9

MODELS FOR GENERALIZED LOCATION SYSTEMS

*Carter T. Butts**

A formal framework is introduced for a general class of assignment systems that can be used to characterize a range of social phenomena. An exponential family of distributions is developed for modeling such systems, allowing for the incorporation of both attributional and relational covariates. Methods are shown for simulation and inference using the location system model. Two illustrative applications (occupational stratification and residential settlement patterns) are presented, and simulation is employed to show the behavior of the location system model in each case; a third application, involving occupancy of positions within an organization, is used to demonstrate inference for the location system. By leveraging established results in the fields of social network analysis, spatial statistics, and statistical mechanics, it is argued that sociologists can model complex social systems without sacrificing inferential tractability.

1. INTRODUCTION

Social systems take many forms and may be studied at multiple levels. Despite more than a century of effort by sociologists to subsume

This paper is based upon work supported by National Institutes of Health award 5 R01 DA012831-05, subaward 918197. I would like to thank Miruna Petrescu-Prahova, Mark Handcock, Garry Robins, Pip Pattison, John Skvoretz, and several anonymous reviewers for their input and advice. Direct correspondence to Carter T. Butts, Department of Sociology, SSPA 2145, University of California-Irvine, Irvine, CA 92697-5100, buttsc@uci.edu.

*University of California, Irvine

this diversity under a single unifying framework—the grand theoretical narratives of Spencer (1896), Sorokin (1957), Parsons (1951), and Coleman (1990) being classic examples—the goal of theoretical unification continues to prove elusive. While the possibility—or even the desirability—of a unified framework for studying social phenomena remains an open question, theorists such as Fararo and Skvoretz (1987) and Fararo (1989) persuasively argue that more modest progress can be made by identifying and exploiting connections among particular classes of social systems. Where common features are present, it may be possible to identify model families that are broad enough to permit cross-fertilization of findings while remaining narrow enough to be deployable in practical settings. This "middle range" approach (Merton 1957) is as much a methodological as it is a theoretical endeavor: to be of scientific use, even a limited unified framework requires a common set of formal representations, deductive methods, and measurement techniques. An important challenge, then, is to identify classes of social phenomena that admit such methodological and conceptual unity and to construct formal systems to facilitate their measurement and analysis.

One promising candidate for a more unified treatment is the broad class of social systems that involve the arrangement of social entities (here referred to generically as "objects") with respect to "locations" of one sort or another. The notions of "object" and "location" employed here are intended to be quite general— for example, encompassing individuals, households, and organizations in the former case and social, economic, or physical positions in the latter. For instance, individuals may hold particular jobs in a market economy (Kelso and Crawford 1982; Sattinger 1993), households may occupy particular housing units (Massey and Denton 1993; Benenson 2004), and firms may elect to site their facilities in particular locales (Sweeney and Feser 2003; Feser, Sweeney, and Renski 2005). At a more abstract level, we may even think of relational structures as involving the allocation of individual entities to structural positions (Lorrain and White 1971; Doreian, Batagelj, and Ferlioj 2005), and discrete choice behavior as individuals' "allocation" of preference to options from a choice set (Luce 1959). What these otherwise disparate systems have in common is that they can be represented (to a first approximation, at least) in terms of a discrete set of objects, a discrete set of locations, and a mapping that assigns the elements of the first set to elements of the second set. Although the location of one object may be independent of the locations of others

in the system, this is not generally the case: in many systems, the objects' location assignments depend on each other in complex ways. In the case of residential settlement, for instance, local tendencies toward homophily or xenophobia affect households' selection of housing units, leading in many cases to racial and/or ethnic segregation at large scales (Schelling 1969; Sakoda 1971; Fossett 2006). Likewise, compositional variations in the populations of jobs and job seekers interact to affect the entry of younger workers into high-ranking positions and can combine with discrimination by race or gender to maintain robust patterns of stratification within and across organizations (see Stewman [1988] for a review). Capturing the behavior of these systems thus requires modeling not only baseline tendencies for certain objects to occupy certain locations but also the impact of interdependencies and occupancy constraints.

The focus of this paper is the development of a general framework for modeling and analysis of systems that can be specified in terms of the arrangement of a finite set of objects with respect to a finite set of locations. As motivated by the examples mentioned above, the behavior of these "generalized location systems" (as we shall call them) may involve dependencies among objects and/or locations as well as constraints on which (and how many) objects can occupy particular locations at any given point in time. To deal with this challenge, we employ a core formalism (the discrete exponential family) that allows us to leverage the large literature on the stochastic modeling of systems with nontrivial dependence structures. This formalism also allows us to construct models that are applicable across a wide range of substantive contexts; that scale well to large social systems; that are readily simulated; that are specifiable in terms of directly measurable properties; and that support likelihood-based inference using (fairly) standard methods. Although framed holistically in terms of system-level behavior (an approach advocated by Mayhew [1980, 1981], among others), models generated under this framework can also be interpreted as arising from certain types of microlevel processes; where the appropriate assumptions are met, therefore, model parameters may be understood in terms of object-level behavior. While procedures for parametrization and efficient simulation of models for generalized location systems occupy the bulk of what is treated here, basic methods for likelihood-based inference on location system parameters from cross-sectional data will also be presented. These methods provide a direct mechanism for empirical

evaluation of competing theoretical claims and are hence an important benefit of this approach. Finally, three illustrative examples will be shown, which demonstrate how processes of occupational stratification and residential settlement can be modeled within the location system framework.

1.1. *Modeling Location Systems: Some Prior Approaches*

As we might expect from the substantive diversity of generalized location systems, a number of distinct modeling approaches have been suggested for location systems in particular empirical contexts. Most commonly, researchers have modeled specific properties of occupied positions (or sets thereof) without attempting to capture the behavior of the system as a whole. Much of the literature on income and educational attainment is in this vein (e.g., Beck, Horan, and Tolbert 1978; Budig and England 2001; Joy 2003; Huffman and Cohen 2004), as is much of the work on residential segregation (Tauber and Tauber 1965; Massey and Denton 1993) and the compositional properties of neighborhoods or other regions (Galster 1982; Frey and Farley 1994). A rather different approach may be found in the family of *stochastic choice models* (Luce 1959; McFadden 1973), which in our terms model the allocation of a decision (object) to a set of possible options (locations). In addition to its applications in economic contexts (Loehman and De 1982; Corstjens and Gautschi 1983; Eckstein and Wolpin 1989), this approach is widely employed in geography and transportation engineering to model route selection (Bovy and Stern 1990; Oppenheim 1995). Common to all of these approaches are severe limits on the types of dependence among system elements that can be modeled. Regression-based approaches (e.g., income attainment models) do not generally treat the assignment process directly and they treat observations as independent conditional on object (and sometimes location) covariates. Spatial autocorrelation models (Anselin 1988) may be employed to capture certain types of dependence but still neglect factors such as occupancy constraints. Standard stochastic choice models, on the other hand, treat decisions as independent while accounting for the composition of the location set (though not limits on total occupancy). A much richer range of dependence can be captured by *permutation* or *assignment models* (Butts, this volume, building on the work of Hubert [1987]), although these are limited to cases where the object/location mapping is 1:1. We shall have

more to say about this family later, since it can be viewed as a special case of the models studied here.

In addition to these statistically oriented families, various dynamic and agent-based approaches have also been used to study particular location systems. Some of the best known of these include Schelling's (1969) and Sakoda's (1971) models of residential settlement, which have their modern incarnation in models such as those of Benenson (2004) and Fossett (2006). White's (1970) classic treatise on "vacancy chains" can also be seen as proposing a number of dynamic models for location systems arising in organizational contexts, and it has spawned a large literature of its own (see Stewman [1988] for a review). A closely related literature in economics deals with *matching models*, which capture strategic interactions in contexts (e.g., labor markets) in which transactions involve discrete matches between traders (Shapley and Shubik 1971; Kelso and Crawford 1982; Roth and Sotomayor 1990). Although most research on matching models has focused on either game-theoretic solutions or large-population behavior in the deterministic limit, recent work by Zhang (2004) employs a boundedly rational stochastic choice model that recasts the Schelling model in terms of a family of *potential games* (Monderer and Shapley 1996; Young 1998). Such games have the property that the differences in any actor's utilities for unilateral strategy changes can be expressed as differences in a real-valued function that depends only on the change being considered and the strategies being employed by other actors (and not, for example, on which actor is involved). Potential games also have the appealing property of having well-defined equilibrium dynamics under stochastic choice (Young, 1998), which Zhang employs to deduce long-run system behavior given a particular family of utility functions. Although Zhang's (2004) discussion is limited to a specific case, the model in question belongs to a much larger family of stochastic processes that will be presented here; thus, it is possible to carry out simulation and inference for the Zhang model (and extensions of it) using the methods discussed in this paper.

Overall, then, prior efforts to understand location systems have generally been domain-specific and have broadly represented a tradeoff between models with well-understood inferential properties (e.g., regression models) and models that capture more complex interactions among elements (e.g., agent-based models). Although specialized models have merit in many situations, there is also value in pursuing a more unified approach; likewise, it is scientifically desirable to have standard

methods for both deduction and inference within the same framework. A successful example of this unifying strategy can be found in the field of social network analysis, where a common formalism (the discrete exponential family) has been used to model a wide (and growing) range of structural phenomena. Because this approach is closely related to that described here, we briefly consider it before continuing to an in-depth discussion of the generalized location system framework.

1.2. *Statistical Models for Dependent Systems: The Case of Network Analysis*

Social networks pose substantial modeling challenges, due to the interdependence of their component parts. Formally, we may think of a network as consisting of a set of *vertices* (representing individual actors) that are connected by *edges* (representing ties between actors); such a structure is generically referred to as a *graph*. While the simplest models of network structure assume all ties to be independent (e.g., the famous random graph models of Erdös and Rényi [1960]), this is at best a loose approximation. Typical social networks display properties such as reciprocity or asymmetry (in which an edge from i to j depends on the state of the corresponding (j, i) edge), transitivity bias (in which the existence of an (i, j) edge is affected by the existence of a path from i to j through some intermediary, k), and even complex biases such as the avoidance of odd-length cycles (e.g., as in predominantly heterosexual networks ([Bearman et al. 2004]) or balanced negative-valence relations [Harary 1953]). Such properties may arise as an artifact of unobserved dyadic effects (such as homophily [McPherson, Smith-Lovin, and Cook 2001]; see Hoff, Raftery, and Handcock [2002] for a statistical discussion), or from intrinsic dynamics (e.g., see Carley [1991]; Hummon and Doreian [2003]). Regardless of how they occur, however, these properties reflect potentially complex patterns of dependence among edges and advances in network modeling have thus required researchers to cope with this problem.

Over the past quarter-century, a great deal of progress has been made within the social network field toward developing practical models for social systems with complex dependence structures. This work began in earnest with the log-linear models of Holland and Leinhardt (1981) and Fienberg and Wasserman (1981), which were

extended by various researchers (e.g., Fienberg, Meyer, and Wasserman 1985; Holland, Laskey, and Leinhardt 1983) to cover more complex cases. In a series of important developments (starting with the foundational work of Frank and Strauss [1986]), this approach was generalized to incorporate processes with at first local and then general dependence among edges (Strauss and Ikeda 1990; Wasserman and Pattison 1996; Pattison and Wasserman 1999; Robins, Pattison, and Wasserman, 1999). What united these various efforts was the use of discrete exponential families as a formalism for representing general classes of distributions on graphs. Given a random graph G drawn from a finite set of possible graphs $\mathbb{G}$, we may write the distribution of G as $\Pr(G = g) \propto \exp(\theta^T t(g))$, where θ is a vector of real-valued parameters and t is a vector of real-valued functions on $\mathbb{G}$. t may be chosen with very few constraints, although most common models involve counts of structural features such as mutual dyads, star formations, or triangles (some of the technical reasons for this are sketched in Wasserman and Robins [2005]). Since the log-probability of a graph is proportional to the weighted sum of its statistics, t may be intuitively understood as describing structural features that are either enhanced (where θ is positive) or suppressed (where θ is negative) by the model. Models represented in this way are known as *exponential random graph (ERG)* or $p*$ models, although this is more properly understood as a description of their parametrization rather than their content; indeed, any fixed distribution on a finite $\mathbb{G}$ can be written in this manner (albeit not always parsimoniously). By providing a general framework for the parametrization of graph distributions with nontrivial dependence (via choice of t) and via its associated inferential theory (Barndorff-Nielsen 1978; Brown 1986), the discrete exponential family formalism has been central to progress in the network modeling area.

Recent work has expanded on these innovations with improved inferential strategies (Crouch, Wasserman, and Trachtenburg 1998; Snijders 2002; Hunter and Handcock 2006), new parameterizations for structural effects (Snijders et al. 2006), and an expanded understanding of the models themselves (Handcock 2003b; Robins, Pattison, and Woolcock, 2005). While these models have often been couched in purely methodological terms, it has become increasingly apparent that they can be employed to capture theoretically relevant local influences on structure formation (Robins et al. 2005; Robins and Pattison 2005), as well as (in some cases) mechanisms of structural evolution (Robins and

Pattison 2001). Although much of this work is still in a fairly early stage of development, the foundations have arguably been laid for a minor revolution in structural analysis.

While time will tell if this promise is realized, the successes that have so far been obtained underscore the aforementioned value of synthesis in scientific research. Rather than arising in isolation, they have resulted from cross-application of work in fields as diverse as spatial statistics (e.g., Besag 1974, 1975) and statistical physics (Strauss 1986; Swendsen and Wang 1987), as well as innovations in computing technology and simulation methods (Geyer and Thompson 1992; Gamerman 1997). Although frequently motivated by substantive concerns (e.g., the desire to model balance-theoretic influences [Heider 1958]), modelers have also attempted to work with general formalisms that can be deployed on networks arising within many different substantive contexts. By drawing on results obtained by researchers studying structurally similar problems in other substantive areas, then, network researchers have been able to greatly accelerate development in their own field.

Cross-application of concepts and methods has led to great strides in network analysis, but there is more that can be done. As promising as the developments cited above have been, few if any attempts have been made to extend them to problems other than network formation and diffusion. It has already been noted that processes such as stratification, settlement patterns, migration, firm siting, and occupational segregation pose similar challenges of complex dependence, but they are currently studied through a variety of (generally incompatible) modeling frameworks. A more generic approach would facilitate the cross-application of findings and techniques, thereby laying the groundwork for cumulative theoretical development and, ultimately, unification (Fararo and Skvoretz 1987; Fararo 1989). As we shall see in Section 3.1, the tools for creating such a unifying framework can be found in the same modeling techniques now being employed for social networks; this paper is intended as a first step in this direction.

1.3. *A Brief Comment on Notation*

We here outline some general notation, which will be used in the material that follows. A graph, G, is defined as $G = (V, E)$, where V is a set of vertices and E is a set of edges on V. When applied to sets, $|\cdot|$ represents

cardinality; thus $|V|$ is the number of vertices (or order) of G. In some cases (particularly when dealing with valued graphs), it will be useful to represent graphs in adjacency matrix form, where the adjacency matrix $\mathbf{X}$ for graph G is defined as a $|V| \times |V|$ matrix such that $\mathbf{X}_{ij}$ is the value of the (i, j) edge in G. By convention, $\mathbf{X}_{ij} = 0$ if G contains no (i, j) edge. A tuple of graphs $(G_1, \ldots, G_n)$ on common vertex set V may be similarly represented by a $n \times |V| \times |V|$ adjacency array, $\mathbf{X}$, such that $\mathbf{X}_{i..}$ is the adjacency matrix for G_i.

When referring to a random variable, X, we denote the probability of a particular event x by $\Pr(X = x)$. More generically, $\Pr(X)$ refers to the probability mass function of X (where X is discrete). Expectation is denoted by the operator $\mathbf{E}$, with subscripts used to designate conditioning where necessary. Thus, the parametric pmf $\Pr(X|\theta)$ leads to the corresponding expectation $\mathbf{E}_\theta(X)$. (Likewise for variance, written $\mathrm{Var}_\theta(X)$.) When discussing sequences of realizations of a random variable X, parenthetical superscript notation is used to designate particular draws—for example, $(x^{(1)}, \ldots, x^{(n)})$.

On occasion, some specialized vector notation will also be employed. For vector $\mathbf{x}$, $\mathbf{x}_{-i}$ refers to all elements of $\mathbf{x}$ *other than* the ith. Thus, $\Pr(\mathbf{X} = \mathbf{x} \mid \mathbf{X}_{-i} = \mathbf{x}_{-i})$ refers to the probability that random vector $\mathbf{X}$ is equal to $\mathbf{x}$, conditional on the non-ith values of $\mathbf{X}$ being equal to $\mathbf{x}_{-i}$. In addition to the above, we will also at times need to refer to a vector for which specific values have been replaced (all others remaining unchanged). To this end, the expression ${}^{i,j}\mathbf{x}$ is used to denote a vector whose ith element has been fixed to j and whose other elements are equal to $\mathbf{x}$.

2. GENERALIZED LOCATION SYSTEMS

Our focus here is on what we shall call *generalized location systems*, which represent the allocation of arbitrary entities (e.g., persons, objects, organizations) to "locations" (e.g., physical regions, jobs, social roles). While our intent is to maintain a high level of generality, we will limit ourselves to systems for which both entities and locations are countable and discrete and for which it is meaningful to treat the properties of entities and locations as relatively stable (at least for purposes of analysis). Relaxation of these constraints is possible but will not be pursued here; as will be shown, the present framework still allows for a great deal of flexibility.

