Cosponsorship in the U.S. Senate: A Multilevel Approach to Detecting the Subtle Influence of Social Relational Factors on Legislative Behavior∗

Justin H. Gross† and Cosma Shalizi‡

April 23, 2007

Abstract

Why do members of the United States Congress choose to cosponsor legislation proposed by their colleagues and what can we learn from their patterns of cosponsorship? These are questions that have attracted increasing interest among legislative scholars over the past several years, and are, fundamentally, questions about relationships. Unfortunately, most methods of statistical inference with which researchers, even methodologists, are likely to be familiar tend to be ill-suited for the analysis of relational data, in which observations are typically interdependent. Previous empirical research on cosponsorship in the House and Senate has suffered from two principal limitations. First, it has used statistical tools that ignore the systematic clustering of observations, leading to incorrect inferences. Second, too much emphasis has been placed on large-scale influences such as party

∗Committee: S.E. Fienberg (CMU), David Krackhardt (CMU), Cosma Shalizi (CMU), and Jennifer Victor (Pitt). Additionally, I would like to thank James Fowler for his generosity in giving me access to his cleaned data and Jonathan Woon for his helpful suggestions.
†H. John Heinz III School of Public Policy & Management and Department of Statistics, Carnegie Mellon University
‡Department of Statistics, Carnegie Mellon University
and state or region at the expense of more subtle social factors that operate at a lower order of magnitude.

In the current paper, we show how carefully chosen random effects might be included within a generalized linear model, in order to better handle network-type patterns of dependence. By explicitly modeling the multilevel structure of the data, we can more confidently investigate whether various dyad-specific properties predict senators’ tendency to support one another’s proposals. In addition to being better suited to a network setting, this approach allows us enormous flexibility in the choice of covariates we may incorporate and the degree to which we pool observations from different levels of analysis. To illustrate, we examine whether a number of potential social factors, capturing homophily, proximity, and role are associated with varying odds of cosponsorship among senators.

1 Introduction

Despite the fact that the act of bill cosponsorship is relational by its very nature, with observations made on pairs or sets of actors and their legislative proposals, empirical research has, for the most part, neglected to employ suitable tools in analyzing these data from a social relational point of view. Previous scholarship in the area has focused primarily on possible motivations for cosponsorship behavior, but has largely avoided addressing the sticky statistical issues that inevitably arise in a careful consideration of relational data. Some have simply confined their research to questions which treat individual legislators as the units of observation. Those who have studied the directed dyad (ordered pair of legislators) as observational unit have generally restricted themselves to descriptive techniques or have applied models that mistakenly assume dyads are independent, leading to wild underestimates of standard errors and a misrepresentation of key features of the data. This should not be surprising, as good methods for dealing with the sorts of dependency that arise in social networks are not at all obvious; indeed development of such methodology remains a vibrant area of research. Nonetheless, recent contributions in the literatures of statistics, sociology and social psychology provide us with fairly accessible methods that are particularly applicable to exactly these sorts of investigation. One such

\[Please note, in the current draft, double brackets [[...]] are inserted as a self-reminder to finish a thought that I have had difficulty articulating or retrieve reference information.\]
approach, the multilevel mixed effects model, is especially well suited to the study of bill cosponsorship, as I shall demonstrate.

There are additional reasons, apart from statistical challenges, for an apparent reluctance on the part of researchers to seek out social explanations for intra-chamber legislative behavior. If we are to look at systematic patterns in behavior, it is far from clear which relational variables might be predictive of patterns in cosponsorship, apart from the obvious ones, such as shared party, ideology, and state or region. It might be helpful to ask ourselves what information we would collect if we had complete and unfettered access to senators and their staff, or—better yet—if we could be omnipresent observers in the halls and offices of Congress. What we would look for? We might wonder whether, all else being equal, friendship or feelings of affection increase the likelihood of cosponsorship. But how would we even measure that? Would we sit down with each member of Congress (MC) and ask them for which colleagues they feel the greatest affection? If we could be guaranteed a response, how genuine would it be? If we could be flies on the wall, omnipresent observers of word and gesture, what would we look for? Smiles? Informal conversation? In other settings, for example, social network analysis has looked at children’s interactions for clues to their preferences [cite], or (in an experimental setting) the percentage of time pairs spent in conversation with one another [double-check source] [36]. An analogous study carried out in the halls of Congress, were it feasible, might well yield misleading results. After all, these are professional politicians, not children in a schoolyard; it should come as no surprise that their interpersonal inclinations are not terribly transparent.

Politicians themselves recognize that personal affection can even translate into political support among colleagues who might not necessarily see eye-to-eye on policy matters. Representative Barney Frank (D-MA) recalls how the outcome of his attempt in 1984 to get a bill, sponsored by Morris Udall (D-AZ) and granting power subsidies to certain western states, voted down hinged on interpersonal dynamics:

"Mo’s a public power guy, a westerner, and it was his committee handling the bill," Frank recalled. "The environmentalists were with us; the old system was wasting water. It was bad economics. I talked to some guys on the floor. I said, ‘Look, on the merits of the bill you should be with us.’ And they said, 'But how can we vote against Mo?' It was a fairly close vote but we lost. We were
opposing Mo, and we love Mo. We lost because of Mo—a good example of personality affecting politics.”[[insert original source]], cited in [35]

How might we make the transition from suggestive, but anecdotal, treatments to a more systematic investigation, given the constraints mentioned above? Research on interpersonal dynamics among lawmakers has existed for quite some time, largely limited to state legislatures; perhaps access is more easily obtained and the scale more manageable in such settings. Samuel C. Patterson, in particular, has carried out ground-breaking research in this area for decades. As far back as 1959, Patterson examined interpersonal relations among legislators of the Wisconsin Assembly were for clues to the inner workings of the decision-making process [33]. He and others have, over the years since, conducted research on friendship, communication, and political respect among state legislators [30, 6, 5, 9]. While it would seem natural to, in a similar fashion, consider the possible social aspects of bill cosponsorship, the social dimension has, with only a few of notable exceptions [3, 14, 8, 13], been ignored.

If we wish to identify the sorts of actor and dyad attributes associated with higher than typical cosponsorship activity, we might frame the issue in the following manner. Suppose we select a bill at random, learn the identity of the sponsor, and must place a bet on who might be a likely cosponsor of that bill, how would we proceed? A priori, without any further information about the bill, we might like to put a wager on someone from the sponsor’s own party, with similar ideology, and/or someone from the the same region as the sponsor, perhaps the other senator from the same state. Now imagine the same scenario, but we are asked to wager on whether a particular colleague will cosponsor the bill. How would we decide what odds to demand? What if we could ask one question about the cosponsor or the bill, what might give us the greatest edge? The questions yielding the most information would correspond to the types of variables we wish to identify in our research.

Whether or not senators are from the same party or geographic region has, not surprisingly, a great impact on the likelihood that they will cosponsor one another. Yet if we take such large-scale predictors of cosponsorship as benchmarks and insist that for other traits to be deemed important, they must explain variance on the same order of magnitude, we will find ourselves

---

2Note that we are concerned with contextual rather than statistical significance here.
completely missing many interesting factors. Still, only a few scholars have
dug deeper in examining the possible impact of common traits among MCs;
Swers, for example considers whether women legislators are more likely than
men to cosponsor bills concerning health, family, or education [37]. Other
articles that examine how homophily among MCs corresponds to cosponsor-
ship tend to limit themselves to party and ideology [4, 39]. This makes little
sense, however. Suppose that instead of Congressional networks, we were
interested in agent and dyad variables that serve as predictors of romantic
attraction on a college campus, and were to drop items like shared inter-
ests, backgrounds, and physical attributes from our model simply because
their predictive value paled in comparison to the dyadic variable opposite
sex. Of course we would never do this, recognizing that gender composition
of couples operates at a completely different level from the nuanced factors of
interest to sociologists or psychologists. Instead we might restrict our study
to opposite-gender or same-gender couples, pool the two and look at overall
trends, or compromise in some way. In the case of Congress, we would argue
that certain main drivers of cosponsorship behavior such as party and geog-
raphy should not enter into a model in the same manner as more subtle (and
potentially more interesting) factors do. We thus should distinguish between
the principal drivers of cosponsorship and social relation predictors.

