Building Inferentially Tractable Models of Complex Social Systems: a Generalized Location Framework

Carter T. Butts

8/5/2005

Abstract

Here, a class of structures called "generalized location systems" is introduced, which can be used to characterize a range of social processes. 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. 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.

Keywords: statistical exponential families, Markov chain Monte Carlo, social networks, assignment models, spatial processes

1 Introduction

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 loglinear models of Holland and Leinhardt (1981) and Fienberg and Wasserman (1981), which were extended by various researchers (e.g. Fienburg et al., 1985; Holland et al., 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 et al., 1999). Recent work has expanded on these innovations both in terms of improved inferential strategies (Crouch et al., 1998; Snijders, 2002; Hunter and Handcock, 2004) and an expanded understanding of the models themselves (Handcock, 2003; Robins et al., 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 which have so far been obtained underscore the 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). By drawing on results obtained by researchers studying structurally similar problems in other substantive areas, network researchers have been able to greatly accelerate development in their own field.

Cross-application of concepts and methods has lead to great strides in network analysis, but there is more which 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. Processes such as stratification, settlement patterns, migration, firm siting, and occupational segregation pose similar challenges of complex dependence, but 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). 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.

This paper presents a family of models for social phenomena which can be described in terms of the arrangement of various (possibly related) objects with respect to a set of (again, possibly related) locations. This family is designed so as to leverage the large literature on the stochastic modeling of systems with nontrivial dependence structures. It is also constructed so as to be applicable across a wide range of substantive contexts; to scale well to large social systems; to be readily simulated; to be specifiable in terms of directly measurable properties; and to support likelihood-based inference using (fairly) standard methods. Although the focus in this paper is on foundational issues and simulation, this should not be seen as diminishing the importance of the last two items above. On the contrary, a vital aspect of the framework is the fact that it is built on observables "from the ground up," avoiding the problems of operationalization which plague most sociological theory. While the illustrative examples shown here are stylized for purposes of exposition, even these can be mapped to measurable properties of real social systems.

The structure of the paper is as follows. After a brief comment on notation, we present the core formalism of the paper (the generalized location system). Given this, we turn to a discussion of modeling location systems, including both conceptual and computational issues. Finally, we illustrate the use of the location system model to examine two classes of processes (occupational stratification and residential settlement patterns), before concluding with a brief discussion.

1.1 Notation

We here outline some general notation, which will be used in the material which 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 $\mathbf{X}_{i\cdot\cdot}$ 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 (e.g., $(x^{(1)}, \ldots, x^{(n)})$). Distributional equivalence is denoted by $\sim$ (read: "is distributed as"), so $X \sim Y$ implies that X is distributed as Y. For convenience, this notation may also be extended to pmfs, such that $X \sim f$ (for random variable X and pmf f) should be understood to mean that X is distributed as a random variable with pmf f.

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 which consists of n identifiable *objects*, $O = (o_1, \ldots, o_n)$, each of which may reside in exactly one of m identifiable *locations*, $L = (l_1, \ldots, l_n)$. 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 which are realizable is said to be the set of *accessible configurations*, and is denoted $\mathbb{C}$. One very important parameterization of $\mathbb{C}$ with which we will deal is in terms of *occupancy constraints*. We define the *occupancy function* of a location system as

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

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 ℓ 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 which are of special interest. Feature vectors provide ways of assigning numerical values to individual objects or locations, e.g., 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 ℓ, 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 which 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, out 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 which 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 which satisfies these constraints. The core of this framework is a discrete exponential family of distributions which 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 will 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 which incorporates a range of substantively important effects in a reasonably simple fashion. We then consider some of the statistical mechanical properties of the location system model, which provide insight into the behavior of the location system by means of analogy to related physical systems. Some simple inferential properties of the location system are also discussed, and 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 – given a set of accessible configurations, $\mathbb{C}$– the system will be found to occupy any particular configuration, ℓ, with some specified probability. 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

$$\Pr(S = \ell) = I_{\mathbb{C}}(\ell) \frac{\exp\big(\mathcal{P}\left(\ell\right)\big)}{\sum_{\ell' \in \mathbb{C}} \exp\big(\mathcal{P}\left(\ell'\right)\big)} \tag{3}$$

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

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

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 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 from both physics and mathematical statistics, facilitating new theoretical development (see, e.g., Barndorff-Nielsen, 1978; Brown, 1986). However, the most important property of the exponential family framework is perhaps the third: given an appropriate parameterization of $\mathcal{P}$, there are existing results which permit principled inference from empirical data (Johansen, 1979). While such inference is not the focus of this paper, the availability of viable inferential tools is a major motivation for our approach; models which 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 which incorporates a number of features of known substantive importance for a variety of social systems. This form is introduced in the following section.

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 $\Pr(S = \ell) \propto \mathcal{P}(\ell)$. Substantively, 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 *parameterized* so as to allow differential weighting of effects. Ideally, the social potential family should also be easily computed, and its structure easily interpreted.

An obvious initial solution to this problem is to construct $\mathcal{P}$ from a linear combination of sufficient statistics (i.e., deterministic functions of 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.

Even if a linear form is supposed, however, we are left with a more important question: what effects should be included in 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

Table 1: Elements of the Social Potential

Obviously, these effects must be parameterized as functions of the location and object features. Further, both location and object features (as we are using the term) can 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 which are dyadic (i.e., defined on pairs) and single-mode (i.e., which 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 which draw objects towards 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 which express direct attraction or repulsion between particular objects and locations, based on their attributes. In the second (upper right) cell are effects which 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. Homophily in career choice – where careers are interpreted as "locations" – serves as an example of a location homogeneity effect. 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 sub-potentials (one for each category) such that

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

We now consider each of these functions in turn.

3.2.1 Attraction/Repulsion Potential

The first class of effects which 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 which 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 which 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[1], 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 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 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 $\beta \in \mathbb{R}^b$ a parameter vector. Then we 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 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 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 $\mathbf{B}_{i\cdot\cdot}$. 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.

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

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 $\mathbf{B}_{i\cdot\cdot}$ 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 $\mathbf{B}_{i\cdot\cdot}$ equal to the identity matrix allows β_i to directly parameterize 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. Then $\mathcal{P}_\gamma$ is defined as follows:

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

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

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 $\mathbf{A}_{i\cdot\cdot}$ 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 $R_{\cdot i}$). For $\gamma_i > 0$, $\mathcal{P}_\gamma$ then places more weight on job allocations which increase the within-couple wage rate differences (ceteris paribus). $\gamma_i < 0$ 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, 2005), 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 $\mathbf{W}_{i\cdot\cdot}$. Where $\delta_i < 0$, the kinship network is propinquitous; that is, actors tend to reside (ceteris paribus) in locations which 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 who are tied to one another via production of similar products.)

Another important alignment effect is density dependence. Density dependence occurs as a tendency for objects to cluster (positive dependence) or disperse (negative dependence) with respect to locations. To model density dependence, we create an object relation $\mathbf{W}_{i\cdot\cdot}$ representing a complete graph, and employ the identity matrix for $\mathbf{D}_{i\cdot\cdot}$. 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.

3.2.5 Combined Linear Potential

We are now ready to form the combined linear social potential. Substituting the quantities of Equations 7–13 into Equation 5 gives us

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

in terms of sufficient statistics, or

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

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 will now proceed to a consideration of some of the properties of this model family, before turning to the problem of simulation.

