Online Appendix for:
Persistent Policy Pathways: Inferring Diffusion Networks in the American States

Bruce A. Desmarais* Jeffrey J. Harden† Frederick J. Boehmke‡

November 2, 2014

Forthcoming, *American Political Science Review*

Latent Network Inference

The derivation of the `NetInf` algorithm begins with the definition of a probabilistic model describing how attributes would cascade through a diffusion network. To clarify application to state policy diffusion, we refer to the units and attributes in the model as states and policies, respectively. Denote a single policy cascade—the years in which states adopted a given policy—as c. The model is derived in three steps. First, we construct the probability that state u spreads a policy to state v: $P_c(u,v)$. Second, given these dyadic spread probabilities, we build the probability that a policy spreads through the states in a given *tree* pattern $P(c|T)$, where T specifies which states influence which other states. Third, we define $P(c|G)$, which is the probability of cascade c given the diffusion network (i.e., graph) connecting the states G. With these three quantities defined, we can define a proper likelihood of the policy cascades given a proposed diffusion network by evaluating the probability of each cascade on that diffusion network.

The `NetInf` algorithm assumes that diffusion occurs in continuous time and that diffusion

*Assistant Professor, Department of Political Science, University of Massachusetts–Amherst, 420 Thompson Hall, 200 Hicks Way, Amherst, MA 01003, desmarais@polsci.umass.edu.

†Assistant Professor, Department of Political Science, University of Colorado Boulder, 136 Ketchum, UCB 333, Boulder, CO 80309, jeffrey.harden@colorado.edu.

‡Professor, Department of Political Science, and Director of the Social Science Program in the Public Policy Center, University of Iowa, 341 Schaeffer Hall, Iowa City, IA 52242, frederick-boehmke@uiowa.edu.

time has an exponential distribution. If state u adopts a policy at time t_u and state v adopts a policy at time t_v ($t_v \geq t_u$) and u spreads the policy to v, then the probability of the diffusion time ($t_v - t_u$) is given by

$$P_c(u,v) = \lambda \exp\left(\frac{-(t_v - t_u)}{\lambda}\right), \tag{A.1}$$

where λ is the rate parameter of the exponential distribution. Given this, the probability of observing a cascade that propagates in a given pattern over the states, represented by the tree T that encodes (i,j) pairs listing which states were influenced by which other states, is

$$P(c|T) = \prod_{(i,j)\in T} P_c(i,j). \tag{A.2}$$

The diffusion network G places a constraint on the possible tree structures T along which the policy can spread. That is, a policy cannot spread from i to j if there is not a diffusion pathway from i to j in G. Thus, to build the probability of a cascade c given the diffusion network G, we average the probability of the cascade c over all possible tree structures in G, denoted $\mathscr{T}(G)$.

$$P(c|G) = \frac{1}{|\mathscr{T}(G)|} \sum_{T\in\mathscr{T}(G)} P(c|T), \tag{A.3}$$

where $|\mathscr{T}(G)|$ is the number of tree structures that can be constructed from G. Given a set of policy cascades (C), the likelihood of the cascade data given a proposed diffusion network G is:

$$P(C|G) = \prod_{c\in C} P(c|G). \tag{A.4}$$

Inferring the Network

With the probabilistic model of diffusion along a diffusion network defined, the task of inferring a diffusion network is to find a network structure G under which we would have been highly likely to observe the set of policy cascades C. Ideally, we would identify the network structure that maximized the likelihood of observing C. Likelihood maximization in this case, however, turns out

to be a computationally intractable task. Among the 50 states, there are 2×2^{1225} possible network structures. Moreover, Gomez-Rodriguez, Leskovec, and Krause (2010) show that every network structure would need to be evaluated to assure that the optimal network had been identified.

As a more computationally tractable alternative, Gomez-Rodriguez, Leskovec, and Krause (2010) derive an approach to approximation of the optimal G. They also demonstrate analytically and through simulations that this method is capable of inferring a very-close-to-optimal network structure inference within feasible compute times. Their departures from exhaustive optimization are two. First, instead of computing the likelihood of a cascade given a network structure by enumerating all possible propagation trees represented by that network structure, they simply focus on the most likely propagation tree for each cascade within a given network structure—a shortcut which they refer to as *lazy evaluation*. Second, they adopt a greedy (i.e., local) optimization approach that iteratively adds diffusion ties to the network structure G such that the k^{th} diffusion tie added to the network improves the likelihood function more than any other tie that could be added to the network, given the $k-1$ ties already in the network.

Network Inference: Empirical Conditions

In this section we present ancillary information regarding the application of `NetInf` to the state policy diffusion data. Specifically, we present the complete model fit results from our tuning exercise as well as descriptive data regarding the number of policies and adoption instances used to draw inferences in each year.

`NetInf` Parameter Tuning

We set three parameters in the network inference procedure. First, we need to define the number of preceding years of adoptions (denoted k) that will be used to infer the network for time t. Second, we need to define the number of edges (E) we want to infer in each time period. Third, we need to tune a rate parameter λ of the exponential distribution used by `NetInf` to calibrate how long it takes for policies to diffuse from one state to another. A policy can only diffuse from i to j if there is an edge from i to j in the inferred network. The exponential distribution gives the distribution

of diffusion times between states, provided that there is an edge connecting them. Higher rates place a higher penalty on the addition of edges to the network along which it takes a long time for policies to diffuse. This prevents any given adoption by one state that happens to fall later in time than adoption by another state from contributing to the formation of a tie between the two states.

We take a data-driven approach to finding optimal values of these parameters. We use the conventional discrete-time event history modeling methodology to evaluate the performance of the network in predicting future adoptions measured at different parameterizations. For each unique combination of parameters $\{k, E, \lambda\}$, we fit a pooled (across all policies in the data) logistic discrete-time event history model predicting policy adoption. The model contains three classes of regressors. For state s still in the data at time t for policy p, the regressors are:

(1) *States Adopting*: The number of other states that have adopted by time $t-1$,
(2) *Sources Adopting*: In a network inferred on all adoptions between $t-k$ and $t-1$, the number of s's sources in the network that have adopted p.
(3) *Policy Area:* An indicator variable that models the unique rate of adoption for each policy.

In this design, all of the adoptions used to infer the network used to predict adoptions at time t occurred prior to t. We use a simple grid search to find best-fitting values of $\{k, E, \lambda\}$. We search over $\lambda \in \{0.125, 0.25, 0.5, 1\}$, which corresponds to mean diffusion times of 8, 4, 2, and 1 years, respectively, $k \in \{5, 10, \ldots, 50\}$, and $E \in \{100, 200, \ldots, 1000\}$. We use the Bayesian Information Criterion (BIC) to evaluate the fit of each combination of parameters and search for the combination of parameters that best fits the data (i.e., results in the lowest BIC). Figure A.1 depicts the BIC values for all of the parameter combinations that we consider.

