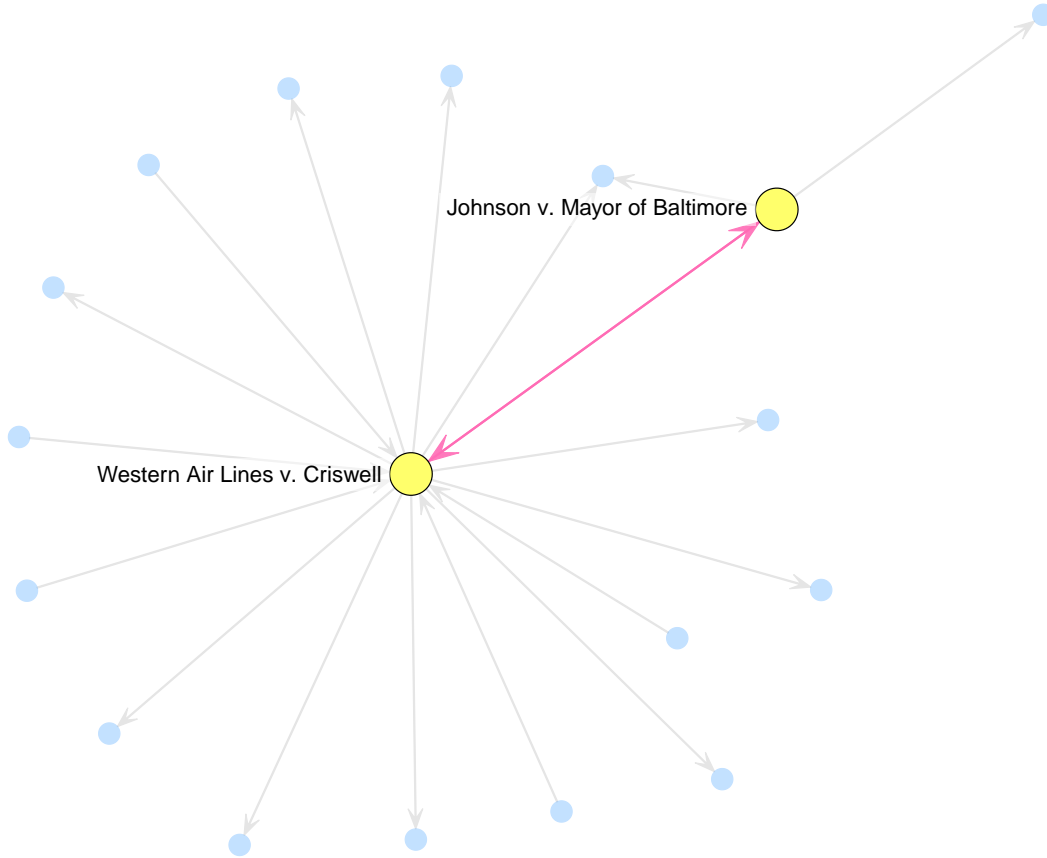

Figure 2: Illustration of a reciprocal tie between two US Supreme Court opinions.

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 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..."

The fourth network property we consider in the context of Supreme Court citations is activity. Activity, or sender activation, is the tendency for ties sent to beget more ties sent (Howard et al., 2016). Structurally, this means that there are few nodes that send a moderate number of ties—the number of ties sent is either small or fairly large. In terms of Supreme Court opinions, for each opinion discussed, that discussion is likely to raise other tangential issues on which the justices will want to draw upon past opinions. Furthermore, for each opinion that applies to and is cited by the current opinion, there is often a case/opinion 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. The example passage that illustrates activity is from Kush v. Rutledge (460 U.S. 719, 1983)—a case addressing witness intimidation and due process protections. In
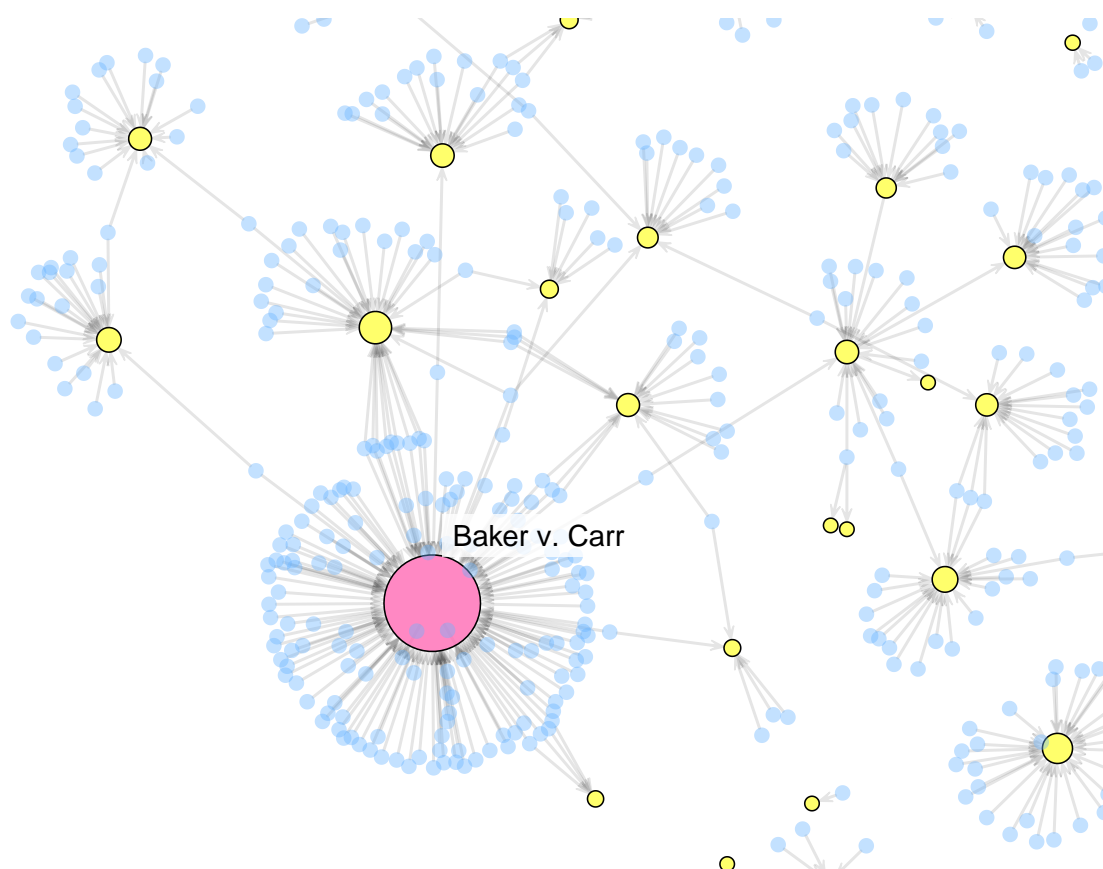
Figure 3: Illustration of ties sent to a landmark Supreme Court opinion via citations.

the Kush opinion, Griffin v. Beckenridge (i.e., Griffin) is discussed at length in terms of its lack of applicability. It is common that opinions distinguish among a variety of legal rules in terms of their applicability to a case, which means that for each rule that is established in an applicable precedent, one or more need to be discussed in terms of why they do not apply. We do not present a separate visualization related to activity, as it would look very similar to the visualization for popularity.

