Considering Network Effects in the Design and Analysis of Field Experiments on State Legislatures

February 8, 2018

Abstract

Recent work on legislative politics has documented complex patterns of interaction and collaboration through the lens of network analysis. In a largely separate vein of research, the field experiment—focused largely on state legislatures—has emerged as an important approach in establishing causal identification in the study of legislative politics. The stable unit treatment value assumption (SUTVA)—the assumption that a unit's outcome is unaffected by other units' treatment statuses—is required in conventional approaches to causal inference with experiments. When SUTVA is violated, a condition termed interference, as in networked social interaction, treatment effects spread to control units through the network structure. We review recently developed methods that can be used to account for interference in the analysis of data from field experiments on state legislatures. The methods we review require the researcher to specify a spillover model, according to which legislators influence each other, and specify the network through which spillover occurs. We discuss these and other specification steps in detail. We find evidence for spillover effects in data from two previously published field experiments. These replication analyses illustrate how researchers can use recently developed methods to test for interference effects, and support the case for considering interference effects in experiments on state legislatures.

1 Introduction

Two recent streams of innovative research in legislative politics include the study of legislative networks and field experiments on legislatures—-state legislatures in particular.

These two emerging approaches have evolved largely separate from one another, but we argue that they should be integrated due to the interdependence that arises between legislators based on processes such as cue-taking. In a study of cue-taking in roll call votes in the California Assembly, Masket et al. (2008) aptly summarizes the importance of understanding sources of interdependence between legislators in accounting for legislative outcomes. Masket (p. 302) notes that,

> "[..] there is a great deal of cue-taking in a legislature. Members defer in their judgment to trusted colleagues with expertise in particular issue areas."

Masket et al. (2008) finds that a connection as informal as two legislators being desk mates in the legislative chamber increases the rate at which two legislators vote in agreement. Legislative networks research, which has grown significantly in recent years, has documented complex forms of interconnectedness that can be observed in patterns of cosponsorship (Kirkland 2013, 2011; Fowler 2006), shared campaign staff (Nyhan and Montgomery 2015), collaborative press events (Desmarais, Moscardelli, Schaffner and Kowal 2015), and caucus co-membership (Victor and Ringe 2009). Any of these networks, and other forms of connections discussed below, could serve as conduits of interdependence between legislators. What the legislative networks literature has been lacking is an approach to research design that its causally valid. Legislative networks literature provides theoretical justification for testing for interdependence, but the extent of interdependence between legislators is still an open question due to the challenges in identifying influence in networks with observational data (Shalizi and Thomas 2011).

Field experiments on state legislatures have emerged as a standard approach to causally valid research design in the study of legislators. Bergan (2009, p. 331) notes the value of experimentation for exactly this case, "Random assignment of legislators to treatment and control can eliminate the potential bias that results from groups strategically choosing whom to lobby." Field experiments have explored the relationship between constituency opinion and roll call voting (Butler, Nickerson et al. 2011), racial conditioning in legislator communications (Broockman 2013), and the effects of lobbying on roll call voting (Bergan and Cole 2015).

Despite the separate insights offered by legislative networks scholarship and legislative field experiments, there is a degree of incompatibility in the assumptions underlying approaches in these two literatures. The interdependence between actors that represents a central concept in legislative networks research poses a challenge to the use of field experiments to identify causal effects. Network-based interdependence (i.e., influence, contagion) violates the stable unit treatment value assumption (SUTVA)— the assumption that a unit's outcome is unaffected by other units' treatment statuses. SUTVA is a bedrock assumption in the conventional approach to causal identification via randomized experiments (Sekhon 2008). If we take recent research on the role of networks in legislative decision-making seriously, simple randomization to treatment is likely not a robust method, as networked interdependence between legislators poses a high likelihood of interference. As Sekhon (2008, p. 5) notes, "When SUTVA is violated, an experiment will not yield unbiased estimates of the causal effect of interest."

Virtually all research on legislative networks is based on observational data, lacking in design-based causal identification strategies (see Rogowski and Sinclair (2012) for an exception). Due to the interconnectedness of actors, observational research on social networks presents myriad confounding problems, that place considerable limits on the feasibility of causal identification (Shalizi and Thomas 2011). As such, confronting interference in legislative field experiments presents two related research opportunities. First, accounting for interference is a vital step in producing unbiased estimates of treatment effects in the presence of SUTVA violations. Second, studying interference in field experiments on legislators represents an approach to studying networked interdependence in legislatures with a more credible identification strategy than that which is attainable in observational research. A growing body of research seeks to study interference through experimental interventions on networks (e.g., Gerber, Green and Larimer 2008; Paluck 2011; Bond, Fariss, Jones, Kramer, Marlow, Settle and Fowler 2012; Muchnik, Aral and Taylor 2013; Aral and Walker 2014; Bapna and Umyarov 2015; ben Aaron, Denny, Desmarais and Wallach Accepted). These studies follow a variety of approaches to designing the interventions and testing for interference effects. However, it is clear that the field has not, as of yet, converged upon a consistent methodological framework for testing for causal effects in the presence of interference. In this paper we review and illustrate a recently developed method that can be used to test for both direct and interference effects in experiments. This methodology, developed by Bowers, Fredrickson and Panagopoulos (2012), allows the researcher to test for causal effects in experiments while relaxing SUTVA. Beyond the review of this methodology, we offer three contributions in this paper. First, we provide a typology of theoretical considerations that researchers can draw upon when formulating hypotheses regarding interference. Second, we provide a focused review of the networks through which scholars of legislative politics should consider in specifying tests for interference. Third, we apply this methodology by analyzing data from past studies that involved field experiments on state legislatures.

2 A Design-Based Test for Network Effects Models

In this section, we review the methodology introduced by Bowers, Fredrickson and Panagopoulos (2012), which enables the researcher to test for both direct and interference effects, represented by models of effects. The model of effects represents how the vector of treatments allocated to subjects in the experiment effects the outcome under study. In a conventional experimental setting, in which SUTVA is assumed, the model of effects is simply that subject i's outcome depends upon subject i's treatment status, but not the treatment status of any other subject. The model of effects tested with the methodology proposed by Bowers, Fredrickson and Panagopoulos (2012) can include separate parameters for direct causal effects of the treatment and spillover effects that depends on how treatments are allocated across subjects situated in a network.

The testing framework proposed by Bowers, Fredrickson and Panagopoulos (2012) is a randomization test (Basu 2011) in which the hypotheses are defined with respect to the model of effects specified by the researcher. The null hypothesis in the test is the sharp null of no effects—the hypothesis that the outcomes observed in the experiment are what they would have been if every subject were in the control group (i.e., if treatment had not been allocated). The test is defined through selecting (1) a model of effects, and (2) a test statistic to be used in comparing the outcomes in the experiment to what would have been expected under the sharp null of no effects. We discuss these choices in greater detail below.

