Spatial Models of Large-Scale Interpersonal Networks

Carter T. Butts[1]

Department of Social and Decision Sciences

Center for the Computational Analysis of
Social and Organizational Systems

Carnegie Mellon University

and

Kathleen M. Carley

Department of Social and Decision Sciences

H.J. Heinz III School of Public Policy and Management

Center for the Computational Analysis of
Social and Organizational Systems

Carnegie Mellon University

November 9, 2000

[1]This material is based upon work supported under a National Science Foundation Graduate Fellowship, and was supported in part by the Center for the Computational Analysis of Social and Organizational Systems and the Institute for Complex Engineered Systems at Carnegie Mellon University.

Abstract

Empirical studies of human contact networks suggest a strong relationship between physical distance and frequency of tie occurrence; such past studies have been complemented by theoretical work suggesting the importance of space, technology, available energy, and demographic factors in determining human relations. Here, a general family of stochastic network models is considered which predict tie probability from a distance function defined on an embedding of actors in a social and/or physical space. It is shown that these models lead to a set of stochastic equivalence relations on sets of actors, which can be expressed purely in terms of spatial regions. Several measures of relational macrostructure are also introduced, and their expectations are given for several spatial models. Implications of one model sub-family (a gravity model based on physical distance) for tie distributions within several archetypical population structures are considered, and a variety of macrostructural measures are illustrated.

Keywords: spatial networks, network formation, physical distance, gravity models, stochastic equivalence

1 Introduction

It is a common aim of the field of social network analysis to map out and ultimately predict the web of human relations – friendships, acquaintanceships, family ties, organizational affiliations, and the like – in which human actors are inextricably embedded. Much (if not most) work in this area has thusfar focused either on complete analyses of relatively constrained networks of actors within groups, families, or organizations, or else on the sampling of egocentric networks within populations. While these studies have given us a great deal of insight into the general nature of human interaction (not to mention their more immediate subject matter), relatively little progress has as yet been made in the characterization of large relational networks among spatially distributed populations. With the accelerated development of global communication and transportation networks, however, as well as the recent resurgence of interest in past network macrostructure studies (e.g., the "small world" studies (Milgram, 1967; Travers and Milgram, 1969)) a renewed consideration of this problem seems warranted. In this paper, then, we seek to construct a family of stochastic network models based on spatial layout of actors which are applicable to the analysis of relational networks on large, spatially embedded populations. In addition to exploring some basic properties of these models, we introduce a number of space-related measures of macrostructure which can be employed to aid in the characterization and investigation of these networks. Finally, we demonstrate several of these measures on sample population distributions under a gravity model of tie formation, and present some suggestions for additional investigations using simulation analysis. The result is a general theoretical and methodological framework for the treatment of large, spatially embedded social networks. This framework should be of value both in studying the effects of spatial macrostructure on other phenomena of sociological interest (e.g., migration, information transfer, cultural diffusion, stratification) and in understanding the determinants of macrostructure itself.

1.1 Theoretical Motivation

Human beings are social animals, but animals we are still; despite our technological innovations, we and our activities must nevertheless bow to the constraints of nature. To interact with others we must consume time and energy, both of which are in limited supply. To form and maintain relations, in turn, we must interact with particular alters repeatedly over time. Whether these interactions are in person or via the proxy of a transmitted message, they are still costly[1], and there are limits to our capacity to maintain them. When our interaction partners are physically distant from us – or where our social distance requires us to expend more effort in relating to one another – these costs grow, making relations even harder to establish (and more vulnerable to disruption) than they would be otherwise.

It has been suggested by theorists such as Mayhew (1980; 1981; 1984) that these sorts of fundamental constraints on the structures of human action form the most fruitful basis for generating structural theory. While there may be many and varied effects at the subnetwork or individual level which shape many of the fine details of interaction, the broader patterns – particularly in large, spatially distributed populations – will be severely constrained by physical limitations alone. Furthermore, mathematical constraints on the number and types of structures which *can* exist may play a significant factor in narrowing the universe of possibilities which researchers need consider. Graph size and density, for instance, tightly constrain other graph level measures such as hierarchy, centralization, and connectedness (Anderson et al., 1999); aggregate size and population density likewise have a strong impact on the number and frequency of interactions between individuals (Mayhew and Levinger, 1976). Matching constraints (e.g., in marriage ties) and communication or ritual interaction costs (Mayhew et al., 1995) can interact with demographic aspects (or "Blau dimensions" (Blau, 1977)) of socio-geographical space to affect the baseline distribution of groups, ties, and exposure to specific categories of alters (Schelling, 1969)[2]. With such factors at work in structuring human interaction, then, Mayhew's prescription has much to recommend it.

In addition to the simple constraints of linear distance, limited time and energy for interaction, and the set of possible network structures, spatially embedded interactions are further limited by the basic problem of the sheer immensity of area itself. The fact that a circular area expands with the square of its radius is well known, but its impact can be hard to intuit. Consider, for instance, the question of "covering ground": how large is the physical area that a given person can hope to cover in the course of his or her wanderings? Generally, the answer is surprisingly small. To take a rather extreme case, if one were to drive 24 hours a day for an entire year at 60mph, one would be able to cover an area of only 40 miles in radius (even including more than 25 feet on either side of the vehicle as "covered" territory). Of course, even this is rather generous; none of us spend every moment of our lives driving, much less at highway speeds, and such a scenario is hardly conducive to social interaction[3]!

While the above example may seem somewhat facetious, the basic question is not an idle one. In order to form and maintain relations with others, human beings must generally be coterminous in physical space for some period of time[4] (this tends to be the rule rather than the exception even with respect to communication technologies such as email (Carley and Wendt, 1991) in which participants could in principle meet virtually). If it happens that individuals can cover only a small area of physical space, then their opportunities for forming relations with others – particularly others who are far removed in space – are correspondingly diminished. The fact that this limitation is relatively insensitive to changes in transportation technology (one cannot, after all, communicate with persons effectively if one is shooting past them at high speed[5]) suggests that it is an important invariant to consider when building general theories of social structure at the macro level.

Despite the clear importance of physical distance in structuring patterns of human interaction, spatial effects are often neglected in models of interpersonal networks[6]. Indeed, Nowak et al. (1990) sharply criticize network analysis generally for a failure to realistically consider the importance of spatial factors, suggesting that an alternative structural approach is required[7]. At the macro level, efforts have been made to apply network concepts to the analysis of flows of persons, goods, and vehicles between population centers (e.g., Irwin and Hughes (1992)), but formal efforts to build explicitly spatial models of human interaction networks have thusfar been limited. Similarly, while some attempts have been made to study features of large-scale, spatially embedded human relation networks (e.g., Milgram, 1967; Travers and Milgram, 1969; Killworth and Bernard, 1978; Bernard et al, 1989), relatively few conceptual or methodological tools exist for this purpose. In this paper, then, we shall attempt to respond to the critique of network analysis offered by Nowak et. al. (1990) by developing and exploring the implications of a simple set of stochastic network models which integrate previous empirical findings concerning the effect of distance on human interaction. We shall also demonstrate how various measures of macrostructure may be defined in terms of the concepts here discussed, and will provide formulas by which the predictions of the spatial network models for these measures may be numerically evaluated from population data. Finally, we shall provide an example of the application of several macrostructural measures to sample population layouts under a spatial model, and will demonstrate how visualization may be employed to aid in the perception of macrostructural patterns.

1.2 Empirical Observations

The relationship between distance and interaction in various forms has been empirically studied at least since the 1930s (Bossard, 1932), in large part with an emphasis on migration and transportation (Stewart, 1941; Zipf, 1949; Irwin and Hughes, 1992). By and large, the central finding of this body of work has been the common occurrence of an inverse square law for the effect of physical distance; remarkably, this law seems to hold up for a wide range of specific phenomena (e.g., college attendance (Stewart, 1941), communication and transportation between cities (Zipf, 1949), memorable interactions between persons (Latané et al., 1995)) at multiple scales of observation (e.g., individual interactions versus large-scale flows of persons or messages). Further support for the effect of physical space on social influence and interaction comes from Festinger et al. (1950), whose classic post-World War II study indicated that the (exogenous) placement of couples in apartments affected them and their subsequent interactions on a number of dimensions. Interestingly, the distances considered by Festinger et al. were quite small (under 1000ft), providing yet more evidence for the power of spatial effects even in highly constrained areas.

Following up on earlier studies, Latané et al. (1995) set out to investigate the relationship between distance and interpersonal interaction in widely dispersed human groups. In three separate populations (south Floridians, students in Shanghai, and an international sample of social psychologists), Latané et al. found the number of memorable interactions with alters to be inversely proportional to the median alter distance[8]; more generally, the study strongly suggested an inverse square relationship between physical distance between actors and the probability of interaction. The Latané et al. study is particularly interesting, from our perspective, because it considered populations which varied on both cultural and socio-economic dimensions. Were it the case that access to new communication technologies truly disrupted the classical relationships between space and interaction, one would expect to see those populations with such access (the social psychologists, and, to a lesser extent, the Floridians) evince a different pattern of relations from those without access. In fact, however, all three groups were found to have the same basic interaction pattern, varying in the total expanse and number of alters but *not in the relationship of alter to distance.* This is consistent with the Carley and Wendt (1991) finding that email communications tend primarily to build on and maintain pre-existing ties, and with a general physical distance model whose scale – but not form – varies with cultural and economic factors.

In general, then, it would seem that the theoretical intuition that spatial placement of actors should have powerful structuring properties seems to have been borne out by past empirical research. Such effects appear to exist at multiple levels of analysis, and for many different types of human action. Furthermore, a nearly ubiquitous inverse-square relationship between interaction and distance appears in prior empirical studies of spatial effects; this relationship appears robust to technological, temporal (it has appeared in studies ranging over 50 years), and cultural changes, suggesting that it is likely due to very basic constraints, and that it may well characterize most spatial relationships in the human world. Given these observations, we shall now attempt to construct a simple family of network formation models which build upon our empirical knowledge of how interaction varies across space, and to thereby examine the affects of these findings for the spatial macrostructure of human relations.

