

Modeling History Dependence in Network-Behavior Coevolution

Robert J. Franzese Jr.* Jude C. Hays† Aya Kachi‡§

Draft: July 23, 2010

Abstract

Spatial interdependence—the dependence of outcomes in some units on those in others—is substantively and theoretically ubiquitous and central across the social sciences. Spatial association is also omnipresent empirically. However, spatial association may arise from three importantly distinct processes: *common exposure* of actors to exogenous external and internal stimuli, interdependence of outcomes/behaviors across actors (*contagion*), and/or the putative outcomes may affect the variable along which the clustering occurs (*selection*). Accurate inference about any of these processes generally requires an empirical strategy that addresses all three well. From a spatial-econometric perspective, this suggests spatiotemporal empirical models with exogenous covariates (common exposure) and spatial lags (contagion), with the spatial weights being endogenous (selection). From a longitudinal network-analytic perspective, we can identify the same three processes as potential sources of network effects and network formation. From that perspective, actors' self-selection into networks (by, e.g., behavioral *homophily*) and actors' behavior that is contagious through those network connections likewise demands theoretical and empirical models in which networks and behavior *coevolve* over time. This paper begins building such modeling by, on the theoretical side, extending a Markov type-interaction model to allow endogenous tie-formation, and, on the empirical side, merging a simple *spatial-lag* logit model of contagious behavior with a simple *p-star* logit model of network formation, building this synthetic discrete-time empirical model from the theoretical base of the modified Markov type-interaction model. One interesting consequence of network-behavior coevolution—identically: endogenous patterns of spatial interdependence—emphasized here is how it can produce history-dependent political dynamics, including equilibrium *path* and path dependence (Page 2006). The paper explores these implications, and then concludes with a preliminary demonstration of the strategy applied to alliance formation and conflict behavior among the great powers in the first half of the twentieth century.

*Professor, Department of Political Science, University of Michigan. E-mail: franzese@umich.edu; URL: <http://www.umich.edu/~franzese.html>.

†Assistant Professor, Department of Political Science, University of Illinois. E-mail: jchays@illinois.edu; URL: <https://netfiles.uiuc.edu/jchays/www/page.html>.

‡Ph.D. Student, Department of Political Science, University of Illinois. E-mail: akachi2@illinois.edu; URL: <https://netfiles.uiuc.edu/akachi2/home>.

§The authors thank commenters at the Political Networks Conference, May 2010 Duke University, the Path Dependence Conference, June 2010 University of Minnesota, and the Spatial Econometrics Association World Conference, June 2010 Chicago, for helpful comments. All errors are ours alone.

1 Introduction

From a network-analytic perspective: Networks are ubiquitous. Whether speaking of friendship or familiarity relations among individuals, trade or conflict relations among states, predator-prey relations in ecosystems, or any other sets of relations (*a.k.a.*: ties, connections, edges, etc.) among sets of units (*a.k.a.*: agents, actors, nodes, etc.), networks are essentially everywhere. And networks usually matter. *Network effects*, arising from various structural aspects of the network, or from actors' positions in the network, or from other actors through the network of connections, often importantly impinge upon the behaviors, opinions, outcomes, or other characteristics of units. Networks are also commonly endogenous. The units within some particular network typically choose or influence their connections, which are the edges that structure the network. A large challenge empirically for social scientists interested in the theory and substance of *network effects* and *network formation* is that network effects on nodes and the formation of edges between nodes tend to be mutually endogenous and, at the same time, both may be caused by outside factors, i.e., by a third mechanism that we have elsewhere called *common exposure* (Franzese and Hays 2006, 2007b, 2008b,a; Hays et al. 2010). In one archetypal application of social-network analysis, for example, we may observe clusters of smokers and of nonsmokers because smoking is contagious—one acquires the habit from friends or avoids acquisition because one's friends abstain—or because smokers choose to hang with smokers and nonsmokers with nonsmokers: homophily by behavior-type—or we may observe clustering of smokers and nonsmokers because both the behavior of (non)smoking and the connections between mutually (non)smoking behavior-types are caused by actors' common exposure to outside conditions, such as shared sociodemographics that affect both the propensity to smoke and friendship formation. To give a more political example (expanded from Koger et al. (2009, 2010)): representatives who sit together may vote similarly because they sit by party and so have similar constituencies (common exposure), or because they talk and influence each other (contagion), or they may choose to sit together because they know and like each other, which may be in some part because they vote similarly (selection). Or, to give the example from our empirical application, international conflict may be contagious through alliance connections, but nations that have similar conflict-behavior patterns are also more likely to ally (selection), and both alliance and conflict patterns may be affected the same exogenous conditions to which particular nation-state dyads are exposed, such as their contiguity.

We have made these arguments before from a spatial-econometric perspective (Franzese and Hays 2006, 2007b,a, 2008b,a, 2009; Hays et al. 2010): As *Tobler's Law* Tobler (1970) aptly sums: “Everything is related to everything else, but near things are more related than distant things.” Furthermore, as Beck et al. (2006)'s pithy title reminds in corollary: “Space is More than Geography”. The substantive content of the proximity in Tobler's Law, and so the pathways along which interdependence between units may operate, extends well beyond physical distance, contact, and contiguity. Long literatures in regional science, geography, and sociology carefully elaborate from those disciplinary perspectives the multifarious mechanisms by which contagion may arise. Simmons and colleagues (Simmons and Elkins 2004; Elkins and Simmons 2005; Simmons et al. 2006) offer a list for international relations: *coercion*, *competition*, *learning*, and *emulation* (to which one should add *relocation diffusion* (Hägerstrand 1967, 1970)). In fact, as, e.g., Brueckner (2003) showed, strategic interdependence, i.e., contagion, arises any time some unit(s)'s actions affect the marginal utility of other(s)'s actions. Given such externalities, *i*'s utility depends on both its policy and that of *j*. Theoretically, substantively, then, spatial interdependence is ubiquitous. Empirically, the clustering or correlation of outcomes on some dimension(s) of proximity, spatial association, is

also obvious across a vast array of substantive contexts. However, and this is the crux of the great empirical challenge/opportunity represented by the substantive and theoretical ubiquity of interdependence, outcomes may evidence spatial association for at least these three distinct reasons. First, units may be responding similarly to similar exposure to similar exogenous internal/domestic or external/foreign stimuli (*common exposure*), or, second, unit(s)’s responses may depend on others’ responses (*contagion*, one sort of *network effect*). We may find states’ adoptions of some economic treaty, for example, to cluster geographically or along other dimensions of proximity, e.g., bilateral trade-volume, because proximate states experience similar exogenous domestic or foreign political-economic stimuli or because each state’s decision to sign depends on whether proximate others sign. A third possibility arises when the putative outcome affects the variable along which clustering occurs (*selection* or *network formation*). Treaty signatories might also cluster according to some variable on which we observe their proximity (volume of trade between them) because being co-signatories affects that variable (spurs bilateral trade).

Whether from the network-analytic or the spatial-econometric perspective, accurately distinguishing and gauging empirically the role and strength of these alternative processes—common exposure, contagion, and selection; a.k.a., node effects, network effects, and network formation—is difficult, because the processes manifest empirically similarly, but also crucial because the theories and policy-intervention advice supported by any observed spatial-*cum*-network phenomenon hinges critically on whether (or the relative degrees to which) the observed phenomena arise from contagion/network-effects, selection/network-formation, or common exposure. The substance of the situations and how policies might best intervene in them vary critically depending on whether state signatories cluster in pockets of dense trade relations because those states tend to experience similar exogenous conditions that favor signing, or because the signing by some states spurs their trading partners to sign, or because the treaty fosters trade between co-signatories. Likewise, whether (non)smokers/(non)smoking clusters in social networks because having smoking or non-smoking friends spur one to adopt the behavior also, because (non)smokers tends to acquire friends who also (do not) smoke, or because some clustered exogenous internal or external conditions, some sociodemographics for instance, affect both one’s (non)smoking behavior and with whom one becomes friends.

Moreover, as we have also argued (and to some extents demonstrated, analytically, by simulation, and/or in applications) elsewhere, drawing effective distinctions and obtaining accurate estimates empirically of any of these separate processes requires great care and attention to specification (including measurement) of all three of components. That is, regardless of whether one’s interests center on network effects, the contagiousness of smoking for instance, or on network formation, what determines trade or conflict patterns for instance, one must model well both the network-effects/contagion and the network-formation/selection, and also whatever relevant external factors important to either process.¹

¹Nor, generally, will causal-inference strategies based on the potential-outcomes framework and assumptions of SUTVA salvage accurate estimation of any of these causal processes without adequate address of all three. SUTVA, in a nutshell, requires that (i) the probability of one unit receiving/taking treatment, (ii) the (constant) magnitude of the treatment, and (iii) the effect of treatment are independent of each other and of any other unit(s) receiving/taking treatment, the sizes of those units’ treatments, or the effects of those treatments in those others. These precluded situations are the essence spatial-*cum*-network effects. “The two most common ways in which SUTVA can be violated appear to occur when (a) there are versions of each treatment varying in effectiveness or (b) there exists interference between units” (Rubin 1990:, p. 282). The first of these is called spatial heterogeneity, the second is spatial interdependence: i.e., these are merely the network or spatial effects argued in those literatures to be ubiquitous and central to all social phenomena (at least).

Our project here aims to develop a framework for theoretical modeling and empirical specification, estimation, and interpretation of social phenomena with (common exposure and) simultaneous contagion and selection, that is, of mutually endogenous network effects and network selection, i.e., of the coevolution of actor behavior and their network connections. Identically from a spatial-econometric perspective, this means models with exogenous covariates reflecting common exposure, with spatial-lag contagion, and with patterns of spatial connectivity (spatial weights), i.e., a network, which are (at least in some part) endogenous to behavior. Our theoretical model of such processes builds from Markov type-interaction models in the extant literature, which explain evolving and steady-state profiles of actor types based on probabilities of type-switching that depend on the previous-period distribution of actor types depending on some set of exogenously given (possibly exogenously varying) connections between actors. These models parallel from the theoretical side extant empirical spatial-lag models of spatial econometrics, notably in the exogeneity of the connectivity matrix, i.e., of the network of connections between actors.² In other words, these models expressly disallow network-formation/selection. Accordingly, we extend these extant Markov type-interaction models to incorporate endogenous determination of the connections between units, connections being made or broken endogenously (to an extent that we can vary with parameters of the model) by the previous behavior-types of those units. Likewise, empirically, we merge extant spatial-lag models of interdependent behavior—specifically, a very simple version of the *spatial-lag logit* model³—which have typically maintained exogenous connections between units, with extant models of network formation, *p-star* models—specifically the simplest *p-star* model of independent ties⁴—which have typically maintained exogenous unit characteristics, including behaviors, as explanators of network ties. Theoretically and empirically, the emergent models are ones of network-behavior *coevolution*.

The combination of network effects, specifically of behavioral contagion,⁵ and of network formation with self-selection of actors into networks, specifically of actors’ self-selection of the ties between actors according to some (dis)similarity or other function of the actors’ behaviors or types (e.g., *homophily*), implies that networks and behavior *coevolve* over time. Both network effects and network formation, i.e., both contagion and selection, are ubiquitous and frequently important across the social sciences (as are the usual plethora of exogenous conditions relevant to both). Therefore, in longitudinal-network or spatiotemporal analyses, scholars must take seriously the modeling of all three processes—common exposure, contagion, and selection—if they are to understand the nature of and properly model and estimate the structure of the coevolutionary dynamics in their data, i.e., if they are to explain accurately the network formation and dynamics and the behavioral decision-making evidenced in their data. In this paper, we emphasize that one of the more interesting consequences of such network-behavior coevolution is that it can produce history-dependent political dynamics, including what (Page 2006) defines as *phat*, path, and/or equilibrium depen-

²In practice, most theoretical Markov type-interaction models have employed very simple, uniform and universal, and therefore anonymous, connections between actors; i.e., all actors are equally or equiprobably connected to all others and therefore are anonymously exchangeable in this sense. The extension of our model to endogenous selection of ties on the basis of past behavior-types must forego this anonymity; reformulating the model to keep track of these individuals and dyads itself proved a nontrivial extension.

³For now, to start, we employ only a time lag of the spatial-lag dependent variable, and assume this an adequate model of the spatiotemporal dynamics of contagion, thereby evading the multidimensional-integration complications of simultaneous spatial-lags in latent-variable models (see, e.g., Franzese and Hays (2009); Hays et al. (2010)).

⁴Given the assumed conditional independence of the network ties across ties and actors in that simplest model—again, for now, to start—this amounts very simply to a set of $\frac{1}{2}N(N-1)$ simple logits.

⁵As previously mentioned, network effects more generally include (i) effects of the network structure on nodes, (ii) effects of nodes’ positions within the network on nodes’ behaviors, and effects through the network connections of *alter* on *ego*. We focus, for now at least, on the last: *contagion*.

dence.⁶ We first establish theoretically that systems with coevolution can easily generate multiple equilibria (i.e., multiple steady states of the system), using the aforementioned modified Markov type-interaction model extended to allow endogenous tie-formation. The potential of multiple equilibria raises a very difficult empirical question—how sensitive are equilibrium distributions (over types) to the past states? Nor are the stakes in this question merely academically nontrivial. What can be achieved by potential policy interventions today and how we should design policy interventions for the future depend critically on whether and how history matters in path- or path-dependent ways for the equilibrium attained in the society. To evaluate the empirical magnitude and substantive and statistical significance of coevolutionary dynamics, therefore, we combine as just noted the spatial-lag logit and *p-star* logit models to develop discrete-time Markov models that can estimate the empirical magnitude and significance of any coevolutionary dynamics in the data. One strength of this empirical approach lies in its direct connection with the theoretical Markov type-interaction model, which allows us to assess the full substantive content of history dependence in observed data and which can provide a foundation for developing statistical tests for history dependence generated by coevolution.

The most-developed extant approach to network-behavior coevolution—also, to our knowledge, the *only* extant approach—is Snijders and colleagues’ (Snijders 1997, 2001, 2005; Steglich et al. 2006; Snijders et al. 2007; Ripley and Snijders 2010) *stochastic actor-oriented models* for longitudinal social network-analysis. In this paper, we introduce and briefly explain Snijders and colleagues’ impressive coevolutionary actor-oriented longitudinal-network models and estimation techniques, *Siena* (*Simulation Investigation for Empirical Network Analysis*), and compare them to our own, much simpler, logistic-model estimation-strategy. *Siena*, originally implemented in the stand-alone Windows software, *Siena*, and now also available as an R package, *RSiena*, estimates by simulated method-of-moments certain continuous-time Markov models of network and/or behavior change (and of the actors chosen to make these changes and the inter-observational rates of these changes, if and as desired). We compare our simple estimation strategy, designed with our discrete-time model of endogenous coevolution in mind, to *RSiena* estimation of Snijders and colleagues’ continuous-time stochastic actor-oriented coevolution model in Monte Carlo analysis of data generated from code that we wrote to replicate the exact process corresponding to the *RSiena* model we estimate. To our knowledge, these may be the first Monte Carlo evaluations of the *Siena* estimator (simpler coevolutionary models have been evaluated to some extent in, e.g., Leenders (1995, 1997)), and these seem more certainly to be the first comparison of *Siena* to an alternative (perhaps the first alternative).