Once the parameters of the test are defined, it proceeds via randomization inference. Random permutations of the treatment vector are used to construct the sampling distribution of the test statistic under the sharp null. In each permutation, a new treatment assignment is drawn from the randomization distribution used in the experiment. Based on the re-randomized treatment vector, the hypothesized parameters and model of effects are used to remove the effect of the treatment on all of the subjects in the experiment. Bowers, Fredrickson and Panagopoulos (2012) refer to the outcome vector derived by removing the hypothesized effect of the experiment as the uniformity trial. The test statistic is then calculated to assess the differences across experimental conditions.

A p-value for testing the hypothesized parameter values in the model of effects is calculated as the proportion of test statistics under permutation that exceed the observed test statistic value (i.e., the test statistic value when evaluated on the uniformity trial given the observed treatment vector). Given the setup of the test, higher p-values indicate greater support for the hypothesized model and parameter values. The intuition for this reversal in the conventional direction of the p-value is that the correct hypothesis will be more effective than any other hypothesized model at removing the differences across experimental categories.

3 Considerations in Testing for Interference

In this section, we offer a novel set of recommendations regarding theoretical considerations that can be drawn upon by researchers when they design experiments in which they plan to test for interference and/or specify tests to be conducted on data from field experiments that have already been conducted on legislatures. One of the virtues of controlled experiments, in which treatment allocation is randomized, is that the randomization design can be used as the basis for inference in statistical tests (i.e., design-based or randomization-based inference) (Little and Rubin 2000). Testing using the Bowers et al. framework still relies on design-based inference, as the stochastic nature of the outcomes is assumed to arise from the distribution based on which the treatment was randomized. However, the hypothesis being tested is formulated as a model of causal direct and spillover effects. As these models of effects are more complicated than the conventional form of effects considered in experiments, researchers must put more thought into the functional forms that describe the relationship between the treatment and outcome vectors. It is not possible to enumerate all of the choices available in specifying the model of effects, but we discuss a few salient dimensions below.

3.1 Network selection

The methodology introduced by Bowers, Fredrickson and Panagopoulos (2012) is applicable in any domain of experimental political science research in which interference is suspected, and the networks through which interference might occur can be measured. There are two features of legislative politics that render methodology for testing interference particularly useful. First, since legislatures operate according to explicitly majoritarian reward systems, and it is feasible for any legislator to bargain with his or her colleagues to achieve a legislative goal, legislators face particularly strong incentives to influence each other (Matthews 1959; Ferejohn 1986; Bernhard and Sulkin 2013). Second, we have an active literature on legislative networks that offers several options to consider when testing for interference effects (Kirkland and Gross 2014; Desmarais et al. 2015). Example legislative networks that have been studied include similarity in roll call voting (Kim and Barnett 2012), bill cosponsorship (Fowler 2006), overlapping committee membership (Porter, Mucha, Newman and Warmbrand 2005), collaboration in press events (Desmarais et al. 2015), co-membership in caucuses (Victor and Ringe 2009), the proximity of members of Congress' DC offices (Rogowski and Sinclair 2012), follower-followee connections among members of Congress on Twitter (Peng, Liu, Wu and Liu 2016), the similarity of campaign contributions received by candidates for state legislature (Masket and Shor 2015), a survey to measure collaboration and social networks among members of the Brazilian national legislature (Wojcik 2017), demographic similarity between legislators' constituencies (Bratton and Rouse 2011), and connections between legislative staffers (Ringe, Victor and Gross 2013). In Table **??**, we list the different networks that researchers might consider when investigating interference in legislative networks. This list is drawn directly from the literature. Given a set of prospective networks, such as these, researchers must consider which single network, or combination of networks, through which spillover will occur.

Networks in Legislative Politics
Roll call voting similarity
Bill cosponsorship
Overlapping committee membership
Collaboration in press events
Ideal point similarity
Co-membership in legislative caucuses
Legislative staff sharing
Spatial proximity of legislative offices
Relationships in online social networks (e.g., Twitter)
Similarity in legislators' campaign contributions
Social network surveys administered to legislators
Similarity in constituency demographics

Table 1: List of legislative networks drawn from past research.

The determination regarding which network(s) to consider in any particular application is, of course, best made by the researchers carrying out the application. Selecting which network(s) to test is much like selecting which variables to include when specifying a model—researchers should use a combination of theory and exploration, being careful to adjust for multiple testing bias in hypothesis tests (e.g., via Bonferroni correction (Napierala 2012)). We discuss two dimensions of interference dynamics—exposure and uptake—that should help to inform this determination. Exposure refers to the degree to which the network governs legislators' awareness regarding each others' beliefs or behaviors. Uptake refers to the role of the network in determining which legislators' would adopt each others' beliefs or behaviors if exposed to them. Consider a legislator's position on a major policy issue. It is likely that each legislator in a chamber is aware of each other legislator's opinion on a major issue, so the network does not need to play a major role in exposure to govern interference. However, in order to influence each other on a major policy issue, legislators may need to see each other as closely aligned ideologically. For interference dynamics that do not require exposure through the network, but require uptake, researchers should look for networks that signal ideological similarity such as co-voting on bills. On the other hand, some interference dynamics for which uptake might be highly likely, such as re-use of issue framing in legislators' public statements (Lin, Margolin and Lazer 2016), or the adoption of strategies in responding to constituent requests (**?**), would require legislators to be exposed to each other through explicit communication channels. In applications where the network needs to play an important role in signaling exposure, networks such as twitter follower networks and caucus co-membership may be more appropriate. We can also think of networks that would signal both ideological alignment and explicit communication ties, such as co-participation in press events and frequent bill co-sponsorship (especially early-stage, or original cosponsorships). Note that there are two categories of processes through which interference can occur—spread of the treatment through a network (e.g., an influential lobbying communication is sent to a legislator, and that legislator forwards the communication to others in their network) and spillover of effects (e.g., a lobbying communication influences a legislator's vote, and others in that legislator's network take cues from their vote). A useful thought experiment in selecting networks to use in tests of interference would be to consider which networks would facilitate the spread of treatments, and which networks would facilitate the spillover of effects.

3.2 Interference Model Formulation

The interference model is a function that takes as its input a treatment regime (i.e., a vector that indicates the control/treatment status of each node (e.g., legislator) in the network), a network structure, and the outcomes under the uniformity trial (i.e., the outcome values in the case where each node is assigned to control), and outputs a vector of node outcomes that are conditioned on the treatment regime via the network. In other words, the interference model transforms the uniformity trial into a vector of outcomes using the network and treatment regime. For a given focal node the two components of the model that shape the change that results from the experiment include (1) the set of other nodes whose treatment status could influence the focal node via the network, and (2) the mathematical form of the function through which those other nodes' treatment statuses affect the focal node. Given these two components, it is possible to calculate how any given treatment regime would affect a focal node's outcome. We discuss two important considerations in formulating the interference model to be tested. First, we discuss the specification of the neighborhood, as defined on the network structure, of nodes whose treatment status may affect a focal node (e.g., a node's outcome depends on the treatment statuses of all nodes that are at most two hops away). Second, we discuss the specification of the functional form through which neighbors affect a focal node (e.g., the outcome of a node is a linear function of the proportion of neighbors allocated to treatment).