> 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"

3 The Citation Temporal 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 model covariate and dependence effects in longitudinal network data—the temporal exponential random graph model (TERGM) (e.g., Cranmer, Desmarais and Menninga, 2012; Clark and Caro, 2013; Masket and Shor, 2015; Graif, Lungeanu and Yetter, 2017). The TERGM, as it is currently implemented (Leifeld, Cranmer and Desmarais, 2017; Krivitsky and Handcock, 2017), is designed to model longitudinal network data in which ties among a relatively stable set of nodes

are modeled at multiple time points (e.g., the countries that are and are not defensively allied at time t, the legislators who have and have not cosponsored each others' legislation in the current session). The longitudinal network structures for which TERGMs are currently designed are insufficient to account for the structure of citation networks.

3.1 c-TERGM definition

We develop the citation TERGM (c-TERGM), 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 longitudinal networks for which TERGMs 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 TERGMs, 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-TERGM specification.

The likelihood function of the c-TERGM is given by

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

where $\boldsymbol{\theta}$ is a vector of real-valued parameters, and $\boldsymbol{h}(C_t, C_{<t})$ is a vector of scalar-valued functions
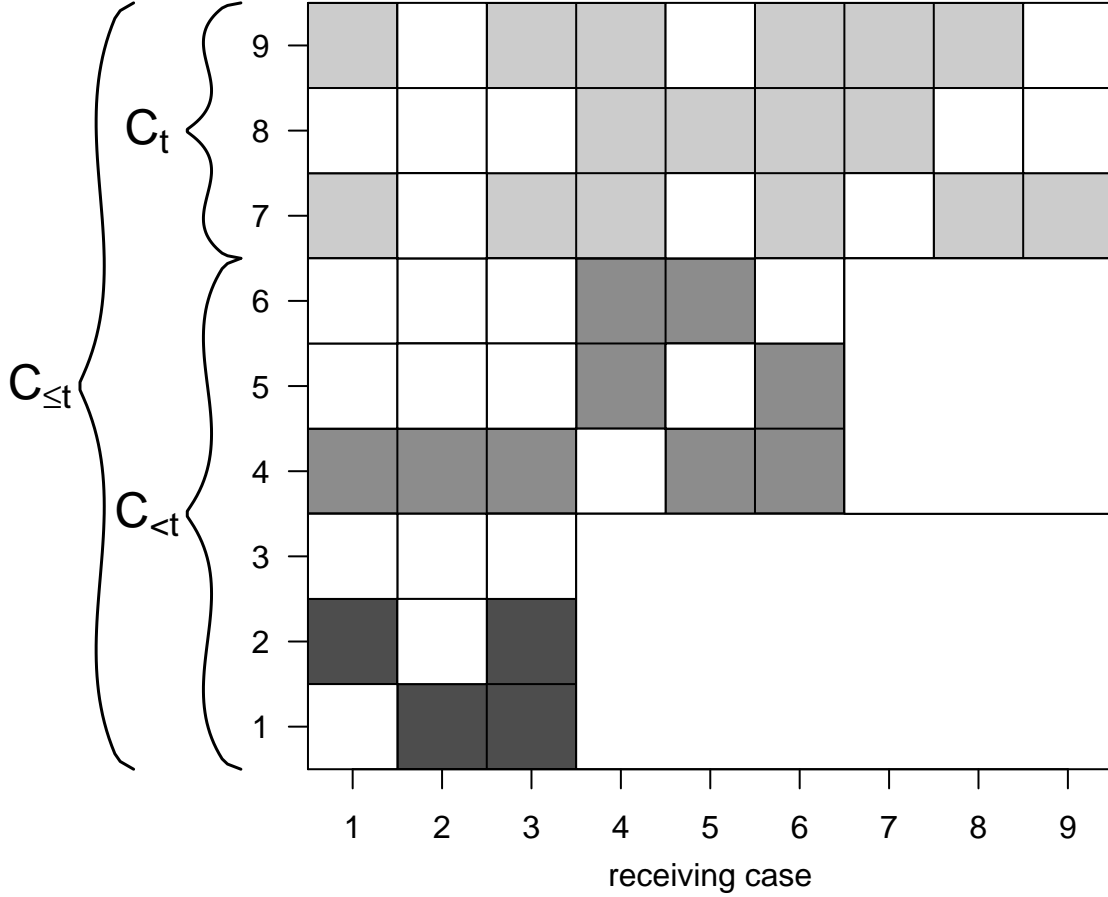
Figure 4: Illustration of temporal structure of the Supreme Court Citation Network. $C_{\leq t}$ is the entire set of citations (and non-citations) on which citations and non-citations at time t (i.e., C_t) depend. C_t are conditioned on the citations and non-citations established before time t (i.e., $C_{<t}$). 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.

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. The probability of observing the citation network up to time T—the final time point in the dataset—is given by the product over the probability of adding the citations at each time point t, conditional on the citations that were added to the network previously.

Though the likelihood function of the c-TERGM 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 element of $\boldsymbol{h}$ that counts the number of mutual edges in the network is 0.5, then the log odds of observing $C_{ij,t} = 1$ increase 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 not unique to c-TERGM. It is well known for the ERGM family (Goodreau, Kitts and Morris, 2009).

We should note that there is nothing about the c-TERGM (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-TERGM can be specified to account for the dependence in the data.

3.2 c-TERGM 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-TERGM. Methods of Monte Carlo approximation are well established in the ERGM literature, (Hunter and Handcock, 2006; Van Duijn, Gile and Handcock, 2009; Hummel, Hunter and Handcock, 2012), but with a number of nodes on the order of $10,000$, even the methods of Monte Carlo MLE are computationally expensive (Schmid and Desmarais, 2017). However, because the network is large, we can rely on a fast estimation method—maximum pseudo-likelihood estimation (MPLE), which is consistent for discrete exponential family models (Nguyen, 2017), and has long been used with ERGM (Strauss and Ikeda, 1990). The MPLE

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.

One challenge that arises with the use of MPLE is that the usual estimate of the asymptotic covariance matrix (i.e., inverting the negated Hessian of the objective function at the maximum), results in standard errors that are biased downward. This leads to a high Type-1 error rate in hypothesis testing with MPLE if conventional standard errors are used (Van Duijn, Gile and Handcock, 2009). A form of non-parametric cluster bootstrap results in confidence intervals with correct coverage probability with MPLE (Desmarais and Cranmer, 2012, 2010). In the case of c-TERGM, the clusters are defined by t—the time points in which the cases are decided, and T time-specific citation matrices C_t are re-sampled with replacement to form a non-parametric cluster bootstrap sample. Confidence intervals are defined by the quantiles in the empirical distributions of estimates that result from re-estimating the MPLE on non-parametric cluster bootstrap samples.