To return to our imaginary wager, suppose we open to a page of the Con-
gressional Record, blindly selecting a bill from a given year, then place our
bet with no information on sponsor nor bill, having demanded odds no less
than $(1 - p) : p$, where $p$ is the probability that a random senator cosponsors
a random bill. If we are allowed one clue to improve our chances, we will
request information on party or geography. However, if we are denied clues
about these major predictors, what then will we choose? If, on the other
hand, we are told whether the two senators are from the same state and
region, and from the same or different parties. What do we choose for our
next clue? These correspond to two extremes of analysis, marginal probabil-
ity (pooled over other categories), and conditional probabilities (restricted to
the known demographics). These are both interesting questions, but statistic-
ical models reminiscent of the former will tend to be derailed by clustering
and heterogeneity, while models along the lines of the latter will tend to
“overfit” the data. The truth in which we are interested as political scientists
lies somewhere in between, so we will be well-served by methods allowing us
to explore the space in between.

The primary purpose of the current article is to introduce scholars of
legislative behavior (and institutional political behavior, more generally) to statistical methods that yield reliable inferences when the units of observation are dyadic. The multilevel framework, of which our particular modeling strategy is but one example, is highly flexible, and especially useful when observations are unevenly distributed with clearly defined clusters. Furthermore, the mixed effects used to capture network dependence have the virtue that they may be incorporated into any generalized linear model according to the nature of one’s particular data set. We will use logistic regression, as the outcomes of interest are dichotomous, but other researchers may wish to adapt these techniques for use with count data or censored data\textsuperscript{3} via the appropriate link function. Using the 108th Senate as an example, we demonstrate the functionality of the mixed effects approach. Restricting our attention to a single two-year session allows us to focus our attention on the merits of the modeling technique; there are sufficient levels of complexity to occupy us without the added question of how to handle turnover in membership. Furthermore, by treating each potential act of cosponsorship as a data point, two thousand or so pieces of legislation translates into a couple of hundred thousand observations, so computing time is not inconsequential. Thus, we will make no claims about the external validity of our results beyond this particular session of Congress, but offer them as points of departure for further study.

A second contribution of the paper may be of particular interest to those who study cosponsorship. We show that certain social relationships between senators—relations more subtle than the obvious predictors of party/ideology and geography—are in fact associated with increased propensity to support legislative proposals. The particular techniques we employ allow us to look beyond the principal (and obvious) predictors of party/ideology and geography to more subtle potential influences. A rich tradition in sociology and social psychology suggests the importance of homophily (“birds of a feather flock together”) and proximity (opportunities for interaction) in tie formation. McPherson, et al. provide a solid introduction to the former [29], while [21] provides an early reference for the latter. We look at a number of possible social relation predictors, including common religion, common profession, similar state demographics, shared committee membership, and office loca-

\textsuperscript{3}Indeed, procedures for truncated data may be especially fruitful once time-varying components are considered; a bill’s lifecycle may begin at any point in a two-year session of Congress, but if it does not end by formally being voted up or down, it will meet certain death at session’s end.
Figure 1: Graphical representation of underlying dynamics to be explored; in short, common traits and opportunities for interaction among senators and their staffs (observable) may lead to social tie formation (not systematically observable), which in turn increases willingness to cosponsor and likelihood that one will be personally encouraged to cosponsor (unobservable), leading to higher actual cosponsorship rate (observable).

In section 2, we briefly review the literature on cosponsorship in the United States Congress, with particular emphasis on recent work that treats cosponsorship as a signal of support between MCs and examines the resulting social network that emerges from such a perspective. In section 3, we introduce the data and explore the 108th Senate cosponsorship network at multiple levels: individual, dyad, and network as a whole. We construct a sociomatrix of weighted proportional support for each ordered pair of senators, and use this to explore possible patterns in cosponsorship via visualization, and demonstrate the presence of reciprocity in the network using simulations and simple correlation. In section 4, we briefly review the principal types of statistical models for social networks, and explain why multilevel modeling is well-suited to the data at hand and the types of dynamics to be exam-
ined more closely. We formally state our model for cosponsorship, as well as stating the explicitly the patterns to be investigated via model specification. Estimation is carried out in the statistical programming environment R, and results are reported and discussed in section 6. Finally, we briefly mention the implications of these results along with possible avenues for further research on social relations modeling of legislative behavior.

2 A Brief History of Cosponsorship and the Literature on Cosponsorship

By the regulations of Congress, each proposed piece of legislation may have only a single official “sponsor” in each chamber. In reality, a number of members of Congress (MCs) may be involved in crafting and pushing a bill from the earliest stages, but only one can be designated as sponsor. The practice of cosponsorship has been permitted in the U.S. Senate since the 1930s and in the House since the 1960s, identifying those others involved or simply wishing to be on the record as supportive of a measure. Nonetheless, for many years, cosponsorship was all but ignored by Congressional scholars, a legislative behavior without a clear procedural purpose. James E. Campbell, in a 1982 article, called this conventional wisdom into question. Using ordinary least squares regression (OLS) he sought to explain cosponsorship in terms of individual ideology and electoral considerations [7].

Various studies would follow suit, focusing on possible reasons for cosponsorship and debating the relative merits of electoral motivation versus institutional signaling mechanism [27, 39, 28]. Table 1 summarizes the research questions and methods of several of the most prominent of these. Notice that explaining cosponsorship activity typically has meant taking as dependent variable the number of bills cosponsored rather than the degree to which certain dyads are more prone to cosponsorship than others.

Only very recently have a small number of researchers begun to consider cosponsorship patterns in the U.S. Congress as constituting a social network (SNA in the methods column of the table). Tracy Burkett’s unpublished dissertation is the first known example of this [3], with two other papers based on her data to follow [4, 8]. James Fowler, with his two articles published in 2006, has made a significant contribution by using over thirty years of data on some 280,000 pieces of legislation to describe the associated support
networks. He introduces a new measure of network centrality, which he terms *connectedness*, a thoroughly network-theoretic construct, and uses this measure to make predictions concerning the ability of legislators to get their bills passed. In so doing, he takes the important step of considering various types of relationship manifest in the network: institutional, regional, issue-based, and personal [13, 14].

3 Data and Exploratory Analysis

In creating an adjacency matrix to represent support via cosponsorship, the choice of measure is non-trivial. We wish to estimate the matrix $P$, a $100 \times 100$ matrix with 9900 meaningful entries (structural zeros on the diagonal), where each element $P_{ij}$ should represent, in some sense, the propensity of senator $i$ to cosponsor legislation proposed by senator $j$. In related work on coauthorship and scientific collaboration networks, typically $P_{ij} = 1$ if $i$ and $j$ have ever collaborated and zero otherwise [31]. This is well-suited to sparse networks without strict boundaries, but we run into trouble if we try this approach with the U.S. Congress, especially the Senate. The Senate cosponsorship network is much denser than other collaboration networks; in our dataset, 67% of the 9900 directed dyads are non-empty. On the other hand, we should not simply use a raw count of times each senator cosponsors each colleague, as this will overemphasize support of senators who sponsor a lot of legislation. We might think of $n_j$, the number of bills sponsored by $j$, as the number of opportunities for each other senator to show support for $j$; if these trials were independent, we might think of $n_{ij}$, the number of times $i$ cosponsors a bill by $j$, as being distributed $Bin(n_{ij}, P_{ij})$. However, the more cosponsors who appear on a bill, the less informative the act of cosponsorship is. To take this into account, Fowler adopts a measure employed by Mark Newman in his article describing coauthorship of scientific journal articles [32]. Each act of cosponsorship is scored fractionally, in inverse proportion to the number of cosponsors on the bill. The resulting measure may exaggerate ties to senators sponsoring a great deal of legislation; for exploratory purposes, I adapt it slightly, dividing Newman and Fowler’s weighted raw score by the total possible score any colleague of $j$ could accrue if appearing as cosponsor on all of $j$’s bills. This approach is still not ideal, but good enough for exploratory analysis.