[Insert Figure A.1 here]

The network that results in the best predictive fit, across all values of λ is one with 300 edges and defined over 35 years of policy adoptions.[1] The fit is not particularly sensitive to the rate

[1] We also use a network based on 400 edges and 10-year periods for use in two applications to policy diffusion models (see below).

parameter, but the network using a rate of 0.5 results in the best fit. This means that policies diffuse, on average, in two years. An average of approximately 1,900 adoption instances over an average of approximately 120 policies is used to infer the network for each year.

Policies and Adoptions Used in `NetInf` Over Time

Figure A.2 gives the number of unique policies and the total number of adoption instances used to infer the diffusion network in each year. The network inferred toward the end of the time series is generally based on more data than the network earlier on in the series.

[Insert Figure A.2 here]

Checking for Heterogeneity in Diffusion Classes

As the heterogeneity in the results from models of policy diffusion in the state politics literature suggests, there is considerable variation in the processes that drive the diffusion of different policies. It is therefore important to check whether we are inappropriately pooling policies to infer a single diffusion network. Though we know that policies vary in terms of the patterns and predictors of diffusion, we must evaluate whether this variation is policy-specific and idiosyncratic with respect to the underlying diffusion network, or whether there are systematic and consistent cross-policy differences. In other words, we need to check whether there are different classes of policies in terms of the underlying diffusion network.

We use a probabilistic mixture modeling approach (Imai and Tingley 2012) to examine whether there are multiple classes of policies in terms of their effects on the inferred diffusion network. We iteratively remove each policy from the dataset and infer a new network with 300 edges that spans the entire time period in our data. For each policy, we have a network inferred without that policy included. If two (or more) policies affect the diffusion network in the same way, the inferred network should change in systematically similar ways when those two policies are removed from the dataset. Using a policies$\times$potential edges—187$\times$2,450—observation dataset, we fit a Bernoulli

mixture model with the likelihood

$$l(\mathbf{y},\alpha,\pi)=\prod_{i=1}^{50}\prod_{j\neq i}\prod_{p=1}^{187}\sum_{a=1}^{k}\alpha_{ap}\pi_{ija}^{y_{ijp}}(1-\pi_{ija})^{(1-y_{ijp})},$$

where y_{ijp} is an indicator of whether there is a diffusion tie from i to j when policy p is removed from the dataset, k is the number of classes (i.e., mixture components) included in the model, α_{ap} is the probability that policy p is a member of class a, and π_{ija} is the probability that there is an edge from i to j in networks inferred excluding policies in class a.

We estimate models with $k \in \{1,2,\ldots,15\}$. The R package `flexmix` (Leisch 2004) is used to fit the models. Estimation also requires an initial assignment of the component membership probability for each policy. We use k-means clustering to identify initial cluster memberships, then assign the component membership cluster probability for each policy according to $\alpha_{ap}^{0}=\frac{\lambda^{\mathbf{1}(c_p=a)}}{\sum_{i=1}^{k}\lambda^{\mathbf{1}(c_p=i)}}$, where c_p is the initial cluster assignment of policy p and λ is a weight that controls the entropy in the initial component assignment probabilities, with higher values of λ corresponding to lower entropy. We evaluate models with 10 values of λ, varied equally between 1 and 5. The model fit results of the mixture modeling are presented in Figure A.3. Following Fraley and Raftery (1998), we evaluate the fit of each model using the BIC.

[Insert Figure A.3 here]

Across all values of λ, the best fitting model is clearly the one with only one component. This indicates that, insofar as removing individual policies changes the results of the network inference, the network is changed in ways that are idiosyncratic with respect to the other policies. In other words, policies do not appear to affect the network in patterns that can be efficiently grouped into a discrete number of classes, aside from the overall patterns that cut across all policies. These results support our use of a single diffusion network to model the diffusion patterns across all of the policies in our dataset.

Adjusting for Total State Coverage in the LexisNexis Analysis

In Table A.1, we present ordinary least squares regression results in which we regress the number of emulation stories identified in LexisNexis on the mean number of diffusion ties sent in the diffusion networks and the total number of search hits of a state's name in LexisNexis Academic. This analysis adjusts for the influence of overall state coverage on the number of reported diffusion ties.

[Insert Table A.1 here]

Table A.2 reports the news outlets in which we identify emulation stories from the LexisNexis database.

[Insert Table A.2 here]

Applying the Inferred Network to Models of Policy Diffusion

Here we present our replications of the five policy diffusion EHA models. Recall from the main text that we replicated four policy-specific models: lotteries (Berry and Berry 1990), Indian gaming (Boehmke 2005), capital punishment (Boehmke 2005), and restaurant smoking bans (Shipan and Volden 2006).[2] We also replicated Boehmke and Skinner's (2012) "pooled event history analysis" (PEHA) model fit to data on 151 different policies diffusing over the period 1960–1999 (see also Boehmke 2009). This approach stacks the data from different policies and estimates a unified model with a common set of independent variables (including state, year, and policy fixed effects). Pooling the data does result in fewer independent variables than for any single policy, but it provides insight into what factors affect diffusion most broadly across the issue spectrum of American politics. We show below that information from our inferred diffusion network is one of those factors.

[2]Specifically, we replicate the following models: Berry and Berry (1990, 409), Table 1, model 1; Boehmke (2005, 85 and 89), Tables 4.2 and 4.4; Shipan and Volden (2006, 839), Table 3, model 9.

Model Details

We focus on these five models for several reasons. First, the four policy-specific models represent a wide variety of policies, and the pooled model represents an even wider range. Thus, we examine whether the diffusion network has a broad or narrow impact on adoption. Second, the original studies presenting the policy-specific models are well-known in the policy diffusion literature, having each garnered at least 60 citations according to Google Scholar.[3] Finally, the models all use similar EHA empirical specifications, enhancing comparability. The dependent variable in each is coded "1" if a state adopted the policy in a given year and "0" otherwise, with states that have already adopted dropping out of the data beginning in the year after adoption.[4]

The theories underlying our replication models each have their own unique characteristics. To conserve space, we refer readers to the original studies for detailed discussions of each. We focus here on comparing the effect of the diffusion network on adoption to that of a factor that consistently appears in these models: geographic contiguity. Nearly all studies of policy diffusion include in their models either the number of or percentage of neighboring states that have previously adopted the policy. The expectation for this variable is that, due to economic competition and/or policy learning, as more neighbors adopt, the probability of a state adopting increases (see, for example, Berry and Berry 1990, 403–404; Boehmke 2005, chapter 4; Shipan and Volden 2006, 828).