We begin our development by assuming a system that consists of n identifiable *objects*, $O = (o_1, \ldots, o_n)$, each of which may reside in exactly one of m identifiable *locations*, $L = (l_1, \ldots, l_m)$. The current state of this system is given by a *configuration vector*, $\ell \in \{1, \ldots, m\}^n$, which is defined such that $\ell_i = j$ iff o_i resides at location l_j. The set of all such configuration vectors that are realizable is said to be the set of *accessible configurations* and is denoted $\mathbb{C}$. One very important parametrization of $\mathbb{C}$ with which we will deal is in terms of *occupancy constraints*. We define the *occupancy function* of a location system as

$$P(x, \ell) = \sum_{i=1}^{n} I(\ell_i = x), \tag{1}$$

where I is the standard indicator function. The vectors of maximum and minimum occupancies for a given location system are composed of the minimum/maximum values of the occupancy function for each state under $\mathbb{C}$ (respectively). That is, we require that $P_i^- \leq P(i, \ell) \leq P_i^+$ for all $i \in 1, \ldots, m$, $\ell \in \mathbb{C}$, where P^-, P^+ are the minimum and maximum occupancy vectors. If $P_i^- = P_i^+ = 1 \ \forall \ i \in 1, \ldots, m$, then it follows that $\boldsymbol{\ell}$ is a permutation vector on $1, \ldots, n$, in which case we must have $m = n$ for non-empty $\mathbb{C}$. This is an important special case, particularly in organizational contexts (White 1970). By contrast, it is frequently the case in geographical contexts (e.g., settlement) that $P_i^- = 0$ and $P_i^+ > n$ $\forall \ i \in 1, \ldots, m$, in which case occupancy is effectively unconstrained.

In addition to configurations and labels, objects and locations typically possess other properties of scientific interest. We refer to these as *features*, with F_O being the set of object features and F_L being the set of location features. While we do not (initially) place constraints on the feature sets, it is worth highlighting two feature types that are of special interest. Feature vectors provide ways of assigning numerical values to individual objects or locations—for example, age, average rent level, or wage rate. Adjacency matrices can also serve as important features, encoding dyadic relationships among objects or locations. Examples of such relationships can include travel distance, marital ties, or demographic similarity. Because relational features allow for coupling of objects or locations, they play a central role in the modeling of complex social processes (as we shall see).

To draw the above together, we define a generalized location system by the tuple $(L, O, \mathbb{C}, F_L, F_O)$. The state of the system is given

by $\ell \in \mathbb{C}$, which will be of primary modeling interest. Various specifications of $\mathbb{C}$ are possible, but particular emphasis is placed on occupancy constraints, which specify the range of populations that each location can support. With these elements, it is possible to model a wide range of social systems and it is to this problem that we now turn.

3. MODELING LOCATION SYSTEMS

Although many approaches to location system modeling are possible, we will here focus on models for the observation of configuration vectors at arbitrary times. In this sense, our focus is on the stochastic equilibrium behavior of the system: if we take a snapshot of a system at any given instant, what is the probability of observing one configuration rather than another? While this perspective can be expanded upon, it nevertheless allows us to say a fair amount regarding system behavior. In modeling state probabilities, it is also essential that the models be constructed in such a way as to allow for inference from extant data; while this may seem to be a self-evident constraint, even a brief perusal of the sociological literature reveals that this condition is often unsatisfied. Finally, it must be the case that the location model be capable of capturing the sorts of complex dependencies that are known to operate within large-scale social systems. These include homogeneity effects, density dependence, homophily/propinquity, and capacity constraints, in addition to more prosaic attraction/repulsion mechanisms.

In this section, we provide a modeling framework that satisfies these constraints. The core of this framework is a discrete exponential family of distributions that is closely linked to related models employed in spatial statistics, statistical mechanics, and social network analysis. Although our focus will be on the modeling of location systems per se, we frequently draw upon results from these fields. Our treatment of the topic begins with the development of the general location system model and proceeds to a specific family of submodels that incorporates a range of substantively important effects in a reasonably simple fashion. Some simple inferential properties of the location system are also discussed as well as methods for simulating draws from the location system model.

3.1. *The General Model*

We here define a stochastic model for the equilibrium state of a generalized location system. In particular, we assume that the system is ergodic and—given a set of accessible configurations, $\mathbb{C}$—the system will be found to occupy any particular configuration, $\boldsymbol{\ell}$, with some specified probability.[1] Our primary interest is in the modeling of these equilibrium probabilities, although some dynamic extensions are possible.

Given the above, we first define the set indicator function

$$I_{\mathbb{C}}(\ell) = \begin{cases} 1 & \text{if } \ell \in \mathbb{C} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The equilibrium probability of observing a given configuration can then be written as

$$\Pr(\mathrm{S} = \ell) = \mathrm{I}_{\mathbb{C}}(\ell) \frac{\exp(\mathcal{P}(\ell))}{\sum_{\ell' \in \mathbb{C}} \exp(\mathcal{P}(\ell'))}, \tag{3}$$

where S is the random state and $\mathcal{P}$ is a quantity called the *social potential* (defined below). The sum

$$Z(\mathcal{P}, \mathbb{C}) = \sum_{\ell' \in \mathbb{C}} \exp\big(\mathcal{P}\left(\ell'\right)\big) \tag{4}$$

is the *normalizing factor* for the location model and corresponds to the partition function of statistical mechanics (Kittel and Kroemer 1980). Clearly, equation (3) defines a discrete exponential family with support on $\mathbb{C}$ and it is generic in the sense that any distribution on $\mathbb{C}$ can be written in the form of equation (3). There are several benefits to working within such a framework. First, as noted, the framework is complete with respect to the underlying location system. Second, much is known about models with the form of equation (3). We have already seen that a parallel formalism (the ERG) is widely used to construct models of networks with complex dependence and the formalism is similarly common

[1] If this assumption does not hold, the model may still be interpreted in terms of the probability of a single observation arising from a static process with the distribution of equation (3). For purposes of exposition, however, ergodicity (and the equilibrium interpretation) will be assumed.

in both physics (Kittel and Kroemer 1980) and mathematical statistics (e.g., see Besag 1974, 1975; Barndorff-Nielsen 1978; Stoyan, Kendall, and Mecke 1987). However, the most important property of the exponential family framework is perhaps the third: given an appropriate parametrization of $\mathcal{P}$, there are existing results that permit principled inference from empirical data (Johansen 1979; Brown 1986). While its deductive value is also important, the availability of viable inferential tools is a major motivation for our approach; models that can capture complex social processes are of little use if they cannot be evaluated on readily available data.

While equation (3) can represent any distribution on $\mathbb{C}$, its scientific utility clearly lies in the specification of $\mathcal{P}$. Intuitively, the social potential for any given configuration is equal to its log-probability, up to an additive constant. Thus, the location system is more likely to be found in areas of high potential, and/or (in a dynamic context) to spend more time in such states. While any number of forms for $\mathcal{P}$ could be proposed, we begin with a constrained family that incorporates a number of features of known substantive importance for a variety of social systems. This form is introduced in the section that follows.

3.2. *A Family of Social Potentials*

As noted above, we seek a family of functions $\mathcal{P} : \mathbb{C} \mapsto \mathbb{R}$ such that $\Pr(S = \ell) \propto \exp(\mathcal{P}(\ell))$. This family should incorporate as wide a range of substantively meaningful effects as possible; since it is not reasonable to expect effects to be identical in every situation, the family should be *parametrized* so as to allow differential weighting of effects. Ideally, the social potential family should also be easily computed and its structure easily interpreted. For our present purposes, these latter qualities will guide our construction of $\mathcal{P}$. It should be noted, however, that alternative approaches are possible (such as the direct use of dependence graphs [Besag 1974; Pattison and Robins 2002; Wasserman and Robins 2005]) and the potentials so generated may be deployed in the same manner as those discussed here.

An obvious initial solution to this problem is to construct $\mathcal{P}$ from a linear combination of sufficient statistics (i.e., deterministic functions of ℓ, F_L and F_O). Employing such a potential function within equation (3) leads to a regular exponential family on $\mathbb{C}$ (Johansen 1979),

which has a number of useful statistical implications. Of course, such an approach also has the usual virtues of linear families (additivity of effects, nesting, etc.) familiar to most social scientists. The canonical ERG parametrization shown in Section 1.2 is of this form as well, although some recent parameterizations employ nonlinear constraints on the manner in which sufficient statistics are weighted (Hunter and Handcock 2006; Snijders et al. 2006).

Even if a linear form is supposed, however, we are left with a more important question: what statistics should be included in the social potential? Obviously, these statistics must be parametrized as functions of the location and object features. Since the impact of the social potential is invariant up to an arbitrary additive constant, per equation (3), it follows that properties that are invariant over $\mathbb{C}$—i.e., those that depend only on aspects of the populations of objects or locations without regard to the assignment of objects to locations—can be safely ignored. All relevant statistics for the construction of $\mathcal{P}$ must thus involve some interaction between object features and the features of locations to which they are assigned; "main effects" of object or location features, in the usual sense of the term, are not meaningful in this context. With respect to the features themselves, these may include both attributes (features of the individual location or object per se) and relations (features of object or location sets). Here, we will limit ourselves to relations that are dyadic (i.e., defined on pairs) and single-mode (i.e., that do not mix objects and locations). Thus, our effects should be functions of feature vectors, and/or (possibly valued) graphs.

While this may seem to leave innumerable possibilities, we can further focus our attention by noting that the purpose of $\mathcal{P}$ is ultimately to control the assignment of objects to locations. This suggests immediately that the effects of greatest substantive importance will be those that draw objects toward or away from particular locations. Table 1 provides one categorization of such effects by feature type. In the first (upper-left) cell, we find effects that express direct attraction or repulsion between particular objects and locations, based on their attributes. In the second (upper-right) cell are effects that express a tendency for objects linked through connected locations to be particularly similar or distinct. (Spatial autocorrelation is a classic example of such an effect.) The converse family of effects is found in the third (lower-left) cell; these effects represent a tendency for objects to be connected to other objects with similar (or different) locations. A tendency for husbands and wives to make

TABLE 1
Elements of the Social Potential

	Location Attributes	Location Relations
Object Attributes	Attraction/repulsion effects	Object homogeneity/heterogeneity effects (through locations)
Object Relations	Location homogeneity/heterogeneity effects (through objects)	Alignment effects

similar career choices—where careers are interpreted as "locations"—serves as an example of a location homogeneity effect. Note that the essential difference between cells two and three lies in whether the clustering/dispersion of similar objects is being assessed (cell 2) versus the assignment of connected (but not necessarily similar) objects to similar/dissimilar locations (cell 3). This will be discussed in greater detail below. Finally, in the fourth (lower-right) cell we have effects based on the tendency of location relations to align (or disalign) with object relations. Propinquity, for example, is a tendency for adjacent objects to reside in nearby locations.

Taken together, these four categories of effects combine to form the social potential. Under the assumption of linear decomposability, we thus posit four subpotentials (one for each category) such that

$$\mathcal{P}(\ell) = \mathcal{P}_\alpha(\ell) + \mathcal{P}_\beta(\ell) + \mathcal{P}_\gamma(\ell) + \mathcal{P}_\delta(\ell), \tag{5}$$

where $\mathcal{P}_\alpha$ is the potential associated with attraction/repulsion effects, $\mathcal{P}_\beta$ is the potential associated with object heterogeneity effects, $\mathcal{P}_\gamma$ is the potential associated with location heterogeneity effects, and $\mathcal{P}_\delta$ is the potential associated with alignment effects. We now consider each of these functions in turn.

3.2.1. *Attraction/Repulsion Potential*

The first class of effects that must be represented in any practical location system are global attraction/repulsion—also called "push/pull"—effects. Residential locations, potential firm sites, occupations, and the like have features that make them generally likely to attract or repel certain objects (be they persons, organizations, or other entities).

Such effects are naturally modeled via product-moments of attributes. Let $\mathbf{Q} \in \mathbb{R}^{m \times a}$, $\mathbf{X} \in \mathbb{R}^{n \times a}$ be exogenous features reflecting location and object attributes (respectively) and let $\alpha \in \mathbb{R}^a$ be a parameter vector. Then we may define $\mathcal{P}_\alpha$ as

$$\mathcal{P}_\alpha(\ell) = \sum_{i=1}^{a} \alpha_i t_i^\alpha(\ell), \tag{6}$$

$$= \sum_{i=1}^{a} \alpha_i \sum_{j=1}^{n} \mathbf{Q}_{\ell j i} \mathbf{X}_{ji}, \tag{7}$$

where t^α is a vector of sufficient statistics.

The behavior of equation (7) is quite intuitive. For instance, let $\mathbf{Q}_i$ be a location feature and let $\mathbf{X}_i = (1, \ldots, 1)$ be a constant object feature. Then $\alpha_i > 0$ and $\alpha_i < 0$ produce attraction and repulsion effects (respectively) based on $\mathbf{Q}_i$. If the effect in question is stronger or weaker for particular objects, this may in turn be produced by allowing $\mathbf{X}_i$ to vary.

One substantively important case of such an effect is discrimination. Discrimination may be understood as a conditional tendency for individuals with certain features to be placed in (or denied access to) certain positions. In terms of social potential, this is simply a push/pull effect where $\mathbf{Q}_i$ describes the location feature with respect to which discrimination is occurring and $\mathbf{X}_i$ encodes the individual feature or group membership that is the basis of discrimination. Such an approach is operationally similar to the treatment used in conventional regression analyses of wage discrimination (e.g., Huffman and Cohen 2004), although there is an important difference in interpretation. While a wage discrimination effect represents a marginal increase/decrease in wages for persons with certain features,[2] a discrimination effect within the location model represents a conditional tendency for persons with certain features to be differentially assigned to particular positions (or positions with particular features). The difference between the two may be appreciated by contemplating a hypothetical change in which $\mathbf{X}_i$ becomes

[2] Note that this "increase" may be interpreted as a difference in means rather than a causal difference. The latter is commonly employed, however.

identical for all actors. This would lead a conventional wage discrimination effect model to predict a mean shift in population wages, while such a shift need not occur under the location model. This last is because a location system model of wage attainment effectively models the process of competition among workers for a fixed set of potential jobs (not all of which have to be filled), rather than taking wage to be a property that arises from the intrinsic properties of the actors themselves (irrespective of the jobs available). The ability to capture such institutional constraints is an attractive property of the location approach. Of course, discrimination effects need not be confined to wages—any tendency for differential assignment may be included in the same manner.

3.2.2. *Object Homogeneity/Heterogeneity Potential*

A second class of effects concerns object homogeneity/heterogeneity—that is, the conditional tendency for associated locations to be occupied by objects with similar (or different) features. Let $\mathbf{Y} \in \mathbb{R}^{n \times b}$ be a matrix of object attributes, $\mathbf{B} \in \mathbb{R}^{b \times m \times m}$ be an adjacency array on the location set, and $\beta \in \mathbb{R}^b$ be a parameter vector. We then define the object homogeneity/heterogeneity potential by

$$\mathcal{P}_\beta(\ell) = \sum_{i=1}^{b} \beta_i t_i^\beta(\ell), \tag{8}$$

$$= \sum_{i=1}^{b} \beta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{B}_{i\ell_j\ell_k} \left| \mathbf{Y}_{ji} - \mathbf{Y}_{ki} \right|, \tag{9}$$

where, as before, t^β is a vector of sufficient statistics. It should be noted that the form of t^β is closely related to Geary's C, a widely used index of spatial autocorrelation (Cliff and Ord 1973). t^β is based on absolute rather than squared differences and it is not normalized in the same manner as C, but its behavior is qualitatively similar in many respects.

As a simple illustration of $\mathcal{P}_\beta$, let L be a set of disjoint spatial regions with contiguity matrix $\mathbf{B}_{i\cdots}$ Let O represent a population of households and let $\mathbf{Y}_{\cdot i}$ be a vector representing an object feature (e.g., a categorical code for racial self-identification of the primary household informant). Then $\beta_i < 0$ corresponds to a tendency for households with similar features (here, race) to be contiguously located, while $\beta_i > 0$ favors a heterogeneous assignment. Put another way, negative β values

induce homogeneity or segregation, while positive β values induce heterogeneity or supra-random mixing. This situation can be complicated further by allowing $\mathbf{B}$ to take on arbitrary values: the magnitude of $\mathbf{B}_{ijk}$ controls the strength of connection between the j, k locations on the ith feature, while the sign of $\mathbf{B}_{ijk}$ determines whether $\beta_i > 0$ induces heterogeneity ($\mathbf{B}_{ijk} > 0$) or homogeneity ($\mathbf{B}_{ijk} < 0$). Thus, it is possible to model both effects within the same relation. Similarly, a diagonal $\mathbf{B}_{i..}$ matrix can be used to model homogeneity/heterogeneity *within* locations in the absence of cross-location ties. Such a structure may be employed, for instance, when attempting to model occupational segregation; in this case, L represents the set of occupations and setting $\mathbf{B}_{i..}$ equal to the identity matrix allows β_i to directly parametrize the extent of "segregation pressure" within the system.

3.2.3. *Location Homogeneity/Heterogeneity Potential*

The parallel case to $\mathcal{P}_\beta$ is $\mathcal{P}_\gamma$, which models the effect of location homogeneity or heterogeneity through objects. Let $\mathbf{R} \in \mathbb{R}^{m \times c}$ be a matrix of location features, $\mathbf{A} \in \mathbb{R}^{c \times n \times n}$ be an adjacency array on the object set and $\gamma \in \mathbb{R}^c$ be a parameter vector. We then define $\mathcal{P}_\gamma$ as follows:

$$\mathcal{P}_\gamma(\ell) = \sum_{i=1}^{c} \gamma_i t_i^\gamma(\ell) \tag{10}$$

$$= \sum_{i=1}^{c} \gamma_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{A}_{ijk} \left| \mathbf{R}_{\ell_j i} - \mathbf{R}_{\ell_k i} \right| . \tag{11}$$

As implied by the above, t^γ is the vector of sufficient statistics for location homogeneity. t^γ is at core similar to t^β, save in that the role of object and location are reversed: absolute differences are now taken with respect to *location* features and are evaluated with respect to the connections between the objects occupying said locations.

While $\mathcal{P}_\gamma$ may seem less intuitive than $\mathcal{P}_\beta$, its utility is easily demonstrated via a simple example. Consider, for instance, the case of wage rates within married couples. To set up the problem, we begin by letting $\mathbf{A}_{i..}$ be a matrix representing all marital ties among members of the sample; this will consist of a set of isolated symmetric dyads, accompanied by isolates if the sample includes unmarried persons. L is taken in this case to be a collection of jobs, each of which is associated with a wage

rate (contained in $\mathbf{R}_{\cdot i}$). For $\gamma_i > 0$, $\mathcal{P}_\gamma$ then places more weight on job allocations that increase the within-couple wage rate differences (*ceteris paribus*), and $\gamma_i < 0$ produces the opposite effect (i.e., within-couple wage homogeneity). Processes leading to within-couple wage heterogeneity have been postulated by Becker (1991), among others; by turns, several processes identified by social capital theorists (Granovetter 1973; Calvo-Armengol and Jackson 2004) would be expected to lead to within-couple homogeneity in wage rates. Such effects can be modeled directly through $\mathcal{P}_\gamma$, above and beyond other allocative mechanisms.