The rest of the paper proceeds as follows. In the next section, we present a theoretical Markov type-interaction model for N actors, modified to allow endogenous tie-formation. In the third section, we propose our statistical model, specified to reflect the theoretical model proposed in the preceding section, and compare it to *Siena*. The fourth section provides Monte Carlo evaluation and comparison of the performance of *Siena* and our simple logistic alternative, in data generated by the process specified in the *RSiena* estimator. The fifth section conducts an illustrative application that examines the coevolution of the military alliances and conflict behavior of major powers in the first half of the twentieth century, again comparing our proposed model and estimation strategy with Snijders and colleagues’ coevolutionary actor-oriented longitudinal-network model, *Siena* (Snijders 2005; Steglich et al. 2006; Snijders et al. 2007; Ripley and Snijders 2010). The concluding section

⁶For an exploration of the current state of formal-theoretical and empirical exploration of path dependence, a review might include Jackson and Kollman (2007); Jackson (2008); Page (2006, 2007); Walker (2007).

offers discussion and an agenda for important next steps in this project.⁷

2 A Discrete-Time Markov-Chain Theoretical Model of Network-Behavior Coevolution

In this section, we introduce a relatively simple theoretical model of behavior-shaping (network contagion effects) and network-formation (selection) that results in a set of Markov chains. In this model, a group of actors are of certain types and their types change over time as actors are influenced by other actors (and exogenous factors). Such contagion of actors' types occurs only if the actors are *connected*. We employ the explicit notion of networks (or spatial weights) to characterize such *connectedness*. Simultaneously, the connectivity of actors also changes over time, not only due to exogenous factors, but also as a function of types taken by actors in the previous period. We particularly focus on behavior-type *homophily* (or *heterophily*), i.e., where network ties are more (or less) likely to form among actors whose behavior-types are more (or less) similar in the previous period. The key features of and the additional sources of complexity due to this extension of extant type-interaction models are the following: (1) it introduces the details of which actor interacts with which to represent the effects—specifically, the contagion effects⁸—of networks on actors' behavior and (2) it describes how those interaction patterns, i.e., the networks *cum* spatial-weights matrices, change endogenously over time based on actors' types in the previous time period, which reflects homophily by behavior-type.⁹

⁷**NOTE:** THIS IS VERY MUCH A WORK IN PROGRESS! In future work, we intend to compare more fully our empirical model and estimation strategy to Snijders and colleagues' *Siena*, including Monte Carlo simulation of properties related to comparable estimates of *effects*, dynamics, and steady-state *effects*. For now, the fourth section contains only evaluation of *Siena* *parameter*-estimate bias/consistency, efficiency, standard-error and test-size accuracy, and test power; evaluation of our estimator's standard-error accuracy, test-size, and power (because the DGP reflected a different model than our estimator and we have not yet derived the corresponding correct parameter values); and a comparison of relative standard-error accuracy, and test power and size-accuracy. We also intend to explore seemingly-unrelated and multivariate-logit extensions of our simple estimator that should enhance variance-covariance estimates and parameter-estimate efficiency, respectively, at relatively low added complexity. We also hope to add consideration of simultaneous interdependence across in network-tie formation and in behavioral choice, although this will come only at much greater added-complexity. Finally, we see these particular models, with contagion and selection both occurring through binary variables observed by the actors and analysts, as only one case of a family of models much in need of development for the diversity of social-science coevolution contexts, models in which behavior or ties may be binary or continuous quantities, contagion and selection may occur through actors' observed or latent quantities, and analyst may observe or not observe these quantities. Explore the proposed Kalman filter EM algorithm for the hard model. On the theoretical side, we also have yet to characterize fully the forms of history dependence—outcome and equilibrium state, path, path dependency (Page 2006)—that emerge from alternative parameter and starting values and the mappings from the latter to the former. We review this long list of next tasks in the conclusion.

⁸As previously mentioned, network effects more generally include three sorts of effects: (i) effects of the network structure (*density* or *hub-and-spoke* structure for instance) on nodes, (ii) effects of nodes' positions within the network (their *centrality* or *betweenness* for example) on nodes' behaviors, and effects through the network of connections of other nodes' characteristics or behaviors on nodes' behaviors (of *alter* on *ego* in network terminology). We focus, for now at least, at first, on the last of these, commonly labeled *contagion* in much of the relevant literatures.

⁹In general, homophily in the network analysis literature simply refers to the phenomenon in which actors are more likely to be connected (friends) with others who are similar to themselves. Similarity can be in various different characteristics of actors. Earlier discussions of homophily, including one of the earliest by Lazarsfeld and Merton (1954), focused on the similarity in actor-specific attributes, such as race, age, gender and occupation (for more thorough discussions of earlier studies of homophily, see McPherson et al. (2001)). Note that in the studies of behavioral formation, these characteristics are treated as exogenous. However, similarity in the explained behavior

We then demonstrate that the model can produce long-run equilibria that depend both on the starting values *and* the history. By equilibrium, we mean consistency between actors’ behavioral types and their behavior-switching rules (i.e., steady state or fixed point). In coevolution models, multiple equilibrium (distributions of) types are consistent with a single behavior-switching rule. The particular type/behavioral-rule combination that emerges at a given point in time is a function of actors’ prior types. In this way, the evolution of behavioral types is history dependent and may be specifically *phat*, path, or equilibrium dependent.¹⁰

2.1 A Minimal Coevolutionary Model

The complexity of models with endogenous behavior, which is contagious via connections between actors, and endogenous connections, which depend in turn on behavior through actors’ homophiletic preferences, grows rapidly as terms and parameters accumulate. To isolate and clarify the critical source of history dependence with some numerical examples, we focus first on an extremely simple theoretical model, just sufficient to incorporate both contagion in node behavior and behavioral homophily in network-tie formation, and show that and how this suffices to generate equilibrium dependency.

Consider the following discrete-time longitudinal process with N actors. Let $i \in \{1, \dots, N\}$ denote these N actors and $t \in \{1, 2, \dots\}$ denote time periods. We distinguish between the *behavior* of an actor and her *behavioral type* (or simply *type*), understanding the latter as the actors’ probability of taking action 1. In our initial models here, we assume that behavior is observed and dichotomous, whereas behavior-type is continuous and unobserved by analysts but observed by actors, with contagion and selection occurring by type.¹¹ In each period, actors choose between *behavior* 1 and 0, e.g., smoking or not, voting or not, taking an aggressive interstate behavior or not, democratizing or not, and so forth. We denote *behavior* of actor i in period t by $s_{it} \in \{1, 0\}$. *Behavioral type* is the probability that an actor chooses behavior 1. We denote the *behavioral type* of actor i in period t by $\sigma_{it} \in [0, 1]$. The state of the system that actors can observe at the end of period t is therefore an N -dimensional vector of types, $\boldsymbol{\sigma}_t = (\sigma_{1t}, \dots, \sigma_{Nt})$, a corresponding vector of behaviors, $\mathbf{s}_t = (s_{1t}, \dots, s_{Nt})$,¹² and a matrix of latent and observed ties between actors to be described subsequently.

itself could also be an important determinant of network shaping. This endogenous source of homophily (endogenous to the behavior that is to be explained in the model) is the one we emphasize in this paper. Lazer (2001), for example, is among the few studies to our knowledge that addresses this issue of homophily in behavior. Also, we consider a homophiletic dynamic that affects both the forming and the severing of ties among actors. (As Noel and Nyhan (2010) show, ignoring effects in one direction importantly diminishes the consequences of misspecifying coevolutionary processes as only involving selection or contagion).

¹⁰As previously noted, ultimately, we aim to characterize more fully the forms of history dependence—outcome and equilibrium state, *phat*, path dependency (Page 2006)—that emerge from alternative parameter and starting values and the mappings from the latter to the former.

¹¹As previously mentioned, we envision eventually a range of possible models, with continuous or discrete behavior, observed or unobserved by analysts and/or by actors, with contagion or selection by behavior or type. The theoretical, substantive, and empirical appeal of these alternative models would presumably vary with the application context. We begin with the model described here because we believe it an appealing one for the application to be offered in section five, because this theoretical model maps well into the empirical model to be offered in section three and evaluated and compared to *Siena* in section four, and because we must begin somewhere.

¹²The behaviors, \mathbf{s}_t , are less prominent in the theoretical model here than in the empirical models to come since this theoretical model has type observed and the basis of contagion and selection.

2.1.1 The Behavior-Type Markov Chains

We focus first on the Markov chains that explain behavioral type, accounting the contagion effects of networks among the N actors. Equation (1) describes the transition of actor i 's type from period t to $t + 1$. This system results in a set of N Markov chains:

$$\begin{aligned}
 & \begin{pmatrix} Pr(s_{i,t+1} = 1|s_{it}) \\ Pr(s_{i,t+1} = 0|s_{it}) \end{pmatrix}' = \begin{pmatrix} \sigma_{i,t+1} \\ 1 - \sigma_{i,t+1} \end{pmatrix}' \\
 & = \begin{pmatrix} \sigma_{it} \\ 1 - \sigma_{it} \end{pmatrix}' \begin{pmatrix} \sigma_{it} & 1 - \sigma_{it} \\ 1 - [c_{1i}(1 - \sigma_{it}) + (1 - c_{1i})\frac{\sum_{j \neq i}(\delta_{ij}(1 - \sigma_j))}{N-1}] & c_{1i}(1 - \sigma_{it}) + (1 - c_{1i})\frac{\sum_{j \neq i}(\delta_{ij}(1 - \sigma_j))}{N-1} \end{pmatrix} \quad (1) \\
 & = \begin{pmatrix} \sigma_{1t} \\ 1 - \sigma_{1t} \end{pmatrix}' \begin{pmatrix} \pi_{11} & \pi_{12}(= 1 - \pi_{11}) \\ \pi_{21}(= 1 - \pi_{22}) & \pi_{22} \end{pmatrix}.
 \end{aligned}$$

The second matrix on the right-hand side of the second and third lines is the transition-probability matrix. This matrix is premultiplied by the (row) vector of last-period's types to get this period's types. For example, cell (1, 1) of this matrix represents the probability that actor i chooses behavior 1 given that i chose 1 in the past period, $Pr(s_{i,t+1} = 1|s_{it} = 1)$. Note that the probability of transitioning from either one specific state to *some* new state is 1; therefore, the transition matrix is *row* (or *right*) *stochastic*: the elements in each row must sum to 1. For a dichotomous-choice model, therefore, defining any one element of each row, the probabilities of staying in state 1 or 0 given in cells (1, 1) and (2, 2) for instance, suffices to complete the transition matrix.

We separate two component terms that together determine a *switching probability*—the probability that an actor's behavior at $t + 1$ becomes $s_{i,t+1}$ or $1 - s_{i,t+1}$ from $s_{i,t}$ or $1 - s_{i,t}$. First, we have a temporal autoregressive component of the probability, which depends on her time- t latent type. An actor is more likely to maintain behavior 1 at $t + 1$ when her latent type at t , σ_{it} , is (closer to) 1. The first row of the transition matrix—whose columns contain, respectively, the “staying” and “switching” probabilities *from* behavior 1—consists only of this autoregressive term. In the second row, which relates the probabilities of staying or switching from 0, element (2, 2), the probability of staying in state 0, has two terms. The first term, $c_{1i}(1 - \sigma_{it})$, captures the autoregressive effect, with the coefficient $c_{1i} \in [0, 1]$ indicating the relative role of autoregression for determining the (non-)switching probability for actor i to stay with behavior 0 the next period, $t + 1$. We introduce the second term in this “staying” probability to reflect contagion (i.e., spatial autoregression).

Each actor's behavioral decisions (i.e., *ego's* choices, in network-analytic terminology) can also be influenced by others' (*alters'*) types; this is the *contagion* in our terminology. Actors in a given dyad (i and j , where $i \neq j$) influence each other's behavioral type only if they are *connected*. This *connectivity* among actors, the *ties* between them, could be friendship, military alliance, trade, and so on. Let $\delta_{ij,t} \in [0, 1]$ denote the probability that a tie exists between the two actors in dyad ij in period t ; we can also interpret this probability, isomorphically in this model, as the strength of the tie.¹³ Ties are undirected in this model; undirected ties are equivalent to symmetric spatial-weights matrices.¹⁴ In this N -actor system, we express the strength of contagion, the extent to which

¹³These alternatives are mathematically identical in this particular model because contagion and selection occur by the continuous, observed type and strength, not the dichotomous, unobserved behavior and tie. In models with other combinations of these conditions, the alternative interpretations may differ slightly due to the distinction between a behavior or tie of strength 1 with probability p and strength 0 with probability $1 - p$ versus a behavior or tie of constant strength p .

¹⁴Extension to the *directed-network/asymmetric-weights-matrices* case would greatly complicate the notation and

others' types influence i 's type, as $\sum_{j=1, (j \neq i)}^N (\delta_{ij} \sigma_j) / (N - 1)$, where we have weighted alters' (j 's) influence on ego (i) by $(N - 1)^{-1}$ to bound the value of this positive term by 1.¹⁵ Notice how, by this construction, the weight $(1 - c_{1i})$ captures the extent to which the types of others (to whom i is connected) influence i 's behavioral choice in $t + 1$. This gives us parameters within the transition model by which to vary the overall strength of contagion. For instance, in an extreme case where i chooses 1 with probability $c_{1i} = 1$, then $1 - c_{1i} = 0$. (This is exactly the assumption made for the first row, the case where $s_{it} = 1$), meaning that i 's decision will not be affected by any others to whom she is connected: i.e., the strength of contagion is 0. Then, the way we combine these two components of an actor's type assures that the whole expression for σ_i , which is the *probability* the actor chooses behavior, s_i to be 1, is likewise bounded $0 \leq \sigma_i \leq 1$.

Although we assume that contagion only operates on (alters only influence ego's) behavior $s_{it} = 0$ but not on behavior $s_{it} = 1$ to give the simplest possible model, contexts where such contagion operates asymmetrically are plausible: e.g., going to vote in period $t + 1$ when the person didn't vote in period t may be subject to influence from others—negative sanctions for failing one's *duty*, perhaps?—but maintaining a non-voting status may not be (or, *vice versa*, non-voting may be susceptible to *adverse* peer pressures but *positive* peer-presures toward voting may be much weaker. In any case, the lack of an explicit contagion term in the first row of the transition matrix does not indicate “no effect” of others' types. Shifting one period back, note that the probability in cell (1, 1) arises from $\sigma_{i,t} = \sigma_{i,t-1}^2 + (1 - \sigma_{i,t-1}) \left[1 - \left[c_{1i}(1 - \sigma_{i,t-1}) + (1 - c_{1i}) \frac{\sum_{j \neq i} (\delta_{ij}(1 - \sigma_{j,t-1}))}{N-1} \right] \right]$. Even in this stylized model with its asymmetric simplification, an indirect and time-lagged effect of others' types contributes to the propensity of behavior 1. Allowing an asymmetry like this would accommodate instances where we expected one behavior, 1 or 0 to be *stickier*.

Most centrally to our purposes regarding the parameters c : in terms of the three processes that may produce spatial/network association, the c_1 parameters are the theoretical placeholders for the relative weight of exogenous external and internal conditions—for now confined just to the temporal autoregressive factor—to which actors may have *common exposure*.

Note that, with the behavioral choice being dichotomous, the Markov chains for actors' types can also be sufficiently summarized in one line solely by a conditional probability: $Pr(s_{i,t+1} = 1 | s_{it}) = \sigma_{i,t+1}$.

$$\begin{aligned} Pr(s_{i,t+1} = 1 | s_{it}) &= \sigma_{i,t+1} = \sigma_{it} \pi_{11}^\sigma + (1 - \sigma_{it})(1 - \pi_{22}^\sigma) \\ &\Leftrightarrow \sigma_{i,t+1} = \sigma_{it}^2 + (1 - \sigma_{it}) \left[1 - \left[c_{1i}(1 - \sigma_{it}) + (1 - c_{1i}) \frac{\sum_{j \neq i} (\delta_{ij}(1 - \sigma_j))}{N-1} \right] \right]. \end{aligned} \quad (2)$$

2.1.2 The Network-Tie Markov Chains

We focus next on the Markov chains that explain tie-formation probabilities (Equation (3)). The unit of each entry is now a dyad (i, j) . The probability that the two actors i and j in a dyad form a tie (e.g., “are friends”) in period $t + 1$ is denoted by $\delta_{ij,t+1} \in [0, 1]$. In conventional type-interaction

accounting of this theoretical model, but should be otherwise relatively straightforward.