It is unlikely that states can only compete with and learn from states with whom they share a border. Indeed, Berry and Berry (1990) point out that there are many plausible means of state-to-state influence, including shared borders, a shared region, or even shared culture. They further suggest that it would be useful to have a measure of which states a state tends to "follow" in policy adoption. With information on "predesignated leader states" in regions, the authors "would hypothesize that a state's probability of adopting a lottery increases after one or more states with a

[3] In fact, Berry and Berry (1990) is included on the "high impact" list of most influential articles appearing in the *American Political Science Review* (Sigelman 2006).

[4] The Berry and Berry (1990) and Boehmke (2005) models are estimated with probit and the Shipan and Volden (2006) and Boehmke and Skinner (2012) models are estimated with logistic regression.

reputation as a leader within its region adopt it" (Berry and Berry 1990, 403). However, they also acknowledge that they have no means of measuring this concept because there are no "reliable data about which states are perceived. . . to be regional leaders in a policy area" (Berry and Berry 1990, 403).

Including Network Information

Our inferred policy diffusion network provides those data that previous scholars of policy diffusion have not had available. In fact, beyond simply measuring regional leaders, the network gives information on any state that tends to be a leader, or source, of policy innovation for another state. In our replications we incorporate information from the estimated diffusion network by creating a variable on the same scale as *Neighbors Adopting*: the number of a state's sources in a given year that previously adopted the policy. We use the inferred network to produce a list of states that influence the state in a specified time period immediately preceding a given year.[5] This list represents all of that state's sources at that time. Next, to create the variable *Sources Adopting* we count the number of states from that list that have previously adopted the policy. We also computed this measure as a percentage, similar to studies that compute the percentage of *Neighbors Adopting* (e.g., Shipan and Volden 2006). We present both sets of results below (our substantive conclusions do not change).

After creating the *Sources Adopting* variable, we then add it to each of the five replication models. Recall that we avoid endogeneity problems because we only use adoptions that occurred before a given year to measure the network for that year. As such, the adoption of policy j by state i at time t cannot inform the network used in the diffusion models to predict the adoption of policy j by state i at time t. The adoptions used in the network for time t occur prior to t.[6]

[5] As mentioned above, we constructed a version using 35-year periods and one with 10-year periods. Results between the two are generally very similar. For each model we used the version that produced the lowest AIC and BIC values (35-year version for the lotteries, capital punishment, and pooled models; 10-year version for the Indian gaming and smoking ban models).

[6] We include all policies in the construction of the network used to produce *Sources Adopting*, including the policy of interest in the EHA model. We also estimated the models after having removed the policy area of interest and found results that are virtually identical to what we present below.

Estimates and Model Fit

We first examine the extent to which the inclusion of *Sources Adopting*—instead of or in addition to *Neighbors Adopting*—improves model fit.[7] Table A.3 reports coefficient estimates and standard errors for the two variables as well as model fit statistics for three specifications: (1) the original model with *Neighbors Adopting* (plus the authors' other covariates), (2) a model with *Sources Adopting* substituted for *Neighbors Adopting* (plus the other covariates), and (3) a model with both *Neighbors Adopting* and *Sources Adopting* (plus the other covariates). In all cases the coefficients are positive (as expected), though statistical significance varies somewhat across specifications and replications. We assess the substantive impact of these effects in section .

[Insert Table A.3 here]

To compare model fit we compute AIC, BIC, and cross-validated percent correctly classified. We compute this last measure via leave-one-out cross-validation, which involves iteratively dropping one observation, estimating the model, computing an expected probability from that model for the left-out observation, then generating a predicted value of the dependent variable based on a single draw from the Bernoulli distribution with that expected probability. We then compute the percentage of the observations for which the prediction matches the actual dependent variable value. Thus, unlike information-based measures of fit such as AIC and BIC, this measure assesses each specification's capacity to make out-of-sample predictions. In Table A.3, the values in bold indicate the best-fitting model according to each statistic.

The AIC and BIC values support the inclusion of *Sources Adopting* in all but the restaurant smoking ban model, where the original model and the model with *Sources Adopting* produce AIC and BIC values within 2 units of each other, indicating equal fit. The cross-validated percent correctly classified measure also generally supports the inclusion of *Sources Adopting*. In four of the five replication models the percent correctly classified in one or both models with *Sources Adopting* increases from the original model with *Neighbors Adopting* (the restaurant smoking ban model is

[7]The question of whether *Sources Adopting* should replace or complement *Neighbors Adopting* is context-dependent. We focus on model fit here, but theoretical expectations should also be an important guide.

again the lone exception). These improvements are somewhat small in magnitude—ranging from $+1$ to $+3$ percentage points across the different models. Nonetheless, they consistently point to the models that include *Sources Adopting* in the specification as the best fit.

Overall, Table A.3 provides good evidence that *Sources Adopting* can improve the fit of policy diffusion EHA models, either in place of or in addition to *Neighbors Adopting*. Importantly, across the five models, *none* of the fit statistics decisively selects the original model with *Neighbors Adopting* as the better fit. Given this evidence that *Sources Adopting* is a useful addition to diffusion models, our next step is to examine its substantive impact on policy adoption.

Marginal Effects

We examine the substantive implications of including *Sources Adopting* in Figure A.4 by graphing the average marginal effects of *Neighbors Adopting* (top row) and *Sources Adopting* (bottom row) in each model on the probability scale.[8] All estimates are computed from the specifications that include either *Neighbors Adopting* or *Sources Adopting*.[9]

[Insert Figure A.4 here]

The first point to note from Figure A.4 is the effect of the count of *Neighbors Adopting* (lotteries, Indian gaming, capital punishment, and pooled model) and percentage of *Neighbors Adopting* (restaurant smoking bans) is positive. Consistent with the expectation that states react to economic competition and/or policy learning, more neighboring states with the policy corresponds with an increase in the probability of adoption. The magnitude and level of uncertainty varies somewhat across the models, but the effect is consistently in the positive direction.

Moving to the bottom row of Figure A.4, note that when substituted for *Neighbors Adopting*, the effect of *Sources Adopting* is also positive in all five models; as the number of sources adopting the policy increases, so too does probability of a state adopting the policy. From the minimum (0) to the maximum (lotteries: 7, Indian gaming: 10, capital punishment: 10, restaurant smoking bans:

[8] We employ the "observed value" method of Hanmer and Kalkan (2013) in these computations. Rather than setting the other variables in the models to particular values (e.g., their means or modes), we allow them to vary naturally over the observed values for every case in the data, then compute the average expected probability for each observed value of *Neighbors Adopting* and *Sources Adopting*, respectively.

[9] Results with both included in the same model are substantively similar (see below).