3.2.4. *Alignment Potential*

The final element of the social potential is the alignment potential, $\mathcal{P}_\delta$, which expresses tendencies toward alignment or disalignment of object and location relations. Given object and location adjacency arrays $\mathbf{W} \in \mathbb{R}^{d\times n\times n}$ and $\mathbf{D} \in \mathbb{R}^{d\times m\times m}$ (respectively) and parameter vector $\delta \in \mathbb{R}^d$, the alignment potential is given by

$$\mathcal{P}_\delta(\ell) = \sum_{i=1}^{d} \delta_i t_i^\delta(\ell) \tag{12}$$

$$= \sum_{i=1}^{d} \delta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{W}_{ijk} \mathbf{D}_{i\ell_j\ell_k} \tag{13}$$

where, as in the prior cases, t^δ represents the vector of sufficient statistics. The form chosen for t^δ is Hubert's gamma, which is the standard matrix cross-product moment (see Hubert [1987] for a range of applications).

Although the alignment potential has been utilized in prior work on graph comparison (see Butts, this volume), our application is more concerned with modeling the direct impact of relations on location assignment. As the name implies, the alignment potential captures the extent to which relations among objects are mirrored by relations among their associated locations. Consider, for instance, a collection of disjoint spatial regions with travel distance matrix $\mathbf{D}_{i\cdot\cdot}$, and a population of actors whose kinship network is represented by the adjacency matrix $\mathbf{W}_{i\cdots}$ Where $\delta_i < 0$, the kinship network is propinquitous; that is, actors tend to reside (*ceteris paribus*) in locations that are physically proximate. By contrast, $\delta_i > 0$ would indicate a dispersal effect, in which actors who are tied to one another tend to occupy more distant locations. (Such an

effect might be expected, for instance, among firms that are tied to one another via production of similar products.)

Another important alignment effect is density dependence—the tendency for objects to cluster (positive dependence) or disperse (negative dependence) with respect to locations. To model density dependence, we create an object relation $\mathbf{W}_{i..}$ representing a complete graph and employ the identity matrix for $\mathbf{D}_{i...}$ Under this construction, t_i^δ indexes the extent to which objects are clustered in a small number of locations; $\delta_i > 0$ increases this tendency, while $\delta_i < 0$ inhibits it. Replacing the identity matrix with an inverse distance matrix allows for a more general form of spatial dependence, but the general intuition is similar.

As a final point, it may be noted that the alignment effect is "generic" for the potential family employed here in the sense that the attraction/repulsion, object heterogeneity, and location heterogeneity statistics can be written as alignment statistics on suitably transformed input matrices. We may thus usefully characterize the present social potential family as that composed of all potential linear combinations of matrix cross product-moments between object and location features. While this also means that (given appropriate data transformations) $\mathcal{P}$ can be written entirely in terms of $\mathcal{P}_\delta$, we continue to separate the subpotentials throughout the paper. One reason for this is substantive: as Table 1 shows, each subpotential arises from a conceptually distinct combination of object and location features and is most easily understood in this fashion. Another reason for the separation is computational; specifically, there are computational shortcuts that are available for other effects, which cannot be realized in the generic $\mathcal{P}_\delta$ case. While we continue to draw such distinctions, then, it should be borne in mind that they are not essential in character.

3.2.5. *Combined Linear Potential*

We are now ready to form the combined linear social potential. Substituting the quantities of equations (7) through (13) into equation (5) gives us

$$\mathcal{P}(\ell) = \sum_{i=1}^{a} \alpha_i t_i^\alpha(\ell) + \sum_{i=1}^{b} \beta_i t_i^\beta(\ell) + \sum_{i=1}^{c} \gamma_i t_i^\gamma(\ell) + \sum_{i=1}^{d} \delta_i t_i^\delta(\ell) \quad (14)$$

in terms of sufficient statistics, or

$$
\begin{aligned}
&= \sum_{i=1}^{a} \alpha_i \sum_{j=1}^{n} \mathbf{Q}_{\ell_j i} \mathbf{X}_{ji} + \sum_{i=1}^{b} \beta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{B}_{i\ell_j\ell_k} \left| \mathbf{Y}_{ji} - \mathbf{Y}_{ki} \right| \\
&\quad + \sum_{i=1}^{c} \gamma_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{A}_{ijk} \left| \mathbf{R}_{\ell_j i} - \mathbf{R}_{\ell_k i} \right| \\
&\quad + \sum_{i=1}^{d} \delta_i \sum_{j=1}^{n} \sum_{k=1}^{n} \mathbf{W}_{ijk} \mathbf{D}_{i\ell_j\ell_k} \qquad (15)
\end{aligned}
$$

in terms of the underlying covariates. Together with equation (3), equation (15) specifies a regular exponential family of models for the generalized location system. As we have seen, this family allows for the independent specification of attraction/repulsion, heterogeneity/homogeneity, and alignment effects (including differential attractiveness, segregation, homophily/propinquity, and density dependence as special cases). We now proceed to a consideration of some of the properties of this model family, before turning to the problem of simulation.

3.2.6. *Interpreting the Social Potential*

Per equation (3), the social potential is a real-valued function $\mathcal{P}$ on $\mathbb{C}$ such that the probability of observing any given configuration $\ell \in \mathbb{C}$ is proportional to $\exp(\mathcal{P}(\ell))$. As we have seen, $\mathcal{P}$ can be constructed so as to incorporate a variety of effects, ranging from simple attraction/repulsion to the enhancement or suppression of heterogeneity. These effects are parametrized via one or more sufficient statistics, which are weighted by coefficients. Beyond the narrower discussion of Section 3.2, it is useful to consider some of the ways in which the social potential per se (and any associated parameters) may be interpreted. In the discussion that follows, we economize notation slightly by employing the general parameter vector $\theta = (\alpha, \beta, \gamma, \delta)$ in place of the four separate parameter vectors: attraction, object heterogeneity, location heterogeneity, and alignment parameter vectors. We will likewise concatenate the four sufficient statistic vectors as $t = (t^\alpha, t^\beta, t^\gamma, t^\delta)$, leading to the more compact social potential expression $\mathcal{P}(\ell) = \theta^T t(\ell)$. Given this, we now consider the interpretation of $\mathcal{P}$ in terms of total system behavior and of hypothetical microlevel dynamics.

Holistic Interpretation. Since $\Pr(S = \ell) \propto \exp(\mathcal{P}(\ell))$, $\mathcal{P}$ can be properly regarded as providing an expression for the joint behavior of the location system as a whole. In particular, the system is more likely to be observed in configurations of high potential than in configurations of low potential, which leads to an immediate interpretation for a given parameter, θ_i. Where $\theta_i > 0$, configurations for which the associated statistic t_i is large are given higher potential (*ceteris paribus*); thus positive values of θ_i indicate a general tendency of the location system to exhibit configurations with higher t_i values than would be otherwise observed. Similarly, negative values of θ_i imply that configurations with large values of t_i are suppressed. This is easily seen by considering the probability ratio for two hypothetical configurations $\ell, \ell' \in \mathbb{C}$ under potential $\mathcal{P}$:

$$\frac{\Pr(S = \ell')}{\Pr(S = \ell)} = \frac{\exp\big(\mathcal{P}(\ell')\big)}{Z(\mathcal{P}, \mathbb{C})} \frac{Z(\mathcal{P}, \mathbb{C})}{\exp\big(\mathcal{P}(\ell)\big)} \tag{16}$$

$$= \exp\big(\mathcal{P}(\ell') - \mathcal{P}(\ell)\big) \tag{17}$$

$$= \exp\Big(\theta^T\big(t(\ell') - t(\ell)\big)\Big). \tag{18}$$

Thus, every unit change in t_i multiplies the odds of observing ℓ' versus ℓ by a factor of e^{θ_i}. This observation provides a direct quantitative interpretation of potential function parameters, although it should be borne in mind that many statistics may not be fully separable in practice. This phenomenon is well-known in the context of ERG models, for which the intrinsic relationships between statistics can be particularly strong (e.g., see Handcock 2003b); it would be a mistake to view this as a unique property of models for dependent systems, however, since it will arise in any probability model whose sufficient statistics are potentially related to one another (e.g., OLS regression with correlated predictors). When dealing with heavily correlated statistics, it may be more useful to interpret the effects of parameters in batches (much as we might jointly interpret the parameters of a polynomial regression). Alternately, it may in some cases be helpful to reparameterize t so as to produce statistics that are closer to being orthogonal over $\mathbb{C}$. Regardless of any intra-t relationships, however, the relationship between relative state probability and changes in t described in equation (18) remains a valid means of interpreting the effect of θ as a whole.

In keeping with the above, it should be emphasized that equation (3) may be interpreted as a model for the distribution of a single cross-sectional observation of a location system, even where no dynamic equilibrium interpretation is appropriate. In this case, $\mathcal{P}$ simply parametrizes the dependence among system elements within the observed cross-section, irrespective of the putative generating mechanism. Similar "fall-back" interpretations exist for other models in this general class (for example, ERGs) and are not unique to the location system framework.

Microdynamic Interpretation. In building the location system model, it will be noticed that little has been said about the underlying microprocesses that give rise to the configuration vector, or about the detailed evolution of ℓ under equilibrium conditions (other than distributional properties). This omission is deliberate: the location system model may be viewed as *process agnostic*, in that there are many conceivable microprocesses that would give rise to the same equilibrium distribution. Nevertheless, there are some useful statements that can be made about dynamic aspects of location models and we review several of these here.

First, we posit a family of microprocesses whose long-run dynamics give rise to the equilibrium distribution of equation (3). Consider a process in which, at finite (but otherwise arbitrary) time intervals, a random object X is drawn with some fixed distribution such that all objects are selected with positive probability. Let ℓ be the pre-draw system state and let ${}^{X,i}\ell$ for $i \in 1, \ldots, m$ be equal to ℓ save for the assignment of object X to state i. The system then transitions from ℓ to ${}^{X,i}\ell$ with probability $I_{\mathbb{C}}({}^{X,i}\ell)(\sum_{j=1}^{m} I_{\mathbb{C}}({}^{X,j}\ell)\exp(\theta^T(t({}^{X,j}\ell) - t({}^{X,i}\ell))))^{-1}$ (with ℓ to ℓ being an acceptable "transition"), the realization of which becomes the base state for the next transition event. So long as it is possible to transition from any given state ℓ to state ℓ' in a finite series of moves, the states of the above process form an irreducible Markov chain on $\mathbb{C}$. Furthermore, the transition probabilities at each step can be recognized as the conditional distribution of ℓ_X, given θ and the other elements of ℓ. (This follows directly from equation [3], where the set of accessible states is restricted to those that involve changing only ℓ_X.) A Markov chain of this type is commonly known as a *Gibbs sampler* (Gilks, Richardson, and Spiegelhalter 1996a), and its equilibrium distribution is the joint distribution of ℓ; here, this is simply the distribution of equation (3), demonstrating that a process in which each object moves randomly in proportion to the relative exponentiated potential of its possible

locations will generate a global distribution of states that is compatible with the location system model. If the time intervals between transition events are independent of the system state, it also immediately follows that the expected fraction of time spent in each accessible state is similarly proportional to exp $(\mathcal{P}(\ell))$ (where such expectations exist).

One type of microlevel process with this behavior stems from a class of potential games (as alluded to in Section 1.1). Let us consider an n player "assignment game" (with the object set $O = (o_1, \ldots, o_n)$ corresponding to the players), in which each player has m potential strategies (corresponding to the choice among the elements of L). Since each player's strategy corresponds to a choice of location, we can represent an assignment of strategies to all players by ℓ. We assume that the preferences of a given player o_i are represented by the utility $u_i(\ell_i, \ell_{-i})$, where the first argument is trivially the strategy of o_i and the second denotes the strategies of all other players. Now define $\mathbb{C}'$ to be the set of all accessible ℓ_{-i} (irrespective of ℓ_i). Given this, we say that the assignment game belongs to the class of potential games if there exists a function $\rho : (L, \mathbb{C}') \mapsto \mathbb{R}$ such that

$$u_i\left(l_j, \ell_{-i}\right) - u_i\left(l_k, \ell_{-i}\right) = \rho\left(l_j, \ell_{-i}\right) - \rho\left(l_k, \ell_{-i}\right) \tag{19}$$

for all $i \in 1, \ldots, n, l_j, l_k \in L, \ell_{-i} \in \mathbb{C}'$. (This follows immediately from the definition of Young [1998: 36].) Intuitively, the above implies that an assignment game is a potential game if and only if we can posit some function (the potential, ρ) such that the change in utility for a unilateral strategy shift is equal to the change in potential. While assuming identical utilities for all actors will clearly result in such a game, this is not a necessary condition: for instance, utilities that are identical up to an additive constant will also satisfy the above definition.

For our purposes, the above is most relevant when we take $\rho(\ell_i, \ell_{-i}) = \mathcal{P}(\ell)$. In this case, the potential for the assignment game is equal to our social potential and changes in utility for unilateral moves are equal to the corresponding changes in $\mathcal{P}$. Let us further posit that the assignment game is played by boundedly rational actors who select strategies according to a stochastic choice model. Specifically, given an opportunity to move, we posit that actor o_i selects his or her location

as a random variable K_i with probability mass

$$\Pr(K_i = l_j) = I_{\mathbb{C}}({}^{i,j}\ell)\frac{\exp\big(u_i(l_j, \ell_{-i})\big)}{\sum_{k=1}^{m} I_{\mathbb{C}}({}^{i,k}\ell)\exp\big(u_i(l_k, \ell_{-i})\big)} \tag{20}$$

(where ${}^{i,j}\ell$ denotes the assignment formed by adding $\ell_i = j$ to ℓ_{-i}). This can be rewritten as

$$\Pr(K_i = l_j) = I_{\mathbb{C}}({}^{i,j}\ell)\left[\sum_{k=1}^{m} I_{\mathbb{C}}({}^{i,k}\ell)\exp\big(u_i(l_k, \ell_{-i}) - u_i(l_j, \ell_{-i})\big)\right]^{-1} \tag{21}$$

and substitution of utility differences for potential differences then yields

$$= I_{\mathbb{C}}({}^{i,j}\ell)\left[\sum_{k=1}^{m} I_{\mathbb{C}}({}^{i,k}\ell)\exp\big(\mathcal{P}(l_k \cup \ell_{-i}) - \mathcal{P}(l_j \cup \ell_{-i})\big)\right]^{-1}. \tag{22}$$

This is immediately recognizable as the transition mechanism for the Gibbs sampler (as shown above); hence, the state distribution arising from strategy choice in the assignment game will be proportional to $\exp(\mathcal{P}(\ell))$, so long as the irreducibility condition of the Gibbs sampler is met. This may be satisfied in a number of ways, including (1) sequential moves by each actor in turn and (2) sequential moves by randomly chosen actors, where every actor is chosen from a fixed distribution with positive probability (Gilks 1996). Heuristically speaking, the long-run dynamics of such a system are generally insensitive to the details of the movement opportunity process, so long as actions are selected in accordance with equation (20).

As these results imply, the generalized location system model is compatible with certain microprocess interpretations, including potential games like those considered by Zhang (2004) and Young (1998). In the latter circumstance, specific model parameters can be interpreted as reflecting the partial utilities associated with changes in their corresponding statistics. Although it should be emphasized that these are not the only processes that can lead to distributions of the type explored here, they nevertheless serve as useful examples of how such micro/macro connections (strongly endorsed by Coleman [1990]) may be made in practice.

3.3. *Simulation*

For purposes of both prediction and inference, it is necessary to simulate the behavior of the location system model for arbitrary covariates and parameter values. While it is not generally possible to take draws from the location system model directly, approximate samples may be readily obtained by means of a Metropolis algorithm.[3] Given that numerous accessible references on the Metropolis algorithm are currently available (e.g., see Gamerman 1997; Gilks et al. 1996b; Gelman et al. 1995), we will focus here on issues that are specific to the model at hand. Fortunately, the location system model is not especially difficult to simulate, although certain measures are necessary to ensure scalability for large systems.

To review, a Metropolis algorithm proceeds in the following general manner (see Gilks et al. [1996b] for further details). Let S be the (random) system state. We begin with some initial state $\ell^{(0)} \in \mathbb{C}$ and propose moving to a candidate state $\ell^{(1)}$, which is generally chosen so as to be in a neighborhood of $\ell^{(0)}$. (Some additional constraints—for example detailed balance—apply to the candidate distribution, but these do not affect the results given here.) The candidate state is then "accepted" with probability $\min(1, \frac{\Pr(S=\ell^{(1)}|\mathcal{P},\mathbb{C})}{\Pr(S=\ell^{(0)}|\mathcal{P},\mathbb{C})})$. If accepted, the candidate becomes our new base state and we repeat the process for $\ell^{(2)}$. If rejected, $\ell^{(1)}$ is replaced by a copy of $\ell^{(0)}$ and again the process is repeated. This process constitutes a Markov chain whose equilibrium distribution (under certain fairly broad conditions) converges to the target distribution (here, $\Pr(S|\mathcal{P}, \mathbb{C})$). It is noteworthy that this process requires only that the target distribution be computable up to a constant factor; this feature makes Metropolis algorithms (and related MCMC techniques) very attractive to those working with exponential family models (e.g., Strauss 1986; Snijders 2002; Butts, this volume, page x).