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

[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 both popularity and activity dynamics (Strogatz, 2001). Figure 6 displays the complete network. 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 by defining the terms within these two classes.

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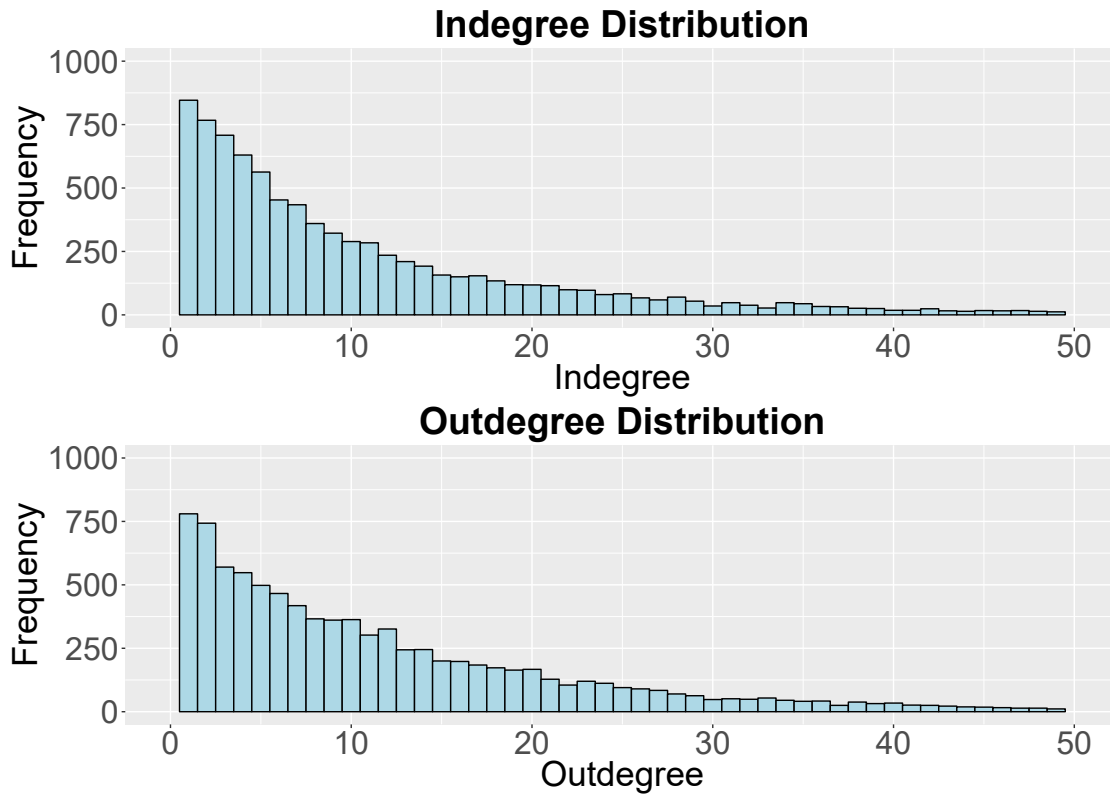
Indegree Distribution

Outdegree Distribution

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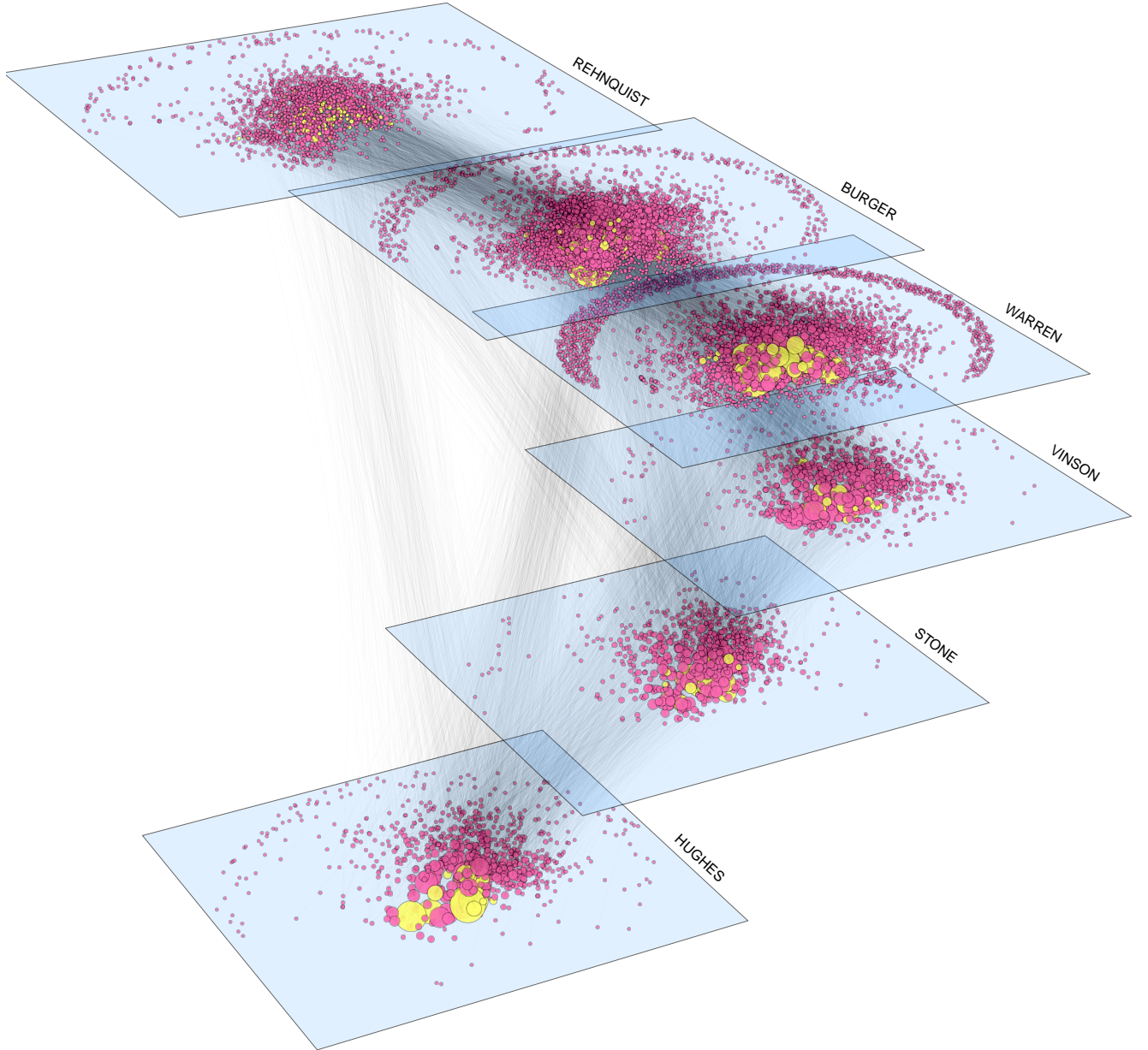
Figure 6: Supreme Court Citation Network, 1937-2001. Nodes are Supreme Court cases, with size based on incoming citations. Salient cases (*Oxford Guide*) are in yellow. Each network layer contains cases decided under a different chief justice. Because our data is temporally bound, cases for the Hughes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