9, pooled model: 14) of *Sources Adopting*, the probability of adoption increases by the following percentage points, on average: 24 (lotteries), 24 (Indian gaming), 50 (capital punishment), 16 (restaurant smoking bans), and 11 (pooled model). As with the effect of *Neighbors Adopting*, the confidence intervals indicate varying degrees of uncertainty around these estimates.[10] Nonetheless, these graphs show that *Sources Adopting* exerts a substantively significant, positive impact on the probability of adoption across many different policies.

Moreover, these positive effects remain even after controlling for *Neighbors Adopting* (see below). In short, these replication results show that information from our policy diffusion network can make a valuable contribution to diffusion studies. We show examples from four specific policy areas and a 151-policy pooled model in which states utilize a persistent set of diffusion sources to guide their policymaking decisions.

Marginal Effects with *Neighbors* and *Sources*

Figure A.5 presents the average marginal effects of *Neighbors Adopting* and *Sources Adopting* on the expected probability of adoption, *controlling for the other* (i.e., from the models in column 3 of Table A.3). Note that results are substantively similar to those in Figure A.4, which presents results from models with one variable or the other.

[Insert Figure A.5 here]

Replication Results with Percentage of *Sources Adopting*

Table A.4 presents the coefficient estimates and model fit statistics for the replication models using the *percentage* of sources (rather than number of sources) adopting the policy as the independent variable of interest. To maintain consistency we use the original authors' operationalization of the neighbors variable, which is a count in all but Shipan and Volden (2006). In an actual analysis we recommend that researchers use the same operationalization (count or percentage) for neighbors *and* sources.

Overall, these results are consistent with the results using the number of sources (see Table

[10]This is at least partially due to the fact that policy adoption models tend to have many independent variables (the median is 19 in the four policy-specific replications).

A.3): including the source variable in the model produces a positive (and often statistically significant) coefficient estimate and improves model fit.

[Insert Table A.4 here]

We do not for advocate one measure over the other, but rather defer to individual researchers in making the choice based on theoretical and empirical considerations. The two approaches represent very different views on the diffusion process. The percentage measure specifies a diffusion process where the non-adopting neighbors (sources) have just as much influence as the adopting neighbors (sources) and the state ends up being pulled between the two. The count-based measure assumes that non-adopting neighbors (sources) do not influence a state's decision to adopt. If a researcher thought that states are only affected by their sources who adopt a particular policy, the count measure would likely make the most sense. In contrast, if states look to both their adopting and non-adopting sources, the percentage-based measure may be more appropriate.

Fixed Effects Logit Results

Table A.5 presents results using a source and follower fixed-effects logit model. Overall, results are consistent with those reported in the main text.

[Insert Table A.5 here]

References

Berry, Frances Stokes, and William D. Berry. 1990. "State Lottery Adoptions as Policy Innovations: An Event History Analysis." *American Political Science Review* 84(2): 395–415.

Boehmke, Frederick J. 2005. *The Indirect Effect of Direct Legislation: How Institutions Shape Interest Group Systems*. Columbus, OH: The Ohio State University Press.

Boehmke, Frederick J. 2009. "Policy Emulation or Policy Convergence? Potential Ambiguities in the Dyadic Event History Approach to State Policy Emulation." *Journal of Politics* 71(3): 1125–1140.

Boehmke, Frederick J., and Paul Skinner. 2012. "The Determinants of State Policy Innovativeness." Presented at the Annual State Politics and Policy Conference, Houston, TX.

Fraley, Chris, and Adrian E. Raftery. 1998. "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis." *The Computer Journal* 41(8): 578–588.

Gomez-Rodriguez, Manuel, Jure Leskovec, and Andreas Krause. 2010. Inferring Networks of Diffusion and Influence. In *The 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*.

Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1): 263–277.

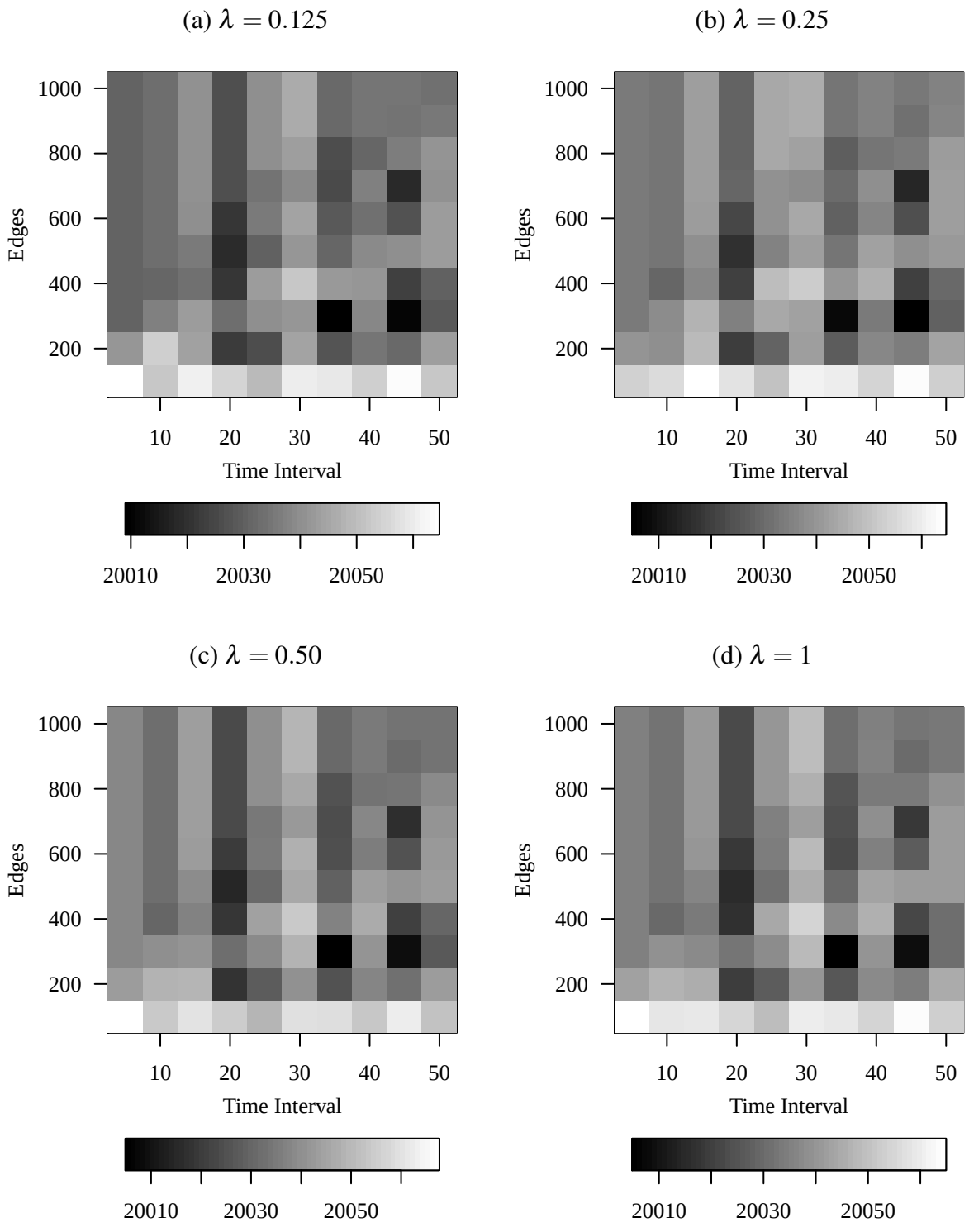
Imai, Kosuke, and Dustin Tingley. 2012. "A Statistical Method for Empirical Testing of Competing Theories." *American Journal of Political Science* 56(1): 218–236.