Taking the lowest level observation to be $Y_{ij(k)} = 1$ if $i$ cosponsors $j$’s $k$’th
<table>
<thead>
<tr>
<th>Author/Paper</th>
<th>Primary research questions</th>
<th>Explanatory variables</th>
<th>Dependent variables</th>
<th>Method(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fowler 2006 [13, 14] 93rd to 108th House &amp; Senate</td>
<td>Most “connected” MCs, does connectedness correspond to influence?</td>
<td>Connectedness (actor in net)</td>
<td>Political Influence (actor in net)</td>
<td>SNA, GLM (neg. binom. for counts)</td>
</tr>
<tr>
<td>Burkett &amp; Skvoretz (unpublished) [4] 93rd to 101st House and Senate</td>
<td>Influence of party &amp; tenure on support &amp; reciprocity through time</td>
<td>Actor &amp; dyad traits</td>
<td>Dyad counts</td>
<td>SNA ((p_1)log-linear, (\chi^2)–test, correlations)</td>
</tr>
<tr>
<td>Faust &amp; Skvoretz 2002 [8] 93rd to 101st Senate</td>
<td>Comparison of diverse networks</td>
<td>NA</td>
<td>Network (graph) components</td>
<td>SNA ((p^*)) and correspondence analysis</td>
</tr>
<tr>
<td>Koger 2003 [28] 96th to 105th House</td>
<td>What traits and goals of MC influence choice to cosponsor (&amp; frequency)</td>
<td>Actor covariates</td>
<td>Number of bills cosponsored</td>
<td>OLS Regression &amp; “semi-structured interviews”</td>
</tr>
<tr>
<td>Garand &amp; Burke [16] 102nd to 105th House</td>
<td>Impact of majority status on cosponsorship</td>
<td>seniority, ideology, etc.</td>
<td>Number of bills cosponsored</td>
<td>GLM (neg. binom. for counts)</td>
</tr>
<tr>
<td>Swers 2005 [37] 103rd &amp; 104th House</td>
<td>Gender and interest in education, health &amp; family legislation</td>
<td>Actor (senator &amp; constituency) traits</td>
<td>Number of bills cosponsored</td>
<td>GLM (neg. binom. for counts)</td>
</tr>
<tr>
<td>Kessler &amp; Krehbiel 1996 [27] 103rd House</td>
<td>Impact of ideology on frequency of cosponsorship, timing of action</td>
<td>Actor characteristics</td>
<td>Timing of cosponsorship</td>
<td>Survival Analysis (hazard rate function)</td>
</tr>
<tr>
<td>Wilson &amp; Young 1997 [39] 99th House &amp; Senate</td>
<td>Compare competing models of success for bills</td>
<td>Number of cosponsors per bill</td>
<td>Legislative success</td>
<td>(\chi^2)and GLM (probit)</td>
</tr>
<tr>
<td>Campbell 1982 [7] 95th House &amp; Senate</td>
<td>Predicting number of bills cosponsored</td>
<td>Actor characteristics</td>
<td>Number of bills cosponsored</td>
<td>OLS Regression</td>
</tr>
</tbody>
</table>
Figure 2: Bills and Amendments with at Least 1 and fewer than 99 Cosponsors (The distributions for the two activities are very similar, as evident from the histograms and by the similar ratios of quartiles and means: $\frac{Q^c_{1,sp}}{Q^c_{1}} = 6.62$, $\frac{Q^c_{2,sp}}{Q^c_{2}} = 5.35$, $\frac{Q^c_{3,sp}}{Q^c_{3}} = 6.08$, and $\frac{\mu^{c,sp}}{\mu^{sp}} = 5.84$. )
bill, and 0 otherwise, we let $n_j =$ the number of bills sponsored by $j$, $n_{ij} =$ the number of bills sponsored by $j$ that are cosponsored by $i$, and $c_{ij(k)} =$ the number of cosponsors on senator $j$’s $k$’th bill. Then the relative weighted propensity to cosponsor of $i$ for $j$ shall be defined as:

$$WPC_{ij} = \frac{\sum_{k=1}^{n_j} \frac{Y_{ij(k)}}{c_{j(k)}}}{\sum_{k=1}^{n_j} \frac{1}{c_{j(k)}}}.$$ 

The dominator represents the total raw score possible if $i$ were to appear as cosponsor on every single piece of legislation proposed by $j$.

Most visualization tools used in social network analysis assume ties to be dichotomous. That is, ties are presumed to either exist or fail to exist. While we lose much information by dichotomizing, and are left to make a somewhat arbitrary decision about the threshold, we can learn a lot from taking “snapshots” as we raise this threshold, noting which sorts of relationships seem to persist at high thresholds. Now the choice of threshold ($t$ such that an arrow will be generated from $i$ to $j$ wherever $Y_{ij} > t$) determines what sort of picture we get, from the extremely dense network for $t = 0$ to a completely disconnected one for $t$ large. This sensitivity to threshold choice does not, however, render the practice of discretizing links useless. In fact, much may be learned by slowly changing the cutoff and noting the evolution of associated “snapshots” as we do so, paying special attention to patterns that persist as weaker connections are removed. Figures 3 through 8 were generated by Netdraw, within the software suite Ucinet [2]. For lower thresholds, the main thing to understand is that relative positions are determined so that those nodes sharing an edge tend to be near one another. Labels are of the form name/party/state. Arrowheads indicate the direction of support, but are not easy to see at lower thresholds. Colors indicate party (red for Republican and blue for Democrats, yellow for Jeffords, an Independent voting mainly with Democrats).

### 3.1 Party

In figure 3, with threshold set to 0.2, approximately the mean of all $WPC_{ij}$, cosponsorship by party dominates the visible pattern. Nodes here are situated by a “spring-embedded” algorithm that attempts to minimize the energy
in the system by placing those with shared edges close together [cite]. Notice that Zell Miller, the conservative southern Democrat is situated far to one side, surrounded by Republicans, and that the most liberal Republicans, Chafee, Collins, and Snowe, all from New England, are swimming in a sea of blue. The density of the network is also clearly greater on the Democratic side, as the proliferation of edges darken the neighborhood around Senators Clinton, Kennedy, and Harkin. Shapes represent different six different regions of the country, NE, SE, MW, SW, NW, plus the non-continental states of Hawaii and Alaska; some clustering by geography is evident within each of the two parties.

At a threshold of 0.20, (Figure 4), a cluster of high-profile Democrats, mostly from the Northeast, remain connected to one another, but most other links are by region and state.
Figure 4: Weighted Propensity to Cosponsor (Threshold = 0.20)
3.2 Geography

It might be easier to identify potential regional patterns if we place the nodes approximately as they would appear on a U.S. Map. In Figure 5, same-state pairs abound, with little remaining of the partisan clustering seen at lower thresholds. Other than the state links, there is not an overwhelming degree of regional alignment here. Instead, we see a number of seemingly random cross-party ties stretched across the map.

It turns out that these transcontinental streaks are hardly random. Once we raise the threshold to 0.35, at which point only relatively few ties persist (around thirty), the pattern becomes striking. In Figure 6, all but a few links are same-state pairs. Of thirteen that do not fall into this category, ten consist of committee leader pairs, with ranking minority member (RMM) sending a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.\footnote{Jeffords had only recently left the GOP, and Chafee would be voted out of office onselling a non-reciprocated tie to the committee chair. Of the remaining three pairs, two are RMM/Chair for subcommittees of the prominent Appropriations Committee. The only edge not linking either same-state pairs or leader pairs is from Jeffords (I-VT) to Chafee (R-RI). They share a committee (Environment and Public Works), but more importantly, perhaps, represent the last of a dying breed—the New England liberal Republican.}.
3.3 Ideology

We can use senators’ vote-based ratings by the American Conservative Union ACU for 2003 (http://www.acuratings.org) to see whether stronger ties appear to connect those senators with similar ideological leanings. In Figure 7, larger nodes correspond to higher ratings by the ACU, and non-reciprocated ties (at threshold 0.25) appear in red.

At this threshold, there no longer appears to be much evidence of ties between ideologically similar senators. As already noted, the ties that persist the longest in this exercise are those linked by state or by shared committee leadership.

3.4 Committee Leadership

Just as we arranged nodes geographically to make that pattern stand out, we can do something similar for committee leaders. In Figure 8, Republican Chairs are situated directly above their Democrat counterparts (RMM) on the same committee.