To implement the Metropolis algorithm, then, our core concern is computation of the probability ratio between states (a problem

[3] Obviously, the Gibbs samplers discussed in Section 3.2.6 can also be used in this capacity; we focus here on the Metropolis algorithm due to its superior scalability.

encountered earlier in Section 3.2.6). Given a current state, $\ell^{(i)}$, the probability of accepting a candidate state, $\ell^{(i+1)}$, is then

$$\frac{\Pr\left(S=\ell^{(i+1)}|\mathcal{P},\mathbb{C}\right)}{\Pr\left(S=\ell^{(i)}|\mathcal{P},\mathbb{C}\right)}=\frac{\exp\left(\mathcal{P}\left(\ell^{(i+1)}\right)\right)}{\exp\left(\mathcal{P}\left(\ell^{(i)}\right)\right)}\frac{Z(\mathcal{P},\mathbb{C})}{Z(\mathcal{P},\mathbb{C})} \tag{23}$$

$$=\frac{\exp\left(\mathcal{P}\left(\ell^{(i+1)}\right)\right)}{\exp\left(\mathcal{P}\left(\ell^{(i)}\right)\right)} \tag{24}$$

$$=\exp(\mathcal{P}(\ell^{(i+1)})-\mathcal{P}(\ell^{(i)})). \tag{25}$$

Thus, the log-probability of a state change is simply the difference in social potentials between the two assignments. In the case of the linear potential, substituting the potential function from equation (14) further gives us

$$\begin{aligned}
&\mathcal{P}\left(\ell^{(i+1)}\right)-\mathcal{P}\left(\ell^{(i)}\right)\\
&\quad=\sum_{j=1}^{a}\alpha_j\left[t_j^{\alpha}\left(\ell^{(i+1)}\right)-t_j^{\alpha}\left(\ell^{(i)}\right)\right]+\sum_{j=1}^{b}\beta_j\left[t_j^{\beta}\left(\ell^{(i+1)}\right)-t_j^{\beta}\left(\ell^{(i)}\right)\right]\\
&\quad+\sum_{j=1}^{c}\gamma_j\left[t_j^{\gamma}\left(\ell^{(i+1)}\right)-t_j^{\gamma}\left(\ell^{(i)}\right)\right]\\
&\quad+\sum_{j=1}^{d}\delta_j\left[t_j^{\delta}\left(\ell^{(i+1)}\right)-t_j^{\delta}\left(\ell^{(i)}\right)\right],
\end{aligned} \tag{26}$$

or, substituting from equation (15),

$$\begin{aligned}
&=\sum_{g=1}^{a}\sum_{j=1}^{n}\alpha_g\mathbf{X}_{jg}\left[\mathbf{Q}_{\ell_j^{(i+1)}g}-\mathbf{Q}_{\ell_j^{(i)}g}\right]\\
&\quad+\sum_{g=1}^{b}\beta_g\sum_{j=1}^{n}\sum_{k=1}^{n}\left[\mathbf{B}_{g\ell_j^{(i+1)}\ell_k^{(i+1)}}-\mathbf{B}_{g\ell_j^{(i)}\ell_k^{(i)}}\right]\left|\mathbf{Y}_{jg}-\mathbf{Y}_{kg}\right|\\
&\quad+\sum_{g=1}^{c}\gamma_g\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{A}_{gjk}\left[\left|\mathbf{R}_{\ell_j^{(i+1)}g}-\mathbf{R}_{\ell_k^{(i+1)}g}\right|-\left|\mathbf{R}_{\ell_j^{(i)}g}-\mathbf{R}_{\ell_k^{(i)}g}\right|\right]\\
&\quad+\textstyle\sum_{g=1}^{d}\delta_g\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{W}_{gjk}\left[\mathbf{D}_{g\ell_j^{(i+1)}\ell_k^{(i+1)}}-\mathbf{D}_{g\ell_j^{(i)}\ell_k^{(i)}}\right].
\end{aligned} \tag{27}$$

Although equation (27) can be used to compute the potential difference directly, equation (26) demonstrates that the same quantity can be expressed in terms of a fixed linear combination of differences in sufficient statistics. For purposes of simulation, then, we need only track such differences. (This process is an exact analog to the "changescore" methods used in ERG simulation tools such as Handcock et al. [2003].) As this implies, we can speed computation by choosing our proposals so as to facilitate difference calculations; an obvious choice in this regard is a proposal mechanism that reassigns a randomly chosen object to a randomly selected location. In addition to simplicity of implementation, this proposal density admits considerable improvement in computational efficiency over the iterated calculation of equation (27). In particular, let ℓ be the current state and ℓ' the proposal formed by assigning object j to location k. Then the respective differences in sufficient statistics are as follows:

$$t_i^{\alpha}(\ell') - t_i^{\alpha}(\ell) = \mathbf{X}_{ji}\left[\mathbf{Q}_{ki} - \mathbf{Q}_{\ell_j i}\right] \tag{28}$$

$$t_i^{\beta}(\ell') - t_i^{\beta}(\ell) = \sum_{g=1}^{n} \left|\mathbf{Y}_{ji} - \mathbf{Y}_{gi}\right| \left[\mathbf{B}_{ik\ell_g} - \mathbf{B}_{i\ell_j\ell_g} + \mathbf{B}_{i\ell_g k} - \mathbf{B}_{i\ell_g\ell_j}\right] \tag{29}$$

$$t_i^{\gamma}(\ell') - t_i^{\gamma}(\ell) = \sum_{g=1}^{n} \left[\mathbf{A}_{ijg} + \mathbf{A}_{igj}\right] \left[\left|\mathbf{R}_{ki} - \mathbf{R}_{\ell_g i}\right| - \left|\mathbf{R}_{\ell_j i} - \mathbf{R}_{\ell_g i}\right|\right] \tag{30}$$

$$\begin{aligned} t_i^{\delta}(\ell') - t_i^{\delta}(\ell) = \sum_{g\neq j, g=1}^{n} \Big[& \mathbf{W}_{ijg}\left(\mathbf{D}_{ik\ell_g} - \mathbf{D}_{i\ell_j\ell_g}\right) \\ & + \mathbf{W}_{igj}\left(\mathbf{D}_{i\ell_g k} - \mathbf{D}_{i\ell_g\ell_j}\right)\Big] \\ & + \mathbf{W}_{ijj}\left(\mathbf{D}_{ikk} - \mathbf{D}_{i\ell_j\ell_j}\right). \end{aligned} \tag{31}$$

Calculation of equation (26) for a single reassignment using equations (28) through (31) is an $\mathcal{O}(n)$ operation. This is a substantial improvement over the $\mathcal{O}(n^2)$ complexity for direct application of equation (27) in the arbitrary case, particularly for large-n systems. As a side note, it should be mentioned that occupancy constraints may not allow single assignments to take place. (The permutation case is a trivial example, since the smallest change possible is the dyadic exchange.) In this

case, the proposal mechanism may need to include multiple reassignments in a single step; however, it is still the case that the above computations can be performed for each such reassignment and the resulting complexity is still linear in n so long as the number of reassignments per step is bounded by a constant. Even fairly complex schemes can thus be reduced to an iterated application of the reassignment calculation.

3.3.1. *Estimating the Partition Function*

Though the above provides the essential elements needed to simulate draws from the location system model, the approach used bypasses calculation of the partition function. This is deliberate: Z is not directly computable in polynomial time and the unevenness of the Boltzmann factor ($\exp(\mathcal{P}(\ell))$) renders simple Monte Carlo strategies hopelessly inefficient. What is to be done, however, when the partition function (or its derivatives) is needed for a specific application (such as the deviance calculations discussed in Section 3.4)? In this case, we employ the fact that we are able to simulate draws from the location system model to produce an *importance sample*, thereby allowing efficient Monte Carlo quadrature of Z.

To begin, we assume that a sample of M draws (denoted $\ell^{(1)}, \ldots, \ell^{(M)}$) have been taken from the location system model with combined parameter vector $\theta = (\alpha, \beta, \gamma, \delta)$ and vector of sufficient statistic functions $t = (t^{\alpha}, t^{\beta}, t^{\gamma}, t^{\delta})$. Our interest is in estimating $Z(\theta', C)$, where θ' is a combined parameter vector that is close to θ (in the sense that $\|\theta' - \theta\|$ is small). Our estimator of the partition function is based on the result that

$$\lim_{M\to\infty} \frac{|\mathbb{C}| \sum_{i=1}^{M} \exp\left(\left(\theta' - \theta\right)^T t\left(\ell^{(i)}\right)\right)}{\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)} = Z(\theta', \mathbb{C}). \tag{32}$$

This result may be shown as follows. First, we note that, from the standard Monte Carlo theorem in the discrete case (Kalos and Whitlock 1986),

$$\lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} f\left(\ell^{(i)}\right) \frac{Z(\theta, \mathbb{C})}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)} = \sum_{\ell' \in \mathbb{C}} f\left(\ell'\right) \tag{33}$$

where convergence is almost sure and in mean square, so long as the function $f : \mathbb{C} \mapsto \mathbb{R}$ has a finite second moment. Setting $f(\ \ell\ ) = \exp(\theta'^T\ t(\ \ell\ ))$ then gives us

$$\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} \exp\left(\theta'^T t\left(\ell^{(i)}\right)\right) \frac{Z(\theta,\mathbb{C})}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)} = \sum_{\ell'\in\mathbb{C}} \exp\left(\theta'^T t(\ell')\right) \quad (34)$$

$$= Z\left(\theta',\mathbb{C}\right). \quad (35)$$

While this gives us an expression for $Z(\theta',\mathbb{C})$ on the right hand side, it requires us to know the value of $Z(\theta,\mathbb{C})$ and thus is of little immediate use. Since the partition function does not depend on ℓ , however, it may be pulled out of the initial summand:

$$\frac{1}{M}\sum_{i=1}^{M} \exp\left(\theta'^T t\left(\ell^{(i)}\right)\right) \frac{Z(\theta,\mathbb{C})}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)}$$

$$= Z(\theta,\mathbb{C})\,\frac{1}{M}\sum_{i=1}^{M} \exp\left((\theta'-\theta)^T t(\ell^{(i)})\right). \quad (36)$$

Thus,

$$\lim_{M\to\infty} Z(\theta,\mathbb{C})\,\frac{1}{M}\sum_{i=1}^{M} \exp\left(\left(\theta'-\theta\right)^T t\left(\ell^{(i)}\right)\right) = Z\left(\theta',\mathbb{C}\right). \quad (37)$$

Dividing through by $Z(\theta,\mathbb{C})$ then gives us

$$\lim_{M\to\infty} \frac{1}{M}\sum_{i=1}^{M} \exp\left(\left(\theta'-\theta\right)^T t\left(\ell^{(i)}\right)\right) = \frac{Z(\theta',\mathbb{C})}{Z(\theta,\mathbb{C})}. \quad (38)$$

To see the value of this, let us return to the left-hand side of equation (32). Immediately, we note that

$$\frac{|\mathbb{C}|\sum_{i=1}^{M} \exp\left((\theta'-\theta)^T t\left(\ell^{(i)}\right)\right)}{\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)} = |\mathbb{C}|\,M\left[\frac{1}{M}\sum_{i=1}^{M} \exp\left(\left(\theta'-\theta\right)^T t\left(\ell^{(i)}\right)\right)\right]$$

$$\times\left[\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)\right]^{-1}. \quad (39)$$

As $M \to \infty$, we have already seen that $\frac{1}{M}\sum_{i=1}^{M} \exp((\theta' - \theta)^T t(\ell^{(i)})) \to \frac{Z(\theta', \mathbb{C})}{Z(\theta, \mathbb{C})}$. Therefore, the above becomes (for large M)

$$\approx |\mathbb{C}| M \left[\frac{Z(\theta', \mathbb{C})}{Z(\theta, \mathbb{C})}\right] \left[\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)\right]^{-1} \tag{40}$$

$$= |\mathbb{C}| Z\left(\theta', \mathbb{C}\right) \left[\frac{1}{M}\sum_{i=1}^{M} \frac{Z(\theta, \mathbb{C})}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)}\right]^{-1}. \tag{41}$$

Now, what of the factor on the right? Returning to equation (33), we simply take $f(\ell') = 1$, which yields

$$\frac{1}{M}\sum_{i=1}^{M} 1 \frac{Z(\theta, \mathbb{C})}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)} \to \sum_{\ell' \in \mathbb{C}} 1 \tag{42}$$

$$= |\mathbb{C}|, \tag{43}$$

and thus, by substitution,

$$\lim_{M\to\infty} \frac{|\mathbb{C}| \sum_{i=1}^{M} \exp\left(\left(\theta' - \theta\right)^T t\left(\ell^{(i)}\right)\right)}{\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)} = Z\left(\theta', \mathbb{C}\right) |\mathbb{C}| \frac{1}{|\mathbb{C}|} \tag{44}$$

$$= Z\left(\theta', \mathbb{C}\right). \tag{45}$$

It therefore follows that we can estimate the partition function directly, given a sample from the location model. Since the approximation works well for any θ' that is close to θ, only a single sample is usually needed to compute numerical derivatives.[4] In the special case for which we are

[4] More properly, the "closeness" of θ' to θ can be assessed in terms of the variance of $\exp((\theta' - \theta)^T t(\ell))$. To the extent that this function is flat, the associated integral will be well-approximated (Kalos and Whitlock 1986).

solely interested in $Z(\theta, \mathbb{C})$ (given an importance sample based on θ), equation (32) further simplifies to

$$\lim_{M\to\infty} \frac{M|\mathbb{C}|}{\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)} = Z(\theta, \mathbb{C}). \tag{46}$$

This last follows immediately by substitution.

3.3.2. *Expected Value Estimates*

An important use for the partition function approximation of equation (32) is the calculation of approximate expected values for the sufficient statistics of the location system model. While these can obviously be estimated directly via simulation (per Section 3.3), it is in some cases more efficient to calculate expected statistics by means of the partition function. In doing so, we exploit the standard result (e.g., Brown [1986]) that, for a regular exponential family with parameter vector θ and sufficient statistics t, the expected value of t under θ is equal to the first derivative of the logarithm of the partition function with respect to θ. Allowing $t = (t^\alpha, t^\beta, t^\gamma, t^\delta)$ and $\theta = (\alpha, \beta, \gamma, \delta)$, this translates to the relation

$$\mathbf{E}_\theta t(\ell) = \frac{\mathrm{d}}{\mathrm{d}\theta} \ln Z(\theta, \mathbb{C}) \tag{47}$$

for the location system model. Since direct computation of Z is infeasible, this would seem to be of little use; however, we have already seen that Z may be approximated from importance sample draws using equation (32). Here, too, it may at first blush seem that using a sample to approximate a samplable quantity provides no particular advantage. However, consider the case in which one must evaluate $\mathbf{E}_\theta\, t(\ell)$ at several different points, all of which are reasonably close to one another. In such situations, equation (32) may be employed to quickly calculate multiple approximations from a single sample. This is particularly useful when seeking to calculate derivatives of the expectations themselves, as for the moment matching method described in Section 3.4.2 below.

Let us begin by assuming that we have drawn an importance sample $\ell^{(1)}, \ldots, \ell^{(M)}$ from a location system model with support $\mathbb{C}$ and concatenated parameter vector θ (e.g., using a Metropolis algorithm).

To calculate the expectation of some statistic t_i under parameter vector θ', it is natural to replace the derivative of equation (47) by a standard finite difference approximation (Press et al. 1992). For some $\epsilon \ll 1$ ($\epsilon = 0.00001$ being a value that gives good precision in most cases),

$$\mathbf{E}_{\theta'} t_i(\ell) = \frac{\mathrm{d}}{\mathrm{d}\theta'_i} \ln Z(\theta', \mathbb{C}) \tag{48}$$

$$\approx \frac{\ln Z({}^{i,\theta'_i+\epsilon}\theta', \mathbb{C}) - \ln Z({}^{i,\theta'_i-\epsilon}\theta', \mathbb{C})}{2\epsilon}. \tag{49}$$

Clearly, this is not directly computable. However, we may substitute the importance sampling estimate of equation (32) to obtain

$$\approx \frac{1}{2\epsilon}\left[\ln\left(\frac{|\mathbb{C}|\sum_{j=1}^{M}\exp\left(\left({}^{i,\theta'_i+\epsilon}\theta' - \theta\right)^T t\left(\ell^{(j)}\right)\right)}{\sum_{j=1}^{M}\exp\left(-\theta^T t\left(\ell^{(j)}\right)\right)}\right)\right.$$
$$\left. - \ln\left(\frac{|\mathbb{C}|\sum_{j=1}^{M}\exp\left(\left({}^{i,\theta'_i-\epsilon}\theta' - \theta\right)^T t\left(\ell^{(j)}\right)\right)}{\sum_{j=1}^{M}\exp\left(-\theta^T t\left(\ell^{(j)}\right)\right)}\right)\right], \tag{50}$$

which simplifies to

$$= \frac{\ln\left(\sum_{j=1}^{M}\exp\left(\left({}^{i,\theta'_i+\epsilon}\theta' - \theta\right)^T t\left(\ell^{(j)}\right)\right)\right) - \ln\left(\sum_{j=1}^{M}\exp\left(\left({}^{i,\theta'_i-\epsilon}\theta' - \theta\right)^T t\left(\ell^{(j)}\right)\right)\right)}{2\epsilon}. \tag{51}$$

Repeating this calculation for each i gives the entire vector of expectations (if desired). Each such calculation requires $\mathcal{O}(M)$ time, whereas drawing the initial sample requires $\mathcal{O}(Mn)$ time (or worse, for example, if subsampling—"thinning"—of a larger set of draws is performed

prior to use); this can be a substantial savings, in practice. Such gains do not occur without cost, of course. The tradeoff here lies in the accuracy of the approximation, which is in turn governed primarily by the accuracy with which $Z(\theta', \mathbb{C})$ is estimated. As noted earlier, accuracy of the partition function estimator is generally high when θ is close to θ', degrading as the distance between θ and θ' grows large. For this reason, it is usually wise to employ equation (51) only when evaluating expectations in the immediate vicinity of the sampled parameters. Comparison of directly estimated expectations with those estimated by this method can be used to evaluate the degree of closeness required for a particular model, where this is a significant concern.

3.4. *Inference*

As noted elsewhere in this paper, an important benefit of the use of discrete exponential families in the modeling of location systems is the ability to leverage existing inferential theory (for example, Barndorff-Nielsen 1978; Brown 1986). Given a location system specified by the tuple $(L, O, \mathbb{C}, F_L, F_O)$ with parametric social potential $\mathcal{P}$, the joint likelihood of observed state ℓ is given by equation (3). Since this quantity is well-defined, principled inference using Bayesian or maximum likelihood methods would seem to be a straightforward affair; the computational expense of directly calculating the required normalizing factor, $Z(\mathcal{P}, \mathbb{C})$ makes this task less trivial than it might be, however. Fortunately, a number of methods exist for circumventing this problem. Here, we will consider two: estimation based on pseudo-likelihoods and maximum likelihood estimation via first moment matching. These two approaches build on each other (with the former providing initial estimates to be refined by the latter) and will hence be considered in the above order. Other alternatives are also available (the Monte Carlo approach of Geyer and Thompson [1992] being an obvious possibility; see Butts [this volume] for a closely related application) and in general it should be possible to apply any method for inference on regular discrete exponential families to the location system model.