Leisch, Friedrich. 2004. "FlexMix: A General Framework for Finite Mixture Models and Latent Class Regression in R." *Journal of Statistical Software* 11(10): 1–18.

Shipan, Charles R., and Craig Volden. 2006. "Bottom-Up Federalism: The Diffusion of Antismoking Policies from U.S. Cities to States." *American Journal of Political Science* 50(4): 825–843.

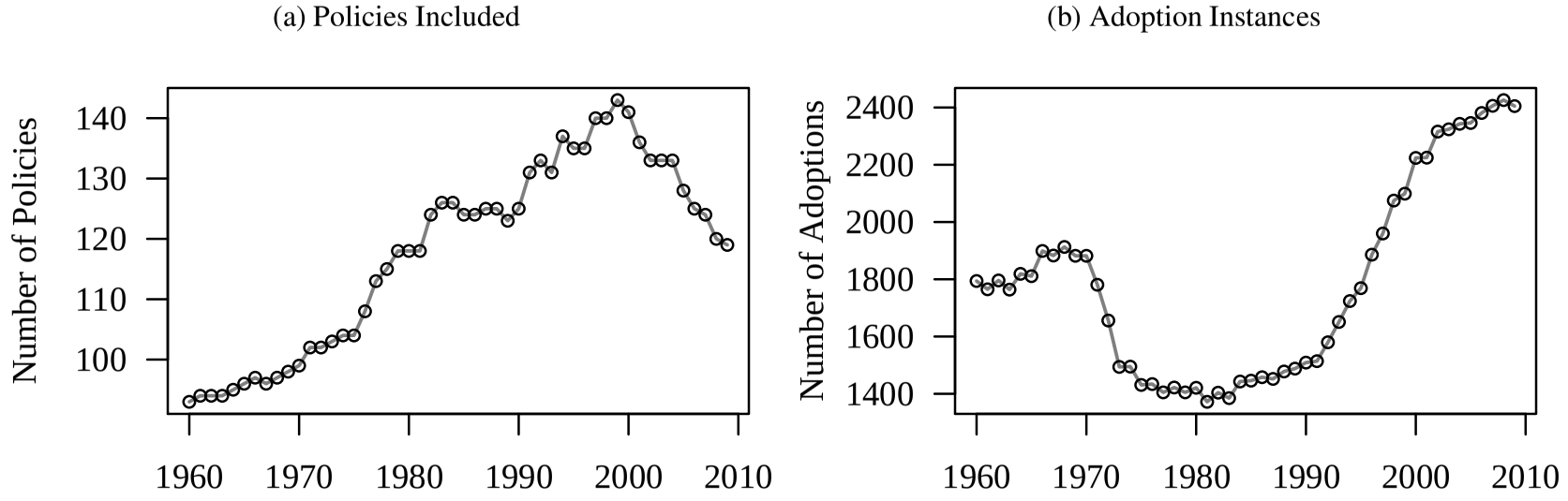
Sigelman, Lee. 2006. "Top Twenty Commentaries." *American Political Science Review* 100(4): 667–687.

Figure A.1: BIC of the Pooled Discrete Time Event History Models



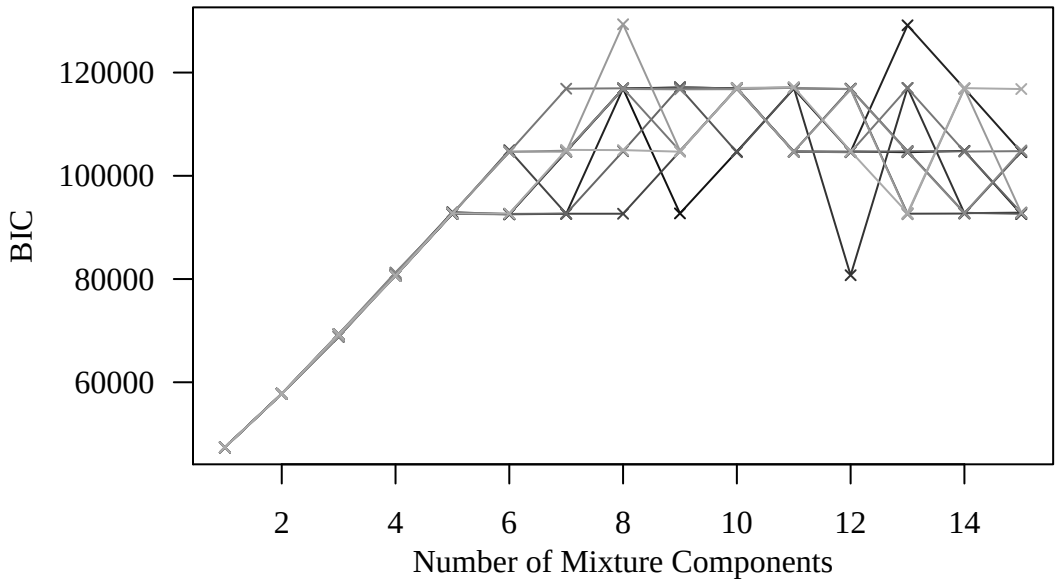
Note: There are 187 policies and 65,885 observations in each model.