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 three sets of dummy variables that account for the issue areas of cases. In the first set of dummy variables—sender intercepts—the variable X^{ij} indicates the issue area of the sending case (i) (e.g., for the Criminal Procedure indicator, $X_{ij} = 1$ if case i is a criminal procedure case, and 0 otherwise). The second set of dummy variables represent receiver issue area intercepts in which X^{ij} indicates the issue of the receiver case. The third set of dummy

variables account for the rate of citation within each issue area and between each combination of issue areas. There is a separate dummy variable for each directed pair of issue areas. For example the A→B dummy variable $X_{ij} = 1$ if case i is issue area A and case j is issue area B, and 0 otherwise. Note that, given the high dimension of these dummy variables, we omit one each of the sender and recipient indicators, and 28 of the 196 mixing rate indicators. Unfortunately, this leaves these indicators largely uninterpretable, but these variables are nonetheless important to absorb variation in citations that is attributable to the legal domains of two cases in a dyad. 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 a set of covariates that account for justice-specific variation in citations. We include two types of terms at the justice level. First, we include an indicator variable that models the rate at which justices cite themselves. 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. Second, we use two sets of dummy variables to account for differences across justices both in terms of their tendencies to include citations in the opinions they write and to be cited by other opinions. The justice sender dummy variables indicate the identity of the justice who authored the majority or plurality opinion of case i. The justice recipient dummy variables indicate the identity of the justice who authored the majority or plurality opinion of case j. This set of variables is motivated by the use of citations by Kosma (1998) to measure the influence of individual Supreme Court justices.

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.

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.

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

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 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).

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) (see, e.g., Makhoul et al., 1999, for discussion of these measures). All three measures are bounded between 0 and 1, with higher scores indicating better performance.

	Independent Model		Full Model	
	mean	range	mean	range
precision	0.6007	(0.5934, 0.6074)	0.8666	(0.8647, 0.8701)
recall	0.1522	(0.1495, 0.1549)	0.5969	(0.5945, 0.6019)
F1 score	0.2429	(0.2389, 0.2462)	0.7069	(0.7048, 0.7107)

Table 3: The predictive performance of the directed dyadic models with linear trends, 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.

	Independent Model		Full Model	
	mean	range	mean	range
precision	0.5938	(0.5844, 0.5996)	0.8668	(0.8653, 0.8693)
recall	0.1551	(0.1515, 0.1575)	0.5969	(0.5937, 0.6005)
F1 score	0.246	(0.2406, 0.249)	0.707	(0.7045, 0.7095)

Table 4: The predictive performance of the directed dyadic models with no trends, over ten 80/20 train/test splits.

	Independent Model		Full Model	
	mean	range	mean	range
precision	0.6209	(0.611, 0.6351)	0.8669	(0.8612, 0.8711)
recall	0.1511	(0.1451, 0.1558)	0.5962	(0.5884, 0.6011)
F1 score	0.243	(0.2347, 0.2486)	0.7065	(0.7011, 0.7103)

Table 5: The predictive performance of the directed dyadic models with quadratic trends, over ten 80/20 train/test splits.

We also conduct a predictive experiment for the model that does not consider any trends and for the model that does consider quadratic instead of linear trends. The predictive performance results are given in tables 4 and 5. By comparing the full model results in tables 4 and 5 with the results of the full linear trend model in table 3 we can conclude all three models perform similarly well.

Comment from Benjamin Kassow: In section 6.2 (predictive performance), I like what is here, but was hoping to see a bit more discussion about the predictive performance of the model, with some explanation of what the predictive performance means. I understand it (I think), but depending on where this gets sent to, I am not sure if all reviewers will. I also think tying this into the overall model explanation or discussion of model performance overall will be helpful for better integrating this section into the rest of the manuscript.

4.3 c-TERGM Results

The results from the c-TERGM are presented in Table 6 and Figures 7 and 8. Table 6 gives the coefficient estimates and 95% bootstrap confidence intervals. In Figures, 7 and 8, we depict coefficient estimates and confidence intervals for all of the effects for which we estimated trends (excluding the intercept—edge count—which is typically not interpreted substantively). 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. We first discuss the dependence effect estimates (panels (a)–(d) in Figure 7). The reciprocity effect is substantial, positive, and statistically significant for most of the period under study. Conditional upon cases i and j being considered within the same term, the log of the odds that opinion i cites opinion j increases by approximately 2 when opinion j cites opinion i. The naive probability that any one case cites another case decided at the same time or previously in the Supreme Court Citation network is approximately 0.0019. This probability increases eight-fold, to approximately 0.015 if the citation would be reciprocal. The transitivity (i.e., triangle closure) effect is also substantial, positive, and statistically significant over the entire period studied. The probability that case i cites case j increases approximately five-fold if the citation would close a triangle involving cases i and j. The activity and popularity effects are both estimated to be positive, but trend downward and are more modest in scale (and in the case of activity, not statistically significant later in the time series). Interpreting the popularity effect—the log odds that i cites j increases by 0.018–0.035 with every additional citation to case j. Based on the in-degree distribution, it appears that the popularity of cases can be classified by multiples of 10. If case j increases from ten to twenty citations from other cases, the probability that case i cites case j increases from approximately 0.0030 to 0.0040. Overall, these results represent evidence of complex dependence in Supreme Court citation formation.

The effects of the exogenous covariates included in the c-TERGM (panels (e)–(k) in Figure 8) exhibit some surprising dynamics. The effects of absolute difference in Martin-Quinn score, and majority size of the recipient case all flip signs throughout the time period under study. This

may be an artifact of the simple linear trends we have estimated, but is interesting nonetheless. The attitudinal model holds that ideological preferences drive decision-making on the US Supreme Court. Ideological selection of precedents would be evidenced by a negative effect of the absolute difference in MQ scores in the c-TERGM. We find this effect is negative and statistically significant only after 1975. The absolute difference in MQ scores variable is on the scale of approximately 0—3. At its most ideological point (i.e. 2001), the probability that case i cites case j decreases from approximately 0.024 to 0.018—by one third—if the ideological distance between the majority coalitions in the two cases increases by 3. The effects of the majority size and ideological breadth of case j are similar in both scale and trend. It is not until later in the time series that they both exhibit their expected positive effects. The effect of same issue is positive and statistically significant, and substantial for the time period under study. The probability that i cites j increases by approximately three-fold over the naive probability of citation if i and j have the same legal issue area.