3.4.1. *Maximum Pseudo-Likelihood Estimation*

While we have seen that it is possible to estimate the normalizing factor for the likelihood of equation (3) using importance sampling, this

process is too computationally expensive to permit direct maximization of the likelihood surface (though see Geyer and Thompson [1992] for an effective ratio-based approach). One alternative approach (originating with Besag [1975] but better known to sociologists from Strauss and Ikeda [1990]) involves approximating the joint likelihood of the data by a product of conditional likelihoods; the parameter vector that maximizes this *pseudo-likelihood* is then used as an estimator of the unknown true parameters. Such a vector is known as a *maximum pseudo-likelihood estimator*, or MPLE. Put more formally, let $\mathbf{Y}$ be a vector of k random variables, with parameter vector ϕ. Then the pseudo-likelihood of realization y of $\mathbf{Y}$ is given by

$$\tilde{L}(\mathbf{y}|\phi) = \prod_{i=1}^{k} p(Y_i = y_i \mid \mathbf{y}_{-i}, \phi) \tag{52}$$

$$\approx p(\mathbf{Y} = \mathbf{y} \mid \phi). \tag{53}$$

Note that the pseudo-likelihood is equal to the true likelihood when the elements of $\mathbf{Y}$ are independent given ϕ. (This follows immediately from the fact that $p(Y_i = y_i|\mathbf{y}_{-i}, \phi) = p(Y_i = y_i|\phi)$ and $p(\mathbf{Y} = \mathbf{y} \,|\phi) = \prod_{i=1}^{k} p(Y_i = y_i|\, \phi)$ under conditional independence.) Where this assumption does not hold, $\tilde{L}$ will depart from the true likelihood to some extent. Nevertheless, the maximum of $\tilde{L}$ with respect to ϕ is often close to the corresponding maximum on the likelihood surface, making it a potentially viable estimator when $\tilde{L}$ is easily calculable. Alternatively, the MPLE can be used as an initial approximation to the true maximum likelihood estimator (MLE), to be subsequently refined using other methods. This last strategy has proved practical in the estimation of ERG models (as implemented, for example, by Handcock et al. [2003]) and is that suggested here.

To define a pseudo-likelihood function for the location system model, it is first necessary to select a partition of the joint distribution into individual elements. The conditional likelihoods of these elements are then multiplied to produce the corresponding pseudo-likelihood. In the present case, the most natural decomposition of the random system

state S is in terms of the individual object assignments. The appropriate conditional probability in this case is then

$$\begin{aligned}&\Pr\left(S_i = \ell_i \,|\, S_{-i} = \ell_{-i}, \mathcal{P}, \mathbb{C}\right)\\ &\quad = I_{\mathbb{C}}(\ell)\left[\sum_{j=1}^{m} I_{\mathbb{C}}({}^{i,j}\ell)\exp\left(\mathcal{P}\left({}^{i,j}\ell\right) - \mathcal{P}\left(\ell\right)\right)\right]^{-1},\end{aligned} \tag{54}$$

leading to the corresponding pseudo-likelihood function

$$\tilde{L}(\ell|\mathcal{P}, \mathbb{C}) = \prod_{i=1}^{n} \Pr\left(S_i = \ell_i \,|\, S_{-i} = \ell_{-i}, \mathcal{P}, \mathbb{C}\right). \tag{55}$$

Note that the evaluation of $\tilde{L}$ requires $\mathcal{O}(nm)$ potential difference computations. Since these can be performed in $\mathcal{O}(n)$ operations (per Section 3.3), the total computational complexity of $\tilde{L}$ is $\mathcal{O}(n^2m)$. This is moderately expensive, but obviously far better than $\mathcal{O}(m^n)$! To form the MPLE, we simply find the parameter vectors that maximize $\tilde{L}$— that is,

$$\left(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}\right) = \arg\max_{(\alpha,\beta,\gamma,\delta)} \tilde{L}(\ell \,|\, \mathcal{P}, \mathbb{C}). \tag{56}$$

This can be done using standard heuristic optimization methods, such as Newton's method, simulated annealing, or the like (Press et al. 1992; Acton 1990).

Although the above partition of S is perhaps most natural, it cannot be applied for certain choices of $\mathbb{C}$. In particular, if there exists at least one location for which $P_i^+ = P_i^- > 0$, then not all objects can be unilaterally assigned. In this case, a reasonable choice for the decomposition of S is in terms of the set of dyadic exchanges on ℓ (i.e., location swaps). In this case, we may treat the ordering of each dyad as our variable of interest, leading to the conditional distribution

$$\begin{aligned}&\Pr\left(S_i = \ell_i, S_j = \ell_j | \{S_i, S_j\} = \{\ell_i, \ell_j\}, S_{-\{ij\}} = \ell_{-\{ij\}}, \mathcal{P}, \mathbb{C}\right)\\ &\quad = I_{\mathbb{C}}(\ell)\left[1 + I_{\mathbb{C}}\left({}^{i,\ell_j}\left({}^{j,\ell_i}\ell\right)\right)\exp\left(\mathcal{P}\left({}^{i,\ell_j}\left({}^{j,\ell_i}\ell\right)\right) - \mathcal{P}\left(\ell\right)\right)\right]^{-1}.\end{aligned} \tag{57}$$

To form the associated pseudo-likelihood, we then take the product of the conditional distributions over all dyads in O:

$$\tilde{L}_p(\ell|\mathcal{P}, \mathbb{C}) = \prod_{i=1}^{n}\prod_{j=1}^{i-1} \Pr(S_i = \ell_i, S_j = \ell_j | \{S_i, S_j\} = \{\ell_i, \ell_j\}, S_{-\{ij\}} = \ell_{-\{ij\}}, \mathcal{P}, \mathbb{C}). \tag{58}$$

(Note that this is directly analogous to the permutation model pseudo-likelihood of Butts [this volume, page x].) The complexity of this calculation is $\mathcal{O}(n^3)$, which will be better than the complexity of the single-state pseudo-likelihood when $m > n$. On the other hand, $\tilde{L}_p$ conditions on the occupancy structure of ℓ and as such may represent less information than $\tilde{L}$. To find the MPLE under dyadic exchange, we simply substitute $\tilde{L}_p$ for $\tilde{L}$ in equation (56) and solve in the same manner as the single move case. This allows us to compute MPLEs for most choices of $\mathbb{C}$, including the important special case of 1:1 matchings (i.e., permutations).

Although the MPLE is not guaranteed to have good frequentist properties in the general case, it is sometimes used directly (e.g., see Wasserman and Pattison [1996]; Pattison and Wasserman [1999]; Robins et al. [1999]; Contractor et al. [2006]). Where this is to be done, it is strongly recommended that draws from the estimated model be simulated (using the method of Section 3.3) and that the first moments of the simulated sufficient statistics be compared to the same statistics on the observed data. To the extent that substantial disparities are observed (e.g., with respect to a t or similar statistic), the MPLE should be employed cautiously (if at all). On the other hand, a close match between the mean simulated statistics and the observed statistics indicates that the MPLE is in fact functionally close to the MLE; we shall see the basis for this conclusion in the next section.

3.4.2. *Maximum Likelihood Estimation*

Maximum likelihood estimation for statistical exponential families has been extensively studied (see Johansen [1979] or Brown [1986] for reviews) and is attractive on both frequentist and approximate Bayesian grounds (the MLE appearing as a noninformative limit of Bayes estimators in many settings [Robert 1994]). Under the linear social potential proposed in Section 3.2, the maximum likelihood estimator for the

parameters of $\mathcal{P}$ given observed state ℓ_{obs} is

$$\left(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}\right) = \arg \max_{(\alpha,\beta,\gamma,\delta)} I_{\mathbb{C}}(\ell_{\text{obs}}) \frac{\exp\left(\mathcal{P}\left(\ell_{\text{obs}} \mid \alpha, \beta, \gamma, \delta\right)\right)}{Z(\mathcal{P}, \mathbb{C})}, \quad (59)$$

where this maximum exists. Under the proposed social potential, equation (3) defines a regular exponential family and it is thus a standard result that the MLE exists (and is unique) if and only if the elements of t are finite and affinely independent and if $t(\ell)$ belongs to the relative interior of the convex hull of t over $\mathbb{C}$ (Barndorff-Nielsen 1978). In practice, nonexistence of the MLE arises where one or more statistics (or a linear combination thereof) are maximally extreme (e.g., all women occupying the highest paid position and all men occupying the lowest paid position in an occupational model with a gender/wage effect). In this case, the MLE effectively diverges (in our example, the apparent gender effect is unbounded) and no finite estimate exists. Fortunately, such extreme arrangements are unlikely to occur in large systems and even then practical approximations (e.g., truncating the diverging parameters at values of very large, but finite, magnitude) will usually permit reasonable estimates of the remaining parameters (see Handcock [2003a] for a discussion).

Due to the expense of approximating Z, direct maximization of the likelihood is generally infeasible. Since (3) is a regular exponential family, however, it is a standard result that

$$\mathbf{E}_{(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})}\left(t^{\alpha}(\ell), t^{\beta}(\ell), t^{\gamma}(\ell), t^{\delta}(\ell)\right) = (t^{\alpha}_{\text{obs}}, t^{\beta}_{\text{obs}}, t^{\gamma}_{\text{obs}}, t^{\delta}_{\text{obs}}), \quad (60)$$

where $(t^{\alpha}_{\text{obs}}, t^{\beta}_{\text{obs}}, t^{\gamma}_{\text{obs}}, t^{\delta}_{\text{obs}})$ is the vector of observed sufficient statistics, provided that the MLE exists. This result motivates a method of moments technique, in which heuristic search is used to equate the (simulated) expected sufficient statistics to their observed values; the parameter vector that gives rise to these values is the MLE. Although sometimes slow, this approach can be quite efficacious and has been successfully employed by Snijders (2002) in the context of exponential random graph families. In practice, convergence can be accelerated by initiating the search procedure with an initial approximation to $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta})$—the MPLE, $(\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta})$, is suggested for this purpose. Simulation costs in estimating $\mathbf{E}_{(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})}\left(t^{\alpha}(\ell), t^{\beta}(\ell), t^{\gamma}(\ell), t^{\delta}(\ell)\right)$ can also be reduced by making use

of the fact that $\mathbf{E}_{(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})}(t^{\alpha}(\ell), t^{\beta}(\ell), t^{\gamma}(\ell), t^{\delta}(\ell)) = \frac{\mathrm{d}}{\mathrm{d}(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})} \ln Z(\mathcal{P}, \mathbb{C})$ as was discussed in Section 3.3.2. Specifically, an initial sample from the current point estimate (using the method of Section 3.3) can be used as an importance sample for the numerical estimation of the derivatives of $\ln Z(\mathcal{P}, \mathbb{C})$ (e.g., by applying the method of finite differences to the logged partition function estimator of Section 3.3.1). This same sample can be reused multiple times to obtain the derivatives of $\mathbf{E}_{(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})}(t^{\alpha}(\ell), t^{\beta}(\ell), t^{\gamma}(\ell), t^{\delta}(\ell))$ in the vicinity of the original estimate, which can then be used to project a new estimate that leads to a better approximation of $(t^{\alpha}_{\text{obs}}, t^{\beta}_{\text{obs}}, t^{\gamma}_{\text{obs}}, t^{\delta}_{\text{obs}})$ (e.g., via Newton's method). A new sample is then drawn from the refined estimate and the process is repeated until the desired degree of convergence is obtained.

Given the resulting MLE $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta})$, the associated deviance can also be estimated by means of the importance sampling method of Section 3.3.1. Specifically, let $\ell^{(1)}, \ldots, \ell^{(M)}$ be draws from $S|\hat{\theta}$ (e.g., taken via the Metropolis algorithm of Section 3.3). Then substitution from the partition function estimator of equation (46) into the likelihood of equation (3) yields

$$D(\ell_{\text{obs}}|\hat{\theta}) = -2 \ln \Pr(S = \ell_{\text{obs}}) \tag{61}$$

$$\approx -2\mathcal{P}(\ell_{\text{obs}}|\hat{\theta}) + 2 \ln |\mathbb{C}| + 2 \ln M - 2 \ln \left(\sum_{i=1}^{M} \exp\left(-\hat{\theta}^T t\left(\ell^{(i)}\right)\right) \right). \tag{62}$$

The estimated deviance may then be used to compare models using standard selection criteria, such as the AIC or BIC (see Bozdogan [2000] and Wasserman [2000] for comparative reviews).

With respect to estimates of uncertainty, it should also be noted that standard asymptotics hold for location model MLEs in the case of independent observations from the same social system. (This is a consequence of the fact that the location model forms a regular exponential family; e.g., see Johansen [1979].) Whether similar asymptotic results can be obtained in the limit of increasing system size is not known. This problem is essentially equivalent to the problem of asymptotics for exponential random graph models, which is also unsolved at this time. Where asymptotic results cannot be relied upon, however, Monte Carlo procedures can be employed to obtain standard errors and p-values for classical tests. (See Hunter and Handcock [2006] for a parallel case involving

ERG models.) Thus, the standard tools of likelihood-based inference avail themselves here. Bayesian treatment of the location model is another possibility, although posterior simulation is greatly complicated by the difficulty of computing the likelihood function. Approximation methods based on curvature of the posterior near the mode (Gelman et al. 1995) would seem to provide an obvious starting point.

4. ILLUSTRATIVE APPLICATIONS

One of the positive features of the location system model is the great number of substantive problems for which it may be employed. Here, we illustrate some of the behaviors of the model by means of two simple applications, one involving economic inequality and the other involving residential segregation. While both are simplified for purposes of exposition, it should be emphasized that slightly elaborated versions can be fit to data from survey or archival sources using the tools of Section 3.4. Simulation studies such as these can thus form the basis for subsequent empirical investigation, without the necessity of adding cumbersome operationalization assumptions. As a third example, we also demonstrate the inferential use of the location system model to study position occupancy in organizations, using a data set of Lazega (2001).

4.1. *Job Segregation, Discrimination, and Inequality*

Our first example demonstrates the use of the location system in modeling occupational stratification. In the interest of clarity, we restrict our analysis to a simplified "microeconomy" of 100 workers (objects) matched with 100 distinct jobs (locations) on a 1:1 basis. Workers are evenly divided by gender and are randomly allocated to (heterosexual) couples such that all members of the set have exactly one partner. Finally, workers are ranked on a unidimensional "human capital" score, with ranks assigned randomly in alternating fashion by gender. (Thus, rank distributions are effectively identical by gender and random with respect to partners.) Jobs are ranked by "wage" and are organized into ten contiguous occupational categories. Thus, the ten highest-paying jobs are in one category, followed by jobs ranked 11 through 20, etc. While this setting is heavily stylized, it nevertheless allows us to capture basic interactions between occupational segregation, household effects,

and factors such as discrimination. Such a model could be elaborated to include hierarchical job categories, distinct unemployed states, additional job/worker attributes, and relaxations of assumptions such as 1:1 matching, as appropriate to the data in hand.

In order to represent the above within the location system framework, we begin by translating selected features into the elements of the social potential of Section 3.1. We here seek to model two attraction/repulsion effects: the tendency of workers of particular genders to be assigned to higher/lower wage jobs ("discrimination"); and the tendency of workers with higher levels of human capital to be assigned to higher/lower wage jobs ("merit"). For this purpose, we set $\mathbf{X}$ such that $\mathbf{X}_{i1}$ and $\mathbf{X}_{i2}$ are respectively the gender (coded dichotomously with male = 1) and the human capital rank (coded from 1 to 100, with 100 representing the highest value) of worker i. The corresponding position feature matrix, $\mathbf{Q}$, is defined such that $\mathbf{Q}_{i1}$ and $\mathbf{Q}_{i2}$ are both equal to the wage rank of job i (from 1 to 100 in ascending order); thus, $t_1^{\alpha}(\ell)$ is the sufficient statistic for the gender/wage interaction, while $t_2^{\alpha}(\ell)$ captures the corresponding human capital effect. For occupational segregation, we posit a single statistic t^{β} arising from a single-column object feature matrix $\mathbf{Y}$ consisting of dichotomous gender codes (as with $\mathbf{X}_{.1}$) interacting with a dichotomous location-location array $\mathbf{B}$ such that $\mathbf{B}_{1ij} = 1$ if job i belongs to the same occupational category as job j. The corresponding location heterogeneity statistic that will be of interest here is couple-level wage heterogeneity (described further below), a property that is parametrized via a statistic t^{γ} formed from a dichotomous object relation array $\mathbf{A}$ such that $\mathbf{A}_{1ij} = 1$ if worker i and worker j are members of the same couple and single-column location attribute matrix $\mathbf{R}$ containing the wage rank of each job (as with the columns of $\mathbf{Q}$). We do not posit any alignment effects for this particular, model and hence this completes our specification of t.

To get a sense of the behavior of the location model, we begin by demonstrating some of the job assignments that can arise under various parameter values. Figure 1 depicts simulated draws from the location model under a variety of conditions. Initially, we shall limit our consideration to configurations arising from manipulation of the discrimination (α_1) and (anti-)segregation (β) parameters, with all other parameters held to 0. For each choice of parameter values, the corresponding panel of Figure 1 shows 250 Metropolis draws from the associated model; these were uniformly thinned (i.e., subsampled) from

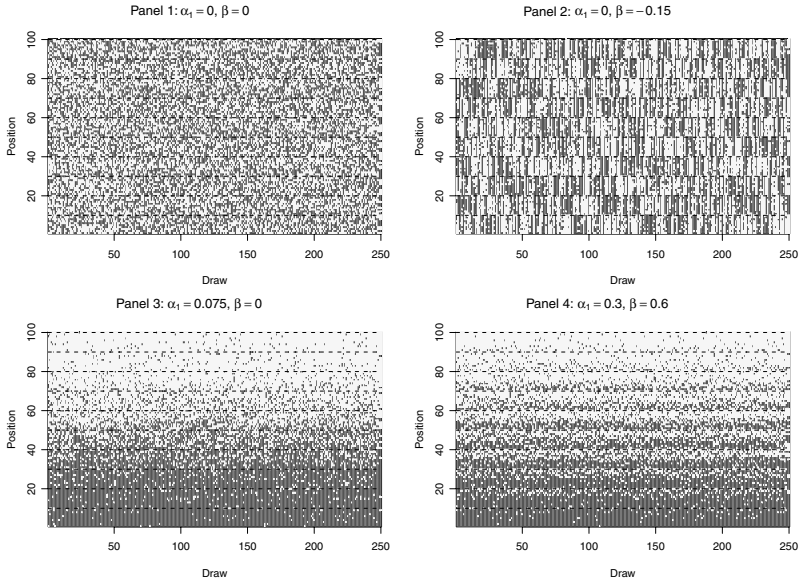

FIGURE 1. Location model draws, job assignment model.

a total of 25 million draws in each case, following a (discarded) burn-in sample of size 200,000. Jobs (ordered by wage rank) are shown on the vertical axis, with occupant gender indicated by color (light corresponds to male). For ease of reference, job category boundaries are shown by horizontal dashed lines; thus the gender composition of each category may be determined by examining the fraction of light versus dark cells between the appropriate lines for a given vertical slice.