¹⁵The row standardization common in spatial econometrics, or the spectral normalization that Kelejian and Prucha (2009) recommend instead, would also serve to bound $0 \leq \sum_{i \neq j} (\delta_{ij} \sigma_j) \leq 1$, but our $1/(N - 1)$ weighting greatly economizes notational and accounting requirements.

models, these tie-formation probabilities are assumed exogenous (and often uniform, in fact). By contrast, the potential for endogenous ties—for instance via behavior-type homophily (heterophily): types seeking like (different) types—is a crucial aspect of our coevolutionary model. In our model, similarity or dissimilarity of actors’ types at t partially determines the tie-formation probabilities in the next time period. In this minimalist model, we allow actors’ preferences for connections to other actors with similar behaviors, call this *behavioral homophily*, to influence the (non-)switching probability from “no tie” at t to “no tie” at $t + 1$. This is the second term, $(1 - c_{2,ij})(\sigma_i - \sigma_j)^2$, of element (2,2) of the transition matrix, $c_{2,ij}(1 - \delta_{ij,t}) + (1 - c_{2,ij})(\sigma_i - \sigma_j)^2$. For homophily, as the distance between two actors’ behavioral types increases, it becomes less likely for their dyad to form a new tie; i.e., $\frac{\partial Pr(d_{ij,t+1}=0|d_{ij,t}=0)}{\partial(\sigma_i - \sigma_j)} > 0$. So, to the degree that $c_2 < 1$, our model exhibits homophiletic tie-formation by behavior-type: two individuals are more likely to form a friendship tie if both smoke, two countries more likely to ally if their conflict behaviors are similar, two members of Congress more likely cosponsor a bill the more similar their political ideologies, etc. As was the case for the behavior transition matrix, a temporal autoregressive term is the other factor—remember: standing in for all *common-exposure* factors—affecting the switching probability. In this case, $\delta_{ij,t}$ in cell (1,1) and $c_{2,ij}(1 - \delta_{ij,t})$ in cell (2,2) contain this information. Analogously to c_1 , the term c_2 reflects the strength of the temporal autoregressive (common-exposure) term relative to the homophily (selection) term in the switching probability. Conversely, $(1 - c_{2,ij})$ reflects the remaining extent to which the distance between the two actors’ types prevents them from establishing a tie. The combinatorial form of the weights on the temporal autoregressive and homophily terms, $c_{2,ij}$ and $(1 - c_{2,ij})$, bounds the value entering cell (2,2) to the $[0, 1]$ range, and the combinatorial form of the elements in (2,1) and (2,2) assure the row-sum is 1.

$$\begin{aligned}
& \begin{pmatrix} Pr(d_{ij,t+1} = 1|d_{ij,t}) \\ Pr(d_{ij,t+1} = 0|d_{ij,t}) \end{pmatrix}' \\
&= \begin{pmatrix} \delta_{ij,t+1} \\ 1 - \delta_{ij,t+1} \end{pmatrix}' \\
&= \begin{pmatrix} \delta_{ij,t} \\ 1 - \delta_{ij,t} \end{pmatrix}' \begin{pmatrix} \delta_{ij,t} & 1 - \delta_{ij,t} \\ 1 - [c_{2,ij}(1 - \delta_{ij,t}) + (1 - c_{2,ij})(\sigma_i - \sigma_j)^2] & c_{2,ij}(1 - \delta_{ij,t}) + (1 - c_{2,ij})(\sigma_i - \sigma_j)^2 \end{pmatrix} \quad (3) \\
&= \begin{pmatrix} \delta_{1,t} \\ 1 - \delta_{1,t} \end{pmatrix}' \begin{pmatrix} \pi_{11}^\delta & \pi_{12}^\delta (= 1 - \pi_{11}^\delta) \\ \pi_{21}^\delta (= \pi_{22}^\delta) & \pi_{22}^\delta \end{pmatrix}.
\end{aligned}$$

As before, we can write the Markov chains for network-tie formation, ties being dichotomous, in a one-line conditional-probability:

$$\begin{aligned}
Pr(d_{ij,t+1} = 1|d_{ij,t}) &= \delta_{ij,t+1} = \delta_{ij,t}\pi_{11}^\delta + (1 - \delta_{ij,t})(1 - \pi_{22}^\delta) \\
&\Leftrightarrow \delta_{ij,t+1} = \delta_{ij,t}^2 + (1 - \delta_{ij,t}) \left[1 - [c_{2,ij}(1 - \delta_{ij,t}) + (1 - c_{2,ij})(\sigma_i - \sigma_j)^2] \right]. \quad (4)
\end{aligned}$$

The systems of differential equations (1) and (3) complete our minimalist theoretical model of network-behavior coevolution, i.e., of jointly endogenous contagion and selection. The long-run steady-state (LRSS) equilibrium of this system consists of a vector of each actor’s type and each dyad’s tie-forming probability, $(\boldsymbol{\sigma}, \boldsymbol{\delta})$. This LRSS equilibrium can be obtained by solving (1) and

(3) for σ by setting $\sigma_{i,t+1} = \sigma_{i,t}$ and $\delta_{ij,t+1} = \delta_{ij,t}$, $\forall i, j \in \{1, \dots, N\}$. In any given social-science context, our interests may lie primarily in the LRSS equilibria and/or the intertemporal dynamic of actors' types, σ , and/or in the LRSS and/or dynamics of tie-formation probabilities, δ . Of course, our interests may also involve all of the above equally. In any case, the states and the equilibria are fully and best characterized by vectors of types and tie-probabilities both, due to the endogeneity generated by homophily and contagion. For compactness only, our exposition will highlight the equilibria of behavior types and (network) ties.

2.1.3 Example: Two-Actor System

To illustrate the existence of multiple equilibria in this highly simplified coevolution model, consider the following example with two actors, ($i = \{1, 2\}$), and so one possible undirected edge, (12). This gives the following system of equations, a set of three equations of motion, one for each of the two actors' behavior-type processes and one for the dyads' tie-formation process:

$$\begin{cases} \sigma_{1,t+1} = \sigma_{1t}^2 + (1 - \sigma_{1t}) \left[1 - \left\{ c_{11}(1 - \sigma_{1t}) + (1 - c_{11})\delta_{12}(1 - \sigma_2) \right\} \right] \\ \sigma_{2,t+1} = \sigma_{2t}^2 + (1 - \sigma_{2t}) \left[1 - \left\{ c_{12}(1 - \sigma_{2t}) + (1 - c_{12})\delta_{12}(1 - \sigma_1) \right\} \right] \\ \delta_{12,t+1} = \delta_{12,t}^2 + (1 - \delta_{12,t}) \left[1 - \left\{ c_{2,12}(1 - \delta_{12,t}) + (1 - c_{2,12})(\sigma_1 - \sigma_2)^2 \right\} \right]. \end{cases} \quad (5)$$

Solving this system for the steady-state σ_i 's and δ_i 's, we get the following equilibria:

$$[\text{Equilibria}] \quad \{\sigma_1 = \sigma_2, \delta_{12} = 1\}.$$

This result demonstrates the system has multiple equilibria: *any* $\sigma_1 = \sigma_2$ and $\delta_{12} = 1$. Depending on the initial conditions, the steady state obtains with the actors connected in any equilibrium but at different $\sigma_1 = \sigma_2$ across the possible equilibria. Although an equilibrium of such form might look unfamiliar in that no explicit function states how exogenous parameters determine the equilibrium values of σ , this equilibrium actually highlights clearly the strong history dependency generated by this particular example of the complex dynamics of contagion and homophily. Rigorous definitions of history dependence (as in Page (2006), e.g.) distinguish the effects of initial conditions, early conditions, and the sequence of past conditions shaping equilibria and, distinctly, outcomes along the way. In this case, the model indicates that the two actors being connected and both choosing the same behavior type, either type, is a steady state. Accordingly, at which type they settle will depend (at least) on where they start.

2.2 Illustration of the History Dependence in Behavioral Types

So far we have shown that N -actor system generates multiple equilibria, given some set of initial values for the endogenous variables and exogenous parameters, and given the two core dynamics of our model: behavior-type contagion and behavior-homophiletic selection. In this section, we further look into the system dynamics and demonstrate that the system is, in fact, *history dependent*.

The concept of *history dependence* refers to the phenomenon where initial conditions or changes in conditions along the system's history alters the course of its history in the future. This broad

concept of history dependence is often conflated with the much narrower concept, *path dependence*, but we follow Page (2006) in defining history dependence as the broadest category and in differentiating *state*, *phat* and *path dependence* as three subcategories of it. Path dependence, which is the most restrictive form of history dependence, means that history of a system depends on the *path*, i.e., the *sequence*, the *order*, of past outcomes. Notice that a situation where specifically when some condition arose, in absolute time or relative to other conditions, matters is an example of path dependence. The less restrictive “phat” dependence implies that a system’s history depends on past outcomes but not their order. This means essentially that the future path of the system depends on the set of things that have happened before, not their sequence. State dependence is the least restrictive form of history dependence wherein a system’s trajectories can be partitioned into a finite number of states that contain all the relevant information that shapes the history of the system regardless of what happened outside the partition of its history. This means, essentially, that the future path of the system depends on its current state, not the path or the set of previous conditions. We have defined *state*, *phat* and *path* dependence as disjoint sets of increasing restrictiveness; path means order matters, phat means set but not order matters, and state means current but not previous set or order matters. We could equivalently have defined them as overlapping, so that each form of history dependence included the less-restrictive subsets. In either case, history dependence is the superset including all three.¹⁶

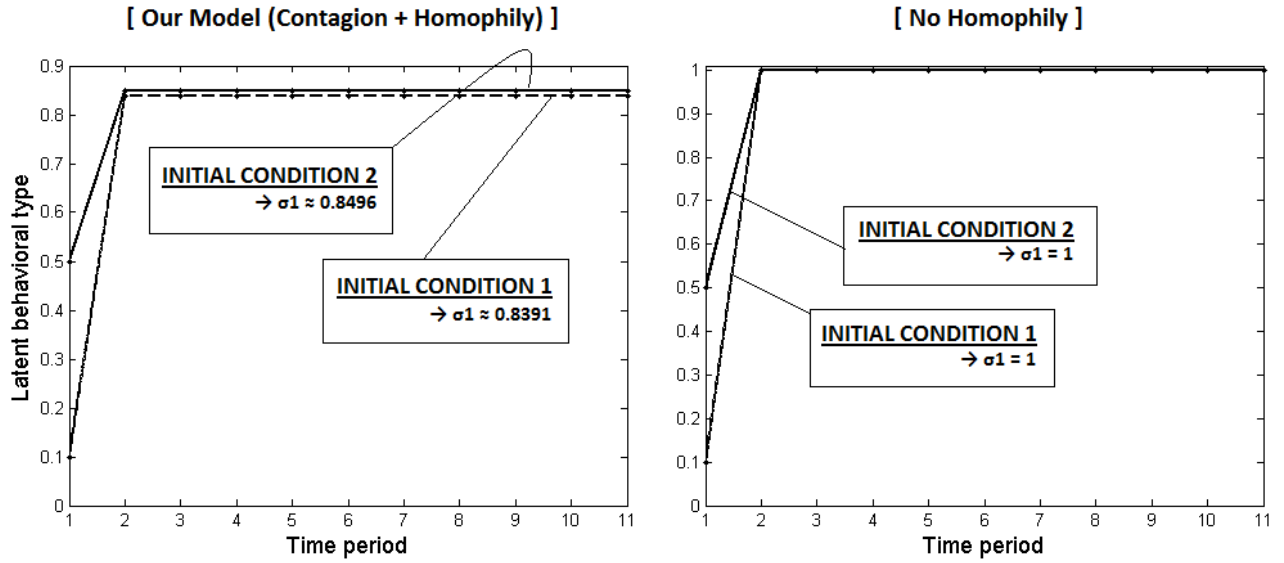
The distinction between *outcome* and *equilibrium* dependence is another important dimension in defining history dependence. A system demonstrates outcome history dependence if the outcome in each time period (e.g., behavioral choice in time t) depends in some manner on the outcome(s) in the past time periods (e.g., behavioral choice at $t-1$ or earlier) or on the time index itself. As can be seen in the difference equations (1) and (3), the process in our model is outcome history-dependent by construction (as is any model with temporal autoregression, of course). More interesting for this paper is equilibrium dependence, which focuses on history (path, phat, or state) dependence of the long-run distribution of outcomes. In our context, this is equivalent to asking whether the long-run steady-states (LRSS) of the latent behavioral-types, σ , and strength-of-ties (δ) depend on their history (sequence, set, or state).

2.2.1 The Steady State of the Two-Actor System: A Numerical Example

We illustrate the forms of history dependence our model may exhibit by a series of numerical exercises given sets of values for the initial behavior types, σ , and tie probability, δ , and for the exogenous parameters, c ’s.¹⁷ First, we present a numerical example of the LRSS equilibrium of our model (Figure 1). This merely demonstrates, for a specific set of starting values for σ and δ , and fixed \mathbf{c} , the solution to the two-actor system presented in Equation (5).

The left-side graph of Figure (1) shows the trajectory of actor 1’s behavioral type (σ_{1t}) from $t = 1$ through $t = 11$ for the case where the Markov chains (in Equation (5)) are solved recursively from two distinct sets of initial values: $\{\sigma_{11} = \sigma_{21} = \delta_{12,1} = 0.1\}$ and $\{\sigma_{11} = \sigma_{21} = \delta_{12,1} = 0.5\}$. The parameter values are fixed at $c_{11} = 0.2$, $c_{12} = 0.9$, and $c_2 = 0.6$ through all 11 periods for both experiments. The LRSS levels of behavioral type are approximately 0.8391 and 0.9496 respectively

Figure 1: Illustration of the Initial-Conditions Sensitivity of the LRSS Equilibrium Behavioral-Type in the “Simplest Coevolutionary System”



		Our Model		No Homophily	
		INITIAL COND. 1	INITIAL COND. 2	INITIAL COND. 1	INITIAL COND. 2
Exogenous parameters (constant over time)	c_{11}	0.2	0.2	0.2	0.2
	c_{12}	0.9	0.9	0.9	0.9
	c_2	0.6	0.6	-	-
Initial values	σ_{11}	0.1	0.5	0.1	0.5
	σ_{21}	0.1	0.5	0.1	0.5
	$\delta_{ij,1}$	0.1	0.5	0.1	0.5
Steady State σ_1		≈ 0.8391	≈ 0.8496	1	1

(note that $\sigma_2 = \sigma_1$ in equilibrium, as summarized in Section 2.1.3).

These numerical examples of how the long-run-steady-state (LRSS) types for our two-actor system vary with particular starting values of the endogenous variables, σ and δ , even when the exogenous parameters (c 's) do not change over time, illustrate the *initial-conditions equilibrium-dependence* of this system. They do not (yet) demonstrate any of the richer forms of history dependence, state, phat, or path dependence. Conventional type-interaction models do not exhibit even this initial-conditions dependence, however. For example, imagine a similar system to our minimal-coevolution system, but without homophily in network-tie formation; i.e., the parameter c_2 is restricted to 0. This system still exhibits connectivity among the actors, explicitly defined with ties δ , and contagion along the pattern given by δ ; but the connectivity pattern, the network, is exogenously fixed. (Spatial econometricians and statisticians will recognize this as the typical spatial-lag model in which \mathbf{W} is taken to be fixed and exogenous.) This “no-homophily, contagion only” model has a unique equilibrium, $\{\sigma_1 = \sigma_2 = 1, \delta_{12} = \text{exogenously given}\}$. Consequently, starting values for σ and δ do not matter. The LRSS behavioral types always converge to 1. This is demonstrated in the right panel of Figure (1).

2.2.2 Early-Condition Sensitivity (*Phat* Dependence): Two-Actor Case

To analyze history dependence, we focus on the effects of changes in the parameters \mathbf{c} over the course of the system’s history. As mentioned previously, these parameters, \mathbf{c} , as coefficients on the temporal autoregressive terms serve as the theoretical placeholders for all the actor or dyad-specific, exogenous conditions to which actors/dyads may have common exposure. In this role, they are analogous to the $\mathbf{X}\beta$ of a regression model. As such, history dependence on these \mathbf{c} is theoretically more interesting and practically more important than dependence on starting values of the types. Scholars may be particularly interested in whether some intervention in an actor/dyad-specific attribute, x —hence a change in some c , which may be possible whereas manipulating initial states is generally not, alters the future path and LRSS characteristics of society because, when the system is not history dependent, such interventions will not change the long-run equilibrium. (Recall that our definition of history dependence follows Page (2006); our conceptualization of history dependence as related to the parameters in a nonlinear system of equations is analogously similar to Jackson and Kollman (2007).)