---
a wave of anti-Bush sentiment in the 2006 elections
Figure 7: Weighted Propensity to Cosponsor, by Ideology (Threshold = 0.25)

Figure 8: Weighted Propensity to Cosponsor (Threshold = 0.20), by Committee Leadership
4 A Multilevel Approach to Statistical Inference for Relational Data

The preceding exploration of the data suggests possible elements to include in specifying a model, but no less important is what this exercise demonstrates in terms of the richness of the data set. Just as a biologist wishing to understand what is going on in some tissue will learn something different by setting her microscope to various levels of magnification, we too stand much to gain by applying inferential tools with the flexibility to allow us to do the same for a complex social structure. Of various methods of statistical inference for social networks, one seems especially well suited to the sort of questions we would like to ask of these particular data.

Statistical modeling of social networks reached new levels of sophistication beginning in the early 1980s, with several key publications relating the distribution of network ties to log-linear models and exponential families [26, 11, 10, 15] Since then, several distinct approaches have developed, each with its own strengths and weaknesses; to choose among them for a given application, one must consider the idiosyncrasies of the particular network and type of inference one wishes to make. For a nice review of relevant methods, see van Duijn and Vermunt’s introduction to a special issue of the journal Methodology devoted to social network analysis [38].

In order to appropriately select a modeling framework for the types of hypothesis we would like to test, we need to think really carefully about the structure of the data. The only analyses that have treated cosponsorship behavior patterns as constituting a social network (or providing clues about an underlying support network) have counted incidents of cosponsorship from each $i$ to $j$ and taken these counts as indication of tie strength [3, 8, 12, 14, 4]. However, the implicit assumption that the probability $p_{ij(k)}$ of one senator cosponsoring legislation by another is constant over bills (i.e. $p_{ij(k)} = p_{ij}, \forall k$) would seem to be too strong. In fact, upon closer examination, what we have is a bipartite graph, containing two distinct components, senators and bills, with the former connected to each other only through the latter. It is common to collapse such a network over one component in order to analyze the connections among actors according to their connections through the other type. In the case of cosponsorship, to do so assumes homogeneity

\footnote{what social network analysts sometimes refer to as a two-mode network}
among bills that is simply not realistic. Moreover, by ignoring the context of each observed opportunity for $i$ to cosponsor a bill by $j$, we lose an important tool to understanding other aspects of such relationships.

To see why a multilevel model is appropriate, one needs to focus on structure rather than form. Consider Figure 9, which pictures the framework of cosponsorship data next to a more familiar application, in educational testing research. Just as a student may take a number of exams, with all test scores grouped by student, so are there a number of opportunities for the link $i$ to $j$ to manifest itself. While students are clustered hierarchically in schools nested in districts, each raw 0-1 observation of $Y_{ij(k)}$ belongs to a directed dyad $(i, j)$, which in turn is nested together in the dyad $\{(i, j), (j, i)\}$. Furthermore, as students may have an one instructor and another exam grader drawn from the same teacher pool, each observation in our data set has cross-classified effects from cosponsor and sponsor, drawn from the same pool of senators.

The greatest advantage for us is that, by acknowledging this structure, we are able to essentially take weighted regressions based on the amount of variance at each level of clustering. So we gain precision in our estimates, and we also get much more realistic standard errors than we would by simply
ignored the structure and pooling results [17].

Another representation suggested by our exploratory analysis may be similarly compared to a typical multilevel setting, the family. Here the role combinations within committees (Chair, RMM, Member, and Non-member) are comparable to family roles. The effect of role-pairings are not taken to be independent and identically distributed. There are only a few different arrangements and we wish to estimate the particular contribution of these, so we treat these effects as if they were fixed parameters.

5 The Model and Hypotheses

Each agent has dual role of actor and partner (or sender and receiver), in our case cosponsor and sponsor in different situations. We must consider the separate impact of sender, receiver, and interaction for each observation. Since we are ultimately interested in the dyad, we necessarily must distinguish between these three components.

If we consider the joint probability of all dyadic relations, taken as a socioarray (\(n_j\) observations for each dyad \(ij\)), the correlated errors translate into a lack of independence among the components of our array,
meaning a crucial assumption of OLS and standard GLM, including logistic regression, does not hold. However, the dependence structure exhibited by these observations is not utterly intractable. To the contrary, there tend to be certain patterns of correlation that arise in networks and relational data in general; if the systematic components of this dependence can be successfully modeled, we may replace independence with conditional independence, so that remaining disturbance terms are no longer correlated. That is, our aim is to incorporate appropriate elements in our model so that we will be able to convincingly claim

\[
Pr(Y) \neq \prod_{i,j,k} Pr(Y_{ij(k)}),
\]

Let us begin with the simplest possible model and build up toward a more realistic one that will induce the sorts of network dependence that tend to be most systematic in this sort of setting. If we begin without any covariate information, and without any random effects whatsoever, we are a single fixed effect, the grand mean \( \mu \), the unconditional probability of cosponsorship. In the spirit of ANOVA-type models, this simple parameter serves as a baseline against which to compare deviations by particular senators and pairs of senators. This parameter, \( \mu \), will become the fixed intercept of our model; in the parlance of social network analysis, \( \mu \) is the density of the network, or proportion of observations where cosponsorship has indeed taken place. Thus, the basic building block of the model would look like this:

\[
g(Y_{ij(k)}) = \mu + \epsilon_{ij(k)},
\]

where \( g(\cdot) \) is some link function, and \( Y_{ij(k)} = 1 \) if Senator \( i \) cosponsors Senator \( j \)'s \( k \)'th bill, and 0 otherwise. Now, as it stands, the error term will have some hopelessly complicated distribution, depending on sponsor, potential cosponsor and particular bill. The goal is to decompose fixed and random effects to the point where any remaining error can be reasonably assumed to have some nice distribution. The setup is essentially Snijders and Kenny's
multilevel social relations model [36], adapted for dichotomous observations in a similar manner to that of Hoff, et al. [22, 24, 23, 25], and to Gill and Swartz [19, 20, 18].

For the moment, we assume $k$s to be exchangeable in the sense that the probability of a given senator cosponsoring any bill with a given sponsor is assumed to be the same. We relax this assumption later.

$logit(Y_{ijk}) = \mu + a_{i}^{sender} + b_{j}^{receiver} + \epsilon_{ij},$

where the actor specific random effects are assumed to be distributed as

$$\begin{bmatrix} a_{i} \\ b_{j} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{a}^{2} & \sigma_{ab} \\ \sigma_{ab} & \sigma_{b}^{2} \end{bmatrix} \right) ,$$

$$\begin{bmatrix} \epsilon_{ij} \\ \epsilon_{ji} \end{bmatrix} \sim MVN \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_{\epsilon}^{2} & r \sigma_{\epsilon}^{2} \\ r \sigma_{\epsilon}^{2} & \sigma_{\epsilon}^{2} \end{bmatrix} \right) ,$$

with $r$ a reciprocity (mutuality) factor [24] and we do not assume the covariance elements for actor effects to be zero, since sender and receiver effects may well be correlated for the same actor. That is, senators’ relative sponsoring activity may covary with level of cosponsorship activity.

Thus, calling the entire random portion of the model $\eta_{ij} = a_{i} + b_{j} + \epsilon_{ij}$, the corresponding first and second moments will be

$$E(\eta_{ij}) = 0,$$

$$E(\eta_{ij}^{2}) = Var(\eta_{ij}) + E^{2}(\eta_{ij}) = Var(a_{i}) + Var(b_{j}) + Var(\epsilon_{ij}) = \sigma_{a}^{2} + \sigma_{b}^{2} + \sigma_{\epsilon}^{2},$$

$$E(\eta_{ij}\eta_{ji}) = Cov(\eta_{ij}, \eta_{ji}) + E(\eta_{ij})E(\eta_{ji}) = 2\sigma_{ab} + r\sigma_{\epsilon},$$

$$E(\eta_{ij}\eta_{im}) = \sigma_{a}^{2},$$

$$E(\eta_{ij}\eta_{mj}) = \sigma_{b}^{2},$$

$$E(\eta_{ij}\eta_{mi}) = \sigma_{ab} = E(\eta_{ij}\eta_{mj}).$$

Finally, $E(\eta_{ij}\eta_{mn}) = 0$, where all four indices are distinct.

Beyond these, all other expected products are assumed zero.