Parameter values for each panel of Figure 1 are interpreted as follows. α_1 effects for all panels parametrize the strength of association between gender and wage, with positive values reflecting a stronger tendency to sort males into high-ranking wage positions. Thus α_1 acts as a discrimination parameter. (Negative α_1 values reverse the sorting direction but produce otherwise identical results; only positive values are considered here.) The object homogeneity/heterogeneity parameter, β, reflects the tendency of jobs within the same occupational category to be occupied by persons of the same gender (i.e., occupational segregation). As per equation (9), $\beta < 0$ indicates a tendency toward segregation (homogeneity), while $\beta > 0$ indicates a tendency toward desegregation (heterogeneity). Zero values for any parameter imply an absence of the

corresponding sorting effect. Thus $(\alpha_1, \beta) = (0, 0)$ results in a null model of uniform assignment.

Examination of the panels of Figure 1 hints at the diversity of configurations that can result from even a small number of interacting regimes. Panel 1 displays the baseline condition of random assignment: men and women appear across the spectrum of wage ranks, in essentially even numbers. The remaining panels show various types of sorting by gender, reflecting an interaction of segregation and discrimination effects. Panel 2, for instance, depicts a "block random" pattern, in which men and women are concentrated into uniform blocks which are otherwise randomly allocated across the wage ordering. Panels 3 and 4, on the other hand, show a clear pattern of stratification, in which men tend to be sorted into higher wage positions. What accounts for these patterns? Clearly, sorting by gender is driven by α_1 and is only observed for samples in which $\alpha > 0$. Heterogeneity within occupational categories is controlled by β, however and hence can act in ways which are distinct of discrimination per se. In Panel 2, for instance, we have pressure toward homogeneity/segregation ($\beta < 0$), which tends to force each occupational category to collapse into a single preferred gender. Since the model is here indifferent to *which* gender occupies any given occupational category, however, there is no net tendency for men or women to be sorted into higher-wage positions. By contrast, Panel 4 illustrates the interaction of a strong wage discrimination effect with a powerful tendency toward heterogeneity/antisegregation ($\beta > 0$). While the former effect seeks to sort men into high-wage positions and women into low-wage positions (as in Panel 3), the latter resists the accompanying necessity of producing gender-homogeneous occupational categories. The result is a structural "compromise," in which an overall gender/wage gradient is somewhat attenuated by the inclusion of both genders within each occupational stratum. Interestingly, the impact of the discrimination parameter on variation *within* occupational categories creates a tendency to reproduce a miniature version of the male/female wage gradient inside each occupational category; thus the well-known tendency for macrolevel stratification patterns to reproduce themselves at multiple levels is shown here to be an emergent property of the interaction of a global sorting process (here, discrimination) with mechanisms favoring local heterogeneity (here, an antisegregation effect).

In addition to the general types of configurations found under different assignment regimes, it is useful to consider the quantitative

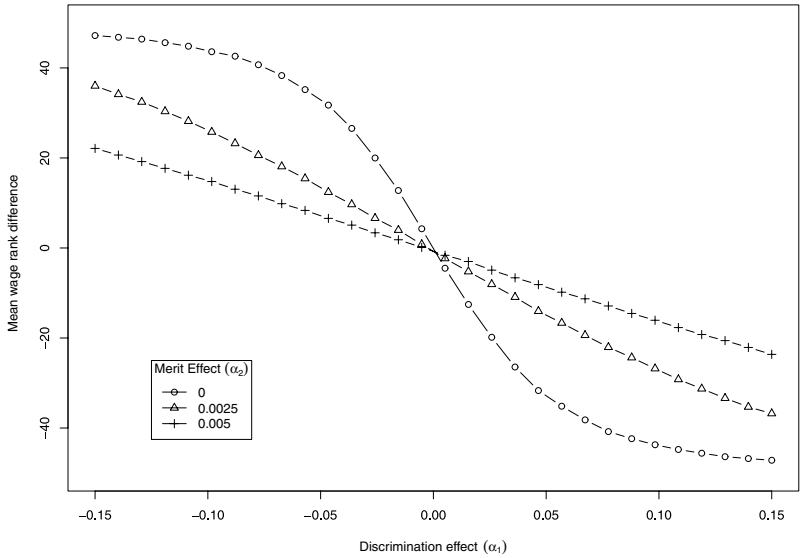

FIGURE 2. Gender difference in mean wage rank, discrimination, and merit effects.

impact of model parameters on outcomes of interest. In the present case, consider the difference in mean wage rank by gender. By construction, discrimination effects must lead to an exaggeration of such differences, but these effects must interact with other processes as well. For instance, in a competitive labor market, we generally expect workers with greater human capital to obtain positions with higher wage rates. Depending on the relationship between human capital and gender, this interaction may strengthen or weaken inequality in wage attainment. An example of this well-known phenomenon is shown in Figure 2, which presents differences in mean wage rank (by gender) for the location model with α effects for discrimination (gender by wage) and merit (human capital by wage).[5] As the figure clearly shows, the impact of discrimination is attenuated by merit effects where human capital is uncorrelated with gender. In addition to weakening the local impact of mild discrimination (i.e., $|\alpha_1|$ small), this attenuation softens the transition from a mixed-wage environment to a "frozen" environment in which wages

[5] For these simulations, $\beta = \gamma = 0$, means are based on 750 Metropolis draws uniformly thinned from samples of 750,000, after a burn-in period of 100,000 draws.

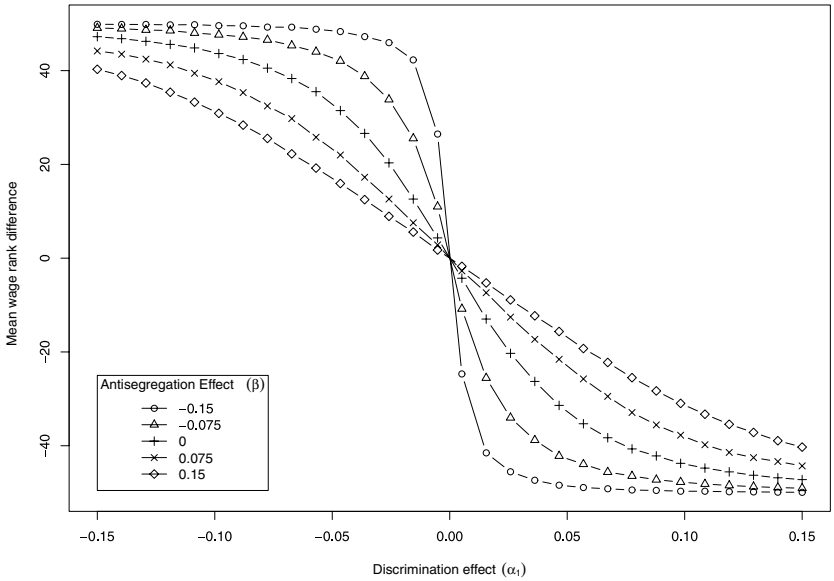

FIGURE 3. Gender difference in mean wage rank, discrimination, and segregation effects.

are strictly stratified by gender. In the absence of competing factors, even a fairly small amount of discrimination is adequate to lock the system into a stratified state; an intervention with the intent of inhibiting stratification by reducing discrimination is thus unlikely to prove effective unless discrimination can be tightly controlled. To draw on a physical analogy, it may be noted that the effect of $1/\alpha$ is directly analogous to a temperature parameter (as pointed out in other exponential family contexts by Strauss [1986] and Robins et al. [2005]). The stratification system "solidifies" at fairly low temperatures ($|\alpha|$ large) and is thus difficult to force into other, less stratified, states by modest changes in discrimination alone. By contrast, an intervention that attempts to inhibit stratification by introducing selective factors uncorrelated with gender could prove effective even with relatively high levels of residual discrimination (by lowering the relevant "melting point"). While it is perhaps intuitive that the introduction of competing selective factors would attenuate discrimination, the quantitative impact of these effects is particularly clear under the location system model.

If human capital effects are inhibitory of stratification in our scenario, what of segregation? Removing the human capital effect and

adding in a β effect for gender segregation by occupational category yields the wage rank difference relationship of Figure 3.[6] The impact of segregation on stratification is both clear and striking: segregation strongly exacerbates discrimination, while desegregation mildly inhibits it. The mechanism involved is a combination of those observed in panels 2 and 3 of Figure 1—namely, the tendency of gender-typed job categories to be "sorted" by wage in the presence of a background discrimination effect. The net effect is a substantially higher degree of stratification than would be obtained by discrimination alone. Desegregation pressure, by contrast, reduces the extent to which high or low wage categories can become male or female dominated, thereby "flattening" the wage distribution (as was seen in panel 4 of Figure 1). Interventions such as affirmative action programs can be understood as acting through mechanisms of this type; interestingly, Figure 2 would seem to suggest that most of the impact of such interventions is likely to come through the elimination of active segregation pressure rather than through pressure for desegregation per se.

While occupational segregation is a factor of obvious importance for stratification outcomes, a less well-studied issue is the impact of within-couple inequality on macroscopic wage differences. Effects as diverse as social influence (Freidkin 1998), homophily on unobserved characteristics (McPherson et al. 2001), and diffusion of opportunity through social ties (Calvo-Armengol and Jackson 2004) can potentially lead to a net tendency for similarity of within-couple wage rates; on the other hand, mechanisms such as market/home production specialization (Becker 1991), normative pressures for intensive parenting (Jones and Brayfield 1997), and the like can generate pressure for heterogeneous wage rates. To explore this within the location system model, we repeat the simulations of Figure 3, replacing the β effect with a γ effect for couple-level wage homogeneity/heterogeneity. The results are shown in Figure 4. As might be expected, heterogeneity pressure exacerbates discrimination. The effect, however, is slight and the marginal impact declines rapidly in γ. By contrast, the impact of couple-level homogeneity is profound: even a small amount of within-couple homogeneity pressure virtually eliminates the impact of discrimination, even when the latter is exceedingly strong. By tying together the fortunes of

[6] All simulation runs for Figures 3 and 4 were conducted in the same manner as those for Figure 2.

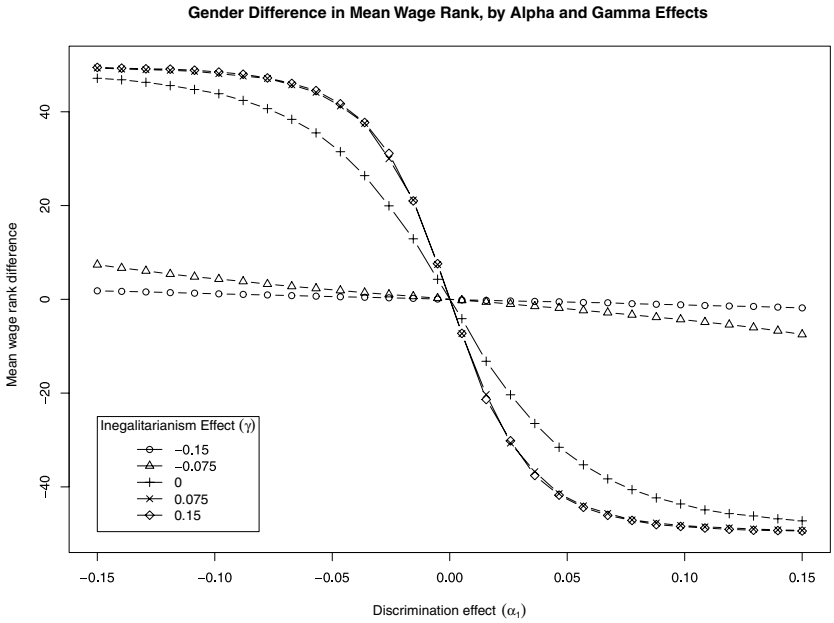

FIGURE 4. Gender difference in mean wage rank, discrimination, and couple-level homogeneity effects.

individual men and women, couples can act to counteract large-scale selection pressures toward wage inequality.

While these simulation results merely scratch the surface of what is possible when using the location system to model occupational inequality, they nevertheless suggest some interesting and nonobvious effects. Of particular import is the relative power of couple-level homogeneity effects in suppressing labor market discrimination, a finding that suggests a stronger connection between processes such as mate selection and marital bargaining with macrolevel stratification than might be supposed. The exacerbation of discrimination effects by segregation is also noteworthy, along with the somewhat less powerful inhibiting effect of active desegregation. These phenomena highlight the importance of considering dependencies—both among individuals and among jobs—when modeling inequality in labor market settings. Such effects can be readily captured by the location system, facilitating a more complete theoretical and empirical treatment of stratification within the occupational system.

4.2. *Settlement Patterns and Residential Segregation*

Another domain of long-standing interest to sociologists, geographers, and economists has been the role of segregation within residential settlement patterns (Schelling 1969; Bourne 1981; Massey and Denton 1993; Zhang 2004). Here, we illustrate the use of the location model on a simplified settlement system involving 1000 households (objects) allocated to regions on a uniform 20 by 20 spatial grid (locations). Unlike the job allocation system described above, this system places no occupancy constraints on each cell; however, "soft" constraints may be implemented via density dependence effects. For purposes of demonstration, each household is assigned a random "income" (drawn independently from a log-normal distribution with parameters 10 and 1.5) and an "ethnicity" (drawn from two types, with 500 households belonging to each type). Households are tied to one another via social ties, here modeled simply as a Bernoulli graph with mean degree of 1.5. The Bernoulli graph is a random structure in which ties between actors are independent and it is commonly used as a simple null model of network structure (Anderson, Butts, and Carley 1999). Regions, for their part, relate to one another via their spatial location. Here, we will make use of both Euclidean distances between regions and Queen's contiguity (for purposes of segregation). To obtain interregional distances, we treat each location as a 1 unit by 1 unit square planar region and take Euclidean distances between centroids; household position is modeled only up to the cell level, in analogy with data observed at the level of areal units such as census tracts. Similarly, two regions are considered to be contiguous under the Queen's rule if they border one another at either a point or an edge. Finally, each region is also assigned a location on a "rent" gradient, which is proportional to the inverse square of centroid distance the region in question to the center of the grid.

With these building blocks, a number of mechanisms can be explored. An obvious attraction/repulsion effect that plays an important role in household settlement patterns is the tendency for low-income households to "avoid" (or be excluded from) high-rent areas. In our case, this is equivalent to an attraction effect between household income and location rent level and hence we model it via a statistic t^{α} based on a single vector object feature matrix $\mathbf{X}$ containing household incomes and a single vector location feature matrix $\mathbf{Q}$ containing location rent levels. Ethnic segregation—a tendency for households to avoid

settling in areas dominated by households of different ethnicity—is a form of object heterogeneity and it is captured here by a single statistic t^{β} formed from the interaction of a single-column matrix of dichotomous ethnicity codes ($\mathbf{Y}$) and a dichotomous $\mathbf{B}$ array such that $\mathbf{B}_{1ij} = 1$ if region i is contiguous with region j (using the Queen's contiguity rule, as noted above). Although we do not posit any location heterogeneity effects in the present case, we consider two alignment statistics. The first is a crowding, or density dependence effect. To define the statistic associated with this effect, we let the object-object relational feature matrix $\mathbf{W}_{1..}$ be a matrix containing only 1s and the location-location relational feature matrix $\mathbf{D}_{1..}$ be the identity matrix (i.e., a matrix with 0s on off-diagonal cells and 1s on the diagonal). The sufficient statistic t_1^{δ} formed by these matrices is then equal to the sum of squared population sizes for each location and serves to parametrize pressures toward (or, for $\delta_1 < 0$, against) crowding. The second alignment statistic, negative propinquity, expresses a tendency for households that are socially tied to one another to reside in spatially distant locations. (We use the term "negative" propinquity here to emphasize that the natural alignment statistic technically measures the inverse of propinquity in its usual sense; $\delta_2 < 0$ thus generates pressure toward propinquity per se.) To form the negative propinquity statistic, we simply let $W_{2i\mathrm{j}} = 1$ if household i is tied to household j (and 0 otherwise), with $\mathbf{D}_{2i\mathrm{j}}$ being equal to the Euclidean distance between the ith and jth regions. This gives us our second t^{δ} statistic and completes our specification of t for this model.

Several examples of configurations resulting from the above mechanisms are shown in Figure 5. Each panel shows the 400 regions comprising the location set, with household positions indicated by circles. (Within-cell positions are jittered for clarity.) Household ethnicity is indicated by color and network ties are shown via edges. While each configuration corresponds to a single draw from the location model, a burn-in sample of 100,000 draws was taken (and discarded) prior to sampling. Configurations shown here are typical of model behavior for these covariates and parameter values.