To illustrate such history dependence in \mathbf{c} , we start with the simplest form of phat (set, but not order) dependency, *early-condition sensitivity*. Figure 2 plots the trajectory of actor 1’s LRSS behavioral type under two different sets of parameters. The system with Condition 1 starts with a parameter set $\{c_{11} = 0.2, c_{12} = 0.9, c_2 = 0.6\}$ and some intervention occurs at the end of $t = 1$, changing the parameters to $\{c_{11} = 0.2, c_{12} = 0.1, c_2 = 0.1\}$. From $t = 2$ on, this set of parameter persists. Note that the system reaches a steady state from its previous state in each period; the graph plots these period-by-period steady states. The system with Condition 2 starts with a different parameter set $\{c_{11} = 0.1, c_{12} = 0.1, c_2 = 0.3\}$, and an intervention also occurs at the end of $t = 1$, again changing the parameters to $\{c_{11} = 0.2, c_{12} = 0.1, c_2 = 0.1\}$. The two systems start with different sets of parameters but shift to the same set at $t = 2$ and stay with that set thereafter. We emphasize that the starting values for the endogenous variables, the behaviors and ties σ and

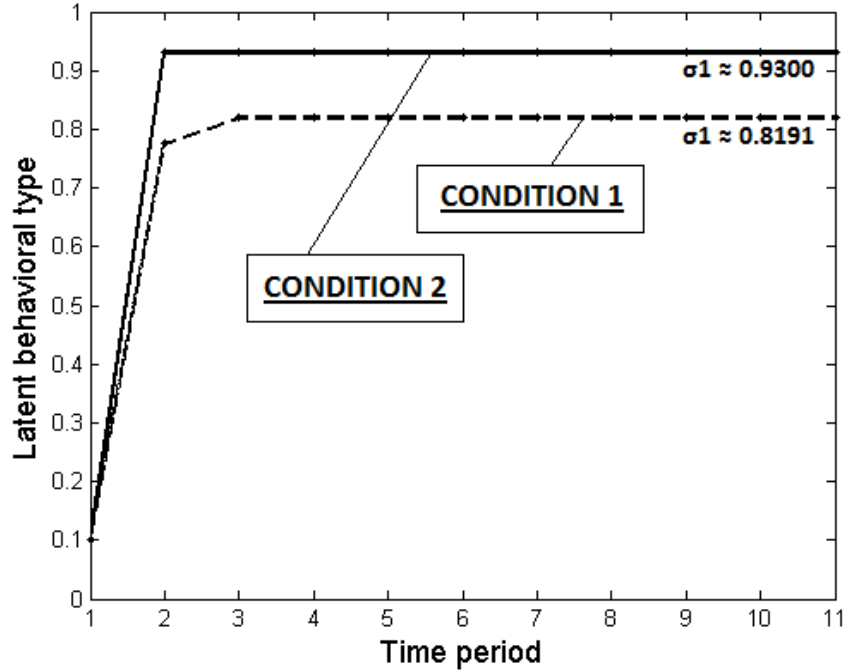
¹⁶In the overlapping sets case, this would make state dependence and history dependence synonyms.

¹⁷We plan to study exactly what assumption(s) of the ergodic theorem is(are) violated in future versions.

δ , are set at the same level for both experiments. The difference in the two trajectories stems solely from the shock in \mathbf{c} .

The result is plain: early conditions in terms of the exogenous parameters \mathbf{c} matter. The system with contagion and behavioral homophily is *phat* dependent, at least.

Figure 2: Illustration of Early-Condition Sensitivity (Equilibrium *Phat*-Dependence) in the “Simplest Coevolutionary System”: Comparison of LRSS Equilibrium Behavior-Types Following 2 Paths



	EARLY CONDITION 1				EARLY CONDITION 2			
Time period→	1	2	3	...	1	2	3	...
c_{11}	0.2	0.2	0.2	...	0.1	0.2	0.2	...
c_{12}	0.9	0.1	0.1	...	0.1	0.1	0.1	...
c_2	0.6	0.1	0.1	...	0.3	0.1	0.1	...

Initial values for the endogenous parameters:
 $\sigma_1 = \sigma_2 = 0.1, \delta_{12} = 0.1.$

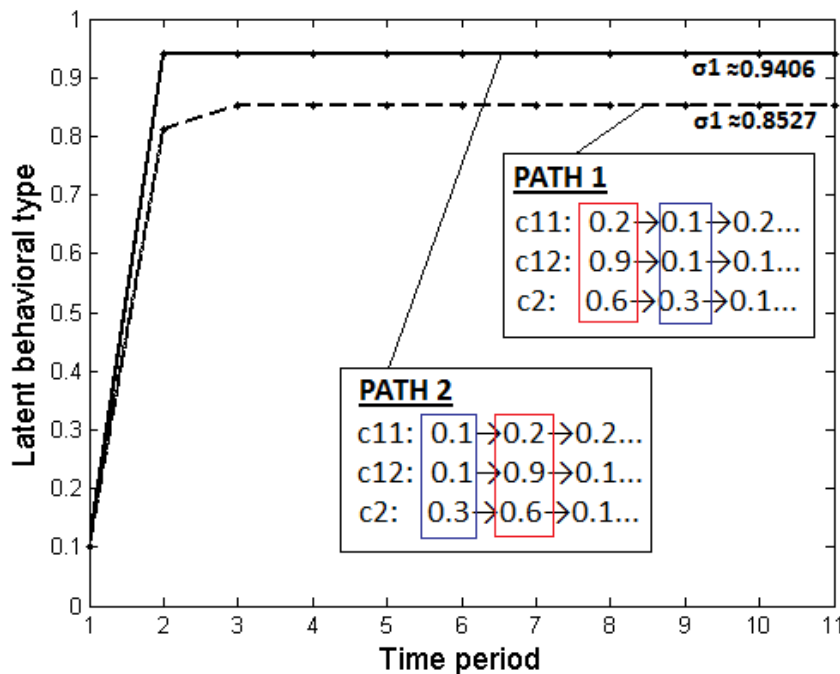
2.2.3 Path Dependency: Two-Actor Case

Finally, Figure 3 illustrates that equilibrium path dependence, the most restrictive form of history dependence (Page 2006), is also possible in our model, given coevolution, i.e., given both behavior-type contagion and behavior-homophiletic selection. In our context, the question of path dependence is equivalent to asking whether the LRSS of the latent behavioral type (σ_i) and strength

of tie (δ) depend on the order of past events. As before, we define history dependence of interest as relating to the set of exogenous parameters, $\{c_{11}, c_{12}, c_2\}$, which reflect actor/dyad-specific attributes in the empirical world related to exogenous factors and the strength of contagion in the case of c_1 and selection in the case of c_2 . The figure illustrates the dynamics and LRSS equilibrium behavior-type for actor 2 in the two-actor system. We consider two scenarios, labeled Path 1 and Path 2, that differ only by the sequence (but not the set) of exogenous values for c_{11} , c_{12} , and c_2 . The early part of the sequence differs—namely: the first two vectors (c_{11} , c_{12} , and c_2) reverse order in the two scenarios—but the history of exogenous factors from period 3 onward is constant within and across both scenarios. Also the starting values for the endogenous variables (σ and δ) are common between the two path experiments: $\{\sigma_1 = \sigma_2 = \delta_{ij} = 0.1\}$.

The two paths (precisely, the order of past events) generate different behavioral-type LRSS for this system: the system is path dependent. The behavioral-type LRSS for the system with Path 1 is $\sigma_1 = \sigma_2 \approx 0.9406$ (with $\delta_{12} = 1$), and for the one with Path 2, it is $\sigma_1 = \sigma_2 \approx 0.8527$ (with $\delta_{12} = 1$).

Figure 3: Illustration of Equilibrium Path-Dependence in the “Simplest Coevolutionary System”: Comparison of LRSS Equilibrium Behavior-Types Following 2 Paths



	PATH 1					PATH 2				
Time period→	1	2	3	4	...	1	2	3	4	...
c_{11}	0.2	0.1	0.2	0.2	...	0.1	0.2	0.2	0.2	...
c_{12}	0.9	0.1	0.1	0.1	...	0.1	0.9	0.1	0.1	...
c_2	0.6	0.3	0.1	0.1	...	0.3	0.6	0.1	0.1	...

Initial values for the endogenous parameters:
 $\sigma_1 = \sigma_2 = 0.1, \delta_{12} = 0.1.$

As one can see, the equilibria differ, and this difference does not fade in time (no matter how far into the future we may have extended the figure). Notice that the starting behavior-type and the set of values of (c_{11} , c_{12} , and c_2) are the same. Only the sequence of (c_{11} , c_{12} , and c_2) differs, so this is true path-dependence, as defined in Page (2006).

2.3 Summarizing the Theoretical Model: Network-Behavior Coevolution and Equilibrium Dependence

To summarize the main aspects of the proposed discrete-time Markov-chain type-switching model of coevolutionary dynamics: Endogenous coevolution of network (spatial) connections, which depend in part on the behaviors of the nodes so connected, and of node behaviors shaped in part by others' behaviors through that network, generates systems of nonlinear differential equations which can easily produce initial-condition, state, path, and path history-dependence in (outcomes and) equilibria. We specified the transition probabilities as a specific example of this sort of system with parameters reflecting temporal autoregression and, implicitly, other exogenous (or predetermined) covariates on the one hand—this hand reflecting what we have previously labeled *common exposure*—and, on the other hand, *contagion* through network connections in the behavioral model and homophily by behavior-type in the network-tie-formation model (*selection*). Our analysis of these sorts of models indicated that the existence of equilibrium history-dependence depended on the presence, jointly, of contagion and selection. Dynamic models of ties and/or behavior without both processes do not, in general, exhibit equilibrium dependence.

Table 1 compares the equilibrium characteristics of our type-interaction model to those from conventional type-interaction models, which do not allow for the coevolutionary dynamic generated by simultaneous homophily and contagion. The table summarizes equilibrium analyses over the four possible combinations of our key dynamics, with behavior-type contagion and homophiletic self selection. Mathematically, we turn these two dynamics on or off by setting the exogenous \mathbf{c} parameters $\neq 1$ or $= 1$: $c_{1i} = 1$ for models *without contagion*, $c_{1i} \neq 1$ for models *with contagion*, $c_{2i} = 1$ for models *without homophily* and $c_{2i} \neq 1$ for model *with homophily*.

As the equilibrium expressions manifestly demonstrate, in the three models that lack either or both of the contagion and/or selection dynamics (cells (1, 1), (1, 2) and (2, 1)), the steady-state values are constant: either at 1 or at the starting values of a given system (or given data). By contrast, the model with both contagion and selection (cell (2, 2)) exhibits steady-state behavioral types that depend both on the starting values *and* the history—specifically, the “path” of the system. The only regularity established by the analytical solution for σ_i is that $\sigma_i = \sigma_j$ must hold in equilibrium.

To repeat the crucial upshot: path dependence is generated *if and only if* the model exhibits *both* behavior-type contagion and behavior-homophiletic selection. This will suggest the form of a possible test for evidence of path dependence in the empirical model proposed below. Minor extensions of the model can produce *separating* equilibria, in which some actor(s) behave differently and/or some dyads connect and some do not, also.

Another way to appreciate the complexity of introduced by endogenous coevolution is to consider Page's *Rule of Six* (Page 2007): the numbers of actors plus choices must sum to at least six for a system to have multiple equilibria. With endogenous coevolution, even this simplest setting of

Table 1: Path-Dependence Comparisons for Type-Interaction Models

	No Contagion ($c_1 = 1$)	Contagion ($c_1 \neq 1$)
Exogenous Tie-Formation ($c_2 = 1$)	<p><i>No contagion in actors' behavior-types, and tie formation is also exogenous to behavior type.</i></p> <p>LRSS: $\sigma_{it}^* = \sigma_{i0}, \forall i, \delta_{ij,t}^* = \delta_{ij,0}, \forall i, j$</p> <p>Path Independent</p>	<p><i>Behavior type is contagious, but tie formation is exogenous to behavior type.</i></p> <p>LRSS: $\sigma_{it}^* = 1, \forall i, \delta_{ij,t}^* = \delta_{ij,0}, \forall i, j$</p> <p>Path Independent</p>
Endogenous Tie-Formation (homophily) ($c_2 \neq 1$)	<p><i>No contagion in actors' behavior-types, but tie formation is endogenous to behavior types, with actors more likely to form ties with similar behavior-types (homophily).</i></p> <p>LRSS: $\sigma_{it}^* = \sigma_{i0}, \forall i, \delta_{ij,t}^* = 1, \forall i, j$</p> <p>Path Independent</p>	<p>[Our Model]</p> <p><i>Behavior type is contagious, AND tie formation is endogenous to behavior type, with actors more likely to form ties with similar behavior-types (homophily).</i></p> <p>LRSS: $\sigma_{it}^* = \sigma_{jt}^*, \forall i, j, \delta_{ij,t}^* = 1, \forall i, j$</p> <p>Path Dependent</p>

Note 1: LRSS = Long run steady state.

Note 2: The time index "0" in σ_{i0} and $\delta_{ij,0}$ mean that these are the starting values. In other words, given a certain system (society, world, country and so on) these values are constant.

three dichotomous choices, two behaviors and one connection, of *two* actors generates multiple equilibria. (Page’s rule is technically not violated, however, since there are really two network-connection choices even if one fully determines the other.)

In the next section, we use this theoretical model, which expresses two kinds of transition *probabilities*, one in behavior and one in network ties, as functions of these three kinds of conditions—exposure to exogenous factors, contagion, and selection—to outline an empirical strategy for estimating models of such social phenomena that can distinguish these inputs as sources of network-*cum*-spatial association, correlation, or clustering. As we stressed in the introduction, connectivity and interdependence are ubiquitous and central across much (all?) of social science, and successful empirical estimation of any of these causal relations requires careful, effective specification and modeling of all three.

3 Empirical Strategies

In this section, we discuss two discrete-time Markov models for empirical analysis. Both combine the *spatial-lag* model from spatial econometrics and the *p-star* model from network analysis. The first is relatively simple to estimate, while the second is not. For the latter we suggest, but do not fully develop, a possible estimation strategy. We also compare these models with the *actor-oriented* continuous-time coevolution model developed by Snijders and colleagues (Snijders 1997, 2001, 2005; Steglich et al. 2006; Snijders et al. 2007; Ripley and Snijders 2010).

3.1 Discrete-Time Markov Models

Our first discrete-time Markov model is one with behavior (not behavior-type) contagion and observed ties (not latent-strengths of ties). We define the behavior-shaping probability in a simple *spatial-lag* logit model, one with only a time-lagged (and not simultaneous) spatial-lag, and define the tie-formation probability in the simplest *p-star* model, one with independent dyads (which likewise reduces the tie-formation model to a set of conditionally independent logits);¹⁸

$$\begin{cases} Pr(s_{i,t} = 1 | \mathbf{s}_{t-1}, \mathbf{d}_{t-1}) = \text{logit}(\beta_0 + \beta_1 s_{i,t-1} + \beta_2 \mathbf{d}_{i,t-1} \mathbf{s}_{t-1}) \\ Pr(d_{ij,t} = 1 | \mathbf{s}_{t-1}, \mathbf{d}_{t-1}) = \text{logit}(\gamma_0 + \gamma_1 d_{ij,t-1} + \gamma_2 \cdot I(s_{i,t-1} = s_{j,t-1})), \end{cases} \quad (6)$$

where $\mathbf{d}_{i,t-1}$ is a row vector of size N that contains the set of $(N - 1)$ dichotomous tie-formation indicators between i and all the other actors at the end of period $t - 1$ (and 0 in element i for *dyad* ii), and $I(s_{i,t-1} = s_{j,t-1})$ indicates whether the behaviors in the given dyad were the same in the previous period. The term captures homophily. In this model, contagion (network effects) and the homophiletic selection (network formation) both operate through observed behaviors rather than through latent behavior type. Estimating (6) is straightforward. The behavior and tie formation equations can be estimated either separately or as a seemingly unrelated system of logit equations.

¹⁸For a similar approach to modeling network-behavior coevolution, see Lazer (2001). Important extensions for next steps in this project include enriching these two models to the simultaneous spatial/network-dependence case, although this may multiply the already computationally demanding system beyond the possible.