Figure A.2: Number of Policies and Adoption Instances by Year



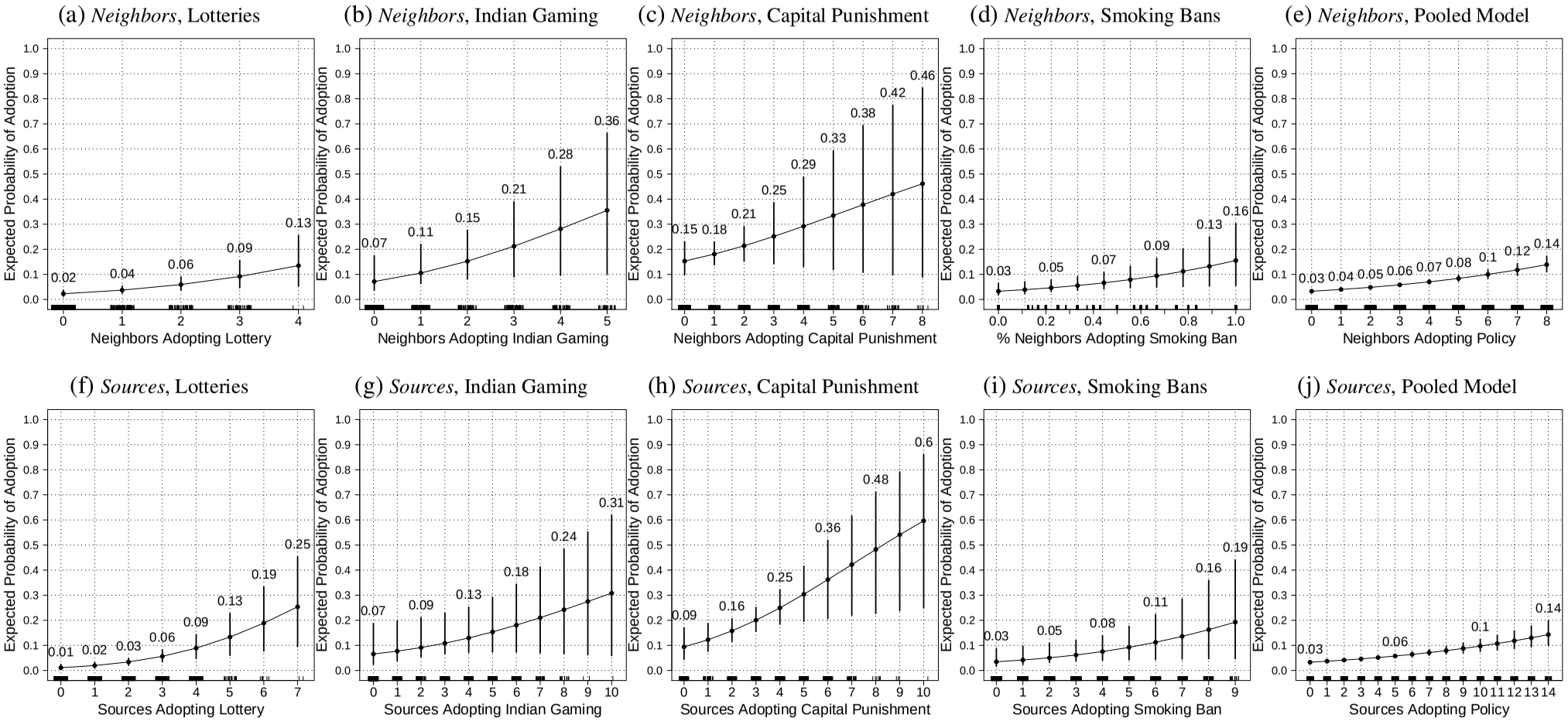
Note: The graphs present the number of policies (panel a) and number of adoption instances (panel b) included in the data used to infer the network in each year.

Figure A.3: Mixture Model Fit



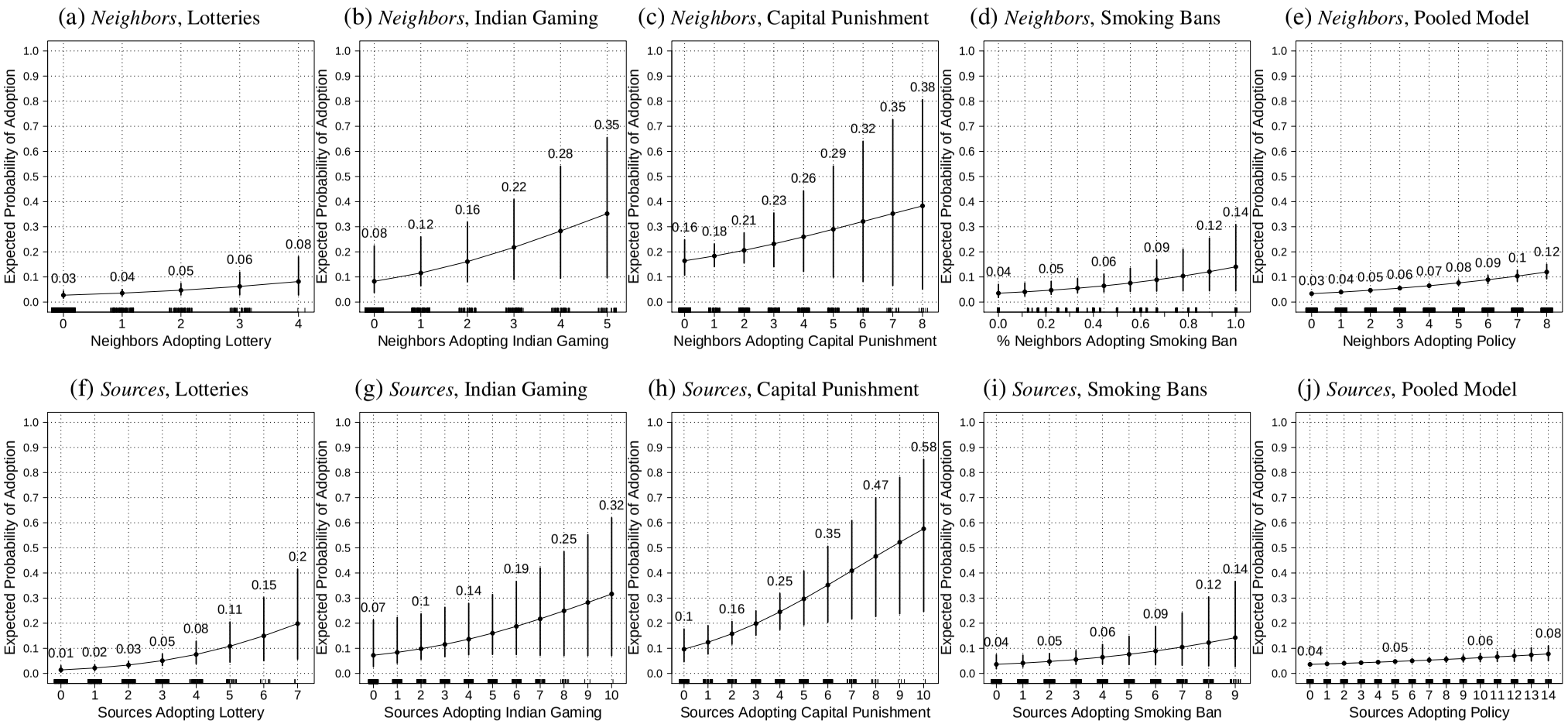
Note: Different line shades correspond to different values of λ, the parameter that controls the entropy in the initial cluster assignment probabilities.