In Table 2 we illustrate how varying the interference model can result in different effects on a focal node. We depict two definitions of the neighborhood—one in which all nodes within two hops of the focal node affect the focal node, and one in which all nodes within three hops of a focal node affect the focal node. We also depict two definitions of the functional form of the interference effects. In one definition, all nodes in the neighborhood affect the node equally. In the other functional form, the effect of neighbors on the focal node decays with the neighbors' distance from the focal node. Combining these two dimensions results in four alternative interference models.

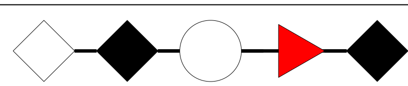
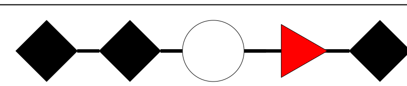
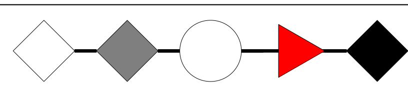
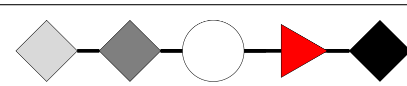
	Two-hop Neighborhood	Three-hop Neighborhood
Constant effect		
Decaying effect		

Table 2: Alternative models of affects, focusing on a single focal node. The red triangle represents the focal node, on which the other nodes have various effects under each model. Square nodes are treated. The circle is a control node. The darker the node's shading, the larger the effect it has on the focal node.

3.2.1 Neighborhood selection

Once the researcher decides which network—or combination of networks—to use in analysis, it is important to determine the neighborhood within which the effects of the treatment can be transmitted. For example, Bond et al. (2012) find that Facebook users' voter turnout, as expressed on their Facebook walls, influences not only their Facebook friends' turnout decisions, but also turnout of the friends of their friends. This means that the effects of a Facebook user's turnout decision spread within a neighborhood of two hops through the friendship network. This specification decision becomes more complicated when the network is weighted (i.e., ties can take on many values rather than just being binary tie/no tie), as in the legislative networks that we consider in our applications. In the weighted network case transmission is likely a function of connection strength, but may also disappear at some threshold (e.g., the level of ideological distance that indicates opposition between two legislators). In our consideration of state legislative networks, we specify the neighborhood in two ways when using the ideological similarity networks:

- Entire network: Treatment effect can propagate through the entire network—proportional to ideological similarity—to affect the outcome of control units.
- K-nearest neighbors: Treatment effects can spread to control units from their K nearest neighbors, varying the value of K.

The definition of neighborhood depends on substantive knowledge about the interaction in a certain network. For example, a state legislature is a relatively small and internally familiar community. As such, everyone may potentially communicate with everyone else regarding major legislative tasks and actions. However in looking at interpersonal political communication networks among regular citizens, even the closest of friends may fail to communicate about an election or other major political event.

3.2.2 Interference effect specification

: The above two specifications—selecting the network and the neighborhood—determine which units play a role in the interference reflected in the hypothesized model. Diffusion model specification involves defining how the treatment effect spreads through the network. We highlight two considerations—the way in which treated and untreated neighbors factor into the interference effects, and the linearity of the interference model.

The first consideration regards whether a control unit influenced by the number of treated units with which it interacts (e.g., as in an epidemic network), or by the balance or proportion of its neighbors that are treated (e.g., as we would assume in a voting or opinion-spreading network). Bowers, Fredrickson and Panagopoulos (2012) specification assumes treatment spreads as a function of the number of treated neighbors. Alternatively, the Voter Model—a classic mode of opinion dynamics in networks—assumes that the proportion of treated neighbors is the relevant quantity (Valentini, Hamann and Dorigo 2014). This specification choice likely comes down to whether the researcher assumes that the treatment and the lack of it are equally powerful forces, or whether change in the outcome can only result from exposure to treated units. If untreated neighbors can offset the effects of treated neighbors, it is likely the proportion that matters. If units are influenced only by treated neighbors, it is likely the raw count of treated neighbors that is relevant.

Though a very familiar consideration in quantitative social science, functional form assumptions are also relevant in the specification of a model of network effects. It is important to determine whether the functional form of the propagation of treatment effect should be linear or non-linear. Does the second treated neighbor have the same effect on a node's outcome as the first treated neighbor, or does the effect diminish? Or, alternatively, is it a threshold effect that only manifests when the number of treated neighbors reaches a critical level (e.g., a model in which a unit adopts the majority opinion among its neighbors)? Coppock (2014) adopts a linear functional form in specifying the way in which legislators learning about their districts' opinions effects the votes of ideologically similar legislators. Alternatively, the classic susceptible-infected-recovered (SIR) model in epidemiology assumes a model in which the probability of transmission increases at a decreasing rate with the number of exposed neighbors to which a unit is exposed (Dodds and Watts 2004).

4 Replication Analyses: Testing for Network Effects

To illustrate testing for effects via network models of effects, we re-analyze results from two field experiments on state legislatures. Our application builds directly on Coppock (2014). Since it is generally infeasible to recruit legislators for lab experiments, field experiments represent the best option for design-based causal identification of effects in research on legislative behavior. The literature offers many recent examples of field experiments in legislative studies (e.g., Bergan 2009; Butler and Broockman 2011; Butler, Karpowitz and Pope 2012; Broockman 2013; Nyhan and Reifler 2015; Bergan and Cole 2015). In these experiments, the researcher introduces a manipulation (e.g., a communication from a constituent, or information about constituent preferences), and then observes legislators' behavior in terms of casting roll call votes or reacting to the communication on an individual basis. Since legislators regularly communicate and collaborate, it is highly possible that SUTVA is violated in a legislative field experiment.

In each of the replications and extensions that follow, we test causal models that include network effects. In order to test these models, we must specify their functional forms and select the data to use in measuring the network. For each replication, we consider multiple definitions of both the network and the neighborhood through which network effects are transmitted, as we do not have strong prior expectations regarding exactly which network or neighborhood should be included in the models of effects.

We make specification choices in terms of both linearity of the model of effects and the effects of control units that are based in theoretical considerations. First, in terms of the linear functional form of the model of effects, we stick with a linear model due to (1) the relatively small datasets with which we are working, and (2) the dichotomous nature of the outcome variables in each experiment. Since the datasets are small there is a significant degrees-of-freedom cost in adding additional parameters to the model of effects, and making the model nonlinear would require adding a parameter that controlled the shape of the curve. The fact that the outcome variable is dichotomous in each experiment also limits the information—in terms of variability—that could be used to identify the functional form of the model of effects. In terms of of the effects of controls (i.e., number vs proportion of treated neighbors), we assume that the number of treated neighbors is the relevant quantity in each application. In each of the experiments we replicate, treated legislators are provided with a form of communication that could, in theory, be passed along to other legislators. Further, the likelihood that this information will be passed along should increase with the number of treated legislators in a legislator's neighborhood. Lastly, in the replications we consider the treatments include informational communications and requests for action, but control units are not provided with either information or requests that counteract that provided to treated units. As such, we do not expect to see any balancing effect between treated and control units.

We present each replication in a separate section below. For each analysis, we visualize the plots of p-values from the test of the model of effects at the specified parameter values. Using the framework of Bowers, Fredrickson and Panagopoulos (2012), the point estimate is the estimate that results in the maximal p-value. We provide tables of point estimates and confidence intervals. The 95% confidence interval is given by the region that includes p-values greater than 0.05.