Our second empirical model is instead a latent-type and strength-of-tie model. This model connects much more closely to the theoretical model presented above, but is also more difficult to estimate. The behavior-switching and tie-formation/dissolution rules in this model take the form:

$$\begin{cases} Pr(s_{i,t} = 1 | \boldsymbol{\sigma}_{t-1}, \boldsymbol{\delta}_{t-1}) = \text{logit}(\beta_0 + \beta_1 s_{i,t-1} + \beta_2 \boldsymbol{\delta}_{i,t-1} \boldsymbol{\sigma}_{t-1}) \\ Pr(d_{ij,t} = 1 | \boldsymbol{\sigma}_{t-1}, \boldsymbol{\delta}_{t-1}) = \text{logit}(\gamma_0 + \gamma_1 d_{ij,t-1} + \gamma_2 (\sigma_{i,t-1} - \sigma_{j,t-1})^2), \end{cases} \quad (7)$$

where $\boldsymbol{\delta}_{i,t-1}$ is a row vector of size N that contains probabilities that i forms ties with, or the proportionate strengths of the ties between i and, each of the others, j , at the end of period $t - 1$ (appropriately zeroed for i 's self-reflexive dyad).

In this model, connections are by continuous, latent strength of tie, $\boldsymbol{\delta}$, not dichotomous tie or indicator of tie, \mathbf{d} , and contagion and selection operate through behavior type, $\boldsymbol{\sigma}$, not dichotomous behavior or indicator of behavior, \mathbf{s} . Behavioral types and the strength of ties are observed by the actors, but may not be observed by the analyst. One appealing way to conceive this substantively is that actors' types (behaviors) are influenced by other actors' underlying types (behavioral tendencies) and not by ephemeral short-run behavioral manifestations. Substantively, this reflects a proposition that, for example, if my friend who rarely votes happens to vote in one election by chance, that behavior is unlikely to influence the rate at which I turn out to vote. Likewise, homophily/heterophily in network-tie formation is driven by type rather than current-behavioral manifestations, which could be attributable to chance. For instance, considering a cosponsorship network in congress, a lifelong liberal senator is much more likely to cosponsor legislation with another committed liberal senator than with a conservative senator who just happened to vote liberal on the previously considered piece of legislation.

Because these types and strength of ties are often unobserved, as in our example application to come, for instance, we need a way to identify and estimate these latent variables. If we assume that types are always in equilibrium given behavior-switching probabilities, and symmetrically that tie strength is always in equilibrium given the Markov chain governing tie-formation and dissolution from period to period, then we know how these variables are likely evolve over time. More formally, we know:

$$\begin{pmatrix} \sigma_{i,t+1} \\ 1 - \sigma_{i,t+1} \end{pmatrix}' = \begin{pmatrix} \sigma_{1,t} \\ 1 - \sigma_{1,t} \end{pmatrix}' \begin{pmatrix} z_t^{11} & 1 - z_t^{11} \\ 1 - z_t^{00} & z_t^{00} \end{pmatrix} \quad (8)$$

and

$$\begin{pmatrix} \delta_{ij,t+1} \\ 1 - \delta_{ij,t+1} \end{pmatrix}' = \begin{pmatrix} \delta_{1,t} \\ 1 - \delta_{1,t} \end{pmatrix}' \begin{pmatrix} w_t^{11} & 1 - w_t^{11} \\ 1 - w_t^{00} & w_t^{00} \end{pmatrix}. \quad (9)$$

Of course, we also know from the theory presented above that these difference equations may have multiple solutions. If so, it seems reasonable to use the solutions that emerge from the Markov chains beginning at the previous period's types and tie strengths. Typically, knowledge of these dynamics, the so-called transition equations in state-space models, allows one to estimate latent variables using a Kalman filter and the EM algorithm. The problem here is that the measurement equation (7)—which, in this case, also happens to be a transition equation—has both latent left-hand-side and latent right-hand-side variables: the latent behavioral and tie switching probabilities at time $t + 1$ are a function of the latent behavioral types and tie strengths at time t , which, using a typical Kalman filter would be predicted initially using the behavior types and tie strengths at

time $t - 1$. We cannot update the type and tie-strength predictions from the transition equations unless we can incorporate relevant and observable time $t + 1$ information. One possibility is to use information about the cross-sectional distribution of observed behaviors and ties, and model the sample averages as a single time series. The sample mean of the distributions of observed ties and behaviors, for example, could be used to estimate the latent behavior and tie probabilities at time t , which should be linked through the measurement equation above to sample average values of the right-hand-side variables, including the latent type and tie strength variables. What is unobserved at the unit-level is estimable at the cross-sectional level. To work well, however, it seems this strategy would require both N and T to be large. Large N to get good estimates of the behavior and tie probabilities. And, with the likelihood written in terms of sample averages, large T to get precise estimates of the parameters. We do not develop this idea further,¹⁹ but the payoff from doing so is potentially large. History dependence, if it exist, is likely to be found in the evolution of these latent type and tie-strength variables.

We can propose also another strategy for estimating this harder model with the behavioral-homophiletic contagion and selection operating through the latent variables, (7). Start again from what we know regarding the transition probabilities assuming types and behaviors are always in equilibrium: (8) and (9). Given these, and given some starting values for σ_{t0} and δ_{t0} , and parameter values for β and γ , we can calculate the probability of observing 1's and 0's at time $t = 1$. We can also update each actor's type and the strength of ties across dyads using the implied transition probability matrices. Again, these difference equations have multiple solutions. We could select, though, the solution that emerges from the Markov chain beginning at the previous period's types and tie strengths. With estimated values for σ and δ at time $t = 1$, we can calculate the probability of observing 1's and 0's at time $t = 2$. By repeating this process, we should be able to calculate the joint likelihood for a given sample.²⁰

3.2 Comparison with *Siena*: Continuous-Time Markov Models

In the network-analytic tradition, Snijders and colleagues (Snijders 1997, 2001, 2005; Steglich et al. 2006; Snijders et al. 2007; Ripley and Snijders 2010) have advanced perhaps furthest in empirical modeling of dynamic, endogenous contagion and selection.²¹ They model the coevolution of networks and behavior thus. N actors are connected by an observed, binary, potentially endogenous, and time-variant matrix of ties, \mathbf{x} , with elements $\mathbf{x}_{ij,t}$. A vector of N observed, binary behaviors, \mathbf{z} , at time t has elements $\mathbf{z}_{i,t}$. Additional exogenous explanators may exist at unit or dyadic level, $\mathbf{v}_{i,t}$ or $\mathbf{w}_{ij,t}$. Opportunities arise for actors to change their network connections, switching at most 1 tie on or off, at continuous-time fixed-rate, $\rho_{i,t}^{net}$, according to an exponential model. At present, ρ_t^{net} is assumed constant for all i but allowed to vary arbitrarily by t , although the model and estimation procedure can accommodate richer parameterizations of ρ . Likewise, opportunities to increment, decrement, or leave unchanged the behavior arise at rate $\rho_{i,t}^{beh}$, again assumed uniform

¹⁹That is, we have not yet been able to explore this idea further, but plan to do so in future work.

²⁰We have written code to implement this second proposed strategy, but we have as yet had no success obtaining estimates in the empirical application below.

²¹Wasserman (1980a,b) and Leenders (1995, 1997) presage. Hoff and colleagues (Hoff et al. 2002; Hoff and Ward 2004; Hoff and Westveld 2007) offer a related, Bayesian latent-space approach for longitudinal-network analysis. Our discussion follows a combination of Snijders (1997, 2001); Steglich et al. (2006); Snijders et al. (2007) and the RSiena 4.0 Manual (Ripley and Snijders 2010) most closely.

across i but allowed to vary arbitrarily across t , although richer models are permissible.²² At present implementation, behaviors are also dichotomous, so the behavior-choice options are also only two: (switch on or off) or leave unchanged. When an opportunity to change network ties arrives for some i , i may choose to alter the status of any one of its $N-1$ ties to *on* or to *off* or to leave all ties unchanged.²³ i makes these choices by comparing the values of some objective function of this form:

$$f_i^{net}(\mathbf{x}, \mathbf{x}', \mathbf{z}) + \varepsilon_i^{net}(\mathbf{x}, \mathbf{x}', \mathbf{z}) = \sum_h \{ \beta_h^{net} \times s_h^{net}(\mathbf{i}, \mathbf{x}, \mathbf{x}', \mathbf{z}) \} + \varepsilon_i^{net}(\mathbf{x}, \mathbf{x}', \mathbf{z}) \quad (10)$$

where \mathbf{x}' is an alternative network under consideration, which can differ from the existing network, \mathbf{x} , only by changing at most one element of (only) row i . $f_i^{net}(\cdot)$ is called the network evaluation function. $s_h^{net}(\cdot)$ is some statistic, i.e., some function of the data, $\mathbf{x}, \mathbf{x}', \mathbf{z}$, that reflects the actor's objectives (ideally, substantively-theoretically derived) regarding the network, \mathbf{x} , and behaviors, \mathbf{z} . The β_h^{net} to be estimated are the relative weights of these objectives. Assuming the ε_i^{net} extreme-value distributed, independently across actors and over time, yields a multinomial-logit model of categorical choice. Similarly, when an opportunity to change behavior arrives, actor i compares the value of an analogous objective function under each of three possible actions: increment or decrement by one or leave unchanged. Formally, i compares \mathbf{z} to \mathbf{z}' given \mathbf{x} and $\mathbf{z}_{j \neq i}$. Again, the behavior evaluation function, $f_i^{beh}(\cdot)$, is the summed product of weights and statistics, β_h^{beh} and $s_h^{beh}(\cdot)$ respectively, and again assuming *i.i.d.* extreme-value stochastic components (ε_i^{beh}), the logistic form emerges once more.²⁴

Importantly for our purposes, the behavior and network objective-functions (and also the rate functions if and as desired) can include any of a number of commonly supposed social-network phenomena. For instance, *covariate-related dissimilarity*, which is "defined by the sum of absolute covariate differences between i and the others to whom he is related" (p. 371):

$$\text{covariate-related dissimilarity: } s_i(\mathbf{x}) = \sum_j \mathbf{x}_{ij} |\nu_i - \nu_j|. \quad (11)$$

This is the mathematical expression of the basis for homophily, or more precisely: heterophily. When the covariates in question are the behaviors of i and j , then the expression gauges the basis for the behavioral homophily that we have been discussing.

²²Since observation occurs at discrete intervals, the freedom to vary these continuous-time rates render effectively inconsequential the assumptions of one actor making one unit-valued change in his/her network ties or behavior at a time. As greater frequency and/or magnitude of changes are observed, estimates of these occurrence rates at this unobserved instantaneous level simply rise to compensate. This does not, however, relax the strong assumption of conditional independence of these actors' choices (which we make as well).

²³In textual descriptions of the model, the options always include doing nothing, but the equations indicated for the multinomial choice of tie to adjust if chosen to act have probabilities summing to one over the $N - 1$ $j \neq i$, meaning the option to do nothing if chosen has zero probability. Our data-generating process, see below, follows the equations in this regard.

²⁴In current implementation, the Siena model actually has three components (on each side: the network-formation and the behavior-choice side, so six total functions, really); Snijders (Snijders 2001) calls them: "the rate function[s], the objective function, and the gratification function" (p. 365). The latter two separate the objective function described in the text into two functions that sum to the total utility to an actor of a considered change and of the existing network and behaviors. The objective function considers these aspects of the network and behaviors in current levels, whereas the 'gratification function' addresses any aspects related to the change in the network or behaviors from current to the considered future. Distinguishing effects of changes from effects of levels in this way facilitates expression in the model of propositions that some explanator(s) may have different effects turning on or incrementing, as compared to turning off or decrementing, the behavior or network connections. The set-up for the gratification function is otherwise identical to that of the objective function, so this merely gives a second multinomial-logit form.

RSiena estimates models of this form by simulated method-of-moments (s-MoM).²⁵ To elaborate, the parameters of the model are $\boldsymbol{\rho}$, the rates of events (or the parameters in those exponential models of the rates), and $\boldsymbol{\beta}$, the parameters of the objective functions. The full parameter vector, $\boldsymbol{\theta}$, has dimensions k . As in general for MoM estimators, one applies a statistic, $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_k)$, such that $\boldsymbol{\theta}$ is the solution of the k -dimensional moment equation:

$$\epsilon_{\boldsymbol{\theta}} \mathbf{Z} = \mathbf{z}, \quad (12)$$

where \mathbf{z} is the sample outcome.²⁶

Minimally, one needs for \mathbf{Z} some statistics that will respond to in known manner the values of the parameters in question. Given such a statistic, we can specify as moment conditions:

$$\frac{\partial \epsilon_{\boldsymbol{\theta}} \mathbf{Z}_k}{\partial \theta_k} > 0, \quad (13)$$

for the MoM estimator. In fact, a quadratic of the moment condition needs to hold also, giving a (presumably more efficient) generalized MoM, G-MoM, estimator:

$$\mathbf{a}' \left(\frac{\partial \epsilon_{\boldsymbol{\theta}} \mathbf{Z}_k}{\partial \boldsymbol{\theta}} \right) \mathbf{a} > 0, \quad \forall \mathbf{a}. \quad (14)$$

For instance, for the rate from period m to $m + 1$, $\rho_{\mathbf{m}}$, a sufficient sample-statistic—sufficient sample-statistics, when known, tend to optimize efficiency of moment estimators—is the number of changes observed from m to $m + 1$, which generally rises in $\rho_{\mathbf{m}}$. For β , the statistics will likewise be the sample value of the objective-function statistics.

Estimated variance-covariance matrices for the vector of parameter estimates are obtained by (numerical computation of) the delta method.

Conditioning on the outcomes rather than using the moment equations involving those statistics reduces the dimensionality of the problem—in brief: by eliminating the rate functions as to be estimated—generally producing a much more stable and efficient optimization procedure. Conditioning thusly becomes especially necessary as the number of observation periods, M , and the event rates, ρ , increase because the dimensionality and complexity of Siena’s optimization problem grows combinatorically in M and ρ .

The moment equations given above may seem simple, but the conditional expectations at their hearts cannot generally be calculated explicitly (except for trivial cases). Accordingly, one applies stochastic-approximation methods, i.e., one simulates random network-behavior outcomes according to the processes of the proposed model and estimates the parameters of that model, as always, by optimizing fit to the observed sample statistics. For further details of, and options in, these estimation procedures, see especially Snijders (2001) and Ripley and Snijders (2010).²⁷

As a theoretical model of and estimation strategy for stochastic, actor-oriented model of network-formation and behavior of actors within networks, Siena is a tremendously impressive construct.

²⁵Description of the estimation procedure follows Snijders (2001).

²⁶These moment equations are further refined by conditioning in various ways on the initial and previous observed-outcomes.

²⁷Various refinements of both the first- and second-moment estimation procedures have been suggested since Snijders (2001) and many are implemented in RSiena. See Ripley and Snijders (2010) for discussion and practical implementation.

Not only state of the art: it *is* the entire corpus of the art. Yet, still, one must also heed several caveats emphasized by Snijders (Snijders 2001):

- “Although in our experience these equations mostly seem to have exactly one solution, they do not always have a solution” (p. 374).
- “[Referring to the moment conditions...] This requirement is far from implying the statistical efficiency of the resulting estimator [which has not been shown], but it confers a basic credibility to the moment estimator and it ensures the convergence of the stochastic approximation algorithm...” (p. 373).
- “This implies that the method proposed here is not suitable for observations...which are too far apart in the sense of the [...total number of changes between observations]. For such observations the dependence of [...this observation on the previous one...] is practically extinguished, and it may be more relevant to estimate the parameters of the process [...without controlling previous observations]” (p. 374).
- “It is plausible that these estimators have approximately normal distributions, although a proof is not yet available” (p. 375).