The typical assumption employed here is that these random effects are normally distributed [36, 24, 19], but this assumption will be more justifiable.
(and potentially verifiable) if we explicitly include the most plausible fixed effects, so that the remaining randomness comes from the sum of many smaller effects, thus likely following a Gaussian distribution, in accordance with the Central Limit Theorem. Of course, the incorporation of fixed senator-specific effects has the added benefit of allowing us to learn (or reaffirm) through model selection the particular attributes tending to be associated with different levels of cosponsorship behavior. This gives us:

\[ \logit \left( Y_{ij(k)} \right) = \mu + (\alpha'S_i + a_i) + (\beta'R + b_j) + \epsilon_{ij}^{(2)}, \]

where \( \alpha \) is a vector of coefficients for sender-specific effects corresponding to \( S_i \), the values of senator covariates for each potential cosponsor \( i \), while \( \beta' \) is a vector of coefficients for receiver-specific effects corresponding to \( R_j \), senator covariate values for the sponsor in question. Notice that these fixed effects are actor-specific only in the sense that the predictors are attributes of particular senators; the role of these predictors is assumed the same for all senators. Thus we will only be estimating \( 2m \) parameters, not \( 200m \). Since most of our dyadic covariates are functions of actor-specific variables, we run into collinearity problems if we try to incorporate fixed effects at the actor level as well. We should only include such coefficients if the covariates in question tell us more about actor-level variability than comparison across actors would tell us about pairs of senators.

Next we incorporate dyadic covariates of interest:

\[ \logit \left( Y_{ij(k)} \right) = \mu^{\text{globalMean}} + \left( (\alpha'_{\text{sender}})' S_i + a_{i_{\text{sender}}} \right) + \left( (\beta'_{\text{receiver}})' + b_{j_{\text{receiver}}} \right) + \delta^{\text{dyadic}} X_{ij} + \epsilon_{ij(k)}, \]

where \( X_{ij} \) is a vector of covariates on ordered senator pair \((i, j)\) and \( \delta \) are the fixed effects of these covariates, which we wish to estimate. The remaining issue is how to model the last term in such a way as to be simple enough to be estimable, but complex enough to capture remaining unspecified network dependence.

Snijders and Kenny point out that when reciprocity is positive, the reciprocity effect is equivalent to a dyad effect; we may think of each \((i, j)\) as being nested in the dyad pair \{((i, j), (j, i))\}. The effects associated with each pair are then modeled as coming from their own distribution, again centered on zero and capturing the variance associated with pairs of senator
without regard to who is in the role of sponsor. We will include the term \( R_{(ij)} \), following the Snijders and Kenny notation including indices parenthetically to indicate ordering as arbitrary. Now, in our data, we have repeated observations on directed senator pairs, so we can also include a term \( D_{ij} \), the random effect of \( i \) as cosponsor and \( j \) as sponsor, beyond what is captured by any mutual propensity to support one another.

\[
\logit \left( Y_{ij(k)} \right) = \mu + \left( (\alpha)' S_i + a_i \right) + \left( (\beta)' + b_j \right) + \delta X_{ij} + D_{ij}^{directed} + R_{(ij)}^{undirected} + \epsilon_{ij(k)}^{(4)}.
\]

Now we may address the lack of exchangeability among repeated observations on the same, \((i, j)\). We will add one last random effect, \( c_{j(k)} \), for the committee of jurisdiction over the bill at hand. Furthermore, and most importantly in testing the opportunities-for-interaction hypothesis, we must include one fixed effect particular to \( i, j, \) and \( k \), the coefficient of covariate \( \text{shared.cmtes.ij}(nok) \), actually an interaction between \( \text{shared.cmtes} \), the number of committees \( i \) and \( j \) have in common, and \( (1 - \text{shared.cmtes.ijk}) \), where \( \text{shared.cmtes.ijk} = 1 \) if \( i \) and \( j \) serve together on the committee having jurisdiction over bill \( k \). Indeed, the nature of the committee-centered legislative process will lead those on committees together to have greater probability of cosponsoring each others’ bills simply by virtue of the fact that members of the committee through which the bill must pass are generally the earliest and most crucial cosponsors. However, if the number of shared committees contributes to support propensity between senators not serving together on the bill’s committee, then the most logical explanation will be that increased opportunities for interaction among these senators and their staffs results in interpersonal connections translating into legislative support.

\[
\logit \left( Y_{ij(k)} \right) = \mu + \left( (\alpha)' S_i + a_i \right) + \left( (\beta)' + b_j \right) + \delta X_{ij} + D_{ij}^{directed} + R_{(ij)}^{undirected} + c_{j(k)} + \epsilon_{ij(k)}^{(5)}.
\]

with \( c_{j(k)} \) the random committee/topic effect for \( j’s k’th \) bill. This can be expected to allow for variation in cosponsorship activity based on committee dynamics, and the bill context. As always, we need to assume a distribution for these random effects, treating them as if committees were drawn from
a hypothetical universe of possible committees. We take $c_{jk} \sim N(0, \sigma_c^2)$, similar to our handling of actor effects $a_i$ and $b_j$.

We hope to have come reasonably close to being able to assume the remaining disturbance terms, $\epsilon_{ij(k)}^{(5)}$, independent. In fact, it would be preferable to add one further layer of interaction typically found in network data, to account for higher-order clustering, including transitivity and balance. Much can be gained from studying these in their own right, and even for our own research goal of interpreting dyadic relations, the neglect of triad properties will likely continue to yield standard errors that seem more precise than is merited. While great strides have been made in addressing higher-order network dependence by the inclusion of an additional term, a function of the latent positions of individual actors, these have yet to be generalized to bipartite networks [22, 23]. Nonetheless, as we shall see, the results under our model will be drastic improvements over standard OLS or GLM. In addressing our particular research questions, we stand to gain more by allowing for heterogeneity of contexts, i.e. of bills themselves, than by sacrificing this in order to capture higher-order correlation. Still, we need to be conservative in our interpretation of estimated standard errors obtained during estimation, as some underestimation is bound to persist.

5.1 Variables

Almost all of the dyadic covariates to be considered, as expressions of homophily or proximity, are functions of individual senator attributes. In some cases, there may be predictive value as actor-specific covariate (e.g. liberal legislators sponsor and cosponsor more often [cite], but for the most part, our focus is on dyadic compositions of these variables. (In fact, we often must choose between including a trait as an actor effect or as a component of dyad effect, to avoid collinearity.

The main data on proposed legislation during the 108th Congress is maintained in the Thomas database of the Library of Congress, and were compiled and cleaned by James Fowler. These include sponsorship and cosponsorship information, committees of jurisdiction, and legislative outcomes. State socioeconomic figures, as well as personal information on profession, veteran status and religious affiliation of senators were obtained from The Almanac of American Politics 2004 [1].
5.1.1 Senator-level variables

Almost all of the dyadic covariates to be considered, as expressions of homophily or proximity, are functions of individual senator attributes. In some cases, there may be predictive value as actor-specific covariate (e.g. liberal legislators sponsor and cosponsor more often [cite]), but for the most part, our focus is on dyadic compositions of these variables. (In fact, we often must choose between including a trait as an actor effect or as a component of dyad effect, to avoid collinearity.

Party During the 108th Congress (January 2003 through December 2004), the United States Senate was nearly evenly split by party. Of the one hundred senators, fifty-one were Republicans, eight were Democrats, and one—Jim Jeffords—a former Republican turned Independent. In the figures of dichotomized support propensities, Jeffords is shown in yellow, but for modeling purposes, he will be treated as a non-Republican, since he chose to align with Democrats on party line votes.

Ideology We use the primary component of Poole’s DW-Nominate scores for the 108th Senate to measure relative ideology (http://voteview.com/dwnomin). These are calculated based on roll-call votes and scored from -1.0 to +1.0 on the liberal-conservative dimension. A number of researchers have pointed out that one dimension is enough to explain most variation in roll-call voting during most periods in American history, and certainly during the past decade or two [34].

State State is even more likely play a role in cosponsorship behavior among Senators than among Representatives, as each is charged with representing the state as a whole, with its mix of regional demographics, rather than just a particular delegation within. Additionally, we might expect that having only two senators from each state invites, if not necessarily a special bond, then at least a common sense of obligation on state-specific concerns.