4.1 Butler, Nickerson et al. (2011)

Butler and Nickerson conducted an experiment on New Mexico legislators to study the effect on legislators of learning public opinion from their constituencies. In 2008, a special session of the New Mexico State Legislature was called to vote on a bill regarding proposed spending plans for a budget surplus—a tax rebate. Butler and Nickerson conducted a large-scale phone survey to gather constituent opinions from across the state. Using matched-pair randomization—matching in terms of political party, 35 out of 70 legislators were assigned to the treatment group. Legislators in the treatment group were sent a letter containing the district-specific support for the proposed spending plan in their own districts. The original paper conducts a direct comparison of outcomes across treatment and control group. They find that the effects of the treatment on legislators' votes were conditioned by the level of support for the measure indicated in the treatment message. In districts with high support for the tax rebate, the treatment had little effect. This is because legislators generally assumed that the tax rebate would be popular, and that constituents would support the measure. In districts with low support for the tax rebate, the treatment had a negative effect on the likelihood of voting in support for the measure.

Coppock (2014) applied the Bowers, Fredrickson and Panagopoulos (2012) methodology to test for propagation of treatment in this experiment. The indirect effect estimates were not statistically significant (Coppock 2016), even when separating the sample into low and high support districts. As detailed in Section 3, Coppock used a network based on ideological similarity, where the legislator's outcome is affected by every other legislator in the network, but with varying weight based on ideological similarity. Coppock used a linear model to represent the direct and indirect effects of the treatment on the outcome. We begin by replicating this analysis. **SAYALI, NEW DESIGN TO DESCRIBE HERE** In the extension, we consider two types of networks— one based on ideological similarity and the other on serving on committees together. We also vary the neighborhood specification to consider an effect from all other legislators—proportional to ideological distance or effect only from k-nearest neighbors. Results of this analysis are presented in section 6.1. The following list summarizes the steps taken to implement Coppock's analysis of interference in the Butler, Nickerson et al. (2011) replication. Note that we re-use these steps in the other replication analyses where the similarity scores are replaced with the respective definition of the network and neighborhood.

1. Calculate W-NOMINATE ideology score for each legislator using roll call vote data

2. Calculate ideological similarity as $Similarity_{i,j} = \frac{2-|ideo_i - ideo_j|}{2}$

3. Calculate raw exposure as $Raw\ exposure_i = \sum_{j=1}^{n} Similarity_{i,j} * z_j,\ j \neq i.$

4. Coppock introduces an adjustment for the expected exposures of legislators. Exposures are simulated under a large number of randomizations. Each randomization where legislator i is in treatment is indexed as k (k = 1, 2, ..., K) and where legislator i is in control is indexed as l (l = 1, 2, ..., L)

$$Expected\ exposure_{i,z_i=1} = \frac{\sum_{k=1}^{K}\sum_{j=1}^{n} Similarity_{i,j} * z_{j,k}}{K},\ j \neq i,\ z_{i,k} = 0$$

$$Expected\ exposure_{i,z_i=0} = \frac{\sum_{l=1}^{L}\sum_{j=1}^{n} Similarity_{i,j} * z_{j,l}}{L},\ j \neq i,\ z_{i,l} = 1$$

5. Using the hypothesized parameter values, remove the model of effects based on a linear regression.

6. Regress hypothesized uniformity trials on direct and indirect treatment, and use the residual sum of squares (RSS) as test statistic. SeeBowers, Fredrickson and Aronow (2016) for discussion of why this is an appropriate test statistic for this framework.

7. The p-value for each hypothesized treatment effect is the proportion of simulated RSS that is lower than the observed RSS (note that smaller RSS indicates more variance explained).

One last detail of implementing the Bowers et al. hypothesis testing framework regards the grid of hypothesized parameters over which p-values are calculated. Since the testing process does not involve an optimization routine, there is no way for the parameter values to be selected automatically. However, standard optimization methods can be used to approximate point estimates around which to expand the grid of hypothesized values. In our applications, the model of effects has a linear form, and we can use linear regression to find the estimates around which to expand the grid. In terms of how far to expand the grid—-there should be enough grid points that none of the confidence interval boundaries lies at a boundary of the grid of hypothesized parameters.

4.1.1 Results for Butler, Nickerson et al. (2011) data

The results of our analyses are presented in the form of heat maps displaying the full range of parameter values considered, and tables in which we present the point estimates and 95% confidence intervals for the estimates. The dependent variable in this application is coded as nay=0, yea=1.The p-value plot for replication of the main analysis in Coppock (2014) is in Figure 1. The p-value is highest (0.997) when the direct effect is -0.25 and indirect effect is -0.15. Negative effects indicate that the treatment reduced the likelihood of voting in favor for those who received the treatment directly as well as those to whom it propagated, due to ideological similarity. The negative effects discovered in this experiment may be attributable to the assumed popularity of the bill. A treatment survey that indicated a high level support had no effect on legislators who were already planning to vote for the bill. However, low or moderate support on a treatment survey may have changed the minds of those legislators who were planning to support the bill. Confidence intervals are drawn using dashed lines. Neither the indirect nor the direct effects are statistically significant at the 0.05 level since each confidence interval contains zero. This is not a surprising result, as the direct effects reported in the original paper were not statistically

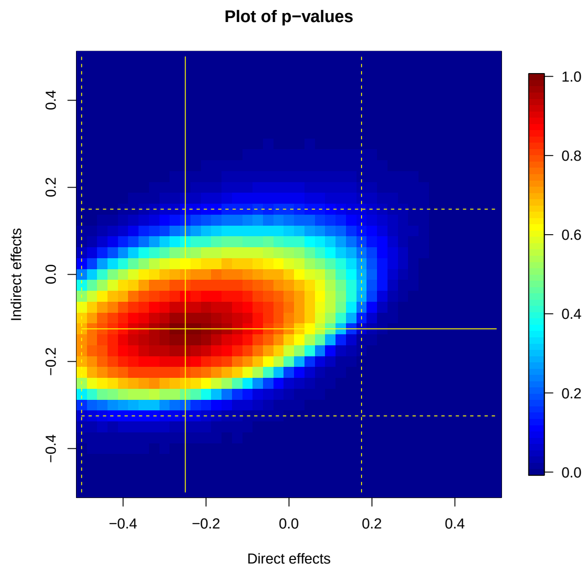


Figure 1: Main analysis for Butler, Nickerson et al. (2011) data

significant at the 0.05 level.

The first extension we consider is a change in the neighborhood specification. Instead of looking at ideological similarity across the network, we consider only the nearest k neighbors at values k = 3,5,8,12. For this network, ideological similarity is replaced with a 1 if j is one of i's k nearest neighbors, and 0 otherwise. In Figure 2, we see that the direct and indirect effects which maximize the p-value, are lower in magnitude than those in the first specification. When we model interference between all legislators in the network, we observe higher spillover than when looking at only nearest neighbors. We see that the observed indirect effect is higher when considering a a broader neighborhood (the entire network being the broadest neighborhood we can consider). Interestingly, we only see a result that is statistically significant, based on the 95% confidence interval, when looking at the indirect effect with a neighborhood defined as the twelve nearest neighbors. Overall, our finding of negative spillover through the ideological network is robust to neighborhood specification, but the estimates are not generally statistically significant.