1.3 A Generalized Inverse Distance Model

To reiterate our basic goal, we here seek to develop a model of large-scale interpersonal networks which is based on previous findings relating interaction frequency to physical (and possibly social) distance. These empirical observations suggest that the effect of distance on tie formation will be an inverse relationship; all other things being equal, the likelihood that one will share a tie with a far-off alter should be lesser than the likelihood that one will share a tie with a nearby alter. To reproduce this relationship, we shall construct a network formation model in which the presence of ties (or edges) are Bernoulli random variables whose probabilities depend upon the locations of the edges' endpoints. Further, we shall here make the additional assumption that these random variables are independent[9], and we shall consider (at present) the expected properties only of single realizations of the potential edge set; as such, then, this is a model of network *formation* (i.e., it suggests which ties are likely to exist) but not of network *evolution*. Some issues involved in extending this model to the intertemporal case will be briefly mentioned, but we do not here treat this topic in detail.

To proceed, then, let us consider a simple random graph, $G = (V, E)$, and a generalized distance function $\mathbf{D}$. Then $\forall v_i, v_j \in V(G) : v_i \neq v_j$,

$$p\left(e\left\{v_i, v_j\right\} \in E\left(G\right)\right) = \frac{p_b}{1 + \mathbf{D}(v_i, v_j)} \tag{1}$$

where p_b is the "base probability" of a tie between two persons zero units apart[10]. This is the generalized inverse distance model of tie formation: the probability of an edge being contained in the network decreases from a given maximum with some function of distance (which, in this case, could refer to social, temporal, or physical distance) between endpoints.

Obviously, the generalized inverse distance model is appropriately considered to be a model framework, rather than a specific model per se. A wide range of distance types and functions may be employed for $\mathbf{D}$, depending on the particular socio-spatial effects being considered. Given this, we shall now move on to consider a specific implementation of this more general model which incorporates additional insights from previous empirical findings.

1.4 The Gravity Model

As we have seen, several empirical studies of human interaction have strongly suggested that interaction not only falls off with physical distance, but that the probability of interaction is proportional to the inverse square of the distance between partners. Because of both the ubiquity and the simplicity of this result, it deserves to be treated directly: here, then, we define a gravity model of tie probability which is a special case of the generalized inverse distance model.

Let us define the function $d_s(v_i, v_j)$ to be equal to the Euclidean distance[11] between the positions of $v_i, v_j \in V$[12]. In general, we shall assume that the distance referred to by d_s is one of physical space, but this need not be the case; like $\mathbf{D}$, d_s could consist of distance on physical, demographic, or other dimensions. (Regardless of the number or content of the dimensions involved, however, the following model is based on a strictly Euclidean framework.) In terms of the general model, then, this amounts to the following assumption:

$$\mathbf{D}(v_i, v_j) = \alpha d_s(v_i, v_j)^2 \tag{2}$$

Note that α is a scaling factor, indicating units[13] for the growth of d_s. If applying a gravity model fitted on x unit data to data on units of size ax (e.g., when converting from miles to kilometers), one should adjust α as follows:

$$\alpha_{ax} = a^2 \alpha_x \tag{3}$$

(Hence, if the appropriate value of α for 1km units is 2, then the corresponding value for 2km units is 8.) It may also be the case that changes in technology – or in wealth – may result in shifts in the scaling parameter. While α shall be assumed to be constant and identical for all actors in our subsequent analyses, the consequences of heterogeneous α values (as, perhaps, in the presence of stratification) are a potentially interesting area for future research.

To complete our derivation of the gravity model, then, we apply the substitution of equation 2 to the general inverse distance model (equation 1) as follows:

$$p_g\left(e\left\{v_i, v_j\right\} \in E\left(G\right)\right) = \frac{p_b}{1 + \mathbf{D}(v_i, v_j)} \tag{4}$$

$$= \frac{p_b}{1 + \alpha d_s(v_i, v_j)^2} \tag{5}$$

Equation 5 above then provides us with the probability of any given edge's existence; we apply this in the sections which follow.

2 Spatial Networks and Stochastic Equivalence

As we have seen, spatial models of interpersonal networks posit that the probability of a relationship existing between any two actors is a function of the distance between them. Furthermore, as we have also seen, it appears to be the case that tie probability falls off quite rapidly with increases in distance. This suggests the possibility that tight-knit groupings of actors which are widely separated from other groupings of actors may share a certain equivalence with respect to those alters: more specifically, their probabilities of interaction with the distant alters may be (approximately) identical. In this section, then, we shall elaborate on this special form of stochastic equivalence between spatially embedded alters, and will define the conditions under which it can be expected to hold for the gravity model.

The intuition behind the equivalence which we will be considering is perhaps best revealed by a simple thought experiment. Let us consider three actors, Atlas, Hyperion, and Theia, and let us imagine that the distance between Hyperion and Theia is much greater than the minimum distance between Atlas and either of the other actors. If we now imagine drawing Atlas yet further away from the other two, what can we say about the probability of ties among them? Clearly, it is far more likely that a tie exists between Hyperion and Theia than between either of them (individually) and Atlas; but it is of more interest to consider each of these latter probabilities separately. If Atlas were "close" to Hyperion and Theia relative to the distance between them, we would expect that the probability of a tie linking Atlas and Hyperion might be much higher or lower than the probability of a tie linking Atlas and Theia. As Atlas is drawn further away, by contrast, the difference in distances – and hence probabilities – between Atlas and his two alters begins to become extremely small relative to the distance they hold in common. If Atlas is thus arbitrarily far away from Hyperion and Theia, the probabilities of his being tied to either become arbitrarily close: in this way, then, Hyperion and Theia become "equivalent" with respect to Atlas. This basic intuition of relative distance is a very simple one; however, it lies at the very core of spatially induced equivalence relations.

As we have suggested, the form of equivalence in which we are interested is a variant of *stochastic equivalence* (Wasserman and Weaver, 1985; Wasserman and Faust, 1994). Stochastic equivalence is itself a generalization of structural equivalence (Lorrain and White, 1971), and can be understood roughly as the notion that actors are interchangeable to the degree to which they have the same probability of interaction with the same alters (see footnote below for a more rigorous characterization). Stochastic equivalence is particularly useful for purely stochastic models of network formation (such as the $p*$ series of models (Wasserman and Iacobucci, 1986; Wasserman and Faust, 1994) or the biased-net models of Rapoport (1949a; 1949b; 1950)) in which tie probabilities by alter can be easily measured and/or unproblematically grouped. In the event that stochastic equivalence classes can thereby be formed, one can often easily calculate the anticipated distribution of ties between classes; in some cases, this may lend itself more readily to empirical verification than classical block relations, particularly in extremely large networks.

The manner in which we employ the notion of stochastic equivalence here, it should be noted, is somewhat different from its standard usage. While we do intend the term to refer to a relationship between vertices such that each has the same probability of interaction with various alters, we apply the additional specification of *which* alters have this probabilistic isomorphism. (This, as we shall see, is a useful generalization of the more strict definition, particularly when the set of alters for to which the equivalence applies is large.) To state the above more formally:

Definition 1 (Relative Stochastic Equivalence) *In a simple graph G, two vertices, $v_i, v_j \in V(G)$, are said to be* stochastically equivalent with respect to *a vertex set* $\mathbf{v} \subset V(G)$ *iff* $p\left(e\left\{v_i, v_k\right\} \in E\left(G\right)\right) = p\left(e\left\{v_j, v_k\right\} \in E\left(G\right)\right) \forall v_k \in \mathbf{v}$. *The maximum set* $SE_{\mathbf{v}} \subset V(G) : v_i$ *is stochastically equivalent to* v_j *with respect to* $\mathbf{v}$ $\forall v_i, v_j \in SE_{\mathbf{v}}$ *is referred to as the* stochastic equivalence class induced by $\mathbf{v}$.

Thus, there may be sets of vertices within a graph which are stochastically equivalent with respect to various alters, but perhaps not with respect to some other set of alters. In the present context, the sets of vertices we will be considering are those bounded by particular regions in physical space. Because of the properties of the distance model, it will be possible to identify regions which designate sets of vertices that are stochastically equivalent with respect to vertices in other regions. This, in turn, may be of theoretical use in providing a clear rationale for treating spatial aggregates as individual nodes at the macrostructural level. Note also that this definition is somewhat less strict than that specified in Wasserman and Faust (1994)[14] in the additional sense that relative stochastic equivalence is here defined solely in terms of edge probabilities vis a vis alters rather than in terms of interchangeability with respect to all events defined on the set of edge variables. In the case of the probability models considered here, these statements are interchangeable (excepting, of course, our limitation that one consider only alters in the equivalence inducing set), but this may not be true in general[15]; those seeking to apply our notion of relative stochastic equivalence elsewhere may wish to use the stricter form.

2.1 Regions of Equivalence

The observation that actors who are physically proximate will have similar probabilities of interaction with distant actors is not in and of itself exceptionally useful. If, however, it happens that actors are distributed in space in such a fashion that almost all actors belong to small, dense population clusters which are spaced widely apart, then things become more interesting. Drawing on the intuition discussed above, if the population clusters in question are small enough relative to the inter-cluster distances, then it will happen that all actors in each cluster will have approximately the same chance of interacting with any actor from a given other cluster (and vice versa). Essentially, in this case, probability of interaction will be determined by cluster membership, and all members of each physical cluster will form an (approximate) stochastic equivalence class.

Consider, for instance, three hypothetical cities: Thrace, Corinth, and Athens[16]. Each city is densely populated, but extremely small relative to the distances between the cities themselves. (For the present example, we shall also assume that no actors live outside of the cities; this is not critical to our argument, but does simplify the intuition somewhat.) Because the denizens of Thrace are relatively close together, it must be the case that they are all approximately the same distance from the (tightly packed) citizens of Corinth and Athens; hence, each must have approximately the same probability of interacting with each resident of these other cities, and they are therefor stochastically equivalent with respect to these alters. The same, of course, must also hold true for the residents of Corinth and Athens, for identical reasons. In this case, then, it is clear that the borders of our three hypothetical cities delineate three stochastic equivalence classes: it then follows that one may reasonably treat these cities as aggregate nodes in a network of cities, where the strength of macro-ties is equal to the expected number of ties between city dyads (again determined by proximity, and by population). Indeed, this could, in principle, be repeated at the regional level, with clusters of cities identified which are stochastically equivalent with respect to other clusters, and so on. (The usefulness of such recursive groupings may be limited, however[17].)