3.3 Thermodynamic Properties of the Location System Model

We have already seen that the stochastic location system model of Equation 3 can be viewed as directly analogous to a standard class of statistical mechanical models. This fact allows us to employ some useful results from the physics literature to elucidate several aspects of model behavior.[2] (Interestingly, many of these results have parallels within the statistical literature, and can be derived in other ways; see, for example, Barndorff-Nielsen (1978).)

As noted above, the normalizing factor, $Z(\mathcal{P}, \mathbb{C})$ is directly analogous to the *partition function* of statistical mechanics. The quantity $F = -\ln Z(\mathcal{P}, \mathbb{C})$, in turn, corresponds to the *free energy* of the location system. In a classical statistical mechanical system, the probability of observing the system in microstate j is given by

$$p_j = \frac{\exp(-\epsilon_j/\tau)}{Z}, \tag{16}$$

where ϵ_j is the microstate energy of j and τ is the temperature. Thus, the log-probability of microstate j is a linear function of the free and microstate energies:

$$\ln p_j = F - \frac{\epsilon_j}{\tau}. \tag{17}$$

Returning to Equation 3, it is immediately apparent that the social potential $\mathcal{P}$ plays the role of ϵ/τ. Indeed, inspecting Equation 14 reveals an even closer correspondence: the realizations of the sufficient statistics associated with the elements of t^α, t^β, t^γ, and t^δ are similar to microstate energies, and the corresponding parameters (α, β, γ, and δ) can be thought of as vectors of inverse temperatures. More precisely, each sufficient statistic is analogous to the energy function (or Hamiltonian) associated with a particular "mode" of ℓ, just as the total microstate energy of a particle system might combine contributions from translational, rotational, and/or vibrational modes. The "energy" associated with a particular microstate, ℓ, in each mode is given by the value of the sufficient statistic for that microstate (i.e., $t^\theta_i(\ell)$). As in the physical case, the log-probability of observing a particular realization of the location system can be expressed as a "free energy" minus a linear combination of microstate "energies" whose coefficients correspond to "inverse temperatures." While one does not conventionally encounter negative temperatures in a physical system[3] (as can arise in our case), we will find that this metaphor is useful in understanding the behavior of the location system. This point was foreshadowed by Mayhew et al.

[2] Except as noted otherwise, the results here are standard and many be found in any text on statistical mechanics, e.g. Kittel and Kroemer (1980).

[3] Although exceptions do exist. See, for instance, Ramsey (1956).

(1995), who invoked an "energy distribution principle" in describing the occurrence of naturally forming groups. The present model implements a notion of precisely this sort, for more general social systems. As exponential family models also have the property of maximizing entropy conditional on their parameters and sufficient statistics, these location system model can also be thought of as a family of *baseline models* in the sense of Mayhew (1984a;b).

In addition to providing insight into system behavior, the above is also helpful in deriving other characteristics. For instance, the average microstate energy of the system described by Equation 16 is given by $dF/d\theta$, where $\theta = \tau^{-1}$. It follows for our purposes that

$$\mathbf{E}_{(\alpha,\beta,\gamma,\delta)} t_i^\theta = -\frac{d \ln Z(\mathcal{P}, \mathbb{C})}{d\theta_i}, \tag{18}$$

where θ_i represents any parameter of the system. Thus, expectations for arbitrary sufficient statistics can be obtained through the partition function. Second moments may be obtained in a similar manner: the Hessian matrix $-\frac{d^2 F}{d\theta^2}$ yields the variance-covariance matrix for all sufficient statistics in the system. (In the physical case, this corresponds to the *energy fluctuation,* or the *variance in energy.*)

Moments of sufficient statistics are useful for a variety of purposes, but other statistical mechanical properties of the location system may also be of value. For instance, the "heat capacity" of the system for parameter θ_i is given by $\mathrm{Var}(t_i^\theta(\ell))\theta_i^2 = -\frac{d^2F}{d\theta^2}\Big|_{ii} \theta_i^2$. In the physical case, heat capacity reflects the capacity of a system to store energy (in the sense of the change in energy per unit temperature). Here, heat capacity for parameter θ_i reflects the sensitivity of the corresponding statistic t_i^θ to changes in the "temperature" $1/\theta_i$. For instance, if θ_i corresponds to a attraction parameter between income and gender, then heat capacity can be used to parameterize the income consequences of a weakening (or strengthening) of the attractive tendency within the larger system.

Using arguments similar to the above, it is possible to derive analogs to various other thermodynamic properties such as pressure, entropy (also obtainable through information-theoretic arguments), and chemical potential. While one must always be careful in interpreting such quantities, they may nevertheless provide interesting and useful ways of describing the properties of location systems. We will see some of the interpretational value of thermodynamic analogy below, when we consider some sample applications of the location system model; before proceeding to this, however, we turn to the question of how location system behavior may be simulated.

3.4 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. Given that numerous accessible references on the Metropolis algorithm are currently available (see, e.g. Gamerman, 1997; Gilks et al., 1996; Gelman et al., 1995), we will focus here on issues which 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., 1996, 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 (e.g., 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, 2005).

To implement the Metropolis algorithm, then, our core concern is computation of the probability ratio between states. Given a current state, $\ell^{(i)}$, the probability of accepting a candidate state, $\ell^{(i+1)}$, is then

$$\frac{\Pr(S=\ell^{(i+1)}|\mathcal{P},\mathbb{C})}{\Pr(S=\ell^{(i)}|\mathcal{P},\mathbb{C})} = \frac{\exp\left(\mathcal{P}\left(\ell^{(i+1)}\right)\right)}{\exp\left(\mathcal{P}\left(\ell^{(i)}\right)\right)} \frac{Z\left(\mathcal{P},\mathbb{C}\right)}{Z\left(\mathcal{P},\mathbb{C}\right)} \tag{19}$$

$$= \frac{\exp\left(\mathcal{P}\left(\ell^{(i+1)}\right)\right)}{\exp\left(\mathcal{P}\left(\ell^{(i)}\right)\right)} \tag{20}$$

$$= \exp\left(\mathcal{P}\left(\ell^{(i+1)}\right) - \mathcal{P}\left(\ell^{(i)}\right)\right). \tag{21}$$