Region Studies partition the country by region in various ways. We group them into Northeast, Southeast, Midwest, Northwest, West, and the non-contiguous Pacific (AK, HI). A number of issues addressed in legislation are of particular regional importance, and senators might additionally feel certain cultural affinity with colleagues from their partic-
ular region, so we examine whether common region may be associated with increased cosponsorship.

**Socioeconomics** Some states may bear demographic similarity to one another despite being far apart geographically. I thus consider urban population percentage, poverty rate, and blue-collar workforce percentage [1] as possible indicators of socioeconomic homophily.

**Class** Senators serve six-year terms, and fall in one of three “classes” according to when they are up for reelection. Since electoral considerations may encourage legislative activity, an indicator for senators in their final session may be included.

**Freshman** Senators serving their first six-year term are indicated as such, as research has indicated that new legislators lacking the clout to successfully sponsor bills tend to utilize cosponsorship as a means of position-taking [[cite]]

**Gender** Fourteen women served as senators during the 108th Congress. Shared identity and concerns may give rise to higher cosponsorship rates among women [37].

**Identity** Shared religious affiliation, professional background, and veteran status [1] are examined as aspects of personal identity that may similarly play a role, if not overtly, then perhaps indirectly by way of common experience and exposure to similar points of view. A few religious denominations are well-enough represented in the Senate to be considered as covariates. We have considered three: Presbyterian (12), Catholic (23), and Jewish (11). Professional background was included in the form of indicators for the two most common: business (13) and legal (54). An additional indicator for veteran status (37), including those in the Reserv es, was also considered.

**Location** Office location is used to gauge potential opportunity for casual contact among senators and staff. For the time being, building and floor are used for this, but this will be replaced by identification of nearest neighboring offices, a more sensitive measure

**Committee** Senators serve on anywhere from two to six committees. Since the greatest opportunity for interaction among senators and staff from
different enterprises exists during committee business, committee membership will be of interest. Committee may also be used as an indicator of power and prestige, but the inclusion of a “top committee” variable for individual senators does not appear to be significant in explaining copresponsorship behavior.

<table>
<thead>
<tr>
<th># committees</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td># senators with this many assignments</td>
<td>4</td>
<td>32</td>
<td>42</td>
<td>16</td>
<td>6</td>
</tr>
</tbody>
</table>

**Leadership** Majority and minority leaders and whips play special roles and may be the intended recipients of signaling via copresponsorship. Furthermore, each committee and subcommittee has a chair and ranking minority member (RMM); these individuals play central roles in controlling the flow of legislation through their respective committees and subcommittees.

### 5.1.2 Dyad-level Variables

The dyadic attributes we are able to consider as covariates are functions of paired attributes of individual agents. In fitting the model, the following dyadic covariates were included at various stages: same party, party pair ($D \rightarrow D, R \rightarrow R, D \rightarrow R, R \rightarrow D$), absolute difference in DW-Nominate score, same region (but not state), difference in state socioeconomic indicator, both female, both Presbyterian, both Catholic, both Jewish, both attorneys, both businesspeople, both veterans, same building and floor, common committee assignments, and committee roles ($RMM \rightarrow Chair, Chair \rightarrow RMM, Member \rightarrow Chair...$)

In order to weigh the possible impact of contact via shared committee membership on copresponsorship behavior, we must decide how to measure increasing opportunities for interaction based on assignments. Our ideal data in this regard would be calendars for each senator and staff member indicating how time spent in committee meetings and working on committee business with others. After all, all committees are not created equal; some are essentially marginal, meeting only occasionally, while others are central and may demand a great deal of attention from members. Since we do not have this information, let us assume for the time being that each has the [same value]. We can potentially sharpen our inference later by looking more closely at relative importance of committees. We still run into the problem of
determining how to value number of shared committees. If we only consider raw counts, we may overvalue dyads where at least one MC serves on five or six committees. Each senator served on from 2 to 6 committees. The mean number was 3.9, median—4, standard deviation—0.9. The exact distribution is shown in the table below. Those serving on committees with the greatest amount of responsibilities tend to serve on the fewest.

Although some dyadic covariates are directed and others are symmetric, the two types can be treated in the same manner. After all, undirected graphs are simply special cases of directed graphs—we can think of every tie as being mutual. If we think of covariates $X_{ij}$ as constituting an array, some levels will be symmetric about the diagonal, while others will not.

5.1.3 Bill-level

For our model, we consider only bills, not amendments, and just those having at least one cosponsor, but fewer than ninety-nine. This gives us 2166 bills, yielding $2166 \times 99 = 214,434$ observations in all, one for each possible cosponsor on every bill.

There are all sorts of bill-specific covariates we might consider. Given the great work being done in text-mining, we are tempted to run some sort of topic-classification algorithm and use the results. Since trying to work bill-specific variability in at all is a challenge, let us maintain a more modest strategy for the moment. At the very least, we wish to have a model that addresses the bill as a source of variance. Thus the simplest approach would be to treat each $Y_{ij(k)}$ as reflecting an underlying propensity for Senator $i$ to support legislation proposed by Senator $j$, plus a random effect $\epsilon_k$.

But we can do better than this and must, if we are to effectively utilize data on committee appointments to explore whether opportunities for senator and staff interaction may increase probability of legislative support.

5.2 Patterns to Examine

**Proximity:** Increased opportunities for interaction between senators and their staffs will be associated with higher propensity to cosponsor. In particular, we examine whether (a) office location of senators, and (b) number of shared committee assignments tend to be predictive of cosponsorship.
Figure 11: Nearly half the 7123 bills and amendments considered by the Senate during the 108th Congress contained no cosponsors at all. Of those with at least one cosponsor, mean = 7.40.

**Homophily:** Common traits based on shared identity and personal history will be associated with higher propensity to cosponsor. Specifically, we attempt to verify this with (a) common religious affiliation, (b) former profession, (c) gender (both female), and (d) veteran status. We also examine the extent to which (e) same state and (f) same region—but not state—drive cosponsorship and compare these to the other principal driving forces, party and ideology.

**Roles:** Leadership roles within party, committee, and subcommittee will be predictive of cosponsorship patterns. Specifically, dyads consisting of committee Chair and Ranking Minority Member (RMM) will be involved in a higher than typical level of cosponsorship, and the direction will be primarily RMM supporting legislation proposed by the Chair.

**Unexplained Affinity:** There are, of course, countless reasons for enterprises (senators and staff) to develop affinity towards one another, constituting a complex web of social and political motivations, whim and happenstance, beyond the understanding of the participants themselves. Although
we cannot directly measure the elements of these varied interpersonal dynamics, we expect their overall impact to carry over to supportiveness on the least ideological legislation, where social inclinations face no opposition from partisan obligations and policy preference. We thus estimate senators’ propensity to cosponsor non-binding, non-controversial resolutions⁶, taken as a proxy for unexplained affinity, and see if this provides any additional explanatory power to our model.  

[[INCOMPLETE: Preliminary results support, but need more satisfactory way for handling missing data, as 11 of 100 senators do not sponsor a single resolution. Add results once this is resolved.]]

6 Estimation and Results

The results were computed with the “lmer” function for linear and generalized linear mixed effects in the R statistical programming environment. Initial estimation was done via penalized quasi-likelihood (PQL), an iterative approximate maximum likelihood procedure, which can suffer in terms of accuracy (as well as possible difficulty converging). Some were checked with the more computationally intensive Laplace Method. The former proceeds by Taylor expansions on estimated data, while the latter uses a trick from Bayesian statistics to approximate the integrand of the likelihood. Another way to proceed would be to take a fully Bayesian approach, and this may actually simplify the estimation process here [23, 24, 25]. For details on the estimation techniques and computational algorithms employed by the lmer class in R, which are used in this paper as a stepping stone toward more thorough Bayesian modeling, see Gelman and Hill, 2007 [17].

In Table 3, we see that the main drivers of cosponsorship behavior, as we have referred to party/ideology and region are clearly significant in either case. Whether a covariate has $|z|$ of 94 or 14 is inconsequential in terms of statistical significance; both are off the charts, as they should be.⁷ The limitations of the basic GLM become apparent upon examination of the homophily and proximity covariates. Two key problems arise at this level. One, nearly everything appears to be significant. Complete pooling treats the observations as if they were 214 thousand independent observations, when in

⁶eg. commending Congressional staff or establishing National Cowboy Poetry Month (seriously!), rather than the occasional partisan resolution.