In practice, of course, even the most severely urbanized societies are unlikely to be perfect. Small towns dot the spaces between large cities, and even in "the middle of nowhere" one is likely to find a habitation of some sort every few miles. That said, this form of regional stochastic equivalence is still applicable, so long as one bears its strict meaning in mind. Thus (to continue our previous hypothetical example) if we add a small town near Thrace it will still be the case that the residents of Thrace will be stochastically equivalent with respect to Corinth, and, indeed, it is likely that the residents of the town will be as well. What may *not* be the case, however, is that the residents of Thrace *and* the town will collectively be stochastically equivalent with respect to the other cities; though the townsfolk have approximately the same tie probabilities with the Corinthians as each other (a condition which also holds among the Thracians), this probability may differ significantly from that of those in Thrace. Likewise, the Thracians may differ significantly in their probabilities of having ties with persons in the nearby town, depending on where they (and the relevant townsfolk) are located, thus implying that the Thracians may not be stochastically equivalent with respect to the townsfolk. None of this is problematic per se (especially if we are not interested in the smaller towns), but it does bear consideration, particularly when identifying stochastic equivalence classes. If it is not possible to partition the population into a single set of classes such that each also induces equivalence among the other classes, then it is important to recognize this fact, and to work only with specific regional equivalences.

2.2 Stochastic Equivalence for the Gravity Model

Let us now reframe the above in a more formal fashion, using the specific formulation of the gravity model[18]. Assume that there exist two vertices, v_i and v_j, such that $d_s(v_i, v_j) = \delta$. Now, let Δ be the minimum of $d_s(v_i, v_k)$ and $d_s(v_j, v_k)$ for some third vertex, v_k. If we assume that the distances $d_s(v_i, v_k)$ and $d_s(v_j, v_k)$ are maximally different[19] (i.e., v_i, v_j, and v_k lie along a straight line with v_k as an endpoint) and let v_i be the more distant of the two vertices from v_k, then

$$p_g\left(e\left\{v_i, v_k\right\} \in E\left(G\right)\right) = \frac{p_b}{1+\alpha d_s(v_i, v_k)^2} \tag{6}$$

$$= \frac{p_b}{1+\alpha(\delta+\Delta)^2} \tag{7}$$

and

$$p_g\left(e\left\{v_j, v_k\right\} \in E\left(G\right)\right) = \frac{p_b}{1+\alpha d_s(v_j, v_k)^2} \tag{8}$$

$$= \frac{p_b}{1+\alpha\Delta^2} \tag{9}$$

Now, clearly $p_g(e\{v_i,v_k\}\in E(G)) \to p_g(e\{v_j,v_k\}\in E(G))$ as $\delta \to 0$ [20]. At this point, obviously, v_i and v_j would have equal tie probabilities for *any* v_k, and would be truly stochastically equivalent. Of course, this equivalence would be a fairly trivial one, since it would require v_i and v_j to reside at identical points in space; surely, this would not be useful for most purposes!

The above, however, does provide us with an important clue towards a more general argument. What if, instead of $\delta \to 0$, it is the case that $\Delta \gg \delta$? If this is so, then we would expect Δ's effect on tie probability to dominate, and hence for v_i and v_j to be *approximately* stochastically equivalent with respect to all v_k such that Δ is sufficiently large.

For a more precise criterion, we might say that if $p_g(e\{v_i,v_k\}\in E(G))$ and $p_g(e\{v_j,v_k\}\in E(G))$ are similar enough that the difference between them is small compared to their own magnitudes, then the two are approximately stochastically equivalent with respect to v_k. In particular, let $\gamma = \frac{p_g(e\{v_i,v_k\}\in E(G))}{p_g(e\{v_j,v_k\}\in E(G))}$. If this is sufficiently close to 1 (say, $\pm 5\%$) then v_i is approximately equivalent to v_j with respect to v_k [21]. Thus,

$$\gamma(v_i,v_j,v_k) = \frac{p_g(e\{v_i,v_k\}\in E(G))}{p_g(e\{v_j,v_k\}\in E(G))} \tag{10}$$

$$= \frac{p_b}{1+\alpha(\delta+\Delta)^2}\frac{1+\alpha\Delta^2}{p_b} \tag{11}$$

$$= \frac{1+\alpha\Delta^2}{1+\alpha(\delta+\Delta)^2} \tag{12}$$

Recalling that $\delta \geq 0$, we can define a "tolerance threshold", $\tau < 1$, such that $\tau \leq \gamma(v_i,v_j,v_k)$, and then find the minimum value of Δ (given δ and α) needed to satisfy τ. Alternately, if Δ is taken as fixed, we can find the maximum δ such that the above is satisfied. (This last amounts to asking how close v_i and v_j would have to be, in the worst case, to be approximately equivalent with respect to v_k at level τ.) Solving for this maximum v_i,v_j distance in terms of the other relevant variables, then, gives us

$$\delta_\tau = -\frac{\tau\alpha\Delta \mp \sqrt{-\tau\alpha(\tau-1-\alpha\Delta^2)}}{\tau\alpha} \tag{13}$$

Of the two solutions, only the second is meaningful. Evaluating this expression for a range of α, Δ, and τ values gives us the values of Table 1. These thresholds give us some basic intuition about the sorts of relative distances required for approximate equivalence, and gives us some additional hints as to the ways in which the model parameters affect δ_τ. First of all, as can be seen, the ratios involved are non-degenerate (and therefore interesting); for instance, a distance of approximately 50km[22] or less between two actors would be sufficient to ensure that their respective probabilities of interaction with an alter 1000km away would differ by no more than 1%. As this is considerably greater than the diameter of most cities, such a distance is certainly large enough to encompass a large number of actors (and 1000km is well under the distance between many US population centers). Secondly, it is clearly the case that δ_τ is relatively insensitive to the scaling parameter, α: changing α by a factor of two away from 1 in either direction does not appear to have a strong impact on δ_τ values[23]. This is somewhat less true for very small Δ values, but even in these cases one does not see extremely substantial effects.

(Insert Table 1 Here)

The third, and perhaps most interesting, aspect of the above is that it can be quite easily – and fairly accurately – approximated. Inspection of the computed values reveals that $\delta_\tau = \frac{1}{2}\Delta(1-\tau)$ provides excellent estimates for $\Delta \geq 10$, though here again we note some anomalies for very small Δs. (These are in part due to the aforementioned effect of large Δ values in damping scaling effects, as well as to nonlinearities which become more significant when Δ is small.) With this expression, we can cast the relationship in yet another way: for some τ (say, 0.95) we can think of this as giving the fraction of Δ which δ_τ must be to satisfy the equivalence requirement (in this case, 0.025). While not absolutely perfect, this rule of thumb is quite reasonable for modest to large values of Δ. Similarly, one can also reverse the rule to find an approximate lower bound on Δ given a particular δ_τ (from the above, $\Delta = \frac{2\delta_\tau}{1-\tau}$, or $40\delta_\tau$ in the present case).

Having established a general relationship between the distance between two actors and the minimum radius[24] such that the two will be stochastically equivalent (to a given degree) to any alter at or beyond it, and having reduced this relationship to a fairly simple linear approximation, we have laid the groundwork which is required to examine spatially induced stochastic equivalence under the gravity model. Because equivalence is here a matter of regions in space rather than individual actors, we do not require data at the individual level to make predictions regarding stochastic equivalence classes among spatially-embedded populations. All that is needed is to identify clusters of actors within some δ_τ: all actors within such clusters will be stochastically equivalent with respect to all actors outside the corresponding threshold radius Δ. Testing this hypothesis merely requires sampling actors from within the clusters and examining their relations with alters beyond the radius; spatial subgroups drawn from within a given cluster should have the same number of ties, on average, to the same (Δ or greater distance) locations.

3 Spatial Measures of Macrostructure

As we have seen, networks of relations on spatially embedded populations may be highly structured, and in some cases these networks may even contain equivalences at the macro level which permit reduction and analysis at the level of population centers. In other cases – or when considering smaller geographic regions – it may be necessary to look to alternative approaches for the measurement of macrostructure. As has been mentioned, the sheer size of even a modest regional population prohibits use of traditional techniques on such networks; indeed, even if such were possible, it is uncertain whether or not this would be theoretically apt. When measuring macrostructures, we are no more interested in the properties of individual positions than a survey analyst is interested in the attitudes of a particular actor. Similarly, spatial macrostructure lacks the properties of boundedness which make many traditional graph-level measures useful in other areas of research (e.g., organizational analysis or group behavior). To analyze spatial macrostructure, we seek measures which allow us to characterize the features of such structure across space and population, which relate to fundamental macrostructural processes such as cultural diffusion and migration, and which can be readily measured using the standard instruments of population analysis (in particular, sampling techniques). In this section, then, we present a variety of such measures, explaining how each can be computed and providing specific formulations for their expectations under the gravity model of spatially embedded interpersonal networks.

3.1 Spatial Subgraph Measures

When examining populations in space, one of our first questions is nearly always "how many people live in any given region?". Though a simple concept, population density implies much about any number of social, environmental, and economic processes; it is an essential building block of spatial population analysis. When we turn to the consideration of macrostructure, we may seek to ask a similar sort of question: "how many ties exist among persons living in a given region?". Knowing the answer to this question gives us a sense of the social density of a particular area, provides insight into the strength of "local" influences on residents, and allows us to estimate the number of conduits within which information regarding matters of local interest may readily flow. Our first measures of macrostructure, then, concern the cardinality of the edge sets of subgraphs which are induced by choosing actors from particular regions of physical or social space.