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

$$\begin{aligned}\mathcal{P}\left(\ell^{(i+1)}\right)-\mathcal{P}\left(\ell^{(i)}\right)=&\sum_{i=1}^{a}\alpha_i\left[t_i^{\alpha}\left(\ell^{(i+1)}\right)-t_i^{\alpha}\left(\ell^{(i)}\right)\right]+\sum_{i=1}^{b}\beta_i\left[t_i^{\beta}\left(\ell^{(i+1)}\right)-t_i^{\beta}\left(\ell^{(i)}\right)\right]\\&+\sum_{i=1}^{c}\gamma_i\left[t_i^{\gamma}\left(\ell^{(i+1)}\right)-t_i^{\gamma}\left(\ell^{(i)}\right)\right]+\sum_{i=1}^{d}\delta_i\left[t_i^{\delta}\left(\ell^{(i+1)}\right)-t_i^{\delta}\left(\ell^{(i)}\right)\right],\end{aligned} \tag{22}$$

or, substituting from Equation 15,

$$\begin{aligned}=&\sum_{i=1}^{a}\sum_{j=1}^{n}\alpha_i\mathbf{X}_{ji}\left[\mathbf{Q}_{\ell_j^{(i+1)}i}-\mathbf{Q}_{\ell_j^{(i)}i}\right]\\&+\sum_{i=1}^{b}\beta_i\sum_{j=1}^{n}\sum_{k=1}^{n}\left[\mathbf{B}_{i\ell_j^{(i+1)}\ell_k^{(i+1)}}-\mathbf{B}_{i\ell_j^{(i)}\ell_k^{(i)}}\right]\left|\mathbf{Y}_{ji}-\mathbf{Y}_{ki}\right|\\&+\sum_{i=1}^{c}\gamma_i\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{A}_{ijk}\left[\left|\mathbf{R}_{\ell_j^{(i+1)}i}-\mathbf{R}_{\ell_k^{(i+1)}i}\right|-\left|\mathbf{R}_{\ell_j^{(i)}i}-\mathbf{R}_{\ell_k^{(i)}i}\right|\right]\\&+\sum_{i=1}^{d}\delta_i\sum_{j=1}^{n}\sum_{k=1}^{n}\mathbf{W}_{ijk}\left[\mathbf{D}_{i\ell_j^{(i+1)}\ell_k^{(i+1)}}-\mathbf{D}_{i\ell_j^{(i)}\ell_k^{(i)}}\right].\end{aligned} \tag{23}$$

Although Equation 23 can be used to compute the potential difference directly, Equation 22 demonstrates that the same quantity can be expressed directly in terms of a fixed linear combination of differences in sufficient statistics. For purposes of simulation, then, we need only track such differences. 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 which 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 23. 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:

$$t_i^\alpha(\ell') - t_i^\alpha(\ell) = \mathbf{X}_{ji}\left[\mathbf{Q}_{ki} - \mathbf{Q}_{\ell_j i}\right] \tag{24}$$

$$t_i^\beta(\ell') - t_i^\beta(\ell) = \sum_{g=1}^{n} \left|\mathbf{Y}_{ji} - \mathbf{Y}_{gi}\right| \left[\mathbf{B}_{ik\ell_g} - \mathbf{B}_{i\ell_j\ell_g} + \mathbf{B}_{i\ell_g k} - \mathbf{B}_{i\ell_g\ell_j}\right] \tag{25}$$

$$t_i^\gamma(\ell') - t_i^\gamma(\ell) = \sum_{g=1}^{n} \left[\mathbf{A}_{ijg} + \mathbf{A}_{igj}\right] \left[\left|\mathbf{R}_{ki} - \mathbf{R}_{\ell_g i}\right| - \left|\mathbf{R}_{\ell_g i} - \mathbf{R}_{\ell_g i}\right|\right] \tag{26}$$

$$\begin{aligned} t_i^\delta(\ell') - t_i^\delta(\ell) = & \sum_{g\neq j, g=1}^{n} \left[\mathbf{W}_{ijg}\left(\mathbf{D}_{ik\ell_g} - \mathbf{D}_{i\ell_j\ell_g}\right) + \mathbf{W}_{igj}\left(\mathbf{D}_{i\ell_g k} - \mathbf{D}_{i\ell_g\ell_j}\right)\right] \\ & + \mathbf{W}_{ijj}\left(\mathbf{D}_{ikk} - \mathbf{D}_{i\ell_j\ell_j}\right) \end{aligned} \tag{27}$$

Calculation of Equation 22 for a single reassignment using Equations 24-27 is an $\mathcal{O}(n)$ operation. This is a substantial improvement over the $\mathcal{O}(n^2)$ complexity for direct application of Equation 23 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.

3.4.1 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? 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 which is close to θ (in the sense that $\|\theta' - \theta\|$ is small). Our estimator of the partition function is based on the result that

$$\lim_{M\to\infty} \frac{|\mathbb{C}| \sum_{i=1}^{M} \exp\left((\theta' - \theta)^T t\left(\ell^{(i)}\right)\right)}{\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)} = Z\left(\theta', \mathbb{C}\right). \tag{28}$$

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),

$$\lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} f\left(\ell^{(i)}\right) \frac{Z\left(\theta, \mathbb{C}\right)}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)} = \sum_{\ell' \in \mathbb{C}} f\left(\ell'\right) \tag{29}$$

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 $f\left(\ell\right) = \exp\left(\theta'^T t\left(\ell\right)\right)$ then gives us

$$\begin{aligned} \lim_{M\to\infty} \frac{1}{M} \sum_{i=1}^{M} \exp\left(\theta'^T t\left(\ell^{(i)}\right)\right) \frac{Z\left(\theta, \mathbb{C}\right)}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)} &= \sum_{\ell' \in \mathbb{C}} \exp\left(\theta'^T t\left(\ell'\right)\right) &(30)\\ &= Z\left(\theta', \mathbb{C}\right). &(31) \end{aligned}$$

While this gives the desired result, it requires us to know the value of $Z\left(\theta, \mathbb{C}\right)$ and thus is of little immediate use. However, since the partition function does not depend on ℓ, it may be pulled out of the initial summand:

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

Thus,

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

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

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

To see the value of this, let us return to Equation 28. Immediately, we note that

$$\frac{|\mathbb{C}| \sum_{i=1}^{M} \exp\left(\left(\theta' - \theta\right)^T t\left(\ell^{(i)}\right)\right)}{\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)} = |\mathbb{C}| M \left[\frac{1}{M} \exp\left(\left(\theta' - \theta\right)^T t\left(\ell^{(i)}\right)\right)\right] \times \left[\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)\right]^{-1} . \tag{35}$$

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

$$\approx |\mathbb{C}| M \left[\frac{Z\left(\theta', \mathbb{C}\right)}{Z\left(\theta, \mathbb{C}\right)}\right] \left[\sum_{i=1}^{M} \exp\left(-\theta^T t\left(\ell^{(i)}\right)\right)\right]^{-1} \tag{36}$$

$$= |\mathbb{C}| Z\left(\theta', \mathbb{C}\right) \left[\frac{1}{M} \sum_{i=1}^{M} \frac{Z\left(\theta, \mathbb{C}\right)}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)}\right]^{-1} . \tag{37}$$

Now, what of the factor on the right? Returning to Equation 29, we simply take $f(\ell') = 1$, which yields

$$\frac{1}{M} \sum_{i=1}^{M} 1 \frac{Z\left(\theta, \mathbb{C}\right)}{\exp\left(\theta^T t\left(\ell^{(i)}\right)\right)} \to \sum_{\ell' \in \mathbb{C}} 1 \tag{38}$$

$$= |\mathbb{C}|, \tag{39}$$

and thus, by substitution,

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

$$= Z\left(\theta', \mathbb{C}\right) . \tag{41}$$

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' which is close to θ, only a single sample is needed to compute numerical derivatives. In the special case for which we are solely interested in $Z(\theta, \mathbb{C})$, Equation 28 further simplifies to

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

This last follows immediately from substitution.