In the next extension we change the network to look at co-committee membership as defining the ties through which interference occurs.[1] An undirected edge exists between legislators who served on standing committees together. We define the network at two thresholds—serving on at least one committee together and serving on at least two committees together. Results indicate that the committee network does not carry the effect of treatment to control units, as the point estimate is zero. The point estimate for the direct effect is still negative with the committee network.

Table 3: Results from Coppock (2014) data

Model	Direct effect		Indirect effect	
	Estimate	95% CI	Estimate	95% CI
Ideology: full network	-0.25	(-0.5, 0.175)	-0.15	(-0.325, 0.15)
Ideology: 3nn	-0.175	(-0.35, 0.025)	-0.075	(-0.175, 0.025)
Ideology: 5nn	-0.175	(-0.4, 0.05)	-0.075	(-0.175, 0.05)
Ideology: 8nn	-0.175	(-0.45, 0.10)	-0.10	(-0.275, 0.05)
Ideology: 12nn	-0.15	(-0.4, 0.1)	-0.125	(-0.3, 0)
Committee: >0	-0.15	(-0.325, 0)	0	(-0.125, 0.075)
Committee: >1	-0.15	(-0.375, 0.05)	0	(-0.1, 0.1)

4.2 Bergan and Cole (2015)

The second dataset we work with comes from an experiment on the Michigan state legislators. This experiment was conducted on legislators from both houses, in the con-

[1]Records of standing committee membership in the 16 standing committees in place during the 2008 regular session was obtained by email correspondence with the New Mexico Legislative Council Service Librarian.

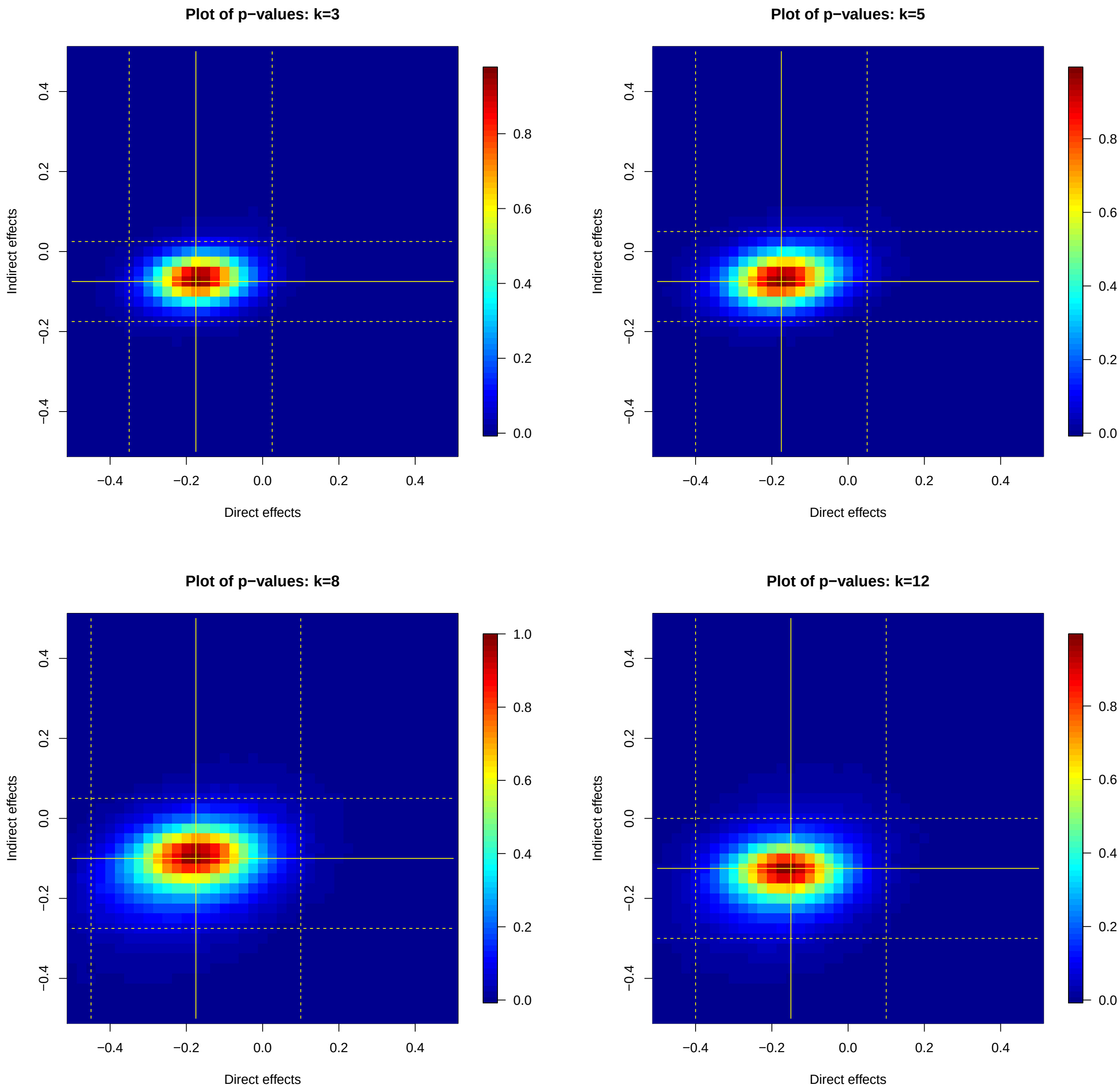


Figure 2: k-nearest ideological neighbors for Butler, Nickerson et al. (2011) data