The basic approach which will be pursued here is one of treating dyads as independent Bernoulli random variables (as per the model described in equations 1 and 5) which are aggregated to form various measures of expected tie frequency. The fundamental method of selecting dyads which will be employed is spatial: we shall consider sets of vertices by reference to the areas in space in which they reside. With the proper assumptions regarding population distribution, this allows us to transform problems of counting dyads into integrals[25] over regions in space; these easily lend themselves to numerical (and, in some cases, symbolic) evaluation using a variety of well-known techniques.

To characterize populations of persons (or, for our purposes, vertices) in space, we obviously cannot hope to keep track of each and every individual (nor would we wish to do so, given that our objective is not to study individual network positions). Instead, we employ a *population density function* which, for any given region in space, returns the number of persons "residing" there. It is important to note that, for our present purposes, we shall treat any given individual as being associated with exactly one spatial position, and we shall treat these positions as both exogenous and fixed. This should not, however, be interpreted as requiring that actors are necessarily immobile per se; rather, we assume that there exists some (knowable) position associated with each actor such that his or her tie probabilities can be determined from this position (and those of other actors)[26]. While we shall generally presume that the position in question corresponds to place of residence, alternatives such as the centroid of actor positions across a given time interval could easily be substituted without doing violence to the structure of the model. Given this, we define the population density function as follows:

Definition 2 (Population Density Function) *Let $P_{\mathbf{A}}$ represent the number of persons residing within a given spatial region $\mathbf{A}$. A function $f_p(\mathbf{v})$ is referred to as a* population density function *if $\int_{\mathbf{A}} f_p(\mathbf{v})d\mathbf{v} = P_{\mathbf{A}}\ \forall \mathbf{A}$.*

Again, we emphasize that $P_{\mathbf{A}}$ in the above definition will be considered exogenous and fixed for purposes of the analyses which follow. Generalization of the above to dynamic or stochastic population models is certainly possible, but would raise various additional issues which would complicate our analyses without adding a great deal of insight[27]. (One simple exception to this, however, is the case in which populations are deterministically dynamic with a relaxation time much greater than that of the implicit network formation process. In this particular case, these results may be applied directly by "slaving" them to the population change process.)

As we have already seen, our assignment of nodes to spatial locations implies that any set of points (or area more generally) implies a node set in the larger network; hence, we can unproblematically refer to induced subgraphs of the total graph (denoted by $G[A]$, where $A \subset V(G)$) in terms of regions within physical space. Let us define the *internal tie volume* of a given area $\mathbf{A}$ (denoted $\mathcal{V}^i(\mathbf{A})$) to be the cardinality of the edge set of the subgraph induced by $\mathbf{A}$. It follows, then, that the expectation of $\mathcal{V}^i$ must be:

$$\begin{aligned} \mathbf{E}\left(\mathcal{V}^i(\mathbf{A})\right) &= \mathbf{E}\left(\left|E\left(G[\mathbf{A}]\right)\right|\right) && (14) \\ &= \frac{1}{2}\int_{\mathbf{A}}\int_{\mathbf{A}} f_p(\mathbf{v_1})f_p(\mathbf{v_2})p\left(\{\mathbf{v_1},\mathbf{v_2}\} \in E(G)\right)d\mathbf{v_1}d\mathbf{v_2} && (15) \end{aligned}$$

(where $\mathbf{v_1}$ and $\mathbf{v_2}$ are vectors in $\mathbf{A}$). Under the spatial model, we make the appropriate substitution for the probability of an edge's existence; the result is simply

$$\begin{aligned} \mathbf{E}\left(\mathcal{V}^i(\mathbf{A})\right) &= \mathbf{E}\left(\left|E\left(G[\mathbf{A}]\right)\right|\right) && (16) \\ &= \frac{1}{2}\int_{\mathbf{A}}\int_{\mathbf{A}} f_p(\mathbf{v_1})f_p(\mathbf{v_2})\frac{p_b}{1+\mathbf{D}(\mathbf{v_1},\mathbf{v_2})}d\mathbf{v_1}d\mathbf{v_2} && (17) \end{aligned}$$

The same procedure holds for the gravity model, giving us

$$\begin{aligned} \mathbf{E}\left(\mathcal{V}^i(\mathbf{A})\right) &= \mathbf{E}\left(\left|E\left(G[\mathbf{A}]\right)\right|\right) && (18) \\ &= \frac{1}{2}\int_{\mathbf{A}}\int_{\mathbf{A}} f_p(\mathbf{v_1})f_p(\mathbf{v_2})\frac{p_b}{1+\alpha d_s(\mathbf{v_1},\mathbf{v_2})^2}d\mathbf{v_1}d\mathbf{v_2} && (19) \end{aligned}$$

Observe that the above amounts to nothing more than summing the expectations of all possible dyads within the subgraph, and then dividing by two (to reflect the fact that the double integral counts each dyad twice). In general, this is a fairly straightforward operation, which may be undertaken numerically for any arbitrary region given a suitable population density function. Note also that the empirical value of this measure can be estimated by surveying a random sample of persons within a given area using name generation techniques and employing the reported numbers of local alters to derive the area's internal tie volume. This measure of macrostructure, then, is well-suited both to theoretical and to empirical study[28].

Often, we may be interested in the tie volume within a particular space (e.g., within a given city, region, or nationality) for theoretical and/or empirical reasons. In such instances, the above measures may be unproblematically deployed via standard numerical estimation procedures, and can be useful in addressing a variety of questions of interest. In some cases, however, we may not have a particular region in mind a priori; indeed, we may be less interested in the value of local tie volume within a given area than in the way in which this volume *varies* as one moves across a region of interest. One means of approaching this problem is to describe the local tie volumes within arbitrarily small regions about particular points: these volumes can then be estimated for a grid of sample points over the area of interest, providing a sense of the more general macrostructure.

To accomplish this analysis, we take the limit of the expectation of $\mathcal{V}^i(\mathbf{A})$ as $\mathbf{A}$ approaches a single vector. This, of course, is only possible if such a limit exists; and, further, it is only useful to the degree that this limit is unique[29]. If the population density function being employed is not suitably well-behaved, this will not be possible, or may yield different expected flow volumes depending on how the limit is taken. Taking f_p to be a smooth curve will generally solve this problem; some fitting may be required, then, prior to computing this measure on empirical population data.

The general expression of the expectation of the instantaneous internal tie volume, then, is

$$\mathbf{E}\left(\mathcal{V}^i(\mathbf{v})\right) = \mathbf{E}\left(\left|E\left(G[\mathbf{v}]\right)\right|\right) \tag{20}$$

$$= \lim_{\mathbf{A}\to\mathbf{v}} \int_{\mathbf{A}} \int_{\mathbf{A}} f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) p\left(\{\mathbf{v_1}, \mathbf{v_2}\} \in E(G)\right) d\mathbf{v_1} d\mathbf{v_2} \tag{21}$$

(This may be reformulated in terms of the specific spatial network models considered here using the above substitutions.)

Another fairly direct application of the internal tie volume measure is the internal tie volume *per actor*. This, which is simply the average degree of a vertex in the spatially induced subgraph, is a useful practical measure of (local) social density, and may be compared to empirical results from the study of egocentric networks. Formally, the measure's expectation is given by

$$\mathbf{E}\left(\frac{\mathcal{V}^i(\mathbf{A})}{P_{\mathbf{A}}}\right) = \frac{\mathbf{E}\left(\left|E\left(G[\mathbf{A}]\right)\right|\right)}{P_{\mathbf{A}}} \tag{22}$$

$$= \frac{\int_{\mathbf{A}} \int_{\mathbf{A}} f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) p\left(\{\mathbf{v_1}, \mathbf{v_2}\} \in E(G)\right) d\mathbf{v_1} d\mathbf{v_2}}{\int_{\mathbf{A}} f_p(\mathbf{v}) d\mathbf{v}} \tag{23}$$

(The fact that population is not itself a random variable allows us to simply divide directly; if this assumption were relaxed, the analysis would be rather more complicated.) Here again, we may substitute the specific probability model as required, using equations 1 and 5.

3.2 Spatial Cut Measures

In the above discussion, we considered subgraphs induced by particular regions with the objective of determining the cardinality of their edge sets (or some statistic thereof). This was motivated by a theoretical concern with being able to assess the number of ties present among actors residing in a particular area; another, complementary question is that of the number of ties between such a set of actors and alters who are not contained within the set. Mathematically, this is a question of cutsets (or cuts): the number of ties between persons residing within a given region and those beyond it is the cardinality of the cut from the set of nodes corresponding to actors within the given region and its complement. The macrostructural measures which follow, then, are concerned with various aspects of the sizes of spatially induced cuts (which, in turn, inform us about the degree to which actors in a given region are exposed to external influences).