3.5 Inference

Although an in-depth discussion of inference for the location model is beyond the scope of this paper, a few basic results are sketched here to give the flavor of what is possible. Given a location system specified by the tuple $(L, O, \mathbb{C}, F_L, F_O)$ with parametric social potential $\mathcal{P}$, the likelihood of observed state ℓ is given by Equation 3. Under the linear social potential proposed in Section 3.2, it then follows that the maximum likelihood estimator for the parameters of $\mathcal{P}$ given ℓ is

$$\left(\hat{\alpha}, \hat{\beta}, \hat{\gamma}, \hat{\delta}\right) = \arg \max_{(\alpha,\beta,\gamma,\delta)} I_{\mathbb{C}}(\ell) \frac{\exp\left(\mathcal{P}\left(\ell \,|\alpha, \beta, \gamma, \delta\right)\right)}{Z\left(\mathcal{P}, \mathbb{C}\right)}. \tag{43}$$

Due to the expense of approximating Z, direct maximization of the likelihood is generally infeasible. Since 3 is a regular exponential family in the case of a linearly separable potential, however, it is a standard result that

$$\mathbf{E}_{(\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{\delta})}\left(t^{\alpha}\left(\ell\right), t^{\beta}\left(\ell\right), t^{\gamma}\left(\ell\right), t^{\delta}\left(\ell\right)\right) = \left(t^{\alpha}_{\text{obs}}, t^{\beta}_{\text{obs}}, t^{\gamma}_{\text{obs}}, t^{\delta}_{\text{obs}}\right), \tag{44}$$

where $\left(t^{\alpha}_{\text{obs}}, t^{\beta}_{\text{obs}}, t^{\gamma}_{\text{obs}}, t^{\delta}_{\text{obs}}\right)$ is the vector of observed sufficient statistics, provided that the MLE exists. It is also the case that the associated MLE is unique, provided that a MLE exists and the sufficient statistics are affinely independent Brown (1986). 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 which gives rise to these values is the MLE. Although slow, this approach can be quite efficacious, and has been successfully employed by Snijders (2002) in the context of exponential random graph families. Alternative methods such as maximum pseudo-likelihood (Besag, 1975; Strauss and Ikeda, 1990) and Markov chain Monte Carlo maximum likelihood estimation (Geyer and Thompson, 1992) may also be employed to the location model, but will not be discussed here. (See, however, an application by Butts (2005) to structural alignment.)

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; see, e.g., 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 (2004) 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 3.5. Simulation studies such as these can thus form the basis for subsequent empirical investigation, without the necessity of adding cumbersome operationalization assumptions.

4.0.1 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 11-20, 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.

To get a sense of the behavior of the location model, we begin by examining how job assignment responds to various effects. Figure 1 depicts simulated draws from the location model under a variety of conditions. For each choice of parameter values, the corresponding panel of Figure 1 shows 250 Metropolis draws from the associated model; these were uniformly thinned from 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 (dark 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 effects for all panels parameterize the strength of association between gender and wage, with positive values reflecting stronger tendency to sort males into high-ranking wage positions. Thus, α acts as a discrimination parameter. (Negative α 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 towards segregation (homogeneity), while $\beta > 0$ indicates a tendency towards desegregation (heterogeneity). Similarly, γ here parameterizes the tendency of couples to hold jobs with similar or different wage rank (a location homogeneity effect). In parallel with β, $\gamma < 0$ implies a tendency towards couple-level homogeneity (similar wages), while $\gamma > 0$ implies a tendency towards couple-level heterogeneity (unequal wages). Zero values for any parameter imply an absence of the corresponding sorting effect. Thus, $(\alpha, \beta, \gamma) = (0, 0, 0)$ results in a null model of uniform assignment.

Examination of the panels of Figure 1 suggests a number of regimes for the occupational sorting model. Panels 2, 3, and 7-10 (counting from the top