The panels of Figure 5 nicely illustrate a number of model features. In Panel 1, a model has been fit with an attraction parameter based on an interaction between rent level and household income ($\alpha = 0.0001$), balanced by a negative density dependence parameter $\delta_1 = -0.01$. Although the former effect tends to pull all households

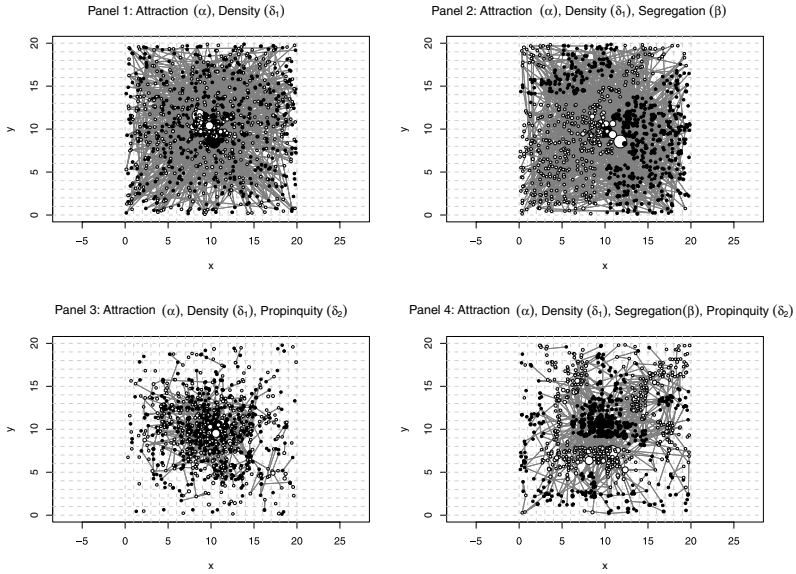

FIGURE 5. Location model draws, spatial model.

toward the center of the grid, the density avoidance effect tends to prevent "clumping." As a result, high-income households are preferentially clustered in high-rent areas, with lower-income households displaced to outlying areas. Note that without segregation or propinquity effects, neither ethnic nor social clustering are present; this would not be the case if ties were formed homophilously, and/or if ethnicity was correlated with income. Clustering can also be induced directly, of course, as is shown in panel 2. Here, we have added an object homogeneity effect for ethnicity through Queen's contiguity of regions ($\beta = -0.5$), which tends to allocate households to regions so as to reduce local heterogeneity. As can be seen, this induces strong ethnic clustering within the location system; while high-income households are still preferentially attracted to high-rent areas, this sorting is not strong enough to overcome segregation effects. Another interesting feature of the resulting configurations is the nearly empty "buffer" territory that lies between ethnic clusters. These buffer regions arise as a side effect of the contiguity rule, which tends to discourage direct contact between clusters. As this suggests, the neighborhood over which segregation effects operate can have a substantial impact on the nature of the clustering that

results. This would seem to indicate an important direction for empirical research.

A rather different sort of clustering is generated by adding a propinquity effect to the original attraction and density model. As noted earlier, we implement propinquity as an alignment effect between the interhousehold network and the Euclidean distance between household locations ($\delta_2 = -1$). As one might anticipate, the primary effect of propinquity (shown in Panel 3 of Figure 5) is to pull members of the giant component together. Since many of these members also happen to be strongly attracted to high-rent regions, the net effect is greater population density in the area immediately surrounding the urban core. Another interesting effect, however, involves households on the periphery: since propinquity draws socially connected households into the core, peripheral households are disproportionately those with few ties and/or which belong to smaller components. The model thus predicts an association between social isolation and geographical isolation. Ironically, this situation is somewhat attenuated by the reintroduction of a residential segregation effect (Panel 4). While there is still a tendency for social isolates to be forced into the geographical periphery, the consolidation of ethnic clusters limits this somewhat. Because ties are uncorrelated with ethnicity, propinquity also acts to break the settlement pattern into somewhat smaller, "band-like" clusters with interethnic ties spanning the intercluster buffer zones. (One would not expect to observe this effect in most empirical settings, however, due to the strong ethnic homophily of most social ties (McPherson et al. 2001).)

Quantitative information on the interaction between propinquity and segregation can be obtained by simulating draws from the location system with systematically varied parameters. Figure 6, for instance, shows the average value of the heterogeneity statistic for ethnicity difference by Queen's contiguity (t^{β}) as a function of (anti-)segregation (β) and (negative) propinquity (δ_2) effects. (Each data point represents a mean of 500 Metropolis draws uniformly thinned from a total sample of size 100,000, with a burn-in of 20,000 draws. All other parameters have been set to 0.) In the absence of a propinquity effect, the location system undergoes a sharp behavioral transition at $\beta = 0$; extreme homogeneity is observed below this threshold, with high levels of heterogeneity immediately above it. While propinquity seems to have little effect on heterogeneity in the segregated regime, its effect in the desegregated regime is uniformly inhibitory. Whether positive or negative,

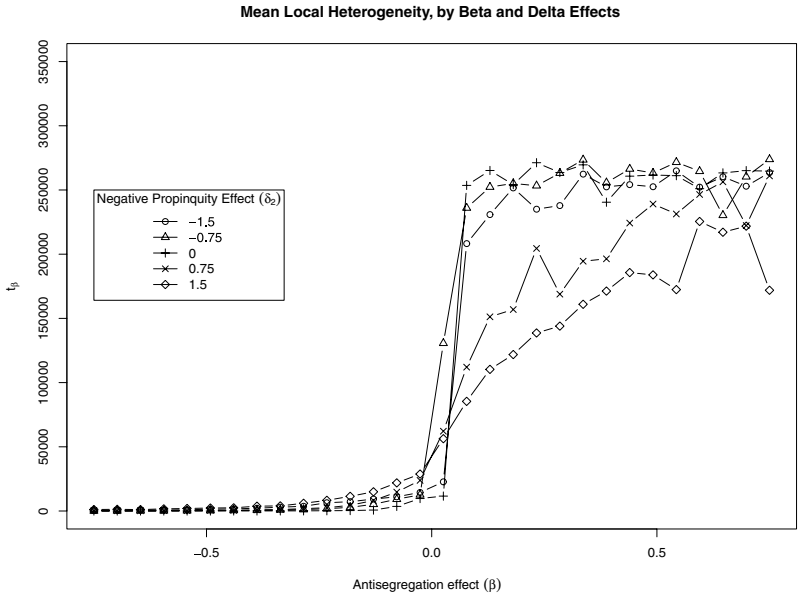

FIGURE 6. Mean segregation statistic, by segregation and propinquity effects.

propinquity effects tend to weaken heterogeneity (with stronger effects being observed in the dispropinquitous case). This phenomenon stems from the fact that propinquity affects *which* actors can be clustered together, with dispropinquitous systems tending to force large numbers of actors to remain in different locations. Such constraints make it more difficult to maximize local heterogeneity, which is most easily accomplished by the formation of dense, ethnically diverse clusters. Although the positive β, δ_2 regime may seem unlikely to arise in most residential contexts, it may still appear in related contexts such as firm siting, in which firms benefit from a heterogeneous market environment but simultaneously seek to avoid being placed too close to competitors. Interactions such as those of Figure 6 may thus have implications for the appearance and survival of locally competitive markets in a spatial context.

In addition to clustering, segregation has implications for population density. This is clearly illustrated by Figure 7, which shows the mean concentration statistic (t_1^δ) formed by the alignment of the identity matrix on locations with the complete graph on objects. (Simulations for this figure are as for Figure 6, with concentration replacing

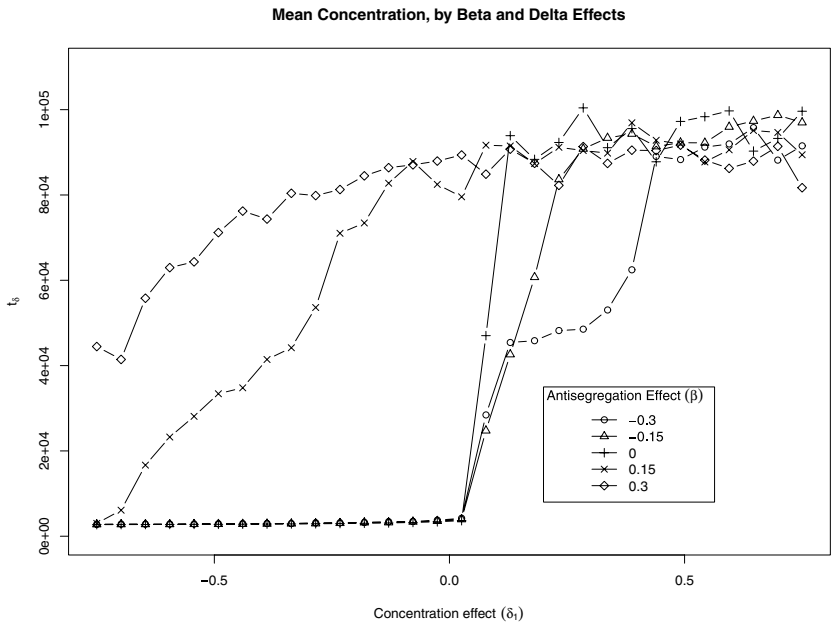

FIGURE 7. Mean concentration statistic, by density and segregation effects.

propinquity.) As the figure shows, population distribution for the location system tends toward one of two regimes: a highly concentrated regime (in which most households are packed into a very small number of regions) and a diffuse regime (in which households tend to be widely dispersed). Transitions between these regimes are moderately abrupt, with some additional consolidation occurring within the high-concentration regime for increasing δ_1. While we might at first imagine that segregation would enhance population concentration, this is not the case. Rather, segregation inhibits population concentration, with *desegregation* actively promoting it. Intuitively, this is due to the fact that the corresponding heterogeneity statistic can be most effectively increased by placing a diverse population within a small area. By contrast, concentrated, segregated population distributions are relatively difficult to produce (since any heterogeneous "incursions" are amplified by the local population level). As this implies, inhibition/promotion does not manifest here through an alteration of the extremity of the two regimes; instead, segregation effects shift the concentration "temperature" at which the transition occurs. Such a result is suggestive of the behavior

of *binary mixtures* (particularly eutectic mixtures), which can solidify at temperatures that differ greatly from those of their constituents (Kittel and Kroemer 1980).

As Schelling (1969) long ago noted, even mild tendencies toward local segregation can result in residential segregation at larger scales. While the location system model certainly bears this out, the model also suggests that factors such as population density and interhousehold ties can interact with segregation in nontrivial ways. Using the location system framework, such interactions are easy to examine and the strength of the relevant parameters can be readily estimated from census or other data sources. It is also a simple matter to introduce objects of other types (e.g., firms) that relate to households and to each other in distinct ways (as represented through additional covariates). In an era in which geographical data is increasingly available, such capabilities create the opportunity for numerous lines of research.

4.3. Empirical Example: Lazega's Lawyers

As a final sample application, we here demonstrate the use of location system models in an empirical context. The data for this example comes from Lazega's (2001) study of a midsized U.S. corporate law firm. The relevant population of the firm consists of 71 lawyers, for whom gender, tenure within the firm, age, specialty (litigation versus corporate), and law school background (here Ivy-league versus non-Ivy) are measured. Lazega also reports on relationships among the lawyers in question, of which we will here use attributions of friendship. Positions within the firm vary on two salient dimensions: seniority level (associate versus partner) and office (three work sites being present). While Lazega considers this case in great depth, the present analysis will be limited to a simple analysis of the factors associated with position occupancy within the firm. Like the other examples provided here, its purpose is more illustrative than substantive.

In studying actors' positions within the firm we will condition on the composition of both the actor and location populations; ℓ is thus taken to be a 1:1 mapping and $\mathbb{C}$ to be the set of permutations on $o_1, \ldots, o_{71}$. We initially hypothesize that the attraction/repulsion of actors to positions is governed by the interaction of the five individual attributes with position seniority. $\mathbf{X}$ is thus composed of these

five attribute vectors (in columns), with dichotomous coding for gender (male = 1), specialty (corporate = 1), and law school (Ivy = 1). Since **Q** involves only one variable, it is composed of five identical columns (with $\mathbf{Q}_{ij}$ being the seniority of l_i, partner coded as 1). In considering potential object heterogeneity effects, one obvious possibility is a tendency toward gender segregation by work site; to test for this, we posit an object heterogeneity term with object attribute matrix **Y** consisting of a column matrix of gender codes and location adjacency matrix **B** coded dichotomously such that $\mathbf{B}_{1ij} = 1$ if positions l_i and l_j are at the same work site. In addition to object heterogeneity effects, we also note the possibility of location heterogeneity within the firm. One obvious effect of this type is the potential for friendships to be stratified by seniority—that is, for friends to occupy the same strata within the firm. Such an effect can be captured by a location heterogeneity term whose object relation matrix, **A**, is the adjacency matrix of the friendship network and whose location attribute matrix **R** is a single vector of position seniority codes. Finally, we conjecture (per Lazega) that there will be a net tendency for friendship to align with work site co-membership. This is modeled via an alignment statistic whose **W** matrix is the adjacency matrix for the friendship network and whose **D** matrix is coded as **B** above.

To fit the location system model to the Lazega data, we employ the moment-matching method of Section 3.4. Initial parameter estimates were obtained using the permutation MPLE and were iteratively refined until the Euclidean distance between the mean simulated statistics and the vector of observed statistics was less than 0.1. Means for convergence testing were taken from a sample of 3000 Metropolis draws, uniformly thinned from a total sample of size 6,000,000, and 20,000 burn-in draws were taken (and discarded) before taking sample draws from each chain. A sample from the (converged) MLE was also used to estimate the deviance for each tested model (per Section 3.3.1). The deviance estimate was then used to compute the corresponding AIC scores, with model degrees of freedom equal to the number of included statistics (and hence parameters).

To assess the general properties of the assignment system, we begin by fitting each of our four effect types to the data as separate blocks. Parameter estimates and goodness-of-fit information for each such model are shown in Table 2, together with equivalent statistics for the null model of random assignment and a model containing all effects

simultaneously. Under random assignment, each assignment vector ℓ has an equal chance of being observed and hence the data likelihood under this model is fixed at $(n!)^{-1}$. Adding parameters improves fit, at the cost of parsimony—thus, models should be compared via the AIC (which penalizes the deviance by the number of parameters) for purposes of model selection. To provide some additional sense of the differences in fit, p-values are also provided for a standard χ^2 test of improvement in deviance versus the null model. Given that the usual asymptotic argument for the χ^2 as a reference distribution when comparing nested regular exponential family models (e.g., Johansen 1979) is not immediate in this case, these p-values should be interpreted heuristically; they nevertheless provide some baseline against which to assess quantitative differences in fit.

An examination of Table 2 reveals that all of the effect blocks other than object heterogeneity appear to have nonnegligible marginal effects, in the sense that each model is preferred to the null model under the AIC. (These differences in fit are also large compared to a χ^2 baseline, as indicated by the associated p-values.) The strongest effect per parameter clearly occurs for the location heterogeneity effect, followed by the alignment and attraction/repulsion effects (respectively). By contrast, object heterogeneity appears to have little impact on the assignment process, resulting in no significant improvement in fit (as assessed either by AIC or the χ^2 reference distribution). This immediately suggests that the dominant relationships within this position system are the tendency for lawyers to be tied to others of similar seniority and for friendships to coincide with shared work sites. Weaker (but still prominent) influences involve general tendencies for lawyers with particular attributes to be assigned to more (or less) senior positions; on the other hand, it does not seem to be the case that assignment within work sites is gender-homophilous.

For a closer look at the properties of the location system, we turn to the parameter estimates of Table 2. Point estimates for each fitted model are shown in two columns, containing the maximum pseudo-likelihood and maximum likelihood estimates for each included parameter (respectively). As expected, the MPLEs are generally in the neighborhood of the MLEs for these models (with sign changes observed only for small-magnitude coefficients). On the other hand, it is immediately evident from inspection of the estimates that the MPLEs are more extreme than the MLEs, being of greater absolute magnitude in 13 of 16

TABLE 2
Estimated Models for the Lazega Data

	Null Model		α Only		β Only		γ Only		δ Only		All Effects	
Effect	MPLE	MLE	MPLE	MLE	MPLE	MLE	MPLE	MLE	MPLE	MLE	MPLE	MLE
Attraction/Repulsion Effects												
Gender × Seniority			2.3916	2.3549							1.3686	0.9780
Tenure × Seniority			0.7913	0.7530							0.3493	0.5277
Age × Seniority			0.1392	0.1264							−0.0854	0.0033
Specialty × Seniority			0.1835	0.2763							−0.3647	−1.2983
School × Seniority			0.3852	0.3564							−0.0453	−0.0117
Object Heterogeneity Effects												
Gender × Same Office					0.0402	0.0284					−0.0409	0.0227
Location Heterogeneity Effects												
Friendship × Seniority							−0.6575	−0.3158			−0.6633	−0.4207
Alignment Effects												
Friendship × Same Office									1.0674	0.7426	1.0120	0.8702
Deviance		469.403		454.245		468.272		438.151		437.271		393.823
Model degrees of freedom		0		5		1		1		1		8
AIC		469.403		464.245		470.272		440.151		439.271		409.823
Null deviance–model deviance		0		15.158		1.131		31.252		32.132		75.580
χ^2 p-value (vs. Null)				0.0097		0.2876		<0.0001		<0.0001		<0.0001

cases. This extremity bias results from the fact that the MPLE considers only the local information involved in pairwise exchanges and hence underweights the impact of parameters on assignment arising from higher order dependence. The phenomenon is a fairly general one, although the effect is exacerbated for systems with stronger dependence among positions.

With regard to the estimates themselves, many of the effects are immediately intuitive. Men and those with long tenure within the firm are more likely to be found in senior positions (even controlling for relational effects). Friendship is homophilous by seniority and propinquitous by work site, as demonstrated by the consistently negative location heterogeneity and positive alignment parameters (respectively). A more subtle set of influences are clearly at work with the effects of age, specialty, and school, whose parameter estimates reverse once the effects of friendship are controlled for. Marginally, workers who are older, who specialize in corporate law, and who attended an Ivy-league school are more likely to be found in senior positions. When controlling for interpersonal relationships, however, we find that the impact of age and school diminish by more than an order of magnitude, contributing little to the relative probability of assignment. Specialty, on the other hand, both reverses sign and increases dramatically: controlling for relational effects, specialists in corporate law are actually less likely to be found in senior positions than would be expected under a "neutral" assignment process. This phenomenon appears to result from the strong tendency of friendship ties to be homophilous by both legal specialty and seniority (QAP $p \leq 0.0001$ for both relationships). We here condition on the former relationship (both specialty and friendship being object features) and controlling for the second (via t^{δ}) leads to a model in which seniority status, *ceteris paribus*, should tend to be heavily clustered by specialty. Given this relationship (and the coincidence of specialization with other attributes), the observed rate at which corporate law specialists occupy senior positions is below that which would be otherwise expected. Whether this is causal is not trivial to determine from cross-sectional data, but the location system model allows us to identify such relationships for possible future examination. As Mayhew and Levinger (1976) famously note, marginal relationships can be very misleading in the presence of strong baseline effects resulting from structural biases; the location system allows us to detect and control for many such interactions.