Figure A.4: Average Marginal Effects of *Neighbors Adopting* and *Sources Adopting*



Note: The graphs present the average marginal effects of *Neighbors Adopting* (top row) and *Sources Adopting* (bottom row) in the five replication models: lotteries (Berry and Berry 1990), Indian gaming (Boehmke 2005), capital punishment (Boehmke 2005), restaurant smoking bans (Shipan and Volden 2006), and the pooled model (Boehmke and Skinner 2012). Points represent expected probability point estimates and vertical lines represent 95% confidence intervals.

Figure A.5: Average Marginal Effects of *Neighbors Adopting* and *Sources Adopting*, Controlling for the Other



Note: The graphs present the average marginal effects of *Neighbors Adopting* (top row) and *Sources Adopting* (bottom row), controlling for the other, in the five replication models: lotteries (Berry and Berry 1990), Indian gaming (Boehmke 2005), capital punishment (Boehmke 2005), restaurant smoking bans (Shipan and Volden 2006), and the pooled model (Boehmke and Skinner 2012). Points represent expected probability point estimates and vertical lines represent 95% confidence intervals.

Table A.1: Effect of Diffusion Ties on Emulation Stories, Adjusting for Total State Coverage

	Estimate	2.5 %-tile	97.5%-tile
Intercept	−6.712	−14.863	0.292
ln(1 + mean ties)	0.779	0.456	1.045
ln(1 + total coverage)	1.030	−0.005	2.240
R^2	0.423		
N	50		

Note: OLS Regression coefficients reported with percentile bootstrap confidence intervals constructed with 10,000 resampling iterations. The dependent variable is ln(1+emulation stories).

Table A.2: News Outlets Reporting Emulation Stories in LexisNexis

The Atlanta Journal-Constitution	11	The Tampa Tribune (Florida)	2
St. Louis Post-Dispatch	10	The Vancouver Sun (12 hour delay)	2
The Denver Post	10	Whittier Daily News (California)	2
Tampa Bay Times	9	Wisconsin State Journal	2
The Philadelphia Inquirer	9	American Banker	1
Deseret Morning News (Salt Lake City)	8	Clean Air Report	1
Lincoln Journal Star (Nebraska)	8	CongressNow	1
Providence Journal	8	Daily News (New York)	1
St. Paul Pioneer Press (Minnesota)	8	El Paso Times (Texas)	1
Omaha World Herald	7	Electric Power Daily	1
Portland Press Herald	7	Electric Utility Week	1
San Jose Mercury News (California)	7	Environmental Policy Alert	1
Tulsa World (Oklahoma)	7	Eureka Times-Standard (California)	1
The Palm Beach Post	6	Finance & Commerce (Minneapolis, MN)	1
The Record (Bergen County, NJ)	6	Global Power Report	1
Topeka Capital-Journal	6	guardian.co.uk	1
Pittsburgh Post-Gazette	5	Idaho Falls Post Register	1
Contra Costa Times	4	Inside EPA Weekly Report	1
South Bend Tribune	4	Investor's Business Daily	1
The Charleston Gazette	4	Legal News Line	1
The Salt Lake Tribune	4	Long Island Business News (Long Island, NY)	1
Discover America's Story	3	Maryland Gazette	1
Herald News (Passaic County, NJ)	3	Michigan Lawyers Weekly	1
News-Journal (Daytona Beach, Florida)	3	Missouri Lawyers Media	1
Richmond Times Dispatch	3	Monterey County Herald (CA)	1
San Gabriel Valley Tribune (San Gabriel Valley, CA)	3	Nanaimo Daily News (12 hour delay)	1
Sarasota Herald-Tribune	3	National Post (12 hour delay)	1
The Austin American-Statesman	3	North Carolina Lawyers Weekly	1
The Bond Buyer	3	North Jersey Community Newspapers	1
The Capital (Annapolis, MD)	3	Ottawa Citizen (12 hour delay)	1
The Toronto Star	3	Pasadena Star-News (California)	1
The Union Leader	3	Pittsburgh Tribune Review	1
Bangor Daily News (Maine)	2	Platts Megawatt Daily	1
Brattleboro Reformer (Vermont)	2	Public Opinion (Chambersburg, Pennsylvania)	1
Charleston Daily Mail	2	Ruidoso News (New Mexico)	1
Chicago Daily Herald	2	San Bernardino Sun (California)	1
Crain's Detroit Business	2	San Mateo County Times (San Mateo, CA)	1
Digital Archives	2	Star Tribune (Minneapolis MN)	1
Dolan Publications	2	Telegraph Herald (Dubuque, IA)	1
Information Bank Abstracts	2	The Baltimore Sun (most recent 6 months)	1
Inland Valley Daily Bulletin (Ontario, CA)	2	The Buffalo News (New York)	1
Inside Bay Area (California)	2	The Calgary Herald (12 hour delay)	1
Journal Record Legislative Report (Oklahoma City, OK)	2	The Capital Times (Madison, Wisconsin)	1
Legal Monitor Worldwide	2	The Columbian (Vancouver, WA)	1
McClatchy Tribune News non-restricted	2	The Decatur Daily (Alabama)	1
Metropolitan News Enterprise	2	The Gazette (12 hour delay)	1
Star-News (Wilmington, NC)	2	The Globe and Mail (Canada)	1
Telegram & Gazette (Massachusetts)	2	The Hamilton Spectator (Ontario, Canada)	1
The Berkshire Eagle (Pittsfield, Massachusetts)	2	The Hill	1
The Bismarck Tribune	2	The Indianapolis Business Journal	1
The Daily News of Los Angeles	2	The New York Post	1
The Daily Oklahoman (Oklahoma City, OK)	2	The Pantagraph	1
The Daily Record (Baltimore, MD)	2	The Patriot Ledger	1
The Florida Times Union	2	The Spokesman-Review	1
The Journal Record (Oklahoma City, OK)	2	The Straits Times (Singapore)	1
The Ledger (Lakeland)	2	The York Dispatch (York, PA)	1
The Santa Fe New Mexican	2	University Wire	1
The State Journal-Register (Springfield, IL)	2	Vallejo Times-Herald (California)	1

Note: Entries report news outlets and number of emulation stories identified in LexisNexis.