Our first spatial cut measure is a simple analog to the expected internal tie volume which we considered previously. Given the cutset on the set of vertices corresponding to the area $\mathbf{A}$ and its complement (denoted $G[\mathbf{A},\bar{\mathbf{A}}] \subset E(G)$), we define the *external tie volume* $\mathcal{V}^e(\mathbf{A})$ to be the cardinality of $G[\mathbf{A},\bar{\mathbf{A}}]$. The expectation of this measure, then, may be straightforwardly expressed as a sum of the expectations of tie variables in the usual manner:

$$\begin{aligned}
\mathbf{E}\left(\mathcal{V}^e(\mathbf{A})\right) &= \mathbf{E}\left(\left|G[\mathbf{A},\bar{\mathbf{A}}]\right|\right) && (24)\\
&= \int_{\mathbf{A}}\int_{\bar{\mathbf{A}}} f_p(\mathbf{v_1})f_p(\mathbf{v_2})p\left(\{\mathbf{v_1},\mathbf{v_2}\}\in E(G)\right)d\mathbf{v_1}d\mathbf{v_2} && (25)
\end{aligned}$$

Of course, under the general distance model we are able to substitute the basic probability model into the above, giving us

$$\begin{aligned}
\mathbf{E}\left(\mathcal{V}^e(\mathbf{A})\right) &= \mathbf{E}\left(\left|G[\mathbf{A},\bar{\mathbf{A}}]\right|\right) && (26)\\
&= \int_{\mathbf{A}}\int_{\bar{\mathbf{A}}} f_p(\mathbf{v_1})f_p(\mathbf{v_2})\frac{p_b}{1+\mathbf{D}(\mathbf{v_1},\mathbf{v_2})}d\mathbf{v_1}d\mathbf{v_2} && (27)
\end{aligned}$$

The same can be performed for the gravity model, as follows:

$$\begin{aligned}
\mathbf{E}\left(\mathcal{V}^e(\mathbf{A})\right) &= \mathbf{E}\left(\left|G[\mathbf{A},\bar{\mathbf{A}}]\right|\right) && (28)\\
&= \int_{\mathbf{A}}\int_{\bar{\mathbf{A}}} f_p(\mathbf{v_1})f_p(\mathbf{v_2})\frac{p_b}{1+\alpha d_s(\mathbf{v_1},\mathbf{v_2})^2}d\mathbf{v_1}d\mathbf{v_2} && (29)
\end{aligned}$$

Note that for population density functions which are sufficiently poorly behaved, the above may not be finite for infinite spaces. While this is clearly not a problem for empirical research (the world being a decidedly finite place), an exceptionally carelessly designed simulation could potentially be distorted by "edge effects" if populations tend to grow rapidly near the bounds of the simulated geography.

Generally, our interest in tie flow across cuts is likely to be motivated by interest in particular regions, population centers, and the like. As with the case of internal tie volume, however, it may be useful in some cases to understand how external tie volume changes across space; this may be facilitated by an instantaneous external tie volume measure. As before, we derive this measure by taking a given area and observing the limit of the expectation of the external tie volume as the area approaches a single point in space. Obviously (as in the case of instantaneous internal tie volume) this measure is an abstraction, assuming a continuous, well-behaved space of population. Nevertheless, it may prove useful in mapping out macrostructural properties, and in identifying areas in which conditions promote an influx of external ties. Following a similar argument to that applied to expected instantaneous internal tie volume, then, we find that the expected instantaneous external tie volume is given by:

$$\mathbf{E}\left(\left|E\left(G[\mathbf{v},\mathbf{S}-\mathbf{v}]\right)\right|\right) = \lim_{\mathbf{A}\to\mathbf{v}} \int_{\mathbf{A}}\int_{\bar{\mathbf{A}}} f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) p\left(\{\mathbf{v_1},\mathbf{v_2}\} \in E(G)\right) d\mathbf{v_1} d\mathbf{v_2} \tag{30}$$

$$= \int_{\mathbf{S}-\mathbf{v}} f_p(\mathbf{v}) f_p(\mathbf{v_1}) p\left(\{\mathbf{v},\mathbf{v_1}\} \in E(G)\right) d\mathbf{v_1} \tag{31}$$

This expression is easily adapted to the general spatial model, and to the gravity model, by substituting the assumed expression for $p\left(\{\mathbf{v_1},\mathbf{v_2}\} \in E(G)\right)$. In practice, rather than actually allowing $\mathbf{A}$ to approach a single point in space, it may be useful to instead consider $\mathcal{V}^e$ for small (but not infinitesimal) cells about a grid of sample points. This approach yields a roughly similar interpretation, and is more likely to be computationally feasible than a strict limit method. (It is also clearly the obvious choice when working with population data which has not been smoothed.)

Just as the instantaneous internal tie volume measure has its external counterpart, so too does the normalized variant. As with internal tie volume, it is a relatively simple matter to identify the external tie volume per capita by simply dividing through by the population within the given area:

$$\mathbf{E}\left(\frac{\mathcal{V}^e(\mathbf{A})}{P_{\mathbf{A}}}\right) = \frac{\mathbf{E}\left(\left|G[\mathbf{A},\bar{\mathbf{A}}]\right|\right)}{P_{\mathbf{A}}} \tag{32}$$

$$= \frac{\int_{\mathbf{A}}\int_{\bar{\mathbf{A}}} f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) p\left(\{\mathbf{v_1},\mathbf{v_2}\} \in E(G)\right) d\mathbf{v_1} d\mathbf{v_2}}{\int_{\mathbf{A}} f_p(\mathbf{v}) d\mathbf{v}} \tag{33}$$

Note that the units of this measure are in ties per person; this is not directly interpretable as a classical density measure. This last is due to the fact that the number of *possible* ties is $P_{\mathbf{A}} P_{\bar{\mathbf{A}}}$, the second factor of which may be extremely large (and which is inversely related to $P_{\mathbf{A}}$, in any event). This measure is intended for use in assessing the total exposure of persons in a given area to external alters...it does not provide information on the degree to which this exposure approaches its theoretical maximum.

3.3 Tie Flow Across Cuts

The families of measures which we have considered thusfar are concerned with counting the expected number of edges in various sets of interest. In some cases, however, we are not only interested in *how many* ties of one sort or another are present, but *where they are going.* Consider taking a hypothetical sample of small communities surrounding a large city. Evaluating this sample, one might find that each of these communities possesses a large external tie volume; clearly, many persons in these communities have ties elsewhere, but *where* exactly? In such a situation, one might suspect that many of these ties lead to the city...but the tie volumes are mute on this point. What is needed for this sort of analysis is a measure not of the *number* of ties leaving a given area, but of the *direction* in which ties from that area tend to go. It is to just such a measure that we now turn.

To see the intuition behind the notion of directionality in our analysis of tie distribution, it is somewhat helpful to think in terms of fluid flow. Consider a relatively complicated, turbulent flow of water under high pressure; as it rushes by, we pick a static region and begin to examine the flow across it. While water might flow in and out at various points (and with various velocities), we can easily imagine characterizing the net direction of flow in terms of a vector which describes the mean direction and velocity of flow crossing the boundary of the region. Clearly, "flows" of ties do not behave quite like flows of water[30], but we may employ this same basic approach to assess the average direction in which ties from persons in a given area are heading.

Let us now proceed to use the above intuition to derive a measure of directional tie flow. Given an area $\mathbf{A}$, let us define the directional external tie volume (which we shall denote $\vec{\mathcal{V}}^e(\mathbf{A})$) as the sum of the direction of the vectors formed by the difference between the endpoints of all edges contained in $G[\mathbf{A}, \bar{\mathbf{A}}]$. Applying our usual procedure, then, we can ascertain that the expected value of $\vec{\mathcal{V}}^e(\mathbf{A})$ is as follows:

$$\mathbf{E}\left(\vec{\mathcal{V}}^e(\mathbf{A})\right) = \int_{\mathbf{A}} \int_{\bar{\mathbf{A}}} \frac{\mathbf{v_2} - \mathbf{v_1}}{\|\mathbf{v_2} - \mathbf{v_1}\|} f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) p\left(\{\mathbf{v_1}, \mathbf{v_2}\} \in E(G)\right) d\mathbf{v_1} d\mathbf{v_2} \tag{34}$$

In terms of the generalized distance model, this is simply

$$\mathbf{E}\left(\vec{\mathcal{V}}^e(\mathbf{A})\right) = \int_{\mathbf{A}}\int_{\bar{\mathbf{A}}} \frac{\mathbf{v_2}-\mathbf{v_1}}{\|\mathbf{v_2}-\mathbf{v_1}\|} f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) \frac{p_b}{1+\mathbf{D}(\mathbf{v_1},\mathbf{v_2})} d\mathbf{v_1} d\mathbf{v_2} \tag{35}$$

and, for the gravity model,

$$\mathbf{E}\left(\vec{\mathcal{V}}^e(\mathbf{A})\right) = \int_{\mathbf{A}}\int_{\bar{\mathbf{A}}} \frac{\mathbf{v_2}-\mathbf{v_1}}{\|\mathbf{v_2}-\mathbf{v_1}\|} f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) \frac{p_b}{1+\alpha d_s(\mathbf{v_1},\mathbf{v_2})^2} d\mathbf{v_1} d\mathbf{v_2} \tag{36}$$

The result of these equations is a vector whose direction is the expected average direction of all ties in the cut induced by **A** and its complement, and whose magnitude grows with the expected number of ties pointing in this direction[31]. Note that this naturally implies that symmetric tie flows (i.e., flows "across" a region) tend to cancel each other out; if one is interested in capturing such flows, it is advisable to reflect all ties across one axis prior to integrating. Further, when interpreting direction it is important to recall that this measure is an average: just because the expected average tie points in a given direction, it does not follow that one should expect no ties in the opposite direction! While $\vec{\mathcal{V}}^e(\mathbf{A})$ can give one an excellent sense of the macrostructural relationships between regions, it should be used in conjunction with other measures to establish a more complete picture.

3.4 Locality of Ties

In analyzing both internal and external tie volumes, one of our basic theoretical goals has been the characterization of tie "locality". In particular, we have been interested in developing some sense of the degree to which ties tend to be distributed purely to physically proximate actors, as opposed to being sent to actors over long distances. To conceptualize the difference, it is perhaps helpful to think of the macrostructure as a sort of social "fabric", a tangled skien of relational threads linking actors across space. Insofar as ties are localized, this fabric is tight and orderly: threads are short and even, connecting neighbor to neighbor across the landscape like components of a wire mesh. Where ties are highly non-local, by contrast, the fabric becomes a tangled mass of overlapping and criss-crossing threads, dense cobwebs stretched over the surface of the earth. Such differences may affect such diverse phenomena as the spread of culture (and the development and/or retention of subcultures), the number and size of collective events, and patterns of migration; they may provide clues as to which regions are likely to ally themselves in times of crisis, and which will be at the forefront of cultural change.

Utilizing tie volume and tie flow measures, we may gain insight into locality; nevertheless, none of these index locality directly, in the sense of characterizing how ties are distributed across space. To rectify this situation, then, we here introduce a direct measure of locality which extends external tie volume to account for distance covered by as well as number of edges.