⁷On the other hand, if we are interested in interpreting the coefficients, we will be mislead into thinking they more precise estimates than they truly are.
Table 2: Random Effects in multilevel GLMM

<table>
<thead>
<tr>
<th>Groups</th>
<th>Var Estimates</th>
<th>St. Err. of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>bill.committee</td>
<td>0.389</td>
<td>TBD</td>
</tr>
<tr>
<td>dyad(_{(ij)})</td>
<td>0.127</td>
<td>TBD</td>
</tr>
<tr>
<td>dyad.directed(_{ij})</td>
<td>0.147</td>
<td>TBD</td>
</tr>
<tr>
<td>receiver(_j)</td>
<td>0.308</td>
<td>TBD</td>
</tr>
<tr>
<td>sender(_i)</td>
<td>0.385</td>
<td>TBD [bootstrap or shift to Bayesian]</td>
</tr>
</tbody>
</table>

fact this is far from the case. Even though few might make the mistake of treating repeated observations on each dyad as independent, simply pooling these by dyad and treating them as proportions or counts would not alleviate the problem encountered here. As we know, the problem runs deeper, for we do not have 9900 independent observations (the number of directed dyads) either.

Note that using ordinary GLM, bothCatholic appears to be significant beyond any doubt, whereas in the mixed effects model it ceases to appear significant at any reasonable level.

While our primary purpose for inclusion of random effects in answering the research questions at hand is so that we can draw reliable inferences about the “fixed” effects of dyadic covariates, an examination of the estimated variance for these random effects is itself useful. On the log scale, the total variance among all 214,434 observations is 1.1, 0.308/1.1 = 28% of which is “accounted for” by variation among receivers and 0.385/1.1 = 35% by that of sender. Beyond the individual actor effects, 13% of observed variation may be attributed to differences among the 9900 particular ordered pairs. Despite the fact that we are working in logodds, the fact that the original observations are binary means we should not try to interpret variance estimates and ratios as we would for continuous data.

Now, as we mentioned earlier, the fact that we did not account for third-order dependence such as transitivity and balance means that we should err conservative on whether fixed effects appear to be non-zero. For example, samefloor does not seem to be different than zero, given the sign does not make sense, but the 0.06 p-value would be borderline anyway. Still, the fact that a number of p-values have come down to earth is heartening. Certainly, covariates with coefficients more than, say, three times their corresponding standard errors, should raise our eyebrows.
The homophily and proximity coefficients that seem to be significant include both female, and shared committees. Each additional shared committee for a dyad corresponds to an increase of 0.6 (not shown) in log odds of cosponsorship, or an increase in probability of no more than 0.15 [17]. However, once the effect is decomposed into the case where both senators sit on a committee with jurisdiction over the bill and the case where they do not, the estimated coefficient for the former is 0.936, while the latter is still significant, but with a coefficient of only 0.075. That is to say, there is some evidence that opportunities for interaction result in greater cosponsorship, though only something on the order of 0.02 extra per shared committee when the committees have nothing to do with the bill considered.

Without addressing the various types of clustering of observations, inferences will be led astray in two principal ways: first, the degrees of freedom will be vastly overestimated, yielding estimated standard errors that are far too small\textsuperscript{8}, and, second, estimates may be biased.

There is a fair amount of anecdotal evidence to suggest that sponsorship serves quite different purposes for members of the minority party than for those in the majority, \textsuperscript{[[find the source]]} and it would not be all that surprising to find that patterns of cosponsorship differ by party combinations of dyads. There are only four possible arrangements of directed dyad by party (DD, RR, DR, and RD); let us run unpooled versions of the above model for each and compare some of the results.

In table 4, significant effects are listed for each of the four ordered party pairings. While it is inappropriate to compare coefficients for ideological difference (based on DW-Nominate scores), once we have subdivided by party, it is interesting that it comes up as statistically significant only in the two cases where Democrat is sponsor, \textit{DD}, and \textit{RD}. This may just be an artifact of the measure’s ability to distinguish among certain types of actors, or it might indicate something deeper, perhaps the fact that the majority Republicans will not compromise on policy preference to cross the aisle and cosponsor as easily as the majority Democrats will. Also interesting is the lack of significance for sameregion.notstate among \textit{DDs}.

Lastly, we run GLMMs for two types of committee role groupings. The Chair sender-effect is insignificant, Chairs cosponsor RMM bills more than would be expected at random, with all else equal.

\textsuperscript{8}This is not only due to the obvious lack of independence among observations on common dyads, but among those with common sender, receiver, and higher order interaction.
Table 3: multilevel GLMM vs. standard GLM results

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Est\textsubscript{GLMM}</th>
<th>S.E\textsubscript{GLMM}</th>
<th>z\textsubscript{GLMM}</th>
<th>z\textsubscript{GLM}</th>
<th>p-val\textsubscript{GLMM}</th>
<th>p-val\textsubscript{GLM}</th>
<th>Estimate\textsubscript{GLM}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-2.507</td>
<td>0.178</td>
<td>-14.07</td>
<td>-94.63</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>-2.359</td>
</tr>
<tr>
<td>ideol.dist</td>
<td>-1.169</td>
<td>0.035</td>
<td>-33.26</td>
<td>-43.03</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>-1.074</td>
</tr>
<tr>
<td>samestate</td>
<td>2.193</td>
<td>0.091</td>
<td>24.11</td>
<td>39.91</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>1.990</td>
</tr>
<tr>
<td>sameregion.notstate</td>
<td>0.272</td>
<td>0.031</td>
<td>8.72</td>
<td>14.40</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>0.301</td>
</tr>
<tr>
<td>poverty.diff*</td>
<td>-0.444</td>
<td>0.067</td>
<td>-0.73</td>
<td>-1.83</td>
<td>0.464</td>
<td>0.06772</td>
<td>-0.597</td>
</tr>
<tr>
<td>urban.diff*</td>
<td>-0.258</td>
<td>0.111</td>
<td>-2.32</td>
<td>3.51</td>
<td>0.205</td>
<td>0.00045</td>
<td>0.228</td>
</tr>
<tr>
<td>seniority.diff*</td>
<td>0.007</td>
<td>0.004</td>
<td>1.58</td>
<td>11.97</td>
<td>0.114</td>
<td>&lt; 2.2e-16</td>
<td>0.007</td>
</tr>
<tr>
<td>both.presbyterian</td>
<td>0.126</td>
<td>0.130</td>
<td>0.96</td>
<td>-0.42</td>
<td>0.333</td>
<td>0.67148</td>
<td>-0.040</td>
</tr>
<tr>
<td>both.catholic*</td>
<td>0.071</td>
<td>0.061</td>
<td>1.17</td>
<td>15.04</td>
<td>0.242</td>
<td>&lt; 2.2e-16</td>
<td>0.449</td>
</tr>
<tr>
<td>both.jewish**</td>
<td>0.204</td>
<td>0.111</td>
<td>1.83</td>
<td>3.71</td>
<td>0.067</td>
<td>0.00021</td>
<td>0.248</td>
</tr>
<tr>
<td>both.lawyer*</td>
<td>-0.037</td>
<td>0.046</td>
<td>-0.80</td>
<td>-3.86</td>
<td>0.424</td>
<td>0.00011</td>
<td>-0.074</td>
</tr>
<tr>
<td>both.veteran*</td>
<td>0.023</td>
<td>0.050</td>
<td>0.462</td>
<td>2.21</td>
<td>0.644</td>
<td>0.02717</td>
<td>0.055</td>
</tr>
<tr>
<td>both.female***</td>
<td>0.285</td>
<td>0.084</td>
<td>3.373</td>
<td>11.41</td>
<td>7.43e-4</td>
<td>&lt; 2.2e-16</td>
<td>0.515</td>
</tr>
<tr>
<td>office.samefloor**</td>
<td>-0.078</td>
<td>0.042</td>
<td>-1.845</td>
<td>-4.58</td>
<td>0.065</td>
<td>4.61e-06</td>
<td>-0.136</td>
</tr>
<tr>
<td>shared.cmte.ijk***</td>
<td>0.936</td>
<td>0.032</td>
<td>29.145</td>
<td>35.72</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>0.863</td>
</tr>
<tr>
<td>sharedcmtes.ij.(no k)***</td>
<td>0.075</td>
<td>0.017</td>
<td>4.456</td>
<td>6.39</td>
<td>8.37e-06</td>
<td>1.69e-10</td>
<td>0.078</td>
</tr>
</tbody>
</table>

*These appear significant under the fixed-effects-only GLM, but are clearly not significant under our model.