There is a small subset of Snijders’ emphasized list of cautions, concerns, and points at which various aspects of the estimation strategy’s performance are unknown or may be problematic. Our point in highlighting these issues could not be further removed from criticism. Siena is, to our knowledge, the most sophisticated and best-developed tool capable of addressing coevolution, which we think is rather common and important in social science. Moreover, in its approach to modeling network formation and behavioral choice more generally, Siena is designed so as to afford address of an empirical challenge for social science, the identification of which as substantively and theoretically central we wholly and wholeheartedly share: the distinction and distinct estimation of common exposure, contagion, and selection in generating social outcomes that ubiquitously exhibit network/spatial association. Our point in this listing is instead simply to underscore that we simply do not know very much about how Siena performs as an estimator. Understandably, given the complexity of the construct, little is known analytically, definitively about its properties, nor has very much about its performance been explored in Monte Carlo analysis, also understandably given its potentially enormous computational demands and the specialized nature of the software for its implementation. We believe the evaluation of the estimator in the next section is among the first conducted (and its comparison to our simple alternative is certainly so).

4 Estimation-Strategy Evaluation and Comparison

In this section, we evaluate and compare the performance of our simple proposed time-lagged spatial-lag logistic-regression strategy and the simulated method-of-moments strategy applied by Siena for estimating models of network-behavior coevolution, models with contagion and selection as well as exposure to exogenous factors.

4.1 Details of the Data-Generating Process

We follow Snijders (2001) to specify a data-generating process (DGP) that exactly matches a Siena model of coevolution where the behavior of N actors that is contagious through a network of connections generated by behavioral homophily. First, the DGP creates a vector for each inter-observational period of the cumulative probabilities of network-change events in *microsteps*²⁸, following a negative-exponential distribution, with hazard rate, $Rate_{net}$. An exactly analogous procedure produces T vectors of cumulative probabilities of behavior events from a negative-exponential distribution with hazard $Rate_{beh}$.

Armed with these probabilities of events at each microstep between observations, the DGP draws randomly the steps in which actors are drawn to possibly make a change in behavior and analogously randomly draws the steps where a network change is made.²⁹ When an event occurs by these random draws, each of the N actors is equally likely (a uniform random-draw in the DGP) to be chosen as the one to consider change. (This uniformity would be replaced by a weighted draw of the actors with weights given by the parameterization of the hazard function, had a substantive model been offered.³⁰)

Given that an actor i has been selected in some step to act on i 's $(N-1)$ network ties, a multinomial-logit form, $\frac{\exp(f(\mathbf{x}_k))}{\sum_{j \neq k, i} \exp(f(\mathbf{x}_j))}$, with $f(\mathbf{x}_k)$ given by i 's network objective-function evaluated for a tie to k , gives the probability that the tie to k is changed. A random draw from a multinomial distribution with this vector of probabilities (given by evaluation of the chosen actor's objective function just described at the current values of the network matrix and behavior vector) then determines which tie of actor i is changed. In our DGP, the objective function is *covariate-related similarity* from the Ripley and Snijders (2010) (monadic covariate effect number 39, p. 66.), defined by the sum of centered similarity scores $simv_{ij}$ between i and the other actors j to whom i is tied as of prior period. Using the behavior of i and j from the previous period, this metric gives the behavioral homophily effect in network-tie formation. The coefficient (in Siena's multinomial-logit form network-tie equation) on covariate-related similarity is set to 1.

The procedure for a behavior event is analogous. The actor drawn to consider a change to its behavior, a logistic form, $\frac{\exp(g(\mathbf{x}))}{1+\exp(g(\mathbf{x}))}$ compares the utility at the current values of the network matrix and behavior vector according to behavioral objective function $g(\mathbf{x})$. In our DGP, the behavioral objective function is given by the *average similarity effect*, defined as the average of centered similarity scores, with behavior again serving as the basis on which similarity is measured, between i and the other actors j to whom i is tied. This makes behavior of i depend more on the behaviors of j with whom i is similar in behavior. The coefficient (in Siena's logit-form behavior-equation) on average similarity is set to 1.

²⁸*Microsteps* is Snijders' term for the simulation periods between observations.

²⁹The asymmetry between possibly changing behavior and certainly changing a connection is intentional, and, as noted before, is in the equations presenting the Siena model even if the text says otherwise.

³⁰Weights would be given in that case by the ratio of the hazard-function value for i to the average hazard-function across actors at that microstep.

4.2 Details of the Monte-Carlo-Simulation Scenarios

We generated 100 trials each of 8 different scenarios.³¹ As noted above, for all scenarios, the DGP’s used covariate (behavior) related similarity in the network-tie equation and average similarity in the behavior equation, setting the respective multinomial-logit and logit coefficients in those specific Siena models to 1 in every case. We varied the number of actors $N \in \{30, 50\}$, the number of observed periods, $T \in \{5, 11\}$,³² and the rates of event occurrence, $Rate_{net} = Rate_{beh} \in \{1, 5\}$. Finally, we set the vector of first-period behaviors to 1 (0) for the first (second) $\frac{1}{2}N$, and the initial network-connection (spatial-weights) matrix such that each actor is connected to the (one) next actor with the same behavior, wrapping at the end. That is, except for the $\frac{1}{2}N^{th}$ and the N^{th} rows, the ones are in the upper first-minor, i.e., elements $(i, i + 1)$. For the $\frac{1}{2}N^{th}$ and the N^{th} rows, where the ones are in the 1^{st} and $(\frac{1}{2}N + 1)^{st}$ columns. That is, 1 connects to 2, 2 to 3, and so on, but the $\frac{1}{2}N^{th}$ connects back to the 1^{st} , and that upper-left block diagonal is then repeated as the lower-right block.

4.3 Results of the Monte-Carlo Simulations

Tables 2 and 3 present the results of our Monte Carlo explorations of *Siena* and our simple spatial-logistic strategy. We must first emphasize in reminder that the models differ—*Siena*’s estimation model mirrors the true DGP, whereas our spatial-logistic simplification is a different logit model, differently parameterized. **The coefficient magnitudes are *not* directly comparable.** In fact, whereas we know that both the contagion and the homophiletic-selection parameters, $\beta_{contagion}$ and $\beta_{homophily}$, are truly equal to 1, we do not know what the “true” value of the parameters of our logistic approximation to the *Siena* DGP should be. We do know they should have the same sign as the true coefficients in the true DGP (here: positive), though.³³ The direct presence of a time-lag regressor and of N unit and $N(N - 1)$ unit-dummies in our simple spatial-logistic strategy, we suspect, should (greatly?) depress these correct magnitudes of the coefficients on spatial-lag behavior and the homophily term in those estimations below the unity of the corresponding $\beta_{contagion}$ and $\beta_{homophily}$ in the true *Siena* model. (Our discussion follows the tables.)

³¹RSiena estimation of this model becomes extremely time-consuming—even on a 64-bit, i7 x920 processor, with 8MB RAM—as T and the *Rates* expand. Exploration with larger numbers of trials, like that standard 1000, must wait later editions of the paper.

³²The odd choices arising from confusion from our DGP code using the number of periods beyond the first whereas RSiena input calls for the total number of periods...

³³We may be able to solve for this though, but we simply have not tried yet. We certainly can, should, and will calculate comparable quantities of interest for these evaluations and comparison. Again, we simply have not had time to do so yet.

Table 2: Monte Carlo Results for $Rate_{net} = Rate_{beh} = 1$ (100 Trials)

Sample: N=30, T=4			Sample: N=30, T=11		
Parameter	Result	SIENA	Parameter	Result	SIENA
Network Selection	Mean	0.998	Network Selection	Mean	0.870
	S.D.	0.561		S.D.	0.305
	RMSE	0.561		RMSE	0.332
	Mean S.E.	0.662		Mean S.E.	0.319
	Overconfidence	0.848		Overconfidence	0.956
Behavior Contagion	Mean	0.744	Behavior Contagion	Mean	0.918
	S.D.	0.770		S.D.	0.437
	RMSE	0.811		RMSE	0.445
	Mean S.E.	0.882		Mean S.E.	0.533
	Overconfidence	0.874		Overconfidence	0.818
Power	0.03	Power	0.30		
Sample: N=50, T=4			Sample: N=50, T=11		
Parameter	Result	SIENA	Parameter	Result	SIENA
Network Selection	Mean	0.996	Network Selection	Mean	0.899
	S.D.	0.400		S.D.	0.237
	RMSE	0.400		RMSE	0.258
	Mean S.E.	0.385		Mean S.E.	0.244
	Overconfidence	1.040		Overconfidence	0.969
Behavior Contagion	Mean	0.942	Behavior Contagion	Mean	0.928
	S.D.	0.599		S.D.	0.466
	RMSE	0.602		RMSE	0.472
	Mean S.E.	0.622		Mean S.E.	0.441
	Overconfidence	0.962		Overconfidence	1.060
Power	0.29	Power	0.57		
Sample: N=30, T=11			Sample: N=50, T=4		
Parameter	Result	SIENA	Parameter	Result	SIENA
Network Selection	Mean	0.464	Network Selection	Mean	0.405
	S.D.	0.197		S.D.	0.1
	RMSE	-		RMSE	-
	Mean S.E.	0.193		Mean S.E.	0.093
	Overconfidence	1.020		Overconfidence	1.070
Behavior Contagion	Mean	0.752	Behavior Contagion	Mean	0.768
	S.D.	0.648		S.D.	0.343
	RMSE	-		RMSE	-
	Mean S.E.	0.604		Mean S.E.	0.347
	Overconfidence	1.070		Overconfidence	0.991
Power	0.19	Power	0.62		

Table 3: Monte Carlo Results for $Rate_{net} = Rate_{beh} = 5$ (100 Trials)

Sample: N=30, T=4			Sample: N=30, T=11		
Parameter	Result	SIENA	Parameter	Result	SIENA
Network Selection	Mean	1.026	Network Selection	Mean	0.424
	S.D.	1.845		S.D.	0.457
	RMSE	1.845		RMSE	0.735
	Mean S.E.	0.907*		Mean S.E.	0.646*
	Overconfidence	2.034		Overconfidence	0.707
Behavior Contagion	Mean	0.383	Behavior Contagion	Mean	0.68
	S.D.	1.802		S.D.	0.698
	RMSE	1.905		RMSE	0.768
	Mean S.E.	1.283*		Mean S.E.	0.881*
	Overconfidence	1.405		Overconfidence	0.792
Power	0	Power	0		
Simple Logit			Simple Logit		
	0.088			-0.007	
	0.096			0.574	
	-			-	
	0.096			0.537	
	1.000			1.070	
	-0.074			0.065	
	0.769			0.055	
	-			-	
	0.684			0.061	
	1.123			0.911	
	0			0.01	
Sample: N=50, T=4			Sample: N=50, T=11		
Parameter	Result	SIENA	Parameter	Result	SIENA
Network Selection	Mean	0.982	Network Selection	Mean	0.504
	S.D.	1.165		S.D.	0.406
	RMSE	1.165		RMSE	0.641
	Mean S.E.	0.916*		Mean S.E.	0.518*
	Overconfidence	1.272		Overconfidence	0.784
Behavior Contagion	Mean	0.778	Behavior Contagion	Mean	0.681
	S.D.	0.827		S.D.	0.747
	RMSE	0.856		RMSE	0.812
	Mean S.E.	1.200*		Mean S.E.	0.810*
	Overconfidence	0.689		Overconfidence	0.922
Power	0	Power	0.01		
Simple Logit			Simple Logit		
	0.085			0.082	
	0.065			0.036	
	-			-	
	0.070			0.044	
	0.933			0.826	
	-0.005			0.023	
	0.621			0.494	
	-			-	
	0.543			0.435	
	1.145			1.135	
	0.02			0.03	

*: Median standard error estimate reported.

4.3.1 Bias

That said, we can see that *Siena*'s coefficient estimates show little to no bias in any of the small- T scenarios on network-selection side of the model regarding the homophiletic-selection parameter, β_h . However, on the behavior side regarding the contagion parameter, β_c , we see a small deflation or negative bias, -6% only in the large- N , small- T , low- ρ scenario. In the higher- ρ , larger- N , smaller- T scenario, we see a larger -22% ; slightly more than that in low-rate, small N and T : -25% ; and this downward bias in β_c estimates explodes to a whopping -62% as T becomes smaller and ρ become larger. Back on the network-selection side regarding β_h now, we notice that glowing low-bias performance of *Siena* dims somewhat, to $-10 - 13\%$ under higher- T , lower-rate scenarios, and explodes to a whopping $-50\% - -58\%$ as both T and ρ become larger.

Trying to summarize the pattern, it seems that larger N generally enhances *Siena*'s bias performance, generally somewhat more so for the behavioral-contagion parameter than for the homophiletic-selection coefficient, since, when β_h can be well estimated, it seems it is so at $N = 30$ or 50 . Larger- T , on the other hand—that is, a greater number of observations over time of the network and behaviors—has more mixed and somewhat paradoxical affects on these biases. On the network-selection side of the model, at the lower-rates, the estimates for β_h go from negligibly biased in the lower- T scenarios to $-10\% - -13\%$ biased at the higher T . On the behavioral-contagion side, we see little effect of T on *Siena*'s biases at higher N and quite notable reductions in the bias at lower N : more what we would expect from more data.

Notice the signs even from this first consideration, of bias, that higher event-rates, ρ , generally wreak havoc on the estimations. Even at lower- T , while bias remains very small for the β_h estimates, we see large (-22%) biases in $\hat{\beta}_c$ at larger- N and enormous (-62%) biases at lower- N . At larger- T , the biases are huge ($-32\% - -58\%$) across both parameters and numbers of units.

As already noted, we do not know what the correct values of the coefficients from the simpler spatial-logit estimation-strategy should be, so we cannot evaluate its bias directly. However, as also explained, we suspect it should be lower, so the notably smaller magnitudes do not alarm us. We cannot evaluate bias precisely, but we do notice that the pattern in relative sizes of the estimates across scenarios and in the direction of changes in those sizes across the scenarios seems to match that of the *Siena* estimator exactly. The proportionate magnitudes of those changes across the scenarios are much smaller though. This suggests to us that these features may be aspects of coevolution-model estimation rather than of the estimators:

- lower T generally hinders estimation of contagion and, paradoxically, aids that of selection, more so in both cases at lower N ;
- lower N generally hinders estimation of both contagion and selection, more so for contagion; and
- higher ρ amplifies these difficulties, tremendously so as T also increases.

Indeed, the difficulties raised by high event-rates—i.e., by large amounts of inter-observational-period change in networks and behaviors—may be debilitating.

We remind the reader here of Snijders' important caveat: "...the method proposed here is not

suitable for observations...which are too far apart in the sense of the [...total number of changes between observations]. For such observations the dependence of [...this observation on the previous one...] is practically extinguished, and it may be more relevant to estimate the parameters of the process [...without controlling previous observations]” (p. 374). This certainly seems true, perhaps more generally than just in the *Siena* method, although, as we shall see, perhaps worse in *Siena* than in our simple spatial-logistic alternative.

4.3.2 Efficiency and Standard-Error Accuracy

Regarding the standard deviation across the trials of the coefficient estimates, we see, first, that event-rates have even more obvious effects here. In almost every N and T scenario, standard-deviations across trials now exceed their trial means. Our simple spatial-logit seems to fare somewhat better in these still mostly-bleak scenarios. At lower rates, we see that, proportionately (i.e., consider ratios of mean parameter-estimate to standard deviation of those estimates), the spatial-logistic approach seems to have an advantage. This is particularly notable on the behavioral-contagion side, where (phrasing the issue incorrectly crudely but amusingly) *Siena* seems to have “greater difficulty obtaining higher ‘ t -ratios’.”

With regard to standard-error accuracy, outside of the low- T , high- ρ scenarios for *Siena*, both estimation strategies seem reasonably free from overconfidence. In fact, *Siena* tends toward underconfidence in these other scenarios. Even in the higher-rate scenarios, where the estimates generally have high to enormous sampling variation, at least the reported standard errors tend to be reasonably honest about that—outside of the low- T , high- ρ scenarios for *Siena*, where misestimation of uncertainty, mostly in the form of overconfidence, is very notable (although even there, the problem seems to fade in N).