Given an area $\mathbf{A}$, let the locality of $\mathbf{A}$ (denoted $\mathcal{L}(\mathbf{A})$) be the expected distance between the endpoints of all ties contained in $G[\mathbf{A}, \bar{\mathbf{A}}]$. It follows, then, that $\mathcal{L}$ may be easily formulated as

$$\mathcal{L}(\mathbf{A}) = \int_{\mathbf{A}} \int_{\bar{\mathbf{A}}} \|\mathbf{v_2} - \mathbf{v_1}\| f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) p\left(\{\mathbf{v_1}, \mathbf{v_2}\} \in E(G)\right) d\mathbf{v_1} d\mathbf{v_2} \tag{37}$$

Following the usual drill, we can re-express this in terms of the generalized distance model:

$$\mathcal{L}(\mathbf{A}) = \int_{\mathbf{A}} \int_{\bar{\mathbf{A}}} \|\mathbf{v_2} - \mathbf{v_1}\| f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) \frac{p_b}{1 + \mathbf{D}(\mathbf{v_1}, \mathbf{v_2})} d\mathbf{v_1} d\mathbf{v_2} \tag{38}$$

or, similarly, of the gravity model:

$$\mathcal{L}(\mathbf{A}) = \int_{\mathbf{A}} \int_{\bar{\mathbf{A}}} \|\mathbf{v_2} - \mathbf{v_1}\| f_p(\mathbf{v_1}) f_p(\mathbf{v_2}) \frac{p_b}{1 + \alpha d_s(\mathbf{v_1}, \mathbf{v_2})^2} d\mathbf{v_1} d\mathbf{v_2} \tag{39}$$

Measured in this way, the locality of a given region suggests the degree to which ties entering or leaving the region tend to come from alters which are spatially proximate, or, alternately, the degree to which such ties tend to span greater distances. (Indeed, while one can infer either from the measure, it should be noted that, despite its name, "locality" is actually a measure of non-locality.) Used in combination with tie volume measures, this can suggest a great deal about the macrostructure of relations within a given area. Consider, for instance, a large region for which the external tie volume is observed to be low; if an examination of subregions of this larger region reveals high values of $\mathcal{L}$ relative to the region's size, then one can readily infer that the region as a whole is both closed to the outside and tightly intraconnected. High values of $\mathcal{V}^e$ combined with low values of $\mathcal{L}$, by contrast, are characteristic of regions with tight connections to the immediate surroundings but few ties to outlying areas. By combining measures, then, we can characterize a wide range of macrostructural configurations with a relatively limited number of observations.

4 An Application of Macrostructural Analysis: Sample Population Structures

To illustrate a few of the measures which we have discussed in the above sections, we shall now apply them (under the gravity model) to several sample population structures. As our purpose is demonstrative rather than substantive, the population structures we shall use here are hypothetical rather than real; each was generated using a simple simulation routine which randomly allocated "population centers" to a continuous landscape, varying population in terms of distance from these centers[32]. This setup allows us to easily compute approximations of instantaneous macrostructural measures, and to manipulate population structure so as to be able to provide fairly simple examples. The basic visualization techniques demonstrated here can then be applied to real-world data relatively easily.

The first population structure which we shall consider is a very simple one: a single population center (maximum density 5000 persons/km^2), placed randomly on a 200x200km square area. Plotted as a three-dimensional surface[33], the population density for this structure appears as shown in Figure 1. Note that population density decays fairly evenly as one moves out from the population center, gradually leveling out at a radius of approximately 50km.

(Insert Figure 1 Here)

Turning to internal tie volume, $\mathcal{V}^i$, we may plot it using the same approach; these results are shown in Figure 2. $\mathcal{V}^i$ values here were calculated numerically for 1x1km cells, using the standard gravity model with $p_b = 0.1$ and $\alpha = 1.0$. As can be seen in the above plot, $\mathcal{V}^i$ follows population density to a great extent; however, the combinatorics of subgraph size tend to cause internal tie volume to grow faster than population (for arbitrarily small cells, $O(f_p^2)$). One result of this is that, for small, uniform regions, small variations in population density across space may be amplified by internal tie volume. This could have implications for the emergence of structural heterogeneity from minor perturbations in population, particularly if internal tie volume is itself related to population growth or retention.

(Insert Figure 2 Here)

Having examined internal tie volume for the simple population structure, let us now consider external tie volume. As can be seen in Figure 3, external tie volume here peaks at the center of population, but diminishes much more slowly with distance; as one moves into outlying areas, one encounters a $\mathcal{V}^e$ "shadow" of the larger population center in the form of high tie volumes drawn towards the peak. This is not an idiosyncratic phenomenon: note that as the area on which one computes external flow volume diminishes (approaching the instantaneous case), $\mathcal{V}^e$ becomes increasingly dominated by nearby population centers. As a result, $\mathcal{V}^e$ will often act much like a smoothed version of the population density function itself when taken over small areas.

(Insert Figure 3 Here)

From simple external tie volume to directional external tie volume: the vectors of the overhead view in Figure 4 indicate the direction of net external tie flow[34] superimposed on isolines of population density. Note how the net external tie flow tends to be perpendicular to population isolines, pointing in the direction of increasing population, and how one can get "turbulence" in uneven areas around peaks. The effect of a large population center in drawing in ties from the immediate area is clearly illustrated here.

(Insert Figure 4 Here)

Considering a simple population form, such as the above, helps illustrate the behavior of several measures of interest without requiring us to disentangle the effects of multiple competing population centers and the like. Now, however, we turn to a somewhat more complex case. Our next population structure was created by placing fifty population centers (maximum population density distributed U(500,1500) persons per km^2) randomly in a 200x200km area (as in the previous case). We show the population density plot in Figure 5.

(Insert Figure 5 Here)

Clearly, this presents a somewhat more rugged population landscape than the single peak we considered earlier. Computing $\mathcal{V}^i$ (again for 1x1km cells) on this population structure, then, gives us the volume "map" of Figure 6. Here again, we see a pattern which closely follows population density itself. Further, the patterns of exaggeration we noted previously are also present: large peaks are somewhat larger but fall off slightly faster, while very small peaks appear to have been damped out relative to their larger neighbors. This is typical of the f_p^2 effect, which has little effect on low populations but which causes substantial $\mathcal{V}^i$ growth once density begins to increase beyond a minimal level.

(Insert Figure 6 Here)

Proceeding to external tie volume, our calculations give us the surface shown in Figure 7. Here, we can see that population "peaks" which are broad, and/or close to other peaks, tend to give rise to $\mathcal{V}^e$ peaks which can be substantially greater in magnitude than would be anticipated from local population densities. This suggests an interesting exception to the earlier smoothing argument: where small, populous regions lie between other populous regions, these central areas may experience disproportionate growth of external tie volume. Insofar as this growth rate is faster than that of $\mathcal{V}^i$, such regions may become dominated by external influences to a degree that their neighbors do not. Interestingly, continued population growth may slow or even reverse this process: if population in the central region becomes sufficiently concentrated, internal tie volumes may begin to outstrip their external counterparts. While this clearly depends on the manner in which population and macrostructure interact, it is interesting to note how these simple combinatorial growth patterns can potentially give rise to fairly subtle effects for population centers with the appropriate local structure.

(Insert Figure 7 Here)

Having considered $\mathcal{V}^e$ for more complex case, now let us look at the directionality of external tie flow; this is presented in Figure 8. (Note that, as before, we have shown only a subsection of the larger space due to graphical constraints.) What is particularly interesting in this case is what does *not* appear: whereas before we saw smooth flow perpendicular to the population isolines, now flow seems to travel across population centers! The explanation for this phenomenon lies in the power of remote population centers to pull ties towards themselves, such that there is a net "overflow" of ties in the direction of large population clusters over the entire landscape. As we saw from the graph of $\mathcal{V}^e$, these local population centers *do* have their share of regional traffic; at the same time, however, the pull of nearby clusters can work subtle distortions in their patterns of ties, increasing the number of paths leading towards the core areas.

(Insert Figure 8 Here)

5 Discussion

The modeling framework which has been presented here is one which incorporates both an important set of theoretical concerns and a body of empirical findings regarding human interaction; nevertheless, it is itself still untested, and may or may not prove to provide a satisfactory portrayal of spatial macrostructure. In particular, the scale which is required to ensure reasonable predictive accuracy is currently unknown. While Festinger et. al. (1950) found effects of physical distance over quite small scales (on the order of 1000ft), our theoretical results regarding stochastic equivalence suggest that larger scales (on the order of 1000km) may be needed for certain applications (e.g., treating cities as single entities). In general, we would expect for the efficacy of the model to be very limited at small scales and to increase substantially as one considers larger areas; the actual speed with which this occurs must be determined from empirical data. On a similar note, it is also the case that the current model is limited by the assumption that the population density function is known with certainty, and that it is static (relative to the speed with which ties are formed). Insofar as the real world deviates from these simplifying assumptions, we can expect reduced performance[35].

Regardless of these limitations, however, the macrostructural measures introduced here should be usable even without the underlying model (though their theoretical significance may in some cases be diminished), and all can be estimated by way of sampling procedures. Because computation of expected values of these measures is relatively simple for models such as the gravity variant, it should be fairly simple to compare observed values to theoretical predictions once population distribution is known[36]. Given a number of such comparisons, it should be possible to evaluate whether or not these models are adequate to explain observed patterns of tie volume in real networks.

Another direction which should be taken to continue the present work is the interface of spatially produced networks with pre-existing network process models. Numerous models of diffusion and influence exist, for instance, which have heretofore been tested largely on either canonical structures (e.g., grids, toroids, etc.) or on random graphs (generally controlled for size and density) (Carley, 1990; Carley, 1991; Butts, 1998a; Latané, 1996; Friedkin and Cook, 1991; Krackhardt, 1997). Do these models exhibit different behaviors on spatially generated networks? If so, this clearly raises questions of validity regarding past studies. Other models which might be investigated include models of exchange (Friedkin, 1995; Cook et al., 1983), spontaneous eruption of panics or fads (Butts, 1998b), cultural evolution (Latané, 1996; Axelrod, 1997), and collective action (Macy, 1991).