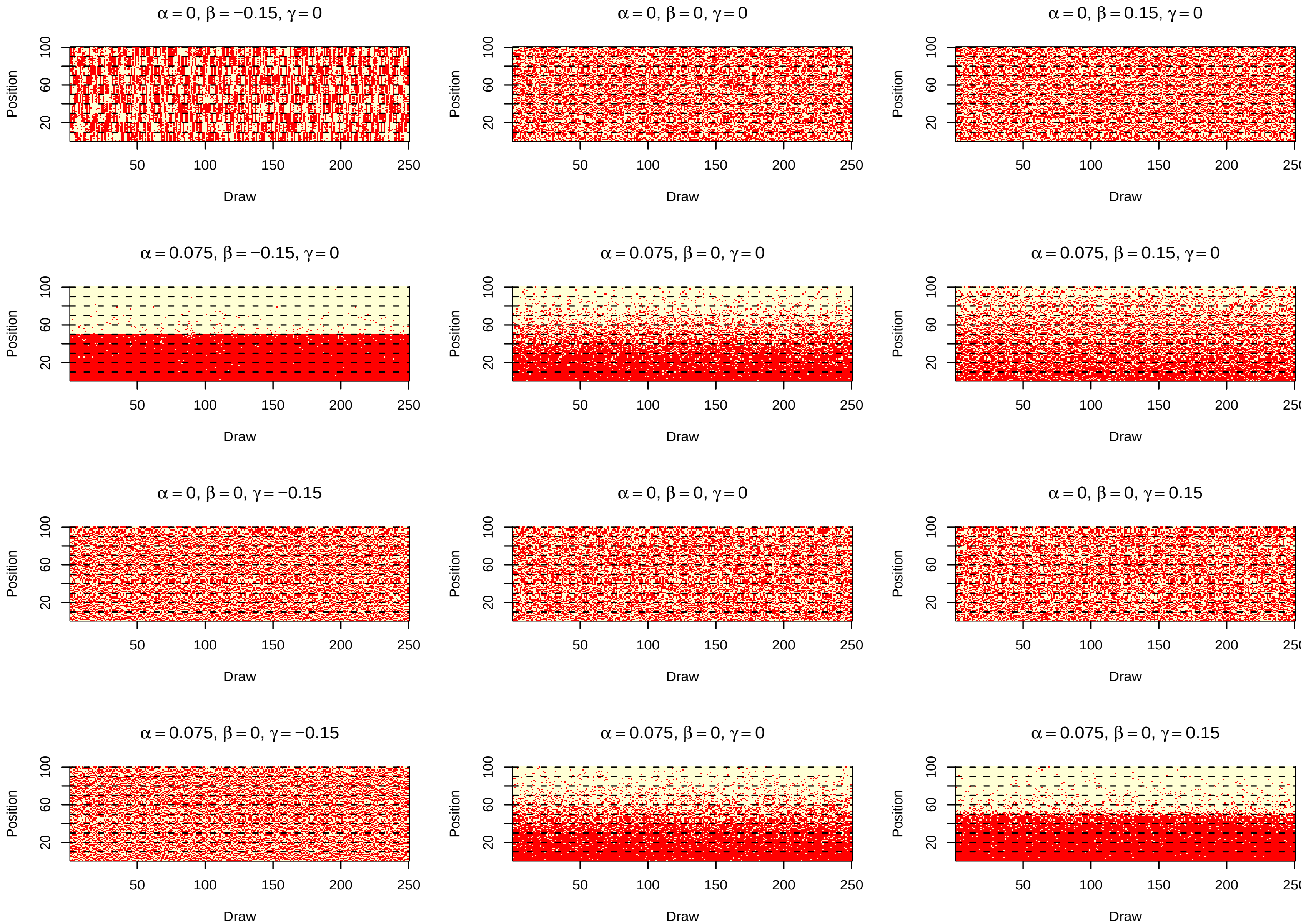

Figure 1: Location model draws, job assignment model

left) display what are effectively random assignments: men and women appear across the spectrum of wage ranks, in essentially even numbers. Panels 4-5 and 11-12 show varying degrees of sorting by gender, with men assigned to higher wage positions (on average) than women. Finally, Panel 1 depicts a "block random" pattern, in which men and women are concentrated into uniform blocks which are otherwise randomly allocated across the wage ordering. What accounts for these patterns? Clearly, sorting by gender is driven by α, and is only observed for samples in which $\alpha > 0$. That said, the extent of inequality clearly depends on other factors. For instance, occupational segregation ($\beta < 0$) tends to lead to a coalescence of men and women into gender-typed occupational categories (Panels 1 and 4). When this tendency is combined with a global discrimination effect, the net result is to "freeze" men into high-ranking blocks and women into low-ranking blocks, thereby exacerbating inequality.[4] By contrast, strong tendencies toward desegregation (Panel 6) interact more weakly with α; while not discernible at the scale of Figure 1, $\beta \gg 0$ ultimately leads to a regime in which the global sorting gradient reproduces itself within each occupational category. While object homogeneity exacerbates inequality, the opposite is clearly true for location heterogeneity (shown by γ). Pressure towards within-couple wage homogeneity clearly tends to place the system within the uniform regime, even under the influence of a sorting gradient (Panel 10). On the other hand, couple-level heterogeneity effects reinforce global inequality (Panel 12) in a fashion which is complementary to that of Panel 4: the heterogeneity effect tends to produce couples with divergent wages, and the α effect then aligns the within-couple wage rank by gender. It should be emphasized that neither occupational segregation nor within-couple heterogeneity creates these effects alone (as is sometimes alleged). Rather, as shown by the location system model, the interactions of these effects act to place the system in regimes which may or may not be globally stratified.

In addition to the general types of configurations found under different assignment regimes, it is useful to consider the quantitative 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, one generally expects workers with greater human capital to

[4]Freezing also occurs for $\beta \ll 0$ in the $\alpha = 0$ case, but the blocks have no net tendency to be ranked by gender. The "institutionalization" of gender-typing at low temperatures/large parameter values has interesting implications for dynamic interpretations of the location system model.

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., $|\alpha|$ small), this attenuation softens the transition from a mixed-wage environment to a "frozen" environment in which wages 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. In the physical analogy, such systems "solidify" at fairly high temperatures ($|1/\alpha|$ large), and are thus difficult to force into an other phases. By contrast, an intervention which attempted 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 that discussed in the context of Figure 1, namely the tendency of gender-typed job categories to be "sorted" by wage in the presence of a background discrimination effect. 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. 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 im-

[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.

[6]All other parameters for these simulation runs are identical to those for Figure 2.

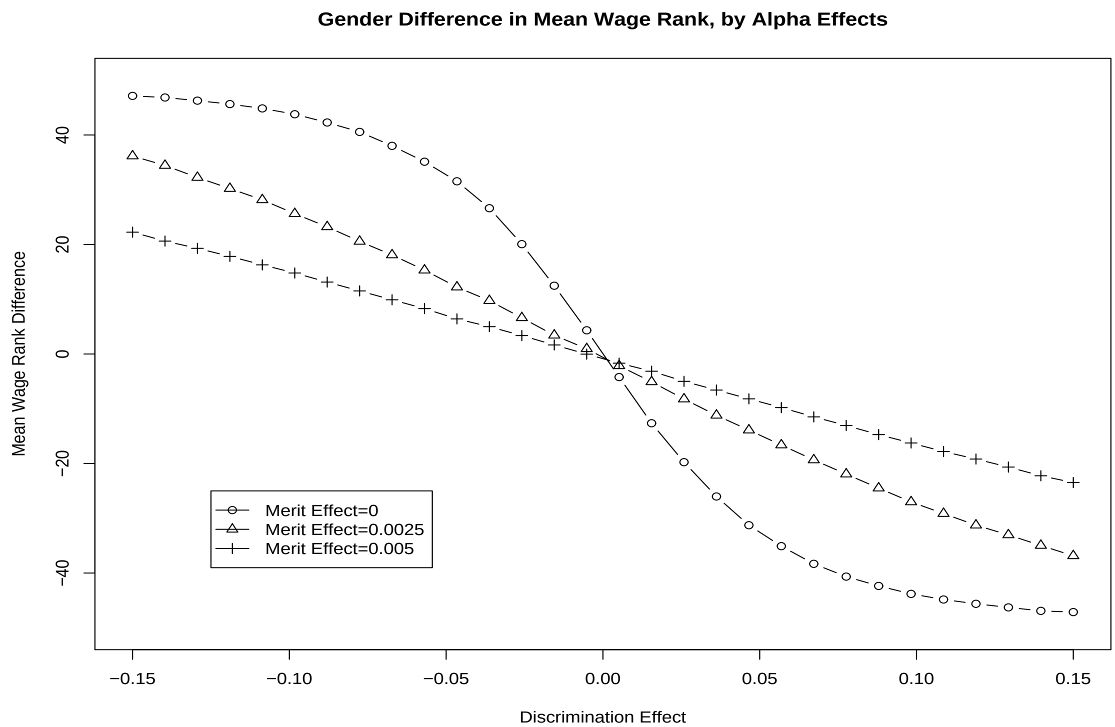

Figure 2: Gender difference in mean wage rank, discrimination and merit effects

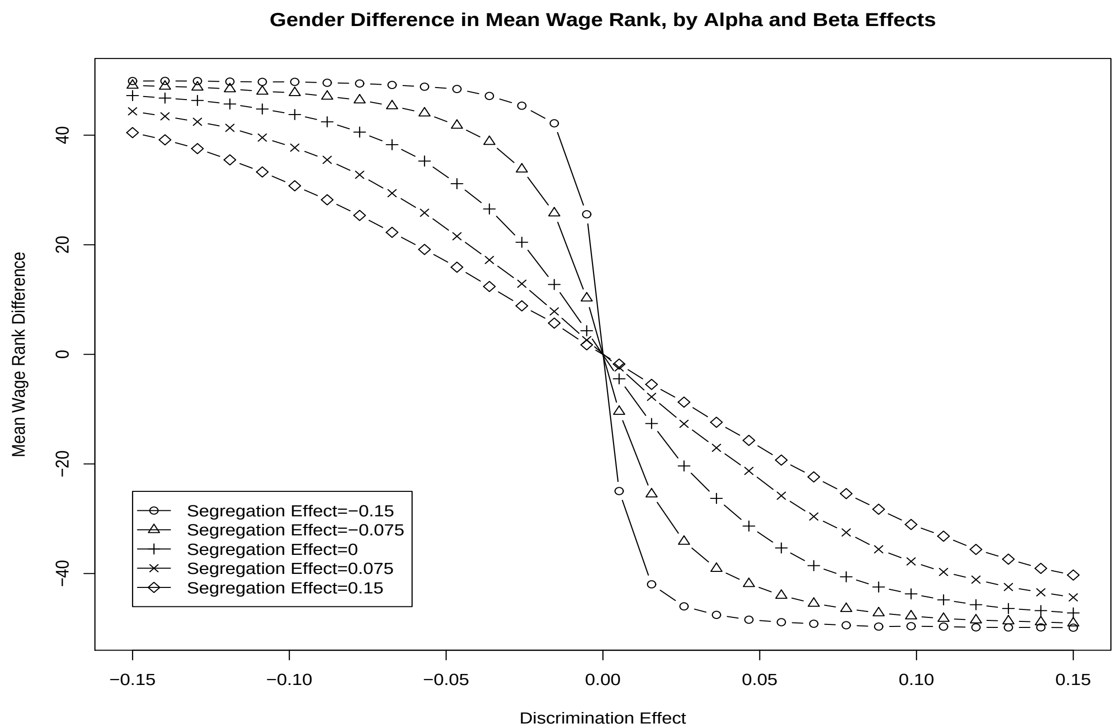

Figure 3: Gender difference in mean wage rank, discrimination and segregation effects

pact 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 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