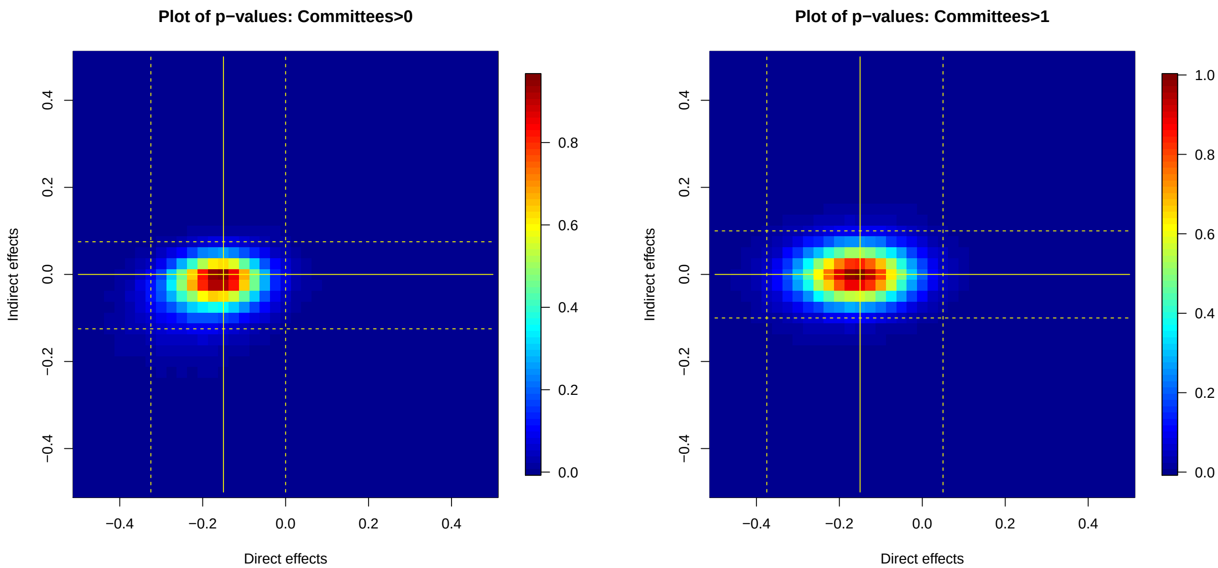


Figure 3: Committee network for Butler, Nickerson et al. (2011) data

text of anti-bullying legislation. Legislators were stratified based on various background variables. The treatments were calls from constituents expressing their support for the proposed bill. Treatment was given in three different doses, which differed in the number of calls places to the given legislator. Once again, the authors conducted an analysis under SUTVA and concluded that this treatment has a significant effect on the final vote on the bill. They observed a 12 percentage point increase in the likelihood of voting in favor of the anti-bullying bill, for those treated. Table 2 summarizes results of the Bergan and Cole (2015).

This data has not been analyzed for indirect effects previously. However, for all the reasons that we would expect to see interference in the Butler, Nickerson et al. (2011) results, we would expect to see them in the Bergan and Cole (2015) results. We conduct an analysis that is analogous to that in Coppock (2014), where a network is constructed

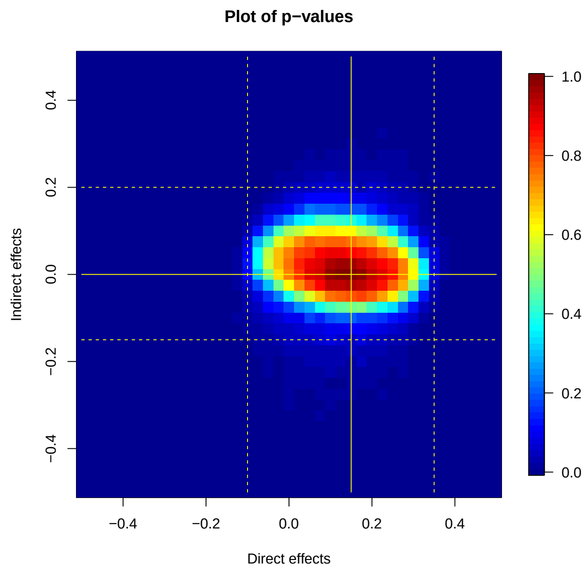


Figure 4: p-values: main analysis for Bergan and Cole (2015) data

based on ideological scores of legislators, using roll call data. We do not find evidence of indirect effects via this network. It is possible that we can attribute this to the nature of the bill. Voting behavior on an anti-bullying bill may not be governed by ideological coalitions. Figure 4 shows the plot of p-values for this analysis.

Figure 5 depicts results of analyzing the Bergan and Cole (2015) data under the ideological network and considers k nearest neighbors (k = 3,5,8,12) based on ideological similarity to constitute the neighborhood. We see that the results regarding indirect effects do not change and the estimate is still zero, indicating no interference effect. Since there is no change in the interference effect estimate, we see no change in the direct effect estimate, which indicates a 15 percentage point increase in the likelihood of voting in favor of the bill in response to being assigned to treatment.

We again consider an extension in which we change the network. In this network, an

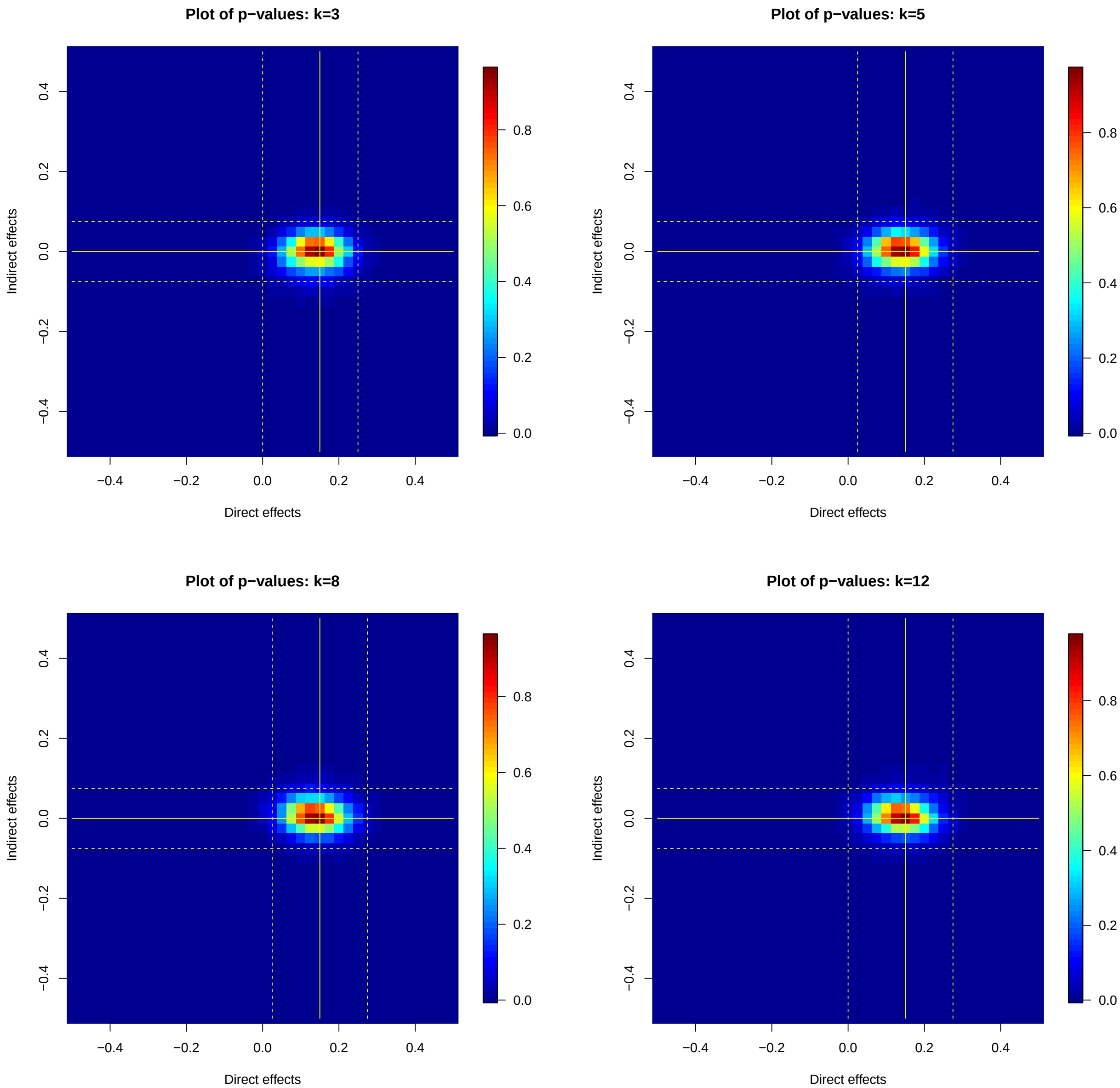


Figure 5: p-values: k-nearest ideological neighbors for Bergan and Cole (2015) data