Perhaps the most obvious extension of the spatial modeling framework presented here is that of permitting treatment of change in population distribution and relations over time. Such innovations would facilitate tie-in with research on migration, mobility, and urban sociology (Irwin and Hughes, 1995), and could potentially lead to development of a unified demographic-relational model of macrostructural evolution. Before the model can be extended in this way, however, more must be known about the relative relaxation times of migratory/demographic and tie formation/disruption processes; otherwise, it is not possible to couple the processes properly. Furthermore, continued development would also be greatly facilitated by a clearer understanding of the underlying processes governing tie formation across space. While a purely static treatment does not necessarily require a great deal of knowledge regarding how or why ties end up exhibiting an inverse square relation to physical distance, a dynamic model has much less flexibility in this regard: if, for instance, ties are formed by chance meetings among actors (whose positions in space are, in turn, given by a probability distribution over locations[37]) then one may be lead to a very different sort of model than one would employ if actors tend to gain ties through "referrals" (à la Granovetter (1973)). (Needless to say, extensions such as these would require the macrostructural measures presented here to be computed in a very different fashion.)

Another promising area for future examination of the spatial modeling framework is the investigation of the distributions of graph-level indices (GLIs) on spatial graphs. While there are a number of challenges involved in the computation of GLIs on macrostructures, increasing availability of inexpensive parallel processing capability should ultimately make a wide range of GLI experiments possible using simulation[38]. Of particular interest in macrostructures are properties such as girth, diameter, and length of the median geodesic, as well as clique frequency and the like; as these GLIs are less-often studied on meso or microstructures than traditional measures such as hierarchy or centralization, there may be room for theoretical and methodological development in this area[39].

6 Conclusion

In this paper we have considered both theoretical and empirical arguments for the importance of space in determining the structure of human relations within large populations, and have presented a stochastic modeling framework which builds on these prior findings. We have shown how one of these models – the gravity model – can induce stochastic equivalence classes in populations which are spatially clustered, and have derived the inter and intra-cluster distances which are necessary to obtain these effects. To characterize spatial macrostructures of human relations, we have introduced a number of measures which may be employed either instantaneously or over areas of interest to determine volumes of tie "flow" across space; expectations of these measures under spatial models of network formation have been given, as well as suggestions for their empirical estimation. Finally, we have illustrated how some of these measures may be visualized on populations of interest, and have indicated some insights into the behavior of the gravity model which these measures reveal. It is hoped that these results will lead to further work on the problem of spatial macrostructure, and on the development of more accurate and extensive models to predict and describe networks generally. As network analysts have shown, a wide range of problems of concern in the modern world – including information flow, innovation, and the spread of disease – are intimately interconnected with patterns of human interaction. By understanding how large-scale structures of relations across entire populations behave, we may gain further insight into these problems as well.

It is perhaps worth reiterating at this point that, as important as the macrostructure of human relations may be for understanding other sociological problems, it is the opinion of the authors that macrostructure is *in and of itself* a valid object for study. To characterize, measure, and ultimately predict the vast webs of interaction which form the fabric of human society is truly a task worthy of the sociological imagination, and is of scientific merit regardless of whether it yields insights at the individual level. As the astrophysicist may study galaxies whose lifespans exceed even the grossest limits of our temporal comprehension, or the biologist may seek to untangle ancient chains of innovation and competition which lead to the emergence of our own human form, so do we attempt to contemplate the dense and ever-changing weave of relationships in which each of us is inextricably embedded. In this spirit, then, it is our hope that the present work may aid us in beginning to understand social macrostructure on its own terms; surely, this is an exciting area for future research.

7 References

Anderson, B., Butts, C., and Carley, K.C. (1999) 'The Interaction of Size and Density with Graph-Level Measures.' *Social Networks*, 21(3), 239-267.

Axelrod, R. (1997) 'The Dissemination of Culture: a Model With Local Convergence and Global Polarization.' *The Journal of Conflict Resolution*, 41, 203-26.

Bernard, H.R., Johnsen, E.C., Killworth, P.D., and Robinson, S. (1989) 'Estimating the Size of an Average Personal Network and of an Event Subpopulation.' In Kochen, M. (ed.), *The Small World*, 159-75. Norwood, NJ: Ablex.

Blau, P.M. (1977) *Inequality and Heterogeneity.* New York: Free Press.

Boorman, S.A., and Levitt, P.R. (1980) *The Genetics of Altruism.* New York: Academic Press.

Bossard, J.H.S. (1932) 'Residential Propinquity as a Factor in Marriage Selection.' *American Journal of Sociology*, 38, 219-44.

Butts, C.T. (1998a) 'A Bayesian Model of Panic in Belief.' *Computational and Mathematical Organization Theory*, 4(4).

Butts, C.T. (1998b) 'Generating Panic in Populations.' ICES Research Report, ICES 88-03-98, Carnegie Mellon University.

Butts, C.T. (2000a) 'An Axiomatic Approach to Graph Complexity.' *Journal of Mathematical Sociology,* in press.

Butts, C.T. (2000b) 'The Complexity of Social Networks: Theoretical and Empirical Findings.' *Social Networks,* in press.

Carley, K.C. (1990) 'Group Stability: A Socio-Cognitive Approach.' *Advances in Group Processes*, 7.

Carley, K.C. (1991) 'A Theory of Group Stability.' *American Sociological Review*, 56(3), 331-354.

Carley, K.C., and Wendt, K. (1991) 'Electronic Mail and Scientific Communication: A Study of the SOAR Extended Research Group.' *Knowledge: Creation, Diffusion, Utilization*, 12(4), 406-40.

Cook, K.S., Emerson, R.M., Gillmore, M.R., Yamagishi, T. (1983) 'The Distribution of Power in Exchange Networks: Theory and Experimental Results.' *American Journal of Sociology*, 89(2), 275-305.

Festinger, L., Schachter, S., and Back, K. (1950) *Social Pressures in Informal Groups.* Stanford: Stanford University Press.

Friedkin, N. (1995) 'The Incidence of Exchange Networks.' *Social Psychology Quarterly*, 58(3), 213-222.

Friedkin, N., and Cook, K. (1991) 'Peer Group Influence.' *Sociological Methods and Research*, 19, 122-143.

Granovetter, M. (1973) 'The Strength of Weak Ties.' *American Journal of Sociology*, 78(6), 1369-80.

Grimmett, G.R., and Stirzaker, D.R. (1992) *Probability and Random Processes.* Oxford: Clarendon Press.

Irwin, M.D., and Hughes, H.L. (1992) 'Centrality and the Structure of Urban Interaction: Measures, Concepts, and Applications.' *Social Forces*, 71(1), 17-51.

Killworth, P.D., and Bernard, H.R. (1978) 'Reverse Small World Experiment.' *Social Networks*, 1, 159-92.

Krackhardt, D. (1997) 'Organizational Viscosity and the Diffusion of Controversial Innovations.' *Journal of Mathematical Sociology*, 22(2), 177-99.

Latané, B. (1996) 'Dynamic Social Impact: The Creation of Culture by Communication.' *Journal of Com-*

munication, 46(4), 13-25.

Latané, B., Liu, J.H., Nowak, A., Bonevento, M., and Zheng, L. (1995) 'Distance Matters: Physical Space and Social Impact.' *Personality and Social Psychology Bulletin*, 21(8), 795-805.

Lorrain, F., and White, H.C. (1971) 'Structural Equivalence of Individuals in Social Networks.' *Journal of Mathematical Sociology*, 1, 49-80.

Macy, M. (1991) 'Chains of Cooperation: Threshold Effects in Collective Action.' *American Sociological Review*, 56, 730-47.

Mayhew, B.H. (1980) 'Structuralism Versus Individualism: Part I, Shadowboxing in the Dark.' *Social Forces*, 59, 335-75.

Mayhew, B.H. (1981) 'Structuralism Versus Individualism: Part II, Ideological and Other Obfuscations.' *Social Forces*, 59, 627-48.

Mayhew, B.H. (1984) 'Baseline Models of Sociological Phenomena.' *Journal of Mathematical Sociology*, 9, 259-81.

Mayhew, B.H., and Levinger, R.L. (1976) 'Size and Density of Interaction in Human Aggregates.' *American Journal of Sociology*, 82, 86-110.

Mayhew, B.H., McPherson, M., Rotolo, Smith-Lovin, L. (1995) 'Sex and Race Homogeneity in Naturally Occurring Groups.' *Social Forces*, 74(1).

Milgram, S. (1967) 'The Small World Problem.' *Psychology Today*, 22, 61-7.

Nowak, A., Szamrej, J., and Latané, B. (1990) 'From Private Attitude to Public Opinion: A Dynamic Theory of Social Impact.' *Psychological Review*, 97(3), 362-76.

Rapoport, A. (1949a) 'Outline of a Probabilistic Approach to Animal Sociology I.' *Bulletin of Mathematical Biophysics*, 11, 183-96.

Rapoport, A. (1949b) 'Outline of a Probabilistic Approach to Animal Sociology II.' *Bulletin of Mathematical Biophysics*, 11, 273-81.

Rapoport, A. (1950) 'Outline of a Probabilistic Approach to Animal Sociology III.' *Bulletin of Mathematical Biophysics*, 12, 7-17.

Schelling, T.C. (1969) 'Models of Segregation.' *American Economic Review*, 59(2), 483-93.

Stewart, J.Q. (1941) 'An Inverse Distance Variation for Certain Social Influences.' *Science*, 93, 89-90.

Tobler, W., Deichmann, U., Gottsegan, J., and Maloy, K. (1995) 'The Global Demography Project.' Technical Report TR-95-6, National Center for Geographic Information and Analysis, University of California at Santa Barbara.

Travers, J., and Milgram S. (1969) 'An Experimental Study of the Small World Problem.' *Sociometry*, 32, 425-43.

Wasserman, S., and Faust, K. (1994) *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.