**These appear significant under GLM, borderline for ours.

***Significant in both, with more reliable estimates in our mixed effects model.
Table 4: Non-pooled data by party

<table>
<thead>
<tr>
<th>Effect</th>
<th>Est. $D \rightarrow D$</th>
<th>s.e. $D \rightarrow D$</th>
<th>Est. $R \rightarrow R$</th>
<th>s.e. $R \rightarrow R$</th>
<th>Est. $D \rightarrow R$</th>
<th>s.e. $D \rightarrow R$</th>
<th>Est. $R \rightarrow D$</th>
<th>s.e. $R \rightarrow D$</th>
</tr>
</thead>
<tbody>
<tr>
<td>dyad directed $ij$</td>
<td>$\hat{\sigma}_{D_0} = 0.0379$</td>
<td>$\hat{\sigma}_{R_0} = 0.2904$</td>
<td>$\hat{\sigma}_{D_0} = 0.1144$</td>
<td>$\hat{\sigma}_{R_0} = 0.2904$</td>
<td>$\hat{\sigma}_{D_0} = 0.3974$</td>
<td>$\hat{\sigma}_{R_0} = 0.2904$</td>
<td>$\hat{\sigma}_{D_0} = 0.2413$</td>
<td>$\hat{\sigma}_{R_0} = 0.2413$</td>
</tr>
<tr>
<td>dyad directed $ij$ -2.00</td>
<td>0.37</td>
<td>0.03</td>
<td>0.27</td>
<td>0.51</td>
<td>0.87</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
</tr>
<tr>
<td>dyad directed $ij$ -0.12</td>
<td>0.31</td>
<td>0.03</td>
<td>0.27</td>
<td>0.51</td>
<td>0.87</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
</tr>
<tr>
<td>dyad directed $ij$ 1.72</td>
<td>0.53</td>
<td>2.45</td>
<td>0.51</td>
<td>0.87</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
<td>0.37</td>
</tr>
<tr>
<td>ideol.dist</td>
<td>-3.00</td>
<td>0.37</td>
<td>0.03</td>
<td>0.27</td>
<td>0.51</td>
<td>0.87</td>
<td>0.20</td>
<td>0.82</td>
</tr>
<tr>
<td>sender $i$</td>
<td>-6.35</td>
<td>0.52</td>
<td>0.59</td>
<td>1.20</td>
<td>3.34</td>
<td>0.53</td>
<td>0.22</td>
<td>0.08</td>
</tr>
<tr>
<td>receiver $j$</td>
<td>-0.03</td>
<td>0.22</td>
<td>0.62</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
<td>0.37</td>
<td>0.20</td>
</tr>
<tr>
<td>ideol.dist</td>
<td>0.001</td>
<td>0.27</td>
<td>0.62</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
<td>0.37</td>
<td>0.20</td>
</tr>
<tr>
<td>shared.cmte.ijk</td>
<td>1.08</td>
<td>0.27</td>
<td>0.62</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
<td>0.37</td>
<td>0.20</td>
</tr>
<tr>
<td>shared.cmte.ijk(no k)</td>
<td>0.20</td>
<td>0.12</td>
<td>0.21</td>
<td>0.11</td>
<td>0.22</td>
<td>0.10</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>same.state</td>
<td>1.72</td>
<td>0.53</td>
<td>2.45</td>
<td>0.51</td>
<td>0.87</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
</tr>
<tr>
<td>same.state</td>
<td>1.72</td>
<td>0.53</td>
<td>2.45</td>
<td>0.51</td>
<td>0.87</td>
<td>0.20</td>
<td>0.82</td>
<td>0.56</td>
</tr>
</tbody>
</table>

35
<table>
<thead>
<tr>
<th>Effect</th>
<th>Est</th>
<th>s.e.</th>
<th>z</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>-2.50</td>
<td>0.083</td>
<td>-30.1</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>ideol.dist</td>
<td>-1.19</td>
<td>0.032</td>
<td>-36.9</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>samestate</td>
<td>2.23</td>
<td>0.074</td>
<td>30.1</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>sameregion.notstate</td>
<td>0.28</td>
<td>0.028</td>
<td>10.0</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>Chair.sender</td>
<td>0.10</td>
<td>0.105</td>
<td>0.9</td>
<td>0.35</td>
</tr>
<tr>
<td>RMM.receiver</td>
<td>-0.35</td>
<td>0.491</td>
<td>-7.2</td>
<td>6.9e-13</td>
</tr>
<tr>
<td>Chair-&gt;RMM</td>
<td>0.76</td>
<td>0.252</td>
<td>3.0</td>
<td>0.003</td>
</tr>
<tr>
<td>shared.cmte.ijk</td>
<td>0.91</td>
<td>0.030</td>
<td>30.3</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>sharedcmtes.ij(no k)</td>
<td>0.09</td>
<td>0.016</td>
<td>6.0</td>
<td>2.1e-09</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Effect</th>
<th>Est</th>
<th>s.e.</th>
<th>z</th>
<th>p-val</th>
</tr>
</thead>
<tbody>
<tr>
<td>intercept</td>
<td>-2.54</td>
<td>0.100</td>
<td>-25.3</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>ideol.dist</td>
<td>-1.18</td>
<td>0.032</td>
<td>-36.7</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>samestate</td>
<td>2.22</td>
<td>0.074</td>
<td>29.7</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>sameregion.notstate</td>
<td>0.27</td>
<td>0.029</td>
<td>10.0</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>RMM.sender</td>
<td>0.10</td>
<td>0.030</td>
<td>30.3</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>Chair.receiver</td>
<td>-0.52</td>
<td>0.0401</td>
<td>-7.2</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>RMM-&gt;Chair</td>
<td>1.90</td>
<td>0.199</td>
<td>9.6</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>shared.cmte.ijk</td>
<td>0.89</td>
<td>0.030</td>
<td>29.8</td>
<td>&lt; 2e-16</td>
</tr>
<tr>
<td>sharedcmtes.ij(no k)</td>
<td>0.09</td>
<td>0.015</td>
<td>5.8</td>
<td>7e-09</td>
</tr>
</tbody>
</table>
[[Focus on most interesting results:]]

1. As predicted, large effect for shared committees (not shown), but when we decompose this into effect of sender, receiver, and bill all being in the same committee, and number of shared committees shared where the bill is not considered, the residual effect is much weaker (but still clearly significant.

2. Sender and receiver from same state is major predictor (more than party, and same region is significant echo of this.

3. RMM->Chair is perhaps the best single predictor. [[Calculate changes in logodds...]]

4. ***If either RMM or Chair of the committee with jurisdiction over the bill is sponsor, the number of cosponsors will tend to be lower (suggesting that extra signatures become unnecessary if a bill has the most important ones. Good lead in to future research on network-based game theory–how is cosponsorship used to send signals to leaders?]]

7 Conclusions

[[Still to complete]]:

Future work:

1. Adapt Hoff’s bilinear (inner product term) based on position in latent space to get at higher-order dependence (transitivity, balance, clustering) To do so, I need to extend his model to include heterogeneous contexts (bills and committees)

2. Simulation using estimates of fixed and random effect parameters. See whether structure is similar to reality.

3. look at actual clustering patterns of “co”-cosponsors, i.e. those who appear together frequently as cosponsors on bills sponsored by someone else. The sponsors become a sort of interaction effect–some grouping tends to co-cosponsor in the presence of j as sponsor or when bill passes through committee f.
4. bill lifespan from birth to death (passage, nay vote, or lack of action by term's end) as a stochastic (censored) process, where cosponsorship takes place as discrete-time event in continuous-time calendar, and potential cosponsors may know the identities of previous cosponsors (allowing bandwagon effect).

5. Game theoretic network evolution, including development of reciprocity patterns over time (within and across sessions of Congress), appearance and disappearance of actors through elections and retirement (not to mention indictment!)

6. Change of majority party, differing dynamics as majority grows

References