The results from the c-TERGM provide evidence that citations between US Supreme Court cases are driven both by exogenous case attributes and complex dependence processes. The scales of covariate and dependence effects are comparable.
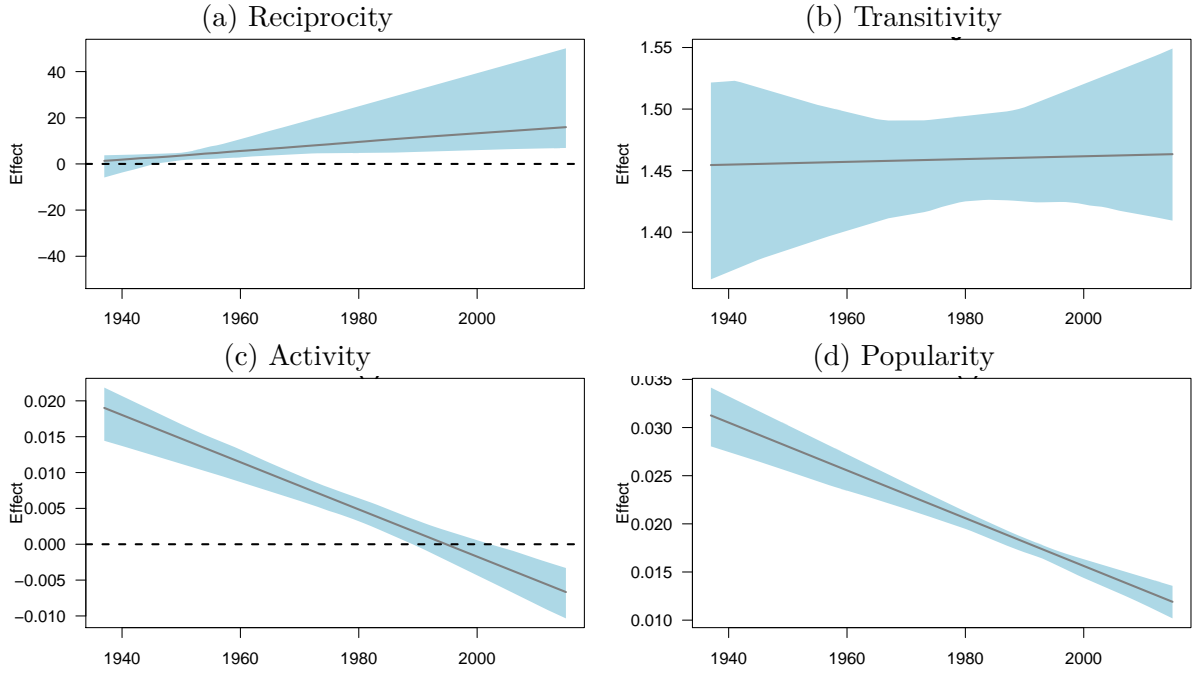
- adjust results/numbers in text

Figure 7: Trending effects for dependence effect estimates from the c-TERGM. The solid lines plot the point estimates, and the shaded areas span 95% confidence intervals.
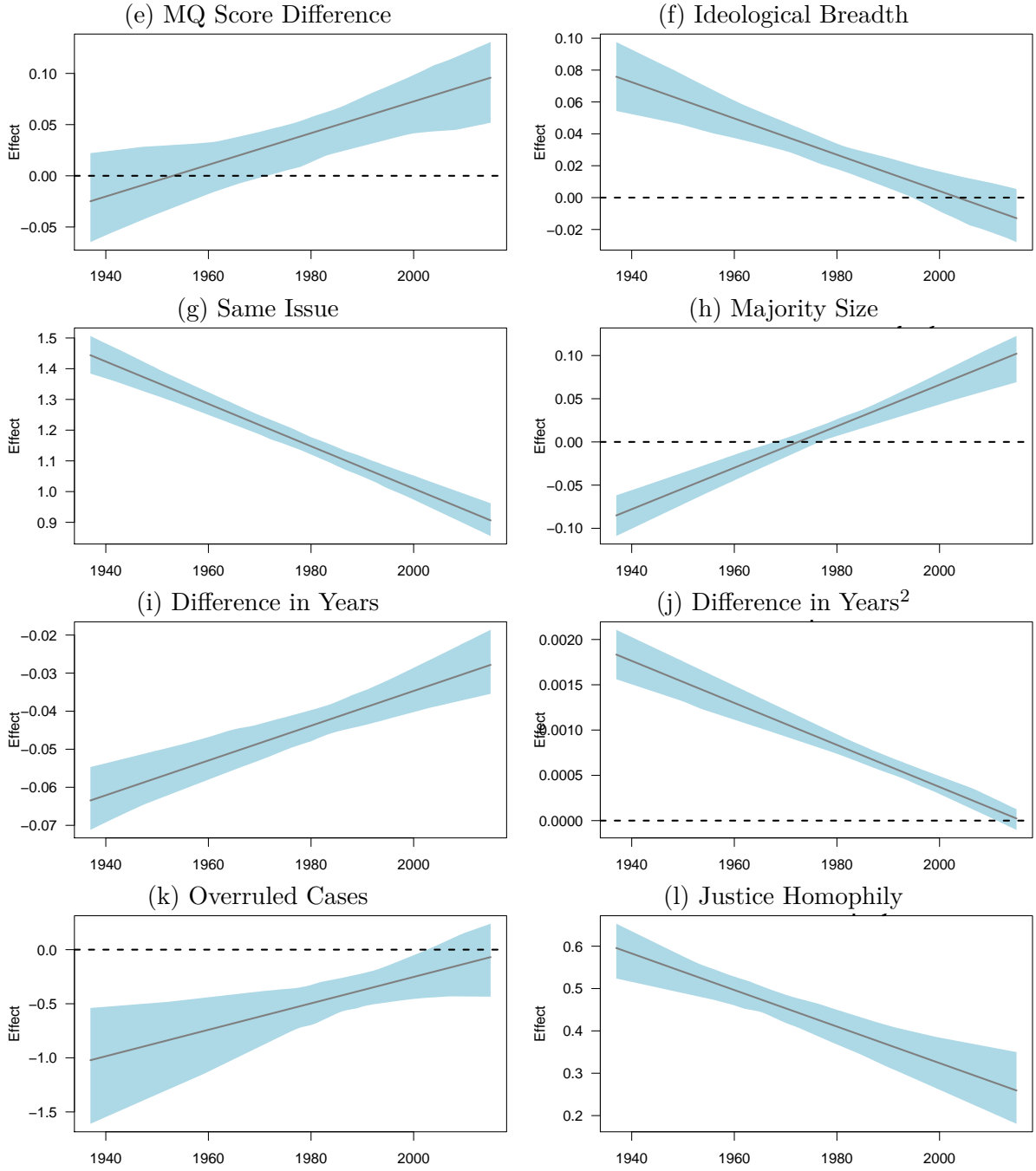
Figure 8: Trending effects for exogenous variables from the c-TERGM. The solid lines plot the point estimates, and the shaded areas span 95% confidence intervals.