4.3.3 Power and Test-Size

Although we do not report them here, we did conduct analyses of test size regarding the tests for homophiletic-network contagion ($H_0 : \beta_{contagion} = 0$) and for behavior-homophiletic selection ($H_0 : \beta_{homophily} = 0$) under some of these scenarios in a slightly different Monte Carlo setup. Notably, the rates were held fixed to the lower 1. Both models seemed to produce accurately-sized tests under these conditions. That is, when we stripped contagion, homophily, or both from the DGP, the models accurately rejected at the 0.05 level that the corresponding coefficients were zero in only 5% of the trials (1000 in those analyses).

In many ways, the results for power—i.e., to be precise and clear: ability of the estimation strategy to detect (by statistically rejecting at the 0.05 level a null of their absence) contagion and homophily of unit magnitude in this *Siena*-style DGP)—reported in Tables 2 and 3 are the most (depressingly) telling. At the higher rates, neither strategy could detect both the contagion and the homophily actually present. (Recall from the substance of the specification, moreover, that the unit magnitudes are rather large substantively and, in fact, are the only systematic source of variation in the respective outcomes!) At lower-rates, power grows with N and with T , which is natural and reassuring, and comfortingly becomes non-negligible although one could certainly wish for better, especially from *Siena*, which, after all, is the exactly correctly specified empirical model

for this DGP. Notwithstanding that fact, however, our simpler proposal of a pair of time-lagged spatial-lag logistic regressions dominates in power, especially at the smaller- T samples.

4.4 Summarizing the Monte Carlo Results

Although the inability to evaluate bias of the simpler spatial-logistic estimation-strategy and to compare it to *Siena* hindered analysis of the performance of the strategies greatly, we can conclude generally on several points. Data from contexts with higher event-rates, i.e., where inter-observational changes in networks and behavior is likely to have been great, seem not amenable to reliable estimation by either strategy of coevolutionary processes, to say the very least. At lower rates, either estimator seems reasonably honest about the certainty of its estimates. *Siena* seems essentially unbiased in lower- T samples but suffers some downward or deflationary bias in its estimates, even in these better-suited rate-conditions, at the larger T . The magnitudes of the spatial-logistic parameter-estimates should and do differ, so we cannot adjudge bias directly, but a roughly parallel (somewhat less in absolute terms, somewhat more proportionately) downward change in estimate-magnitude with T for the spatial-logistic strategy suggests a similar “large- T bias” there. On the other hand, the simpler spatial-logistic strategy has somewhat of an edge in efficiency and, thereby, in power, with this advantage growing more-noticeable with lower T and smaller samples.

The upshot of all this, for now, seems to us to be: neither strategy seems to offer much hope of learning anything reliable in almost any regard about coevolution when event-rates are high (which may be discernable by high amounts of change in networks and/or behaviors between observational periods). At low event-rates, conversely, the strategies seem to offer a very summarizable pattern regarding the quality of their estimates:

- Both produce standard errors with reasonable accuracy, with perhaps a slight edge to spatial-logit, but *Siena* errs only on the high side, which counters;
- The simpler spatial-logistic strategy seems somewhat more efficient and, thereby, also somewhat more powerful; and
- The simpler spatial-logistic strategy yields smaller coefficient-estimates, but both exhibit a similar pattern of coefficient estimates shrinking with T .

Therefore, if the smaller magnitudes of the spatial-logistic coefficient-estimates are ‘unbiased’ (i.e., produce on-average correct substantive-effect estimates), then simplicity and the small efficiency and power advantages would argue strongly for it over *Siena*. If the spatial-logistic estimates are substantively appreciably biased in this sense, then *Siena*’s complexity and weaker power would be worthwhile price. (Determining and studying comparable substantive-effect estimates are obviously extremely high-priority next steps for us.)

5 Illustration: Military Alliances and Conflict Behavior

We illustrate these methods with an empirical analysis of the alliance-formation and conflict behavior of great powers during the first half of the twentieth century (Levy 1981). We suspect that alliance ties and conflict behavior coevolve. States self-select into alliances and these decisions are plausibly driven by preferences homophiletic or heterophiletic in behavior (type). More aggressive (pacific) states may seek likewise aggressive or pacific allies, or the opposite may be true. At the same time, conflict behavior is contagious through alliances. Indeed, that states would be drawn into the conflicts of their allies is usually a key part, if not the core working principle, of alliances.³⁴ We focus on the first half of twentieth century because it was a period with considerable variation in conflict behavior (hardly unique to that period) and of multipolarity during which military alliances were in flux (rarer in other periods).

We present preliminary estimates of our models in Table 4. Model 1 (columns 1A and 1B) is our model with contagion of dichotomous behaviors, with connection and selection occurring through observed dichotomous ties; i.e., the system of equations in (6) above. Model 2 adds covariates. Specifically, in the conflict behavior model we include regime type and national capabilities using Polity and COW CINC scores respectively. In the alliance ties model we include regime similarity measured by one minus the absolute value of the difference in polity scores divided by twenty (i.e., the maximum difference) and the absolute value of of the CINC differences as a proxy for power asymmetry. This is a covariate-similarity measure, but, not in an exogenous regressor (as assumed in this model anyway), unlike the case of our behavioral-homophily regressor. The CINC scores are scaled to sum to one across all countries, so both our regime similarity and power asymmetry measures fall between zero and one.

If the disturbances³⁵ in our discrete-time Markov models are correlated across equations, as they almost certainly are in this application, equation by equation estimation will produce consistent, though inefficient, estimates of the parameter values, and conventional standard-error estimates will be inaccurate. To address the standard-error-inaccuracy issue we report robust standard errors using a systems sandwich estimator of the variance covariance matrix. The sandwich matrix in this formulation, the outer product of the gradients, provides estimates of the parameter covariances across equations, which are incorporated into the variance estimates.³⁶ Models 1 and 2 contain unit indicator variables (country or dyad) as well.

Model 3 (columns 3A and 3B) is the *Siena* continuous-time Markov model of coevolution. To estimate this model we used ‘snapshots’ of the great powers’ alliance networks and conflict behavior taken at five-year increments (i.e., 1900, 1905,...1950). For the network statistic, we used covariate(behavior)-related similarity, and for the behavior statistic, we used the average-similarity effect. The former is defined as $s_i^{net} = \sum_j x_{ij}(\text{sim}_{ij}^z - \widehat{\text{sim}}^z)$, where the similarity scores are $\text{sim}_{ij}^z = \frac{\Delta - |z_i - z_j|}{\Delta}$, Δ being the maximum sample difference, and $\widehat{\text{sim}}^z$ is the mean of all similarity scores. The latter behavior-statistic is defined as $s_i^{beh} = \frac{\sum_j x_{ij}(\text{sim}_{ij}^z - \widehat{\text{sim}}^z)}{\sum_j x_{ij}}$. These are the same

³⁴For a similar argument, see (Kimball 2006).

³⁵I.e., the extreme-value disturbances from the choice model.

³⁶In Stata, obtaining these system sandwich estimates requires one simple post-estimation command, *suest*. Incidentally, this makes the appropriate label for our suggested estimation strategy one of *Maximum pseudo-Likelihood*, although this may already have been the case since, if we maintain the assumption from the preceding section that the Siena DGP is the correct one, then these logit models were only an approximation to begin.

statistics we used in our Monte Carlo simulations. See Snijders (2001) and Ripley and Snijders (2010) for many alternatives and much further discussion.

We find evidence (1) of heterophily—pacific powers are more likely to ally or maintain alliances with aggressive powers—and (2) that conflict behavior is (positively) contagious through alliances. These findings are robust across both the discrete-time and continuous-time Markov models. The Model 1 estimates, for example, imply that the average probability across great powers of engaging in a militarized dispute given no involvement in the previous period rises from .55 to .74 when one’s allies change from pacific to aggressive behavior in the previous period. The effects of heterophily are smaller in size. The average probability that, among great power dyads, an alliance will persist from one period to the next is about .92 when the alliance partners behave dissimilarly in the preceding period. When both partners are either pacific or aggressive, this probability drops to a little less than .89. (Note that the sustaining influence of asymmetry extends beyond behavior to include capabilities as well—see Model 2A. Relatively weak countries are more likely to ally and stay allied with relatively powerful partners.³⁷) One calculates effects from the *Siena* models in terms of odds ratios. The estimate in 3A, for example, implies that a great power is almost thirty-four times more likely to form an alliance with a partner whose conflict behavior, at the decision point in time, was dissimilar.

Overall, our empirical results suggest that the conflict behavior of great powers and their military alliance networks coevolve over time. One significant implication of this is that great power relations may be path dependent. As noted above, our theoretical models suggested that the test of endogenous coevolution, i.e. of contagion in behavior and selection by behavior jointly, were also a test of path dependence. Namely, we test whether $H_0 : \beta_{contagion} \times \beta_{homophily} = 0$. Using a Wald strategy, and the Delta-method asymptotic linear approximation for the estimated variance of $\hat{\beta}_c \times \hat{\beta}_h$, the χ^2 statistics and associated probabilities are 3.56 and 0.0591 for Model 1, 3.25 and 0.0713 for Model 2, and 3.34 and 0.0676 for Model 3; again: suggestive, if not overwhelming, evidence for path dependence.

6 Conclusion

Theoretically, this paper builds a discrete-time Markov type-interaction model in which the behaviors of actors and the networks that connect them coevolve over time. One interesting implication of the model is that it produces history-dependent behavior, possibly including path dependent behavior. It seems likely that there are many areas of inquiry in the social sciences where network-behavior coevolution is important. To evaluate this possibility empirically, we built from the theoretical model a spatial-lag logistic model of coevolution that combines a simple time-lagged spatial-lag model of contagious behavior with a simple *p-star* logit model of behavioral-homophiletic network formation (which is also a time-lagged spatial-lag model). We explored and compared the performance of this proposed simple estimation strategy and/with that of the impressive state-of-the-art model from social-network analysis of node-behavior and tie-formation, *Siena*. Neither strategy seemed capable of gaining traction in environments where a great deal of change in connectivity and behavior occurs between observations, but either seemed capable of doing so in more favorable

³⁷This supports theoretical expectations from the alliance-formation literature regarding power asymmetry and alliance formation; see, e.g. Morrow (1991).

Table 4: Estimation Results: Military-Alliance Ties and Conflict Behavior

(Markov Models>>)	Discrete-Time Ours				Continuous-Time Snijders et al. (by RSiena)	
	(1A) Alliance Networks	(1B) MIDs Behavior	(2A) Alliance Networks	(2B) MIDs Behavior	(3A) Alliance Networks	(3B) MIDs Behavior
Temp lag	4.99*** (0.14)	1.45*** (0.27)	5.04*** (0.14)	1.33*** (0.28)	–	–
<i>Dyad specific</i>						
Previous MIDs similarity (Behavior)	-0.39** (0.15)	–	-0.42*** (0.15)	–	-3.52*** (0.60)	–
Regime similarity	–	–	0.45 (0.28)	–	–	–
Power asymmetry	–	–	6.56*** (1.51)	–	–	–
<i>State specific</i>						
Previous alliance tie (Network)	–	0.85*** (0.31)	–	0.74** (0.33)		1.06** (0.53)
Polity	–	–	–	-0.05 (0.03)		–
National capability	–	–	–	10.96** (4.69)		–
Loglikelihood	-225.08	-179.50	-223.05	-175.34		

Note: The standard errors reported for Model (1) and (2) are SUR-robust standard errors. These models also include fixed unit effects. Coefficients are not reported.

environments, with some analytic work remaining to established the quality of effect-size estimates from the simpler strategy and to compare the two given that and the apparently greater efficiency and power of that simpler strategy. Finally, we demonstrated the feasibility and potential utility of these theoretical and statistical framework by applying them to analyze the patterns of alliance formation and conflict behavior among the great powers during the first half of the twentieth century. A test for path dependence that we derived from these theoretical and empirical efforts suggests that great-power conflict alliance-formation patterns in the first half of the twentieth century likely did contain path dependency.

There is a lot of work left to do. For starters, we need to expand the class of theoretical models, exploring important alternatives such as models in which actors observe dichotomous outcomes and adjust their continuous latent traits. There are multiple other options regarding what diffuses and what forms the basis for selection as well. That is, we see these particular models, with contagion and selection both occurring through binary variables observed by the actors and analysts, as only one case of a family of models much in need of development for the diversity of social-science coevolution contexts, models in which behavior or ties may be binary or continuous quantities, contagion and selection may occur through actors' observed or latent quantities, and analyst may observe or not observe these quantities.

On the theoretical side, we have not yet characterized fully the forms of history dependence—outcome and equilibrium state, *phat*, and path dependency (Page 2006)—that emerge from alternative parameter and starting values of the current class of models (nor, of course, of the not-yet-existing classes) and the mappings from the latter to the former.

Then, we need also to tighten further the connection between the theoretical and empirical models; in particular, we need to develop an empirical strategy that can address the second case considered in that section, where the endogenous selection and contagion are both in the latent variables.

On the empirical side, we need to develop techniques for calculating, testing, and presenting *effects* in this complex (nonlinear system of endogenous equations) context and not merely parameters (see Hays et al. (2010); Franzese and Hays (2009); Hays (2009)). We need to expand the comparison of empirical model and estimation strategy with that of Snijders and colleagues' *Siena* in these regards also, including Monte Carlo simulation of properties related to comparable estimates of these *effects*, dynamics, and steady-state *effects*. We would like also to add consideration of simultaneous interdependence across in network-tie formation and in behavioral choice to the empirical model and estimation strategy, although this will come only at much greater added-complexity.

Much left to do indeed, but we consider this context an extremely important one for social-scientific theoretical and empirical effort. Interdependence and selection are everywhere in society, polity, and economy, and usually quite appreciably strongly so. Spatial-*cum*-network association, correlation, or clustering is at least as omnipresent and sizable empirically across most of social science's substance. Yet, such association can come from *common exposure*, *contagion*, or *selection*, and it matters a great deal for theoretical, scientific and practical, applied purposes which are operating and in what strengths. And, finally, these phenomena tend (as does everything else in social science) to be mutually endogenous. Accordingly, we need theoretical models that can shed light and all these processes simultaneously and empirical strategies that can distinguish their effects and estimate them distinctly well. A tall order, we agree, but we do not view (in fact, we think logic and the contentions just enumerated rule out) as promising any alternative tacks to that taken here:

build theoretical models that incorporate the relevant endogenous processes as well and accurately as possible and use those theoretical models to specify empirical models and estimation strategies that reflect as accurately and fully as possible all of those relevant processes.

Appendix I: Summary of Notations

- Actors: $i \in \{1, \dots, N\}$.
- Discrete time periods: $t \in \{1, 2, \dots\}$.
- Behavior: $s_{it} \in \{0, 1\}$. A dichotomous behavior (behavioral action) that actor i takes in period t . Both actors in the model and analysts (or econometricians) outside the model can observe.
- Behavioral type: $\sigma_{it} \in [0, 1]$. This represents the probability that actor i chooses behavior 1 in period t . We assume that, in this model, actors can observe (know) others' behavioral types.
- Network tie: $d_{ij,t} \in \{0, 1\}$. An undirected indicator of whether the two actors i and j in a certain dyad are connected. The indicator $d_{ij,t} = 1$ if they are connected, $d_{ij,t} = 0$ if not.
- Tie-formation probability: $\delta_{ij,t} \in [0, 1]$, $\forall i, j$ and $i \neq j$. This denotes the probability that a tie forms between the two actors i and j in a given dyad in period t . This probability can also be interpreted as the strength of the tie between i and j .
- State of the system: $(\boldsymbol{\sigma}_t, \boldsymbol{\delta}_t) = (\sigma_{1t}, \dots, \sigma_{Nt}; \delta_{12,t}, \dots, \delta_{N-1,N,t})$. We define the state of the system at the end of period t as a vector of every actor's behavioral type and every dyad's tie-formation probability.
- Exogenous components of transition probabilities: $c_{1i}, c_{2,ij}$. In the “minimal theoretical model” that we present in the main text, $c_{1i} \in [0, 1]$ indicates the relative role of autoregression for determining the (non-)switching probability for actor i to stay with behavior 0 the next period, $t + 1$. Analogous to c_{1i} , $c_{2,ij} \in [0, 1]$ reflects the strength of the temporal autoregressive term relative to the homophily (selection) term in the switching probability.