Table A.3: Estimates and Model Fit Statistics for *Neighbors Adopting* and *Sources Adopting* in the Replication Models

	Only *Neighbors* (Original Model)	Only *Sources*	*Neighbors* and *Sources*
Berry and Berry (1990): Lotteries (Probit, N = 857)			
Neighbors Adopting	0.27*		0.16
	(0.09)		(0.10)
Sources Adopting		0.30*	0.25*
		(0.09)	(0.10)
AIC	195.12	189.96	**189.64**
BIC	233.15	**227.99**	232.42
CV % Correctly Classified	94%	**95%**	**95%**
Boehmke (2005): Indian Gaming (Probit, N = 364)			
Neighbors Adopting	0.42*		0.42*
	(0.20)		(0.21)
Sources Adopting		0.20+	0.21
		(0.12)	(0.13)
AIC	144.25	144.45	**143.54**
BIC	241.68	**237.98**	244.86
CV % Correctly Classified	89%	**91%**	90%
Boehmke (2005): Capital Punishment (Probit, N = 227)			
Neighbors Adopting	0.16		0.14
	(0.14)		(0.14)
Sources Adopting		0.24*	0.23*
		(0.10)	(0.10)
AIC	204.53	**199.66**	200.97
BIC	283.31	**278.43**	283.17
CV % Correctly Classified	75%	**78%**	76%
Shipan and Volden (2006): Restaurant Smoking Bans (Logit, N = 807)			
% Neighbors Adopting	1.92*		1.66+
	(0.86)		(0.94)
Sources Adopting		0.25*	0.19
		(0.12)	(0.14)
AIC	**248.57**	249.96	249.16
BIC	**328.36**	329.75	333.64
CV % Correctly Classified	**94%**	93%	**94%**
151-Policy Pooled Model (Logit, N = 62,290)			
Neighbors Adopting	0.22*		0.19*
	(0.02)		(0.02)
Sources Adopting		0.13*	0.06*
		(0.02)	(0.02)
AIC	17030.64	17089.78	**17021.47**
BIC	19263.41	19322.55	**19263.28**
CV % Correctly Classified	93%	93%	**94%**

Note: Cell entries report coefficient estimates and standard errors (in parentheses) for *Neighbors Adopting* and *Sources Adopting* and AIC, BIC, and cross-validated percent correctly classified values in three specifications of the replication models. All other variables from the original models are included, but those estimates are not shown to conserve space. Numbers in bold identify the best-fitting model for each fit statistic. * $p < 0.05$; + $p < 0.10$ (two-tailed).

Table A.4: Estimates and Model Fit Statistics for *Neighbors Adopting* and Percentage of *Sources Adopting* in the Replication Models

	Only *Neighbors* (Original Model)	Only *Sources*	*Neighbors* and *Sources*
Berry and Berry (1990): Lotteries (Probit, N = 857)			
Neighbors Adopting	0.27*		0.20*
	(0.09)		(0.10)
% Sources Adopting		1.57*	1.30*
		(0.52)	(0.56)
AIC	195.12	193.22	**191.41**
BIC	233.15	**231.25**	234.20
CV % Correctly Classified	94%	**95%**	**95%**
Boehmke (2005): Indian Gaming (Probit, N = 364)			
Neighbors Adopting	0.42*		0.42*
	(0.20)		(0.20)
% Sources Adopting		1.10	1.04
		(0.94)	(0.94)
AIC	**144.25**	145.92	145.11
BIC	241.68	**239.45**	246.44
CV % Correctly Classified	89%	**91%**	89%
Boehmke (2005): Capital Punishment (Probit, N = 227)			
Neighbors Adopting	0.16		0.15
	(0.14)		(0.14)
% Sources Adopting		2.15*	2.13*
		(0.77)	(0.77)
AIC	204.53	**197.84**	198.81
BIC	283.31	**276.62**	281.01
CV % Correctly Classified	75%	**76%**	**76%**
Shipan and Volden (2006): Restaurant Smoking Bans (Logit, N = 807)			
% Neighbors Adopting	1.92*		1.73
	(0.86)		(0.97)
% Sources Adopting		1.04	0.59
		(0.85)	(0.93)
AIC	**248.57**	250.89	249.87
BIC	**328.36**	330.50	334.17
CV % Correctly Classified	**94%**	93%	**94%**
151-Policy Pooled Model (Logit, N = 62,290)			
Neighbors Adopting	0.22*		0.18*
	(0.02)		(0.02)
% Sources Adopting		0.91*	0.54*
		(0.12)	(0.13)
AIC	17030.64	17073.94	**17011.82**
BIC	19263.41	19306.71	**19253.63**
CV % Correctly Classified	93%	93%	**94%**

Note: Cell entries report coefficient estimates and standard errors (in parentheses) for *Neighbors Adopting* and *% Sources Adopting* and AIC, BIC, and cross-validated percent correctly classified values in three specifications of the replication models. All other variables from the original models are included, but those estimates are not shown to conserve space. Numbers in bold identify the best-fitting model for each fit statistic. * $p < 0.05$; + $p < 0.10$.

Table A.5: Fixed Effects Logit Models of State Policy Diffusion Ties

	Coef.	S.E.	Coef.	S.E.
Follower State Characteristics:				
Citizen Ideology			-0.014^{*}	(0.002)
Legislative Professionalism			-0.397	(0.244)
Minority Diversity			0.790^{+}	(0.245)
Per Capita Income			0.699^{*}	(0.097)
Population			0.184^{+}	(0.012)
Unified Democratic Government			0.011	(0.035)
Unified Republican Government			-0.029	(0.041)
Potential Source Characteristics:				
Citizen Ideology			-0.007^{*}	(0.002)
Legislative Professionalism			-0.133	(0.250)
Minority Diversity			-0.147	(0.246)
Per Capita Income			-0.037	(0.093)
Population			0.073^{*}	(0.014)
Unified Democratic Government			-0.038	(0.035)
Unified Republican Government			-0.075	(0.040)
Relative Follower/Source Characteristics:				
Contiguity	0.149^{+}	(0.039)	0.037	(0.040)
Distance Between Capitals	-0.238^{*}	(0.021)	-0.208^{+}	(0.021)
Citizen Ideology (Absolute Difference)			-0.009^{*}	(0.001)
Legislative Professionalism (Absolute Difference)			0.282^{+}	(0.141)
Minority Diversity (Absolute Difference)			-0.119	(0.107)
Per Capita Income (Absolute Difference)			-0.428^{*}	(0.049)
Population (Absolute Difference)			-0.035^{+}	(0.004)
Unified Democratic (Product)			0.123^{+}	(0.046)
Unified Republican (Product)			-0.078	(0.083)
Intercept	-3.611^{*}	(0.202)	-3.133^{*}	(0.384)
N	94,080		94,080	
AIC	62,821		62,048	

Note: Observations are dyadic. The dependent variable indicates whether potential source state is in fact a source for a follower state. We use the network with 300 edges over 35 years of policy adoptions. $^{+}$ indicates statistical significance at the 0.05 level (two-tailed) according to just the parametric p-values from the multilevel logit. * indicates statistical significance at the 0.05 level according to the QAP p-values and the parametric p-values. QAP p-values derived from 1000 network permutations.