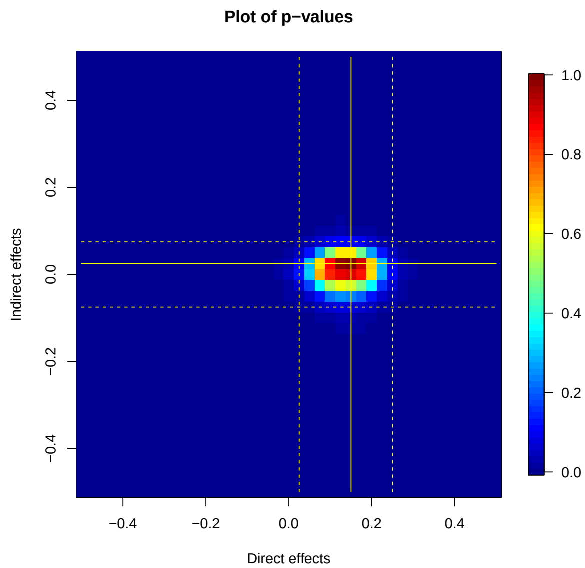


Figure 6: p-values: Cosponsorship network for Bergan and Cole (2015) data

undirected edge represents the number of bills cosponsored by a pair of legislators. The indirect effect estimate with this network is positive. The positive indirect effect indicates that as exposure through cosponsorship neighbors goes up, the likelihood of a legislator voting in favor of the anti-bullying bill goes up. However, the confidence interval indicates that this effect is not statistically significant.

5 Conclusion

In this paper we have advocated that scholars who run field experiments on state legislatures consider testing for interference. We provide guidance in specifying these tests using the methods developed by Bowers, Fredrickson and Panagopoulos (2012). Specifically, we discuss options for specifying the network(s) through which interference occurs,

Table 4: Results from Bergan and Cole (2015) data

Model	Direct effect		Indirect effect	
	Estimate	95% CI	Estimate	95% CI
Ideology: full network	0.15	(-0.1, 0.35)	0	(-0.15, 0.2)
Ideology: 3nn	0.15	(0, 0.25)	0	(-0.075, 0.075)
Ideology: 5nn	0.15	(0.025, 0.75)	0	(-0.075, 0.075)
Ideology: 8nn	0.15	(0.025, 0.75)	0	(-0.075, 0.075)
Ideology: 12nn	0.15	(0, 0.275)	0	(-0.075, 0.075)
Cosponsorship	0.15	(0.025, 0.25)	0.025	(-0.075, 0.075)

selecting the neighborhood of legislators who affect the legislator through the network(s), and specifying the functional form according to which the interference effects manifest. We illustrate this approach with two in-depth replications. We do not find universal evidence for interference effects in our replications. Our mixed findings regarding interference effects are attributable to an actual lack of interference in some contexts, a misspecification of the model of effects (which could include using the wrong network(s)), or some combination of both. Nonetheless, these replications serve to illustrate the variety of choices researchers have to make when testing for interference effects in experiments on state legislatures.

[UPDATE THIS PARAGRAPH WITH RESULTS] The results from our replication of field experiments on legislatures underscore the importance and complexity of accounting for interference. The replications and extensions of Coppock (2014) and Broockman (2013) demonstrate modest evidence of interference effects, and the inferential consequences of choices in specifying both the network and the neighborhood through which treatment is hypothesized to propagate. We did not see evidence for spillover effects in any of the specifications for Bergan and Cole (2015). Our replication study is not intended to provide definitive evidence regarding whether or not state legislative field experiments are subject to interference effects. Rather, we illustrate a broad array of network and neighborhood definitions, and provide evidence that some experiments are characterized by interference effects, and some are not. Given that tools are now available for testing interference effects, researchers have little reason to assume SUTVA in legislative field experiments. Indeed, relaxing SUTVA, and using the methods introduced by Bowers, Fredrickson and Panagopoulos (2012), enables the researcher to explore myriad hypotheses regarding the presence and structure of interdependence in the legislature. In the replication materials for this article, we include an R package that implements functions for carrying out the testing methodology developed by Bowers, Fredrickson and Panagopoulos (2012).[2]

Despite providing preliminary evidence that some state legislative field experiments are characterized by interference, one shortcoming of our replication analyses is that the experiments were designed and data collected with a focus on direct effects, assuming SUTVA. We retrospectively constructed networks to use in testing for interference, relaxing SUTVA, which is not ideal since there are likely to be more appropriate networks for each individual application. In future state legislative field experiments, researchers should consider collecting network data that characterizes the patterns of interdependence between legislators that are most relevant to their experiments. Furthermore, in each of the studies we consider, half of the observations were allocated to treatment, and treatment allocation was uniform-at-random (within blocks). This may not be the optimal randomization design if the researcher is interested in testing for and identifying interference

[2] Shortly after this article is accepted for publication, this package will be submitted to the CRAN network for public distribution.

effects. In experiments designed for testing interference effects, the optimal proportion assigned to treatment is typically much lower than 50% (Bowers, Desmarais, Frederickson, Ichino, Lee and Wang 2016). Furthermore, researchers can use the networks through which they think interference occurs to design higher powered experiments that incorporate the network structure (Bowers et al. 2016). The optimal experimental design depends on the structure(s) of the network(s) through which interference is hypothesized as well as the model of effects. In future work, researchers conducting field experiments on state legislatures should take the networks and models of effects into account at the design stage.

References

Aral, Sinan and Dylan Walker. 2014. "Tie strength, embeddedness, and social influence: A large-scale networked experiment." *Management Science* 60(6):1352–1370.

Bapna, Ravi and Akhmed Umyarov. 2015. "Do your online friends make you pay? a randomized field experiment on peer influence in online social networks." *Management Science* 61(8):1902–1920.

Basu, Debabrata. 2011. Randomization analysis of experimental data: the Fisher randomization test. In *Selected Works of Debabrata Basu*. Springer pp. 305–325.

ben Aaron, James, Matthew J. Denny, Bruce A. Desmarais and Hanna Wallach. Accepted. "Transparency by Conformity: A Field Experiment Evaluating Openness in Local Governments." *Public Administration Review* .

URL: *http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2772975*

Bergan, Daniel E. 2009. "Does grassroots lobbying work? A field experiment measuring the effects of an e-mail lobbying campaign on legislative behavior." *American Politics Research* 37(2):327–352.

Bergan, Daniel E and Richard T Cole. 2015. "Call Your Legislator: a field experimental study of the impact of a constituency mobilization campaign on legislative voting." *Political Behavior* 37(1):27–42.