- visualize Sender/receiver Issue Area Results

- heatplot for issue area nodemix

- visualize Sender/receiver for majority opinion writer

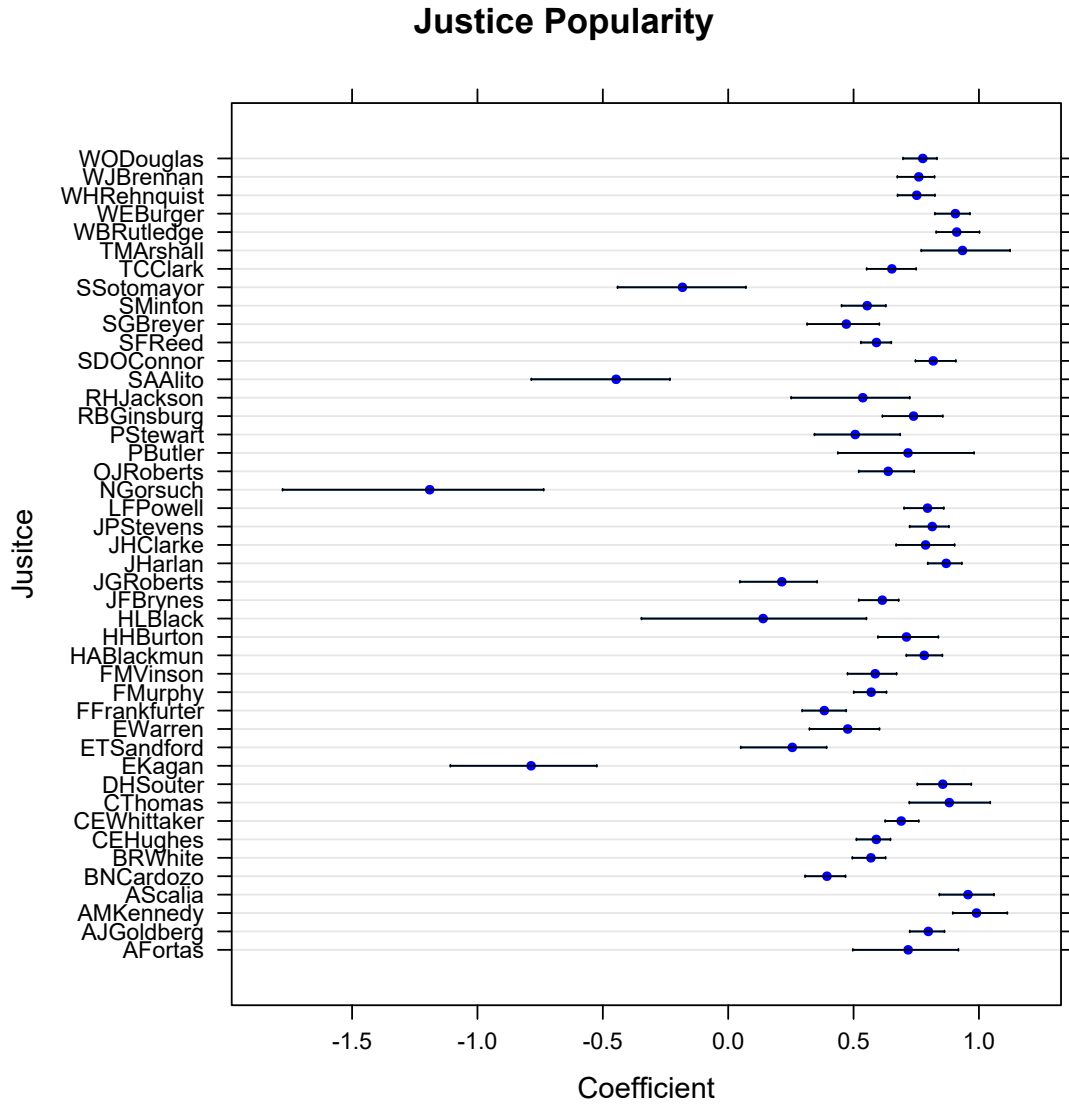- add Network effect plots for model without network statistics (add to existing plots?)

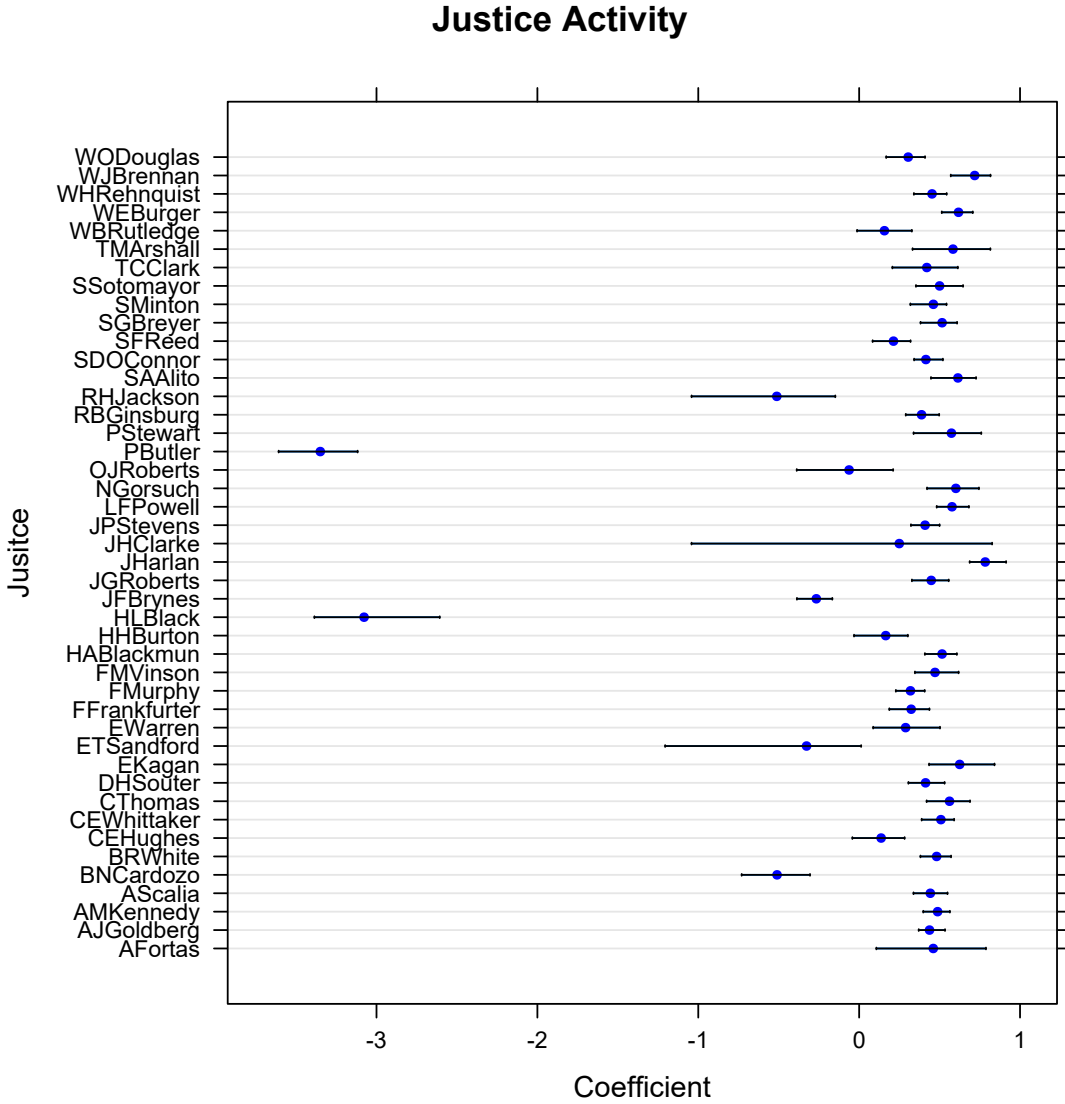Figure 9: Nodeifactor Majority Opinion Writer. Baseline= LDBrandeis