Appendix II: A Slightly Less Minimal Model

A slightly less minimal model of endogenous coevolution, also with contagion via networks whose edges are endogenous to behavioral homophily, removes from the behavioral equation of the minimal model in the text the asymmetry between transition probabilities from current period t type-1 to next period $t + 1$ types being subject to contagion effects while the transition probabilities from current period t type-0 to next period $t + 1$ types not being susceptible to contagion. It still simplifies the tie-formation model to exhibit homophily in only one of its rows. The behavioral model is as follows:

$$\begin{aligned}
& \begin{pmatrix} Pr(s_{i,t+1} = 1|s_{it}) \\ Pr(s_{i,t+1} = 0|s_{it}) \end{pmatrix}' = \begin{pmatrix} \sigma_{i,t+1} \\ 1 - \sigma_{i,t+1} \end{pmatrix}' \\
& = \begin{pmatrix} \sigma_{i,t} \\ 1 - \sigma_{i,t} \end{pmatrix}' \begin{pmatrix} c_{0i}\sigma_{it} + (1 - c_{0i})\frac{\sum_{j \neq i}(\delta_{ij}\sigma_j)}{N-1} & 1 - \left[c_{0i}\sigma_{it} + (1 - c_{0i})\frac{\sum_{j \neq i}(\delta_{ij}\sigma_j)}{N-1} \right] \\ 1 - \left[c_{1i}(1 - \sigma_{it}) + (1 - c_{1i})\frac{\sum_{j \neq i}(\delta_{ij}(1 - \sigma_j))}{N-1} \right] & c_{1i}(1 - \sigma_{it}) + (1 - c_{1i})\frac{\sum_{j \neq i}(\delta_{ij}(1 - \sigma_j))}{N-1} \end{pmatrix} \quad (15) \\
& = \begin{pmatrix} \sigma_{1,t} \\ 1 - \sigma_{1,t} \end{pmatrix}' \begin{pmatrix} \pi_{11} & \pi_{12}(= 1 - \pi_{11}) \\ \pi_{21}(= 1 - \pi_{22}) & \pi_{22} \end{pmatrix}.
\end{aligned}$$

Again, the following conditional probability for actor i to take behavior 1 suffices to describe the Markov chain for behavioral type:

$$\begin{aligned}
Pr(s_{i,t+1} = 1|s_{it}) &= \sigma_{i,t+1} = \sigma_{i,t}\pi_{11}^\sigma + (1 - \sigma_{i,t})(1 - \pi_{22}^\sigma) \\
&\Leftrightarrow \sigma_{i,t+1} = \sigma_{i,t} \left[c_{0i}\sigma_{it} + (1 - c_{0i})\frac{\sum_{j \neq i}(\delta_{ij}\sigma_j)}{N-1} \right] \\
&\quad + (1 - \sigma_{i,t}) \left[1 - \left[c_{1i}(1 - \sigma_{it}) + (1 - c_{1i})\frac{\sum_{j \neq i}(\delta_{ij}(1 - \sigma_j))}{N-1} \right] \right]. \quad (16)
\end{aligned}$$

As before, e.g., cell (1, 1) of the transition-probability matrix represents the probability that actor i chooses behavior 1 given that i chose 1 in the past period;

$$Pr(s_{i,t+1} = 1|s_{it} = 1) = c_{0i}\sigma_{it} + (1 - c_{0i})\frac{\sum_{j \neq i}(\delta_{ij}\sigma_j)}{N-1}. \quad (17)$$

The first term, c_{0i} , is the extent to which i maintains its current behavioral type σ_{it} in the next time period $t + 1$; in other words, it is the coefficient for behavioral-type's autoregressive term. The second term is the weighted average expressing how others' types affect i 's propensity of choosing 1 in period $t+1$. Notice how, by this construction, the weight $(1 - c_{0i})$ captures the extent to which others' types matter for i 's behavioral choice at $t + 1$. This $(1 - c_{0i})$ again gives us parameters within the transition model by which to vary the overall strength of contagion. For instance, in an extreme case where i maintains its behavioral type (σ_{it}) with probability $c_{0i} = 1$ exogenously of the state of the world, then $1 - c_{0i} = 0$, meaning that i 's decision will not be affected by any others to whom she is connected: i.e., the strength of contagion is 0.

The probability that explains the switching of i 's behavior from $s_{it} = 1$ to $s_{i,t+1} = 0$ —cell (1,2) of the transition-probability matrix—can be computed as $Pr(s_{i,t+1} = 0|s_{it} = 1) = 1 - Pr(s_{i,t+1} = 1|s_{it} = 1)$ straightforwardly from the fact that i 's choice is dichotomous and the two action choices are mutually exclusive exhaustive.

The transition probability from $s_{it} = 0$ to $s_{i,t+1} = 0$, cell (2,2), takes the same form as cell (1,1):

$$Pr(s_{i,t+1} = 0 | s_{it} = 0) = c_{1i}(1 - \sigma_{it}) + (1 - c_{1i}) \frac{\sum_{j \neq i} (\delta_{ij}(1 - \sigma_j))}{N - 1}, \quad (18)$$

except the weight for maintaining its behavioral type from the period t , c_{1i} , can differ from c_{0i} . This allows us to address potentially different natures (e.g., sticky behavior or not) of the two behavioral actions. The second row of the transition-probability matrix also has to sum to 1, which gives us the expression for cell (2,1), $Pr(s_{i,t+1} = 1 | s_{it} = 0) = 1 - Pr(s_{i,t+1} = 0 | s_{it} = 0)$.

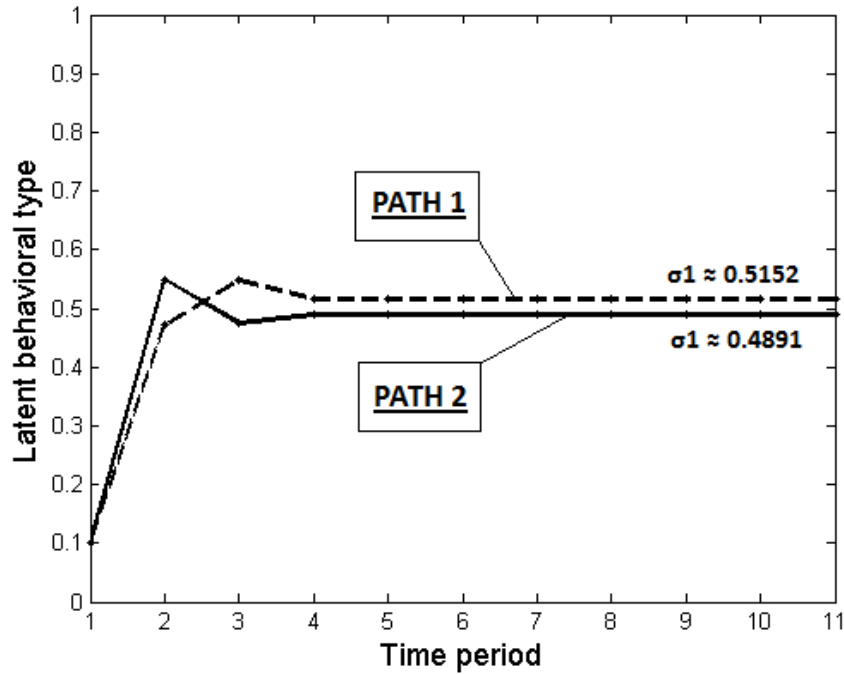
The tie-formation probabilities are identical to those from the minimal model in the text, and give the same one-line conditional-probability expression of those difference equations.

Again, following (Page 2006), the question of path dependency reduces to one of whether the equilibria value of behavior types (we know $a_1 = a_2$) vary across different sequences of past events. Or do they eventually reach a certain fixed value regardless of the sequence of past events?

Again by numerical example: with initial values for behavioral type and tie-formation tendency of $a_1(1,1) = .1$, $b_1(1,1) = .1$, $a_2(1,1) = .1$, $b_2(1,1) = .1$, $t_{12}(1,1) = .1$, and $s_{12}(1,1) = .1$, we find equilibrium at $a_1 = a_2 \approx 0.5152$ and $b_1 = b_2 \approx 0.4891$, respectively, and with $t = 1$ and $s = 1$

Again, the system exhibits path dependence.

Figure 4: Illustration of Path Dependency: Comparison of Actor 1's Long-Run Behavioral Types Following 2 Paths ("Simple System")



Time period →	PATH 1					PATH 2				
	1	2	3	4	...	1	2	3	4	...
c_{01}	0.8	0.1	0.9	0.9	...	0.1	0.8	0.9	0.9	...
c_{02}	0.2	0.6	0.1	0.1	...	0.6	0.2	0.1	0.1	...
c_{11}	0.2	0.1	0.1	0.1	...	0.1	0.2	0.1	0.1	...
c_{12}	0.9	0.1	0.9	0.9	...	0.1	0.9	0.9	0.9	...
c_2	0.6	0.3	0.1	0.1	...	0.3	0.6	0.1	0.1	...

Initial values for the endogenous parameters:

$$\sigma_1 = \sigma_2 = 0.1, \delta_{12} = 0.1.$$

References

- Beck, N., Gleditsch, K., and Beardsley, K. (2006). Space is more than geography: Using spatial econometrics in the study of political economy. *International Studies Quarterly*, 50(1):27–44.
- Brueckner, J. (2003). Strategic Interaction among Governments: An Overview of Empirical Studies. *International Regional Science Review*, 26(2):175.
- Elkins, Z. and Simmons, B. (2005). On Waves, Clusters, and Diffusion: A Conceptual Framework. *The Annals of the American Academy of Political and Social Science*, 598(1):33.
- Franzese, R. and Hays, J. (2006). Strategic interaction among EU governments in active labor market policy-making: Subsidiarity and policy coordination under the European employment strategy. *European Union Politics*, 7(2):167–89.
- Franzese, R. and Hays, J. (2007a). Empirical Models of international Capital-Tax Competition. In Read, C. and Gregoriou, G., editors, *International Taxation Handbook: Policy, Practice, Standards and Regulations*, page 43. Cima Pub.
- Franzese, R. and Hays, J. (2007b). Spatial Econometric Models of Cross-Sectional Interdependence in Political Science Panel and Time-Series-Cross-Section Data. *Political Analysis*, 15(2):140–64.
- Franzese, R. and Hays, J. (2008a). Empirical Models of Spatial Interdependence. In Box-Steffensmeier, J., Brady, H., and Collier, D., editors, *Oxford Handbook of Political Methodology*, Oxford UP. Oxford University Press.
- Franzese, R. and Hays, J. (2008b). Interdependence in Comparative Politics: Substance, Theory, Empirics, Substance. *Comparative Politics Studies*, 41(4-5):742–80.
- Franzese, R. and Hays, J. (2009). The Spatial Probit Model of Interdependent Binary Outcomes: Estimation, Interpretation, and Presentation. *Working Paper (Presented to Public Choice 2009)*.
- Hägerstrand, T. (1967). *Innovation diffusion as a spatial process*. University of Chicago Press.
- Hägerstrand, T. (1970). What about people in regional science? *Papers in Regional Science*, 24(6).
- Hays, J. (2009). Bucking the System: Using Simulation Methods to Estimate and Analyze Systems of Equations with Qualitative and Limited Dependent Variables. In *2nd Annual St. Louis Area Methods Meeting (SLAMM)*, Washington University in St. Louis (https://netfiles.uiuc.edu/jchays/www/Hays_SLAMM09_Revised.pdf).
- Hays, J., Kachi, A., and Franzese, R. (2010). A Spatial Model Incorporating Dynamic, Endogenous Network Interdependence: A Political Science Application. *Statistical Methodology*, 7(3):406–428.
- Hoff, P., Raftery, A., and Handcock, M. (2002). Latent Space Approaches to Social Network Analysis. *Journal of the American Statistical Association*, 97(460):1090–1099.
- Hoff, P. and Ward, M. (2004). Modeling Dependencies in International Relations Networks. *Political Analysis*, 12(2):160–175.
- Hoff, P. and Westveld, A. (2007). A Bayesian Mixed Effects Model for Longitudinal Social Network Data. In *Unpublished*, University of Washington.
- Jackson, J. and Kollman, K. (2007). Models of Path Dependence With An Empirical Application. In *Annual Political Methodology Conference*, State College, PA.
- Jackson, M. (2008). *Social and Economic Networks*. Princeton University Press.

- Kelejian, H. and Prucha, I. (2009). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*.
- Kimball, A. (2006). Alliance formation and conflict initiation: The missing link. *Journal of Peace Research*, 43(4):371–389.
- Koger, G., Masket, S., and Noel, H. (2009). Partisan webs: Information exchange and party networks. *British Journal of Political Science*, 39(03):633–653.
- Koger, G., Masket, S., and Noel, H. (2010). Cooperative Party Factions in American Politics. *American Politics Research*, 38(1):33.
- Lazarsfeld, P. and Merton, R. (1954). Friendship as a Social Process: A Substantive and ethodological Analysis. *Freedom and Control in Modern Society*.
- Lazer, D. (2001). The Co-Evolution of Individual and Network. *Journal of Mathematical Sociology*, 25(1):69–108.
- Leenders, R. (1995). Structure and influence. Statistical models for the dynamics of actor attributes, network structure and their interdependence.
- Leenders, R. (1997). Longitudinal behavior of network structure and actor attributes: modeling interdependence of contagion and selection. *Evolution of Social Networks*, pages 165–84.
- Levy, J. S. (1981). Alliance formation and war behavior: An analysis of the great powers, 1495-1975. *Journal of Conflict Resolution*, 25(4):581–613.
- McPherson, M., Smith-Lovin, L., and Cook, J. (2001). Birds of a Feather: Homophily in Social Networks. 27:415–444.
- Morrow, J. (1991). Alliances and asymmetry: An alternative to the capability aggregation model of alliances. *American Journal of Political Science*, 35(4):904–933.
- Noel, H. and Nyhan, B. (2010). The “Unfriending” Problem: The Consequences of Homophily in Friendship Retention for Causal Estimates of Social Influence. *Working Paper (The Society for Political Methodology Working Paper Archive)*.
- Page, S. (2006). Path Dependence. *Quarterly Journal of Political Science*, 1(1):87–115.
- Page, S. (2007). Type Interaction Models and the Rule of Six. *Economic Theory*, 30(2):223–241.
- Ripley, R. and Snijders, T. (2010). Manual for siena version 4.0.
- Rubin, D. (1990). Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies. *Statistical Science*, 5(4):472–480.
- Simmons, B., Dobbin, F., and Garrett, G. (2006). Introduction: The international diffusion of liberalism. *International Organization*, 60(04):781–810.
- Simmons, B. and Elkins, Z. (2004). The globalization of liberalization: Policy diffusion in the international political economy. *American Political Science Review*, 98(01):171–189.
- Snijders, T. (1997). Stochastic actor-oriented models for network change. *Evolution of social networks*, page 185.
- Snijders, T. (2001). The statistical evaluation of social network dynamics. *Sociological methodology*, 31:361–395.

- Snijders, T. (2005). Models for Longitudinal Network Data. In Carrington, P., Scott, J., and Wasserman, S., editors, *Models and Methods in Social Network Analysis*, chapter 11. Cambridge University Press, New York.
- Snijders, T., Steglich, C., and Schweinberger, M. (2007). Modeling the Co-Evolution of Networks and Behavior. In van Montfort, K., Oud, H., and Satorra, A., editors, *Longitudinal Models in the Behavioral and Related Sciences*, chapter 3, pages 41–71. Lawrence Erlbaum, Mahwah, NJ.
- Steglich, C., Snijders, T., and West, P. (2006). Applying SIENA: An illustrative Analysis of the Co-Evolution of Adolescents' Friendship Networks, Taste in Music, and Alcohol Consumption. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 2(1):48–56.
- Tobler, W. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46(2):234–240.
- Walker, R. (2007). Path, Phat, and State Dependence in Estimating Observation-driven Markov Models. *Working Paper, Presented at the 24th Summer Meeting of the Society of Political Methodology*.
- Wasserman, S. (1980a). A stochastic model for directed graphs with transition rates determined by reciprocity. *Sociological methodology*, 11:392–412.
- Wasserman, S. (1980b). Analyzing social networks as stochastic processes. *Journal of the American Statistical Association*, 75(370):280–294.