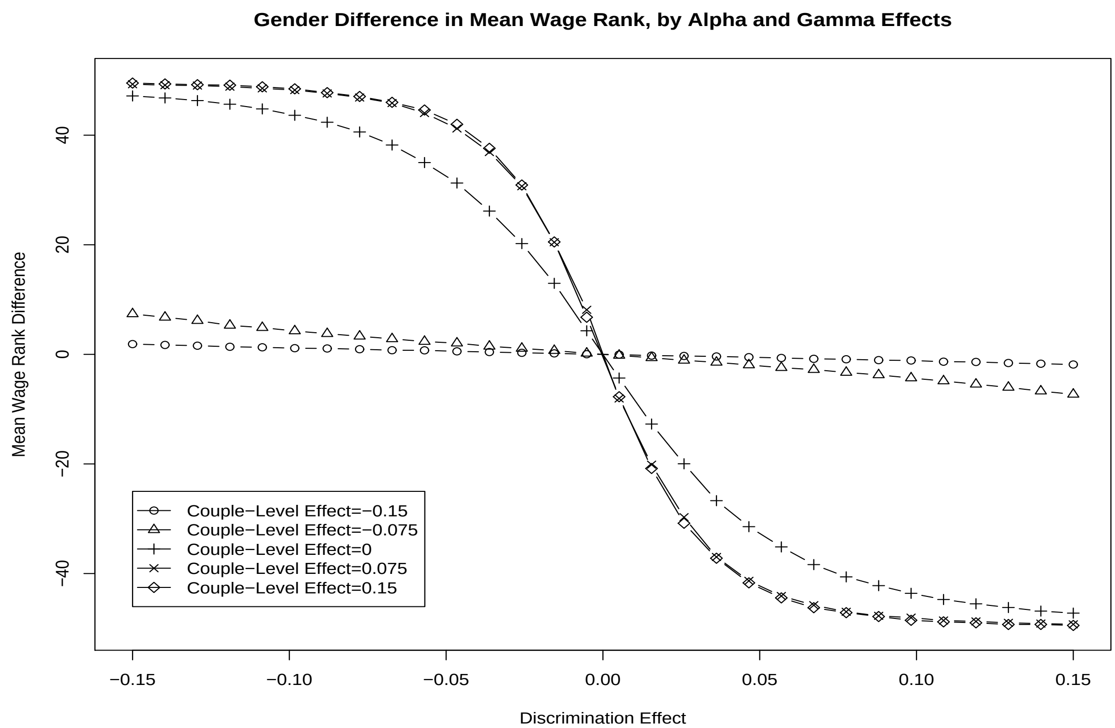

Figure 4: Gender difference in mean wage rank, discrimination and couple-level homogeneity effects

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 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 non-obvious effects. Of particular import is the relative power of couple-level homogeneity effects in suppressing labor market discrimination, a finding which suggests a stronger connection between processes such as mate selection and marital bargaining with macro-level 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.

4.0.2 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, 1992; 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. Regions, for their part, relate to one another via their spatial location. Here, we will make use of both Euclidean distances between regional centroids and Queen's contiguity (for purposes of segregation). Each region is also assigned a location on a "rent" gradient, which scales with the inverse square of centroid distance from the center of the grid.

With these building blocks, a number of mechanisms can be explored. Several examples of configurations resulting from such 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 the first (upper left) panel, a model has been fit with an attraction parameter based on an interaction between rent level and household income ($\alpha = 1e - 4$), balanced by a negative density dependence parameter $\delta = -0.01$. Although the former effect tends to pull all households towards 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 the upper right panel of Figure 5. 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 which 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 which 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. Propinquity is here implemented as an alignment effect between the inter-household network and the Euclidean distance between household locations $\delta = -1$. As one might anticipate, the primary effect of propinquity (shown in the lower-left panel 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.