Bernhard, William and Tracy Sulkin. 2013. "Commitment and consequences: Reneging on cosponsorship pledges in the US House." *Legislative Studies Quarterly* 38(4):461–487.

Bond, R. M., C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle and J. H. Fowler. 2012. "A 61-million-person experiment in social influence and political mobilization." *Nature* 489(7415):295–298.

Bowers, Jake, Bruce A Desmarais, Mark Frederickson, Nahomi Ichino, Hsuan-Wei Lee and Simi Wang. 2016. "Models, Methods and Network Topology: Experimental Design for the Study of Interference." *arXiv preprint arXiv:1601.00992* .

Bowers, Jake, Mark M Fredrickson and Costas Panagopoulos. 2012. "Reasoning about interference between units: A general framework." *Political Analysis* p. mps038.

Bowers, Jake, Mark M Fredrickson and Peter M Aronow. 2016. "Research Note: A more powerful test statistic for reasoning about interference between units." *Political Analysis* p. mpw018.

Bratton, Kathleen A and Stella M Rouse. 2011. "Networks in the legislative arena: How group dynamics affect cosponsorship." *Legislative Studies Quarterly* 36(3):423–460.

Broockman, David E. 2013. "Black politicians are more intrinsically motivated to advance Blacks? interests: A field experiment manipulating political incentives." *American Journal of Political Science* 57(3):521–536.

Butler, Daniel M, Christopher F Karpowitz and Jeremy C Pope. 2012. "A field experiment on legislators? home styles: Service versus policy." *The Journal of Politics* 74(02):474–486.

Butler, Daniel M and David E Broockman. 2011. "Do politicians racially discriminate against constituents? A field experiment on state legislators." *American Journal of Political Science* 55(3):463–477.

Butler, Daniel M, David W Nickerson et al. 2011. "Can learning constituency opinion affect how legislators vote? Results from a field experiment." *Quarterly Journal of Political Science* 6(1):55–83.

Coppock, Alexander. 2014. "Information spillovers: Another look at experimental estimates of legislator responsiveness." *Journal of Experimental Political Science* 1(02):159–169.

Coppock, Alexander. 2016. "?Information Spillovers: Another Look at Experimental Estimates of Legislator Responsiveness??CORRIGENDUM." *Journal of Experimental Political Science* 3(2):206–208.

Desmarais, Bruce A, Vincent G Moscardelli, Brian F Schaffner and Michael S Kowal. 2015. "Measuring legislative collaboration: The Senate press events network." *Social Networks* 40:43–54.

Dodds, Peter Sheridan and Duncan J Watts. 2004. "Universal behavior in a generalized model of contagion." *Physical review letters* 92(21):218701.

Ferejohn, John. 1986. Logrolling in an institutional context: A case study of food stamp legislation. In *Congress and policy change*. Vol. 223 New York: Agathon Press pp. 224–25.

Fowler, James H. 2006. "Connecting the Congress: A study of cosponsorship networks." *Political Analysis* pp. 456–487.

Gerber, Alan S, Donald P Green and Christopher W Larimer. 2008. "Social pressure and voter turnout: Evidence from a large-scale field experiment." *American Political Science Review* 102(01):33–48.

Kim, Jang Hyun and George Barnett. 2012. "Comparing the Influence of Social Networks Online and Offline on Decision Making: The US Senate Case." *E-Politics and Organizational Implications of the Internet: Power, Influence, and Social Change: Power, Influence, and Social Change* p. 198.

Kirkland, Justin H. 2011. "The relational determinants of legislative outcomes: Strong and weak ties between legislators." *The Journal of Politics* 73(3):887–898.

Kirkland, Justin H. 2013. "Hypothesis testing for group structure in legislative networks." *State Politics & Policy Quarterly* 13(2):225–243.

Kirkland, Justin H and Justin H Gross. 2014. “Measurement and theory in legislative networks: The evolvinsg topology of Congressional collaboration.” *Social Networks* 36:97–109.

Lin, Yu-Ru, Drew Margolin and David Lazer. 2016. “Uncovering social semantics from textual traces: A theory-driven approach and evidence from public statements of US Members of Congress.” *Journal of the Association for Information Science and Technology* 67(9):2072–2089.

Little, Roderick J and Donald B Rubin. 2000. “Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches.” *Annual review of public health* 21(1):121–145.

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

Masket, Seth E et al. 2008. “Where You Sit is Where You Stand: The Impact of Seating Proximity on Legislative Cue-Taking.” *Quarterly Journal of Political Science* 3(3):301–311.

Matthews, Donald R. 1959. “The folkways of the United States Senate: Conformity to group norms and legislative effectiveness.” *American Political Science Review* 53(4):1064–1089.

Muchnik, Lev, Sinan Aral and Sean J Taylor. 2013. “Social influence bias: A randomized experiment.” *Science* 341(6146):647–651.

Napierala, Matthew A. 2012. "What is the Bonferroni correction." *AAOS Now* 6(4):40.

Nyhan, Brendan and Jacob M Montgomery. 2015. "Connecting the candidates: Consultant networks and the diffusion of campaign strategy in American congressional elections." *American Journal of Political Science* 59(2):292–308.

Nyhan, Brendan and Jason Reifler. 2015. "The Effect of Fact-Checking on Elites: A Field Experiment on US State Legislators." *American Journal of Political Science* 59(3):628–640.

Paluck, Elizabeth Levy. 2011. "Peer pressure against prejudice: A high school field experiment examining social network change." *Journal of Experimental Social Psychology* 47(2):350–358.

Peng, Tai-Quan, Mengchen Liu, Yingcai Wu and Shixia Liu. 2016. "Follower-followee network, communication networks, and vote agreement of the US members of congress." *Communication Research* 43(7):996–1024.

Porter, Mason A, Peter J Mucha, Mark EJ Newman and Casey M Warmbrand. 2005. "A network analysis of committees in the US House of Representatives." *Proceedings of the National Academy of Sciences of the United States of America* 102(20):7057–7062.

Ringe, Nils, Jennifer Nicoll Victor and Justin H Gross. 2013. "Keeping your friends close and your enemies closer? Information networks in legislative politics." *British Journal of Political Science* 43(03):601–628.

Rogowski, Jon C and Betsy Sinclair. 2012. "Estimating the causal effects of social interaction with endogenous networks." *Political Analysis* 20(3):316–328.

Sekhon, Jasjeet S. 2008. "The Neyman-Rubin model of causal inference and estimation via matching methods." *The Oxford handbook of political methodology* pp. 271–299.

Shalizi, C. R. and A. C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociol Methods Res* 40(2):211–239.

Valentini, Gabriele, Heiko Hamann and Marco Dorigo. 2014. Self-organized collective decision making: The weighted voter model. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. International Foundation for Autonomous Agents and Multiagent Systems pp. 45–52.

Victor, Jennifer Nicoll and Nils Ringe. 2009. "The social utility of informal institutions: Caucuses as networks in the 110th US House of Representatives." *American Politics Research* 37(5):742–766.

Wojcik, Stefan. 2017. "Why Legislative Networks? Analyzing Legislative Network Formation." *Political Science Research and Methods* pp. 1–18.