Justice Activity

Figure 10: Nodeofactor Majority Opinion Writer. Baseline= LDBrandeis

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-TERGM—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-TERGM 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-TERGM as well as theoretical development regarding exactly how citation structure might predict case emergence.

6 Appendix

The following table presents the Bootstrapped MPLE Results for the time period $1937 - 2015$.

- add robustness check: MCMLE for each year -> plot time trend of coefs -> are results similar to the MPLE time trends presented in the paper?

- (investigate which statistic contributes most on recall improvement)

	Estimate	Lower Bound	Upper Bound	Significance
Edges	-6.249	-6.412	-6.110	*
Instar(2)	0.035	0.032	0.038	*
Outstar(2)	0.021	0.019	0.028	*
Mutual	2.071	-4.157	4.072	
Triangle	1.507	1.431	1.559	*
Martin Quinn Score	0.079	0.029	0.130	*
Same Issue Area Homophily	1.447	1.389	1.515	*
Year Difference	-0.068	-0.079	-0.058	*
(Year Differnce)^2	0.0019	0.0016	0.0022	*
Receiver Abs Diff MQ Score in Majority	0.047	0.029	0.063	*
Receiver Number Justices in Majority	-0.075	-0.095	-0.055	*
Receiver Sender Year	0.0039	0.0036	0.0042	*
Overruled Cases	-1.043	-1.624	-0.564	*
Sender Same Issue Area 2	0.142	0.100	0.185	*
Sender Same Issue Area 3	-0.337	-0.403	-0.270	*
Sender Same Issue Area 4	0.498	0.438	0.561	*
Sender Same Issue Area 5	0.182	0.031	0.305	*
Sender Same Issue Area 6	0.536	0.422	0.648	*
Sender Same Issue Area 7	0.370	0.316	0.426	*
Sender Same Issue Area 8	0.107	0.059	0.156	*
Sender Same Issue Area 9	0.278	0.238	0.329	*
Sender Same Issue Area 10	0.371	0.317	0.433	*
Sender Same Issue Area 11	-0.030	-0.151	0.164	
Sender Same Issue Area 12	0.064	-0.014	0.130	
Sender Same Issue Area 13	0.608	0.423	0.792	*
Sender Same Issue Area 14	0.232	0.075	0.393	*
Receiver Same Issue Area 2	0.252	0.199	0.310	*
Receiver Same Issue Area 3	-0.092	-0.194	0.014	
Receiver Same Issue Area 4	0.481	0.397	0.570	*
Receiver Same Issue Area 5	0.374	0.224	0.519	*
Receiver Same Issue Area 6	0.521	0.346	0.652	*
Receiver Same Issue Area 7	0.412	0.339	0.487	*
Receiver Same Issue Area 8	0.174	0.134	0.232	*
Receiver Same Issue Area 9	0.311	0.240	0.361	*
Receiver Same Issue Area 10	0.506	0.429	0.577	*
Receiver Same Issue Area 11	0.299	0.088	0.432	*
Receiver Same Issue Area 12	0.350	0.276	0.445	*
Receiver Same Issue Area 13	0.972	0.780	1.135	*
Receiver Same Issue Area 14	0.517	0.200	0.740	*
Instar(2) × Sender Year	-0.00026	-0.00031	-0.00021	*
Outstar(2) × Sender Year	-0.00031	-0.00045	-0.00027	*
Mutual × Sender Year	0.110	0.010	0.524	*
Triangle × Sender Year	-0.00022	-0.0016	0.0013	
Martin Quinn Score × Sender Year	-0.0019	-0.0030	-0.0009	*
Same Issue Area × Sender Year	-0.0068	-0.0081	-0.0055	*
Year Difference × Sender Year	0.00046	0.00028	0.00065	*
Year Difference^2× Sender Year	-0.000024	-0.000028	-0.000021	*
MQ Score in Majority × Sender Year	-0.00053	-0.00093	-0.0018	*
Justices in Majority × Sender Year	0.0017	0.0013	0.0021	*
Overruled Cases × Sender Year	0.012	0.002	0.023	*

Table 6: Bootstrapped MPLE Results for the time period $1937 - 2015$. A '*' indicates that the 2.5th and 97.5th quantile of the variable does not include '0' and as a result is statistically significant.

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.

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. 2010. Consistent confidence intervals for maximum pseudolikelihood estimators. In *Proceedings of the Neural Information Processing Systems 2010 Workshop on Computational Social Science and the Wisdom of Crowds.*

Desmarais, Bruce A and Skyler J Cranmer. 2012. "Statistical mechanics of networks: Estimation and uncertainty." *Physica A: Statistical Mechanics and its Applications* 391(4):1865–1876.

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* .

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.

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

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

Kosma, Montgomery N. 1998. "Measuring the influence of Supreme Court justices." *The Journal of Legal Studies* 27(2):333–372.

Krivitsky, Pavel N. and Mark S. Handcock. 2017. *tergm: Fit, Simulate and Diagnose Models for Network Evolution Based on Exponential-Family Random Graph Models*. The Statnet Project (`http://www.statnet.org`). R package version 3.4.1.
URL: *https://CRAN.R-project.org/package=tergm*

Leifeld, Philip, Skyler J Cranmer and Bruce A Desmarais. 2017. "Temporal exponential random graph models with btergm: Estimation and bootstrap confidence intervals." *Journal of Statistical Software* .