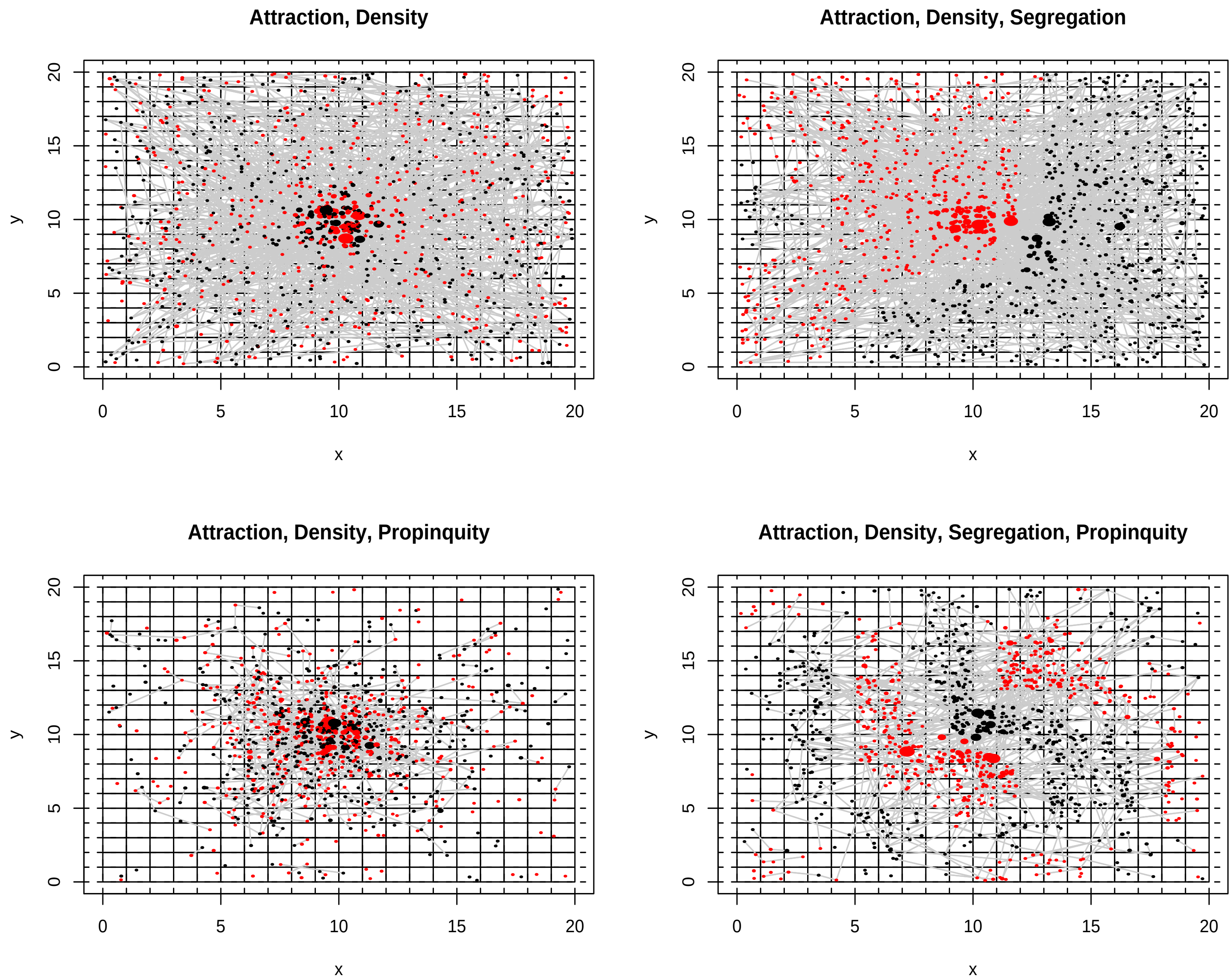

Figure 5: Location model draws, spatial model

Ironically, this situation is somewhat attenuated by the reintroduction of a residential segregation effect (lower-right panel). 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 inter-cluster 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 segregation (β) and propinquity (δ) 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 phase 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, 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 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^{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 propinquity.) As the Figure

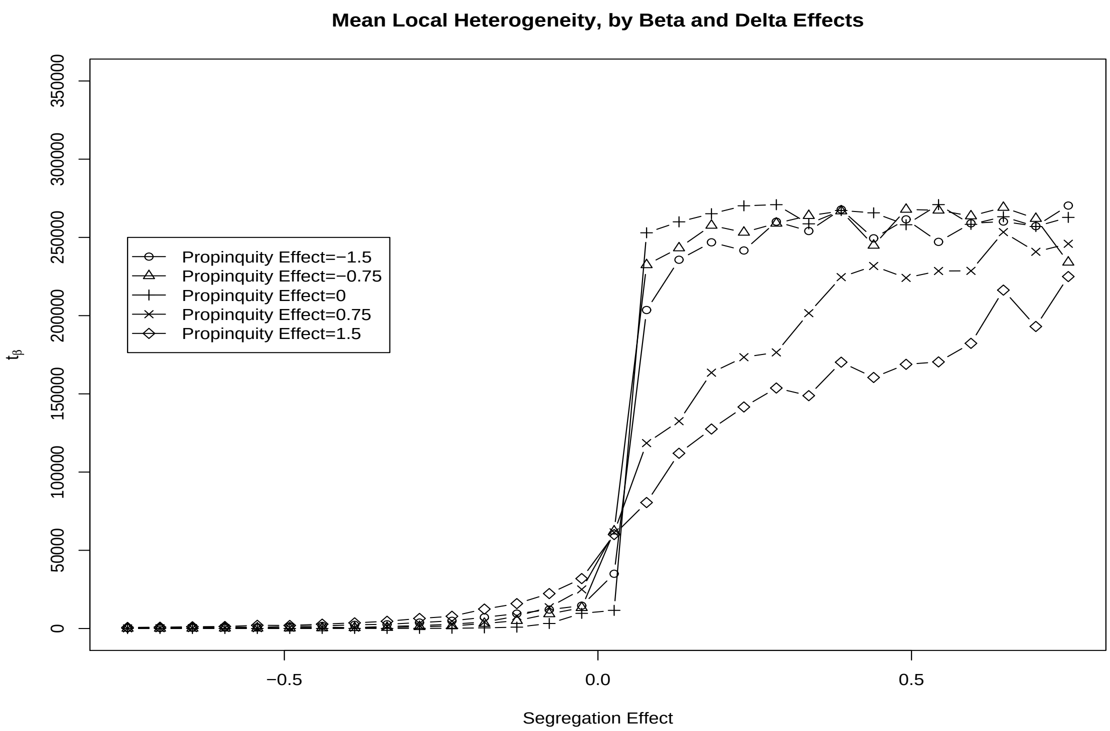

Figure 6: Mean segregation statistic, by segregation and propinquity effects

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. While one 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 phase transition occurs. Such a result is suggestive of the behavior of *binary mixtures* (particularly eutectic mixtures), which can solidify at temperatures which differ greatly from those of their constituents (Kittel and Kroemer, 1980).

As Schelling long ago noted, even mild tendencies towards local segregation can result in residential segregation at larger scales (Schelling, 1969). While the location system model certainly bears this out, the model also suggests that factors such as population density and inter-household ties can interact with segregation in non-trivial 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) which 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.

5 Conclusion

In the foregoing, we have shown a general framework (the generalized location system) which 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 which are well-known in

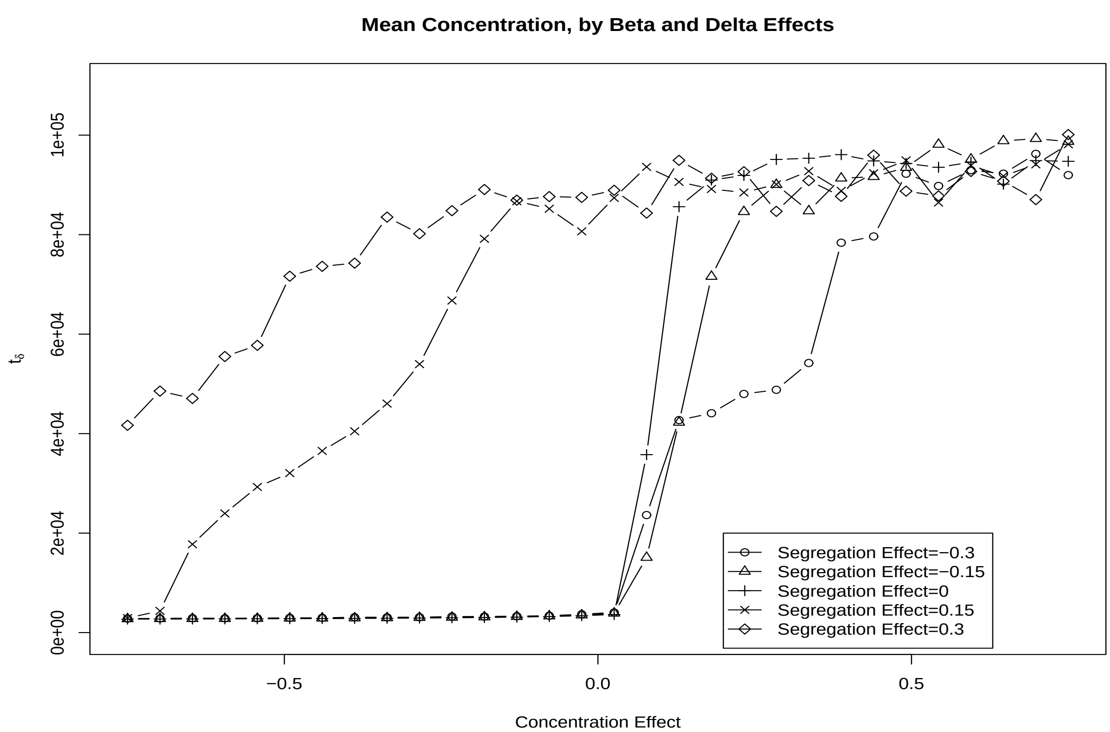

Figure 7: Mean concentration statistic, by density and segregation effects

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. Two illustrative applications (occupational stratification and residential settlement patterns) were presented, and simulation was employed to show the behavior of the location system model in each case. While there are a number of issues which have not been treated here – 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.