As a final comment, it should be noted that the magnitudes of many of the effects estimated here are fairly substantial. Given a pair of lawyers of which one is male and the other female, of whom one occupies a senior position, the odds of finding the male (rather than the female) in the senior position are approximately four times what they would be if both individuals were of the same gender. That this effect is retained even when controlling for firm composition, basic covariates, and relational structure clearly suggests the presence of gender discrimination in hiring or promotion, although a more extensive analysis would be desirable to rule out alternative explanations. Likewise, each year of tenure difference within a pair increases the relative odds of the longer-tenured individual "beating out" a more recent hire for a senior position by approximately 70 percent. Since the average tenure difference within this firm is 10.48 years, this effect can be profound — at the mean tenure difference, the odds of an "upset" given otherwise identical contenders for a senior position are reduced by a factor of over 250! Similarly, our strong estimates for γ and δ suggest that the odds of observing a potential move that increases the net number of mismatches of friendships by work site or seniority versus some baseline configuration are reduced by a factor of 40 percent to 50 percent (respectively) per net discordant edge. Of course, all things are rarely equal and the exact probability of observing any given configuration will depend on the interaction of multiple factors. Nevertheless, it is immediately apparent that the occupancy of positions within this firm is highly structured by both demographic and relational forces. Parameter estimates such as those shown here allow us to determine the strength and direction of those forces by exploiting structural signatures in the observed data. These estimates, in turn, can serve as the basis for subsequent simulation (e.g., to perform "what-if" analyses of firm structure under hypothetical perturbations), theory building (e.g., by mapping the estimated social potential onto a process such a potential game), or theory testing (e.g., by comparing the estimated parameters with those predicted by competing theories).

5. CONCLUSION

We have shown a general framework (the generalized location system) that can be used to characterize a range of social systems. An exponential

family of distributions was developed for modeling such systems, allowing for the incorporation of both attributional and relational covariates. This family belongs to a class of distributions that are well-known in the statistical literature (the regular exponential families) and it also has strong parallels with models developed for physical systems. Drawing on these established results, methods were shown for simulation and inference using the location system model. Three illustrative applications (occupational stratification, residential settlement patterns, and position occupancy within a law firm) were presented and simulation or inference was employed to show the potential utility of the location system model in each case. While there are a number of issues that have been treated here briefly or not at all—including dynamics, compatibility with low-level mechanisms, and endogenization of covariates—the material presented is sufficient to permit deployment of location system models in a wide range of empirical contexts. It is hoped that by cross-applying tools from other domains, the location system will allow for a more thorough and general treatment of complex problems than could be obtained using domain-specific methods.

REFERENCES

Acton, Forman S. 1990. *Numerical Methods That Work*. 3d ed. Washington, DC: Mathematical Association of America.

Anderson, Brigham S., Carter T. Butts, and Kathleen M. Carley. 1999. "The Interaction of Size and Density with Graph-Level Indices." *Social Networks* 21:239–67.

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Norwell, MA: Kluwer.

Barndorff-Nielsen, Ole. 1978. *Information and Exponential Families in Statistical Theory*. New York: Wiley.

Bearman, Peter S., James Moody, and Katherine Stovel. 2004. "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks." *American Journal of Sociology* 110:44–91.

Beck, Elwood M., Patrick M. Horan, and Charles M. Tolbert. 1978. "Stratification in a Dual Economy: A Sectoral Model of Earnings Determination." *American Sociological Review* 43:704–20.

Becker, Gary S. 1991. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.

Benenson, Itzhak. 2004. "Agent-Based Modeling: From Individual Residential Choice to Urban Residential Dynamics." Pp. 67–95 in *Spatially Integrated Social Science*, edited by Michael F. Goodchild and Donald G. Janelle. Oxford: Oxford University Press.

Besag, Julian. 1974. "Spatial Interaction and the Statistical Analysis of Lattice Systems." *Journal of the Royal Statistical Society, Series B* 36:192–236.

———. 1975. "Statistical Analysis of Non-Lattice Data." *The Statistician* 24:179–95.

Bourne, Larry S. 1981. *The Geography of Housing*. New York: V. H. Winston.

Bovy, Piet H. L., and Eliahu Stern. 1990. *Route Choice: Wayfinding in Transport Networks*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Bozdogan, Hamparsum. 2000. "Akaike's Information Criterion and Recent Developments in Information Complexity." *Journal of Mathematical Psychology* 44:61–91.

Brown, Lawrence D. 1986. *Fundamentals of Statistical Exponential Families, with Applications in Statistical Decision Theory*. Hayward, CA: Institute of Mathematical Statistics.

Budig, Michele J., and Paula England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66:204–25.

Calvo-Armengol, Antoni, and Matthew O. Jackson. 2004. "The Effects of Social Networks on Employment and Inequality." *American Economic Review* 94:426–54.

Carley, Kathleen M. 1991. "A Theory of Group Stability." *American Sociological Review* 56:331–54.

Cliff, Andrew D., and J. Keith Ord. 1973. *Spatial Autocorrelation*. London: Pion.

Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Contractor, Nosh, Stanley Wasserman, and Katherine Faust. 2006. "Testing Multitheoretical Multilevel Hypotheses About Organizational Networks: An Analytic Framework and Empirical Example." *Academy of Management Review* 31:681–703.

Corstjens, Marcel L., and David A. Gautschi. 1983. "Formal Choice Models in Marketing." *Marketing Science* 2:19–56.

Crouch, Bradley, Stanley Wasserman, and F. Trachtenburg. 1998. "Markov Chain Monte Carlo Maximum Likelihood Estimation for p^* Social Network Models." Presented at the Eighteenth International Sunbelt Social Network Conference, Sitges, Spain.

Doreian, Patrick, Vladimir Batagelj, and Anuska Ferlioj. 2005. *Generalized Blockmodeling*. Cambridge, England: Cambridge University Press.

Eckstein, Zvi, and Kenneth Wolpin. 1989. "The Specification and Estimation of Dynamic Stochastic Discrete Choice Models." *Journal of Human Resources* 24:562–98.

Erdös, P., and A. Rényi. 1960. "On the Evolution of Random Graphs." *Public Mathematical Institute of Hungary Academy of Sciences* 5:17–61.

Fararo, Thomas J. 1989. "The Spirit of Unification in Sociological Theory." *Sociological Theory* 7:175–90.

Fararo, Thomas J., and John Skvoretz. 1987. "Unification Research Programs: Integrating Two Structural Theories." *American Journal of Sociology* 92:1183–209.

Feser, Edward J., Stuart H. Sweeney, and Henry C. Renski. 2005. "A Descriptive Analysis of Discrete U.S. Industrial Complexes." *Journal of Regional Science* 45:395–419.

Fienberg, Steven E., M. M. Meyer, and Stanley Wasserman. 1985. "Statistical Analysis of Multiple Sociometric Relations." *Journal of the American Statistical Association* 80:51–67.

Fienberg, Steven E., and Stanley Wasserman. 1981. "Categorical Data Analysis of Single Sociometric Relations." Pp. 156–92 in *Sociological Methodology*, vol. 12. San Francisco, CA: Jossey-Bass.

Fossett, Mark. 2006. "Ethnic Preferences, Social Distance Dynamics and Residential Segregation: Theoretical Explanations Using Simulation Analysis." *Journal of Mathematical Sociology* 30:185–274.

Frank, Ove, and David Strauss. 1986. "Markov Graphs." *Journal of the American Statistical Association* 81:832–42.

Freidkin, Noah. 1998. *A Structural Theory of Social Influence*. Cambridge, England: Cambridge University Press.

Frey, William H., and Reynolds Farley. 1994. "Latino, Asian and Black Segregation in U.S. Metropolatin Areas: Are Multi-Ethnic Metros Different?" *Demography* 33:35–50.

Galster, George C. 1982. "Black and White Preferences for Neighborhood Racial Composition." *Real Estate Economics* 10:39–66.

Gamerman, Dani. 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman and Hall.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin. 1995. *Bayesian Data Analysis*. London: Chapman and Hall.

Geyer, Charles J., and Elizabeth A. Thompson. 1992. "Constrained Monte Carlo Maximum Likelihood Calculations" (with discussion). *Journal of the Royal Statistical Society*, Series C, 54:657–99.

Gilks, Walter R. 1996. "Full Conditional Distributions". Pp. 75–88 in *Markov Chain Monte Carlo in Practice*, edited by Walter R. Gilks, Sylvia Richardson, and David J. Spiegelhalter. London: Chapman and Hall.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter. 1996a. "Intoducing Markov Chain Monte Carlo." Pp. 1–20 in *Markov Chain Monte Carlo in Practice*, edited by Walter R. Gilks, Sylvia Richardson, and David J. Spiegelhalter. London: Chapman and Hall.

Gilks, Walter R., Sylvia Richardson, and David J. Spiegelhalter, eds. 1996b. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Granovetter, Mark. 1973. "The Strength of Weak Ties." *American Journal of Sociology*, 78:1369–80.

Handcock, Mark S. 2003a. "Assessing Degeneracy in Statistical Models of Social Networks." CSSS Working Paper No. 39, University of Washington.

———. 2003b. "Statistical Models for Social Networks: Inference and Degeneracy." Pp. 229–40 in *Dynamic Social Network Modeling and Analysis*, edited by Ron Breiger, Kathleen M. Carley, and Philippa Pattison. Washington, DC: National Academies Press.

Handcock, Mark S., David R. Hunter, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2003. "The Statnet Library for R." Software library (http://www.csde.washington.edu/statnet).

Harary, Frank. 1953. "On the Notion of Balance of a Signed Graph." *Michigan Mathematical Journal* 3:37–41.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. New York: Wiley.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. "Latent Space Approaches to Social Network Analysis." *Journal of the American Statistical Association* 97:1090–98.

Holland, Paul W., Kathryn Blackmond Laskey, and Samuel Leinhardt. 1983. "Stochastic Blockmodels: First Steps." *Social Networks* 5:109–37.

Holland, Paul W., and Samuel Leinhardt. 1981. "An Exponential Family of Probability Distributions for Directed Graphs" (with discussion). *Journal of the American Statistical Association* 76:33–50.

Hubert, Lawrence J. 1987. *Assignment Methods in Combinatorial Data Analysis*. New York: Marcel Dekker.

Huffman, Matt L., and Philip N. Cohen. 2004. "Racial Wage Inequality: Job Segregation and Devaluation Across U.S. Labor Markets." *American Journal of Sociology* 109:902–36.

Hummon, Norman P., and Patrick Doreian. 2003. "Some Dynamics of Social Balance Processes: Bringing Heider Back into Balance Theory." *Social Networks* 25:17–49.

Hunter, David R., and Mark S. Handcock. 2006. "Inference in Curved Exponential Family Models for Networks." *Journal of Computational and Graphical Statistics* 15:565–83.

Johansen, Søren. 1979. *Introduction to the Theory of Regular Exponential Families*. Copenhagen: University of Copenhagen.

Jones, Rachael, and April Brayfield. 1997. "Life's Greatest Joy?: European Attitudes Toward the Centrality of Children." *Social Forces* 75:1239–69.

Joy, Lois. 2003. "Salaries of Recent Male and Female College Graduates: Educational and Labor Market Effects." *Industrial and Labor Relations Review* 56:606–21.

Kalos, Malvin H., and Paula A. Whitlock. 1986. *Monte Carlo Methods*. Vol. 1, Basics. New York: Wiley.

Kelso, Alexander S., and Vincent P. Crawford. 1982. "Job Matching, Coalition Formation and Gross Substitutes." *Econometrica* 50:1483–504.

Kittel, Chales, and Herbert Kroemer. 1980. *Thermal Physics*. 2d ed. New York: W. H. Freeman.

Lazega, Emmanuel. 2001. *The Collegial Phenomenon: The Social Mechanisms of Cooperation Among Peers in a Corporate Law Partnership*. Oxford, England: Oxford University Press.

Loehman, Edna, and Vo Hu De. 1982. "Application of Stochastic Choice Modeling to Policy Analysis of Public Goods: A Case Study of Air Quality Improvements." *Review of Economics and Statistics* 64:474–80.

Lorrain, Francois, and Harrison C. White. 1971. "Structural Equivalence of Individuals in Social Networks." *Journal of Mathematical Sociology* 1:49–80.

Luce, R. Duncan. 1959. *Individual Choice Behavior*. New York: Wiley.

Massey, Douglas S., and Nancy A. Denton. 1993. *American Apartheid: Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press.

Mayhew, Bruce H. 1980. "Structuralism Versus Individualism: Part I, Shadowboxing in the Dark." *Social Forces* 59:335–75.

———. 1981. "Structuralism Versus Individualism: Part II, Ideological and Other Obfuscations." *Social Forces* 59:627–48.

Mayhew, Bruce H., and Roger L. Levinger. 1976. "Size and Density of Interaction in Human Aggregates." *American Journal of Sociology* 82:86–110.

McFadden, Daniel. 1973. "Conditional Logit Analysis of Qualitative Choice Behavior." Pp. 105–42 in *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press.

McPherson, J. Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Merton, Robert K. 1957. *Social Theory and Social Structure*. Glencoe, IL: Free Press.

Monderer, Dov, and Lloyd S. Shapley. 1996. "Potential Games." *Games and Economic Behavior* 14:124–43.

Oppenheim, Norbert. 1995. *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium*. New York: Wiley.

Parsons, Talcott. 1951. *The Social System*. London: Routledge.

Pattison, Philippa, and Garry Robins. 2002. "Neighborhood-Based Models for Social Networks." Pp. 301–3 in *Sociological Methodology*, vol. 32, edited by Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.

Pattison, Philippa, and Stanley Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks: II. Multivariate Relations." *British Journal of Mathematical and Statistical Psychology* 52:169–93.

Press, William H., Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. 1992. *Numerical Recipes: The Art of Scientific Computing*. 2d ed. Cambridge, England: Cambridge University Press.

Robert, Christian P. 1994. *The Bayesian Choice: A Decision-Theoretic Motivation*. New York: Springer.

Robins, Garry, and Philippa Pattison. 2005. "Interdependencies and Social Processes: Dependence Graphs and Generalized Dependence Structures." Pp. 192–214 in *Models and Methods in Social Network Analysis*, edited by Peter J. Carrington, John Scott, and Stanley Wasserman. Cambridge, England: Cambridge University Press.

Robins, Garry, Philippa Pattison, and Stanley Wasserman. 1999. "Logit Models and Logistic Regressions for Social Networks, III. Valued Relations." *Psychometrika* 64:371–94.

Robins, Garry, Philippa Pattison, and Jodie Woolcock. 2005. "Small and Other Worlds: Network Structures from Local Processes." *American Journal of Sociology* 110:894–936.

Robins, Garry L., and Philippa Pattison. 2001. "Random Graph Models for Temporal Processes in Social Networks." *Journal of Mathematical Sociology* 25:5–41.

Roth, Alvin E., and Marilda Sotomayor. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge, England: Cambridge University Press.

Sakoda, James M. 1971. "The Checkerboard Model of Social Interaction." *Journal of Mathematical Sociology* 1:119–32.

Sattinger, Michael. 1993. "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature* 31:831–80.

Schelling, Thomas C. 1969. "Models of Segregation." *American Economic Review* 59:483–93.

Shapley, Lloyd, and Martin Shubik. 1971. "The Assignment Game I: The Core." *International Journal of Game Theory* 1:111–30.

Snijders, Tom A. B. 2002. "Markov Chain Monte Carlo Estimation of Exponential Random Graph Models." *Journal of Social Structure* 3 (www.cma.edu/joss/).

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. "New Specifications for Exponential Random Graph Models." Pp. 99–153 in *Sociological Methodology*, vol. 36, edited by Ross M. Stolzenberg. Boston, MA: Blackwell Publishing.

Sorokin, Pitrim A. 1957. *Social and Cultural Dynamics*. Boston: Porter Sargent.

Spencer, Herbert. 1896. *Social Statics*. New York: D. Appleton.

Stewman, Shelby. 1988. "Organizational Demography." *Annual Review of Sociology* 14:173–202.

Stoyan, Dietrich, Wilfrid S. Kendall, and Joseph Mecke. 1987. *Stochastic Geometry and Its Applications*. 2d ed. New York: Wiley.

Strauss, David. 1986. "On a General Class of Models for Interaction." *SIAM Review* 28:513–27.

Strauss, David, and Michael Ikeda. 1990. "Pseudolikelihood Estimation for Social Networks." *Journal of the American Statistical Association* 85:204–12.

Sweeney, Stuart H., and Edward J. Feser. 2003. "Business Location and Spatial Exernalities: Tying Concepts to Measures." Pp. 239–62 in *Spatially Integrated Social Science: Examples in Best Practice*, edited by M. Goodchild and D. Janelle. Oxford, England: Oxford University Press.

Swendsen, Robert H., and Jian-Sheng Wang. 1987. "Non-universal Critical Dynamics in Monte Carlo Simulation." *Physical Review Letters* 58:86–88.

Tauber, Karl E., and Alma F. Tauber. 1965. *Negroes in Cities: Residential Segregation and Neighborhood Change*. Chicago, IL: Aldine.

Wasserman, Larry. 2000. "Bayesian Model Selection and Model Averaging." *Journal of Mathematical Psychology* 44:92–107.

Wasserman, Stanley, and Philippa Pattison. 1996. "Logit Models and Logistic Regressions for Social Networks: I. An Introduction to Markov Graphs and p^*." *Psychometrika* 60:401–26.

Wasserman, Stanley, and Garry Robins. 2005. "An Introduction to Random Graphs, Dependence Graphs and p^*." Pp. 192–214 in *Models and Methods in Social Network Analysis*, edited by Peter J. Carrington, John Scott, and Stanley Wasserman. Cambridge, England: Cambridge University Press.

White, Harrison C. 1970. *Chains of Opportunity: System Models of Mobility in Organizations*. Cambridge, MA: Harvard University Press.

Young, H. Peyton. 1998. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton, NJ: Princeton University Press.

Zhang, Junfu. 2004. "A Dynamic Model of Residential Segregation." *Journal of Mathematical Sociology* 28:147–70.