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.

Makhoul, John, Francis Kubala, Richard Schwartz, Ralph Weischedel et al. 1999. Performance measures for information extraction. In *Proceedings of DARPA broadcast news workshop*. pp. 249–252.

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.

Nguyen, Hien D. 2017. "Near universal consistency of the maximum pseudolikelihood estimator for discrete models." *Journal of the Korean Statistical Society* .

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." *arXiv preprint arXiv:1708.02598* .

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/`.

Van Duijn, Marijtje AJ, Krista J Gile and Mark S Handcock. 2009. "A framework for the comparison of maximum pseudo-likelihood and maximum likelihood estimation of exponential family random graph models." *Social Networks* 31(1):52–62.

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.
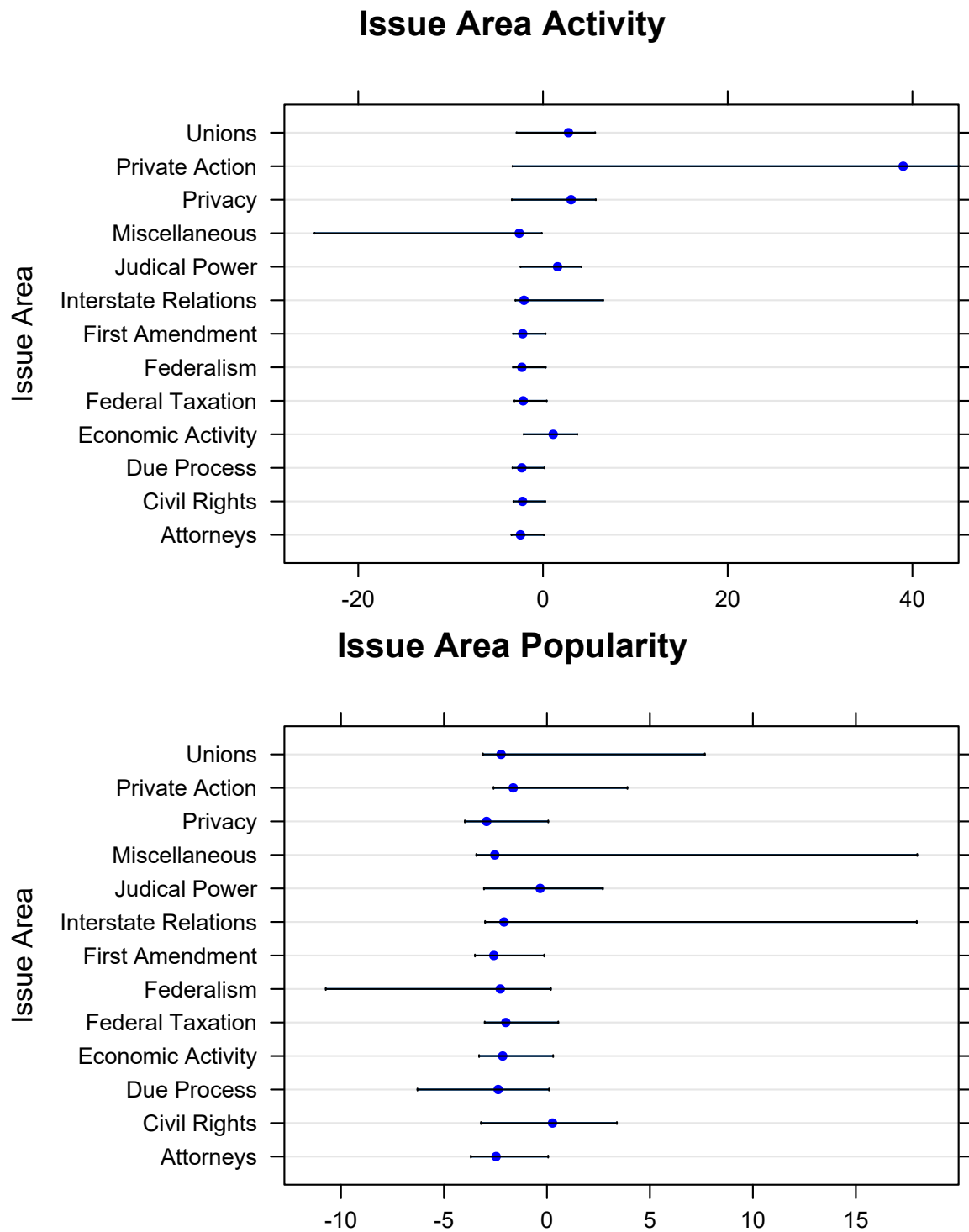
Figure 11: Issue Area Popularity and Activity (nodeifactor and nodeofactor). Baseline is Issue Area 'Criminal Procedure'.
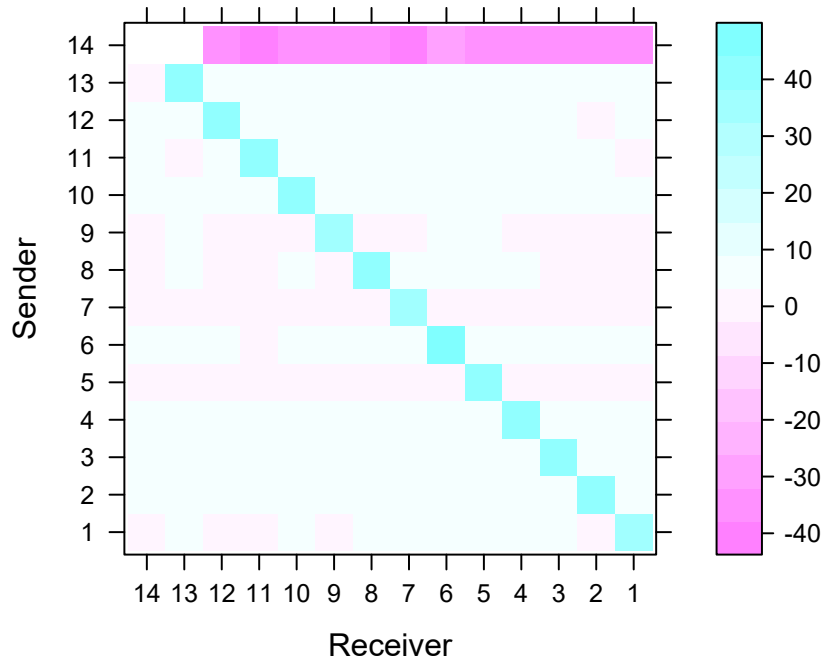
Figure 12: Nodemix Issue Area. Sender14Receiver14 and Sender14Receiver13 couldnt be estimated.