6 References

Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory.* John Wiley and Sons, New York.

Becker, G. S. (1991). *A Treatise on the Family.* Harvard University Press, Cambridge, MA, expanded edition.

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems. *Journal of the Royal Statistical Society, Series B*, 36(2):192–236.

Besag, J. (1975). Statistical Analysis of Non-Lattice Data. *The Statistician*, 24(3):179–195.

Bourne, L. (1981). *The Geography of Housing.* V. H. Winston and Sons, New York.

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

Butts, C. T. (2005). Permutation Models for Relational Data. IMBS Technical Report MBS 05-02, Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA.

Calvo-Armengol, A. and Jackson, M. O. (2004). The Effects of Social Networks on Employment and Inequality. *American Economic Review*, 94:426–454.

Cliff, A. D. and Ord, J. K. (1973). *Spatial Autocorrelation*. Pion, London.

Crouch, B., Wasserman, S., and Trachtenburg, F. (1998). Markov Chain Monte Carlo Maximum Likelihood Estimation for $p*$ Social Network Models. Paper presented at the XVIII International Sunbelt Social Network Conference, Sitges, Spain.

Fararo, T. J. (1989). The Spirit of Unification in Sociological Theory. *Sociological Theory*, 7(2):175–190.

Fararo, T. J. and Skvoretz, J. (1987). Unification Research Programs: Integrating Two Structural Theories. *American Journal of Sociology*, 92:1183–1209.

Fienberg, S. E. and Wasserman, S. (1981). Categorical Data Analysis of Single Sociometric Relations. *Sociological Methodology*, 12:156–192.

Fienburg, S. E., Meyer, M. M., and Wasserman, S. (1985). Statistical analysis of multiple sociometric relations. *Journal of the American Statistical Association*, 80:51–67.

Frank, O. and Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81:832–842.

Freidkin, N. (1998). *A Structural Theory of Social Influence*. Cambridge University Press, Cambridge.

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

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian Data Analysis*. Chapman and Hall, London.

Geyer, C. J. and Thompson, E. A. (1992). Constrained Monte Carlo Maximum Likelihood Calculations (with Discussion). *Journal of the Royal Statistical Society, Series C*, 54:657–699.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors (1996). *Markov Chain Monte Carlo in Practice*. Chapman and Hall, London.

Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6):1369–1380.

Handcock, M. S. (2003). Statistical Models for Social Networks: Inference and Degeneracy. In Breiger, R., Carley, K. M., and Pattison, P., editors, *Dynamic Social Network Modeling and Analysis*, pages 229–240. National Academies Press, Washington, DC.

Holland, P. W., Laskey, K. B., and Leinhardt, S. (1983). Stochastic Blockmodels: First Steps. *Social Networks*, 5(2):109–137.

Holland, P. W. and Leinhardt, S. (1981). An Exponential Family of Probability Distributions for Directed Graphs (with Discussion). *Journal of the American Statistical Association*, 76(373):33–50.

Hubert, L. J. (1987). *Assignment Methods in Combinatorial Data Analysis.* Marcel Dekker, New York.

Huffman, M. L. and Cohen, P. N. (2004). Racial wage inequality: Job segregation and devaluation across U.S. labor markets. *American Journal of Sociology*, 109:902–936.

Hunter, D. R. and Handcock, M. S. (2004). Inference in Curved Exponential Family Models for Networks. Technical Report TR0402, Penn State Department of Statistics.

Johansen, S. (1979). *Introduction to the Theory of Regular Exponential Families.* University of Copenhagen, Copenhagen.

Jones, R. and Brayfield, A. (1997). Life's Greatest Joy?: European Attitudes Toward the Centrality of Children. *Social Forces*, 75(4):1239–1269.

Kalos, M. H. and Whitlock, P. A. (1986). *Monte Carlo Methods, Volume I: Basics.* John Wiley and Sons, New York.

Kittel, C. and Kroemer, H. (1980). *Thermal Physics.* W. H. Freeman and Company, New York, second edition.

Massey, D. S. and Denton, N. A. (1992). *American Apartheid: Segregation and the Making of the Underclass.* Harvard University Press, Cambridge, MA.

Mayhew, B. H. (1984a). Baseline models of sociological phenomena. *Journal of Mathematical Sociology*, 9:259–281.

Mayhew, B. H. (1984b). Chance and necessity in sociological theory. *Journal of Mathematical Sociology*, 9:305–339.

Mayhew, B. H., McPherson, J. M., Rotolo, M., and Smith-Lovin, L. (1995). Sex and race homogeneity in naturally occurring groups. *Social Forces*, 74(1).

McPherson, J. M., Smith-Lovin, L., and Cook, J. M. (2001). Birds of a Feather: Homophily in Social Networks. *Annual Review of Sociology*, 27:415–444.

Pattison, P. and Wasserman, S. (1999). Logit models and logistic regressions for social networks: II. multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52:169–193.

Ramsey, N. F. (1956). Thermodynamics and Statistical Mechanics at Negative Absolute Temperature. *Physical Review*, 103(20).

Robins, G. and Pattison, P. (2005). Interdependencies and Social Processes: Dependence Graphs and Generalized Dependence Structures. In Carrington, P. J., Scott, J., and Wasserman, S., editors, *Models and Methods in Social Network Analysis*, chapter 10, pages 192–214. Cambridge University Press, Cambridge.

Robins, G., Pattison, P., and Wasserman, S. (1999). Logit models and logistic regressions for social networks, III. valued relations. *Psychometrika*, 64:371–394.

Robins, G., Pattison, P., and Woolcock, J. (2005). Small and Other Worls: Network Structures from Local Processes. *American Journal of Sociology*, 110(4):894–936.

Robins, G. L. and Pattison, P. (2001). Random Graph Models for Temporal Processes in Social Networks. *Journal of Mathematical Sociology*, 25:5–41.

Schelling, T. C. (1969). Models of segregation. *American Economic Review*, 59(2):483–493.

Snijders, T. A. B. (2002). Markov Chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3(2).

Strauss, D. (1986). On a General Class of Models for Interaction. *SIAM Review*, 28(4):513–527.

Strauss, D. and Ikeda, M. (1990). Pseudolikelihood Estimation for Social Networks. *Journal of the American Statistical Association*, 85(409):204–212.

Swendsen, R. G. and Wang, J. S. (1987). Non-universal critical dynamics in Monte Carlo simulation. *Physical Review Letters*, 58(86).

Wasserman, S. and Pattison, P. (1996). Logit models and logistic regressions for social networks: I. an introduction to Markov graphs and $p*$. *Psychometrika*, 60:401–426.

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

Zhang, J. (2004). A dynamic model of residential segregation. *Journal of Mathematical Sociology*, 28:147–170.