Wasserman, S., and Iacobucci, D. (1986) 'Statistical Analysis of Discrete Relational Data.' *British Journal of Mathematical and Statistical Psychology*, 39, 41-64.

Wasserman, S., and Weaver, S.O. (1985) 'Statistical Analysis of Binary Relational Data: Parameter Estimation.' *Journal of Mathematical Psychology*, 29, 406-27.

Zipf, G.K. (1949) *Human Behavior and the Principle of Least Effort.* New York: Hafner.

Threshold Parameters		$\alpha = 0.5$	$\alpha = 1$	$\alpha = 2$
τ	Δ	δ_τ	δ_τ	δ_τ
0.9	1	0.155	0.106	0.080
	10	0.551	0.546	0.544
	100	5.410	5.410	5.410
	1000	54.093	54.093	54.093
0.95	1	0.076	0.051	0.039
	10	0.265	0.262	0.261
	100	2.598	2.598	2.598
	1000	25.978	25.978	25.978
0.99	1	0.015	0.010	0.008
	10	0.052	0.051	0.051
	100	0.504	0.504	0.504
	1000	5.039	5.038	5.038
0.999	1	0.002	0.001	0.001
	10	0.005	0.005	0.005
	100	0.050	0.050	0.050
	1000	0.500	0.500	0.500

Table 1: Stochastic Equivalence Thresholds for the Gravity Model

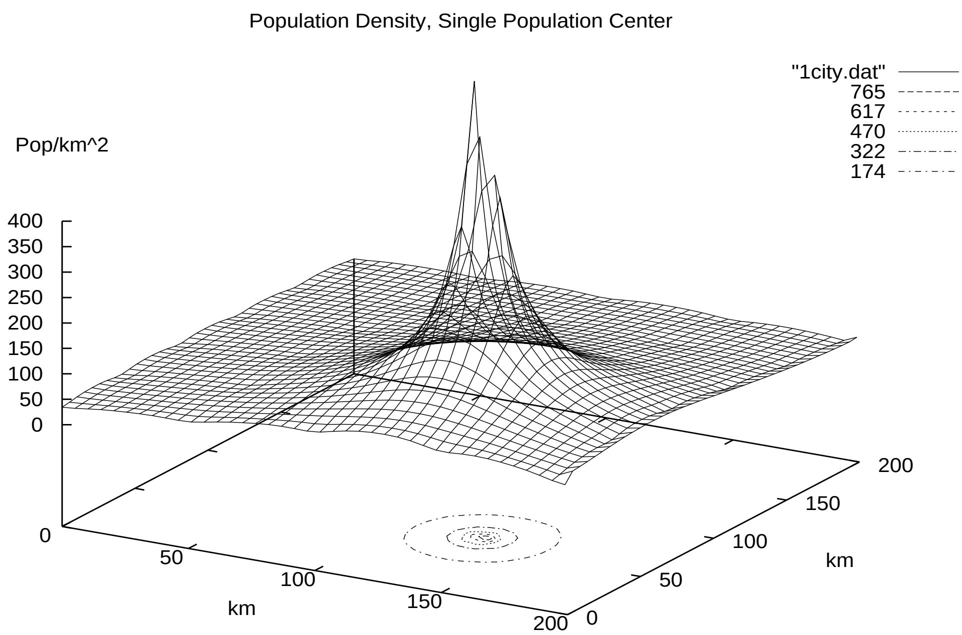

Figure 1: Simulated Population Density, 200km x 200km Region, One Population Center

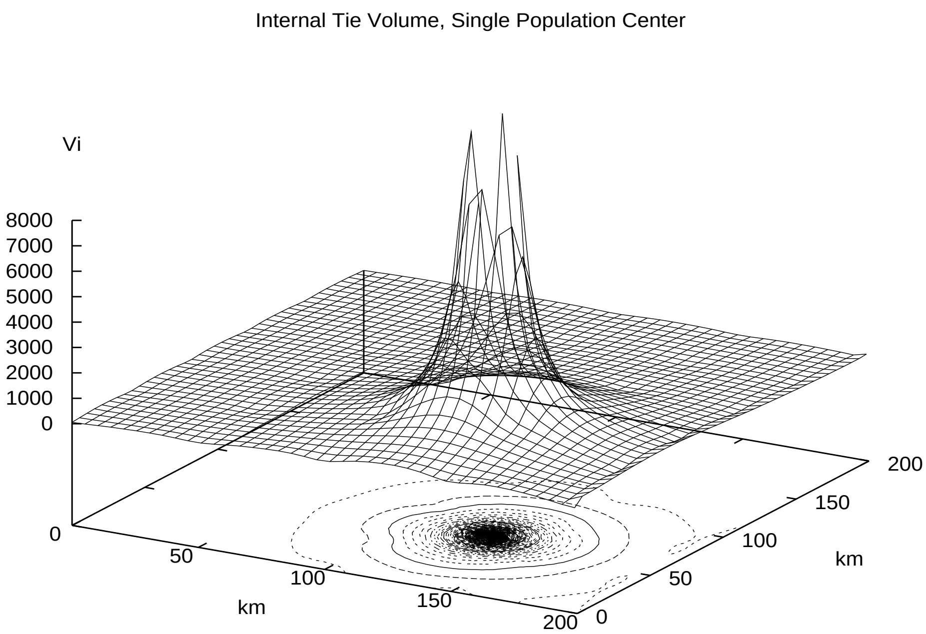

Figure 2: Estimated Internal Tie Volume (1 Pop Center, 1km x 1km Cells)

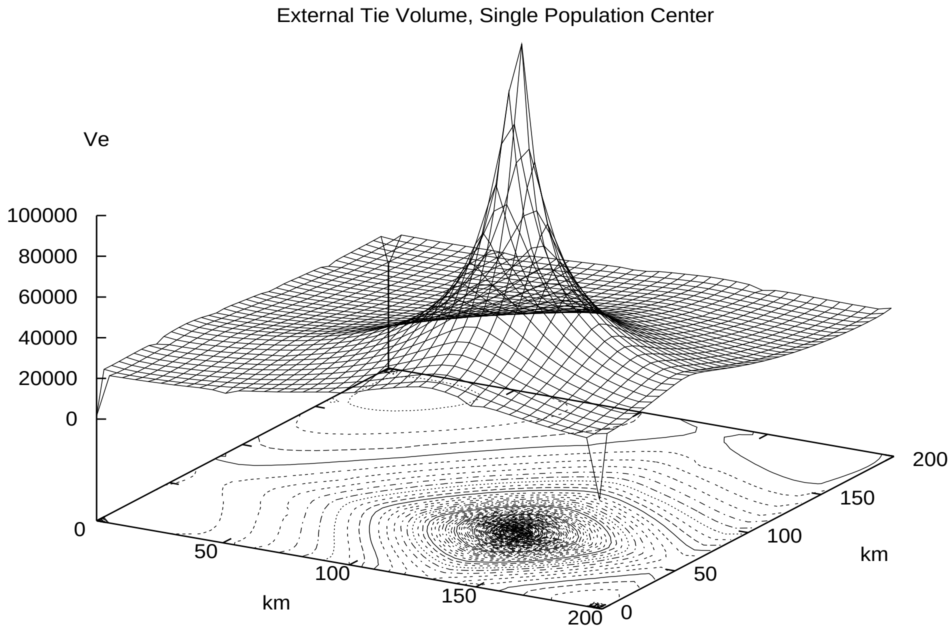

Figure 3: Estimated External Tie Volume (1 Pop Center, 1km x 1km Cells)

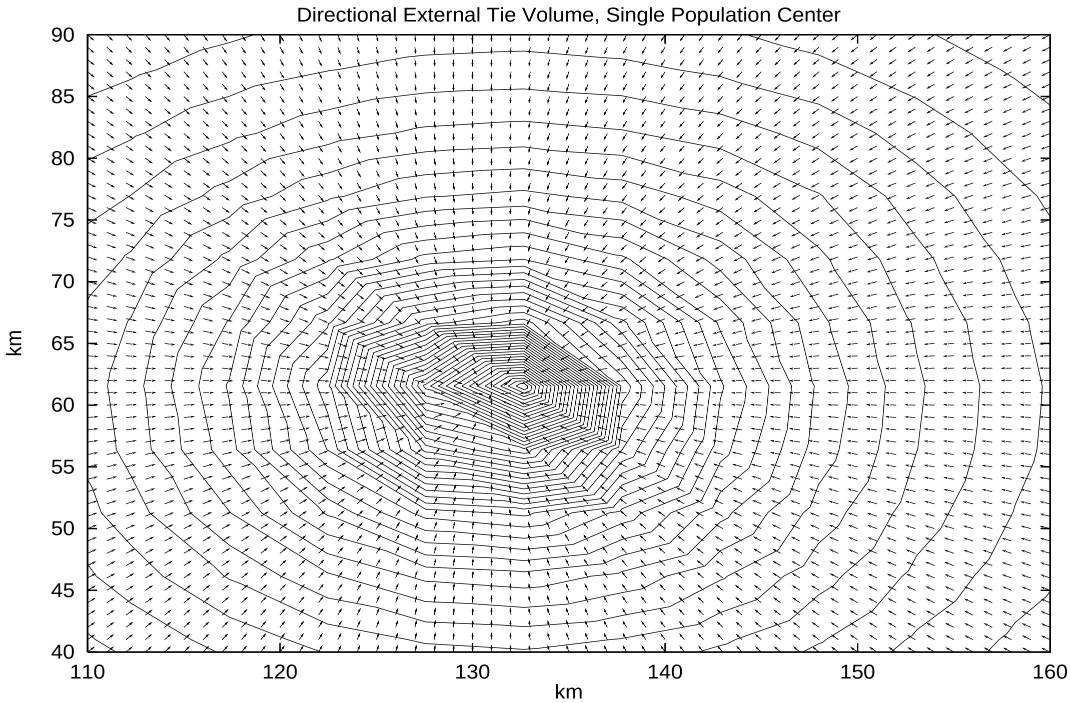

Figure 4: Estimated Directed External Tie Volume (1 Pop Center, 1km x 1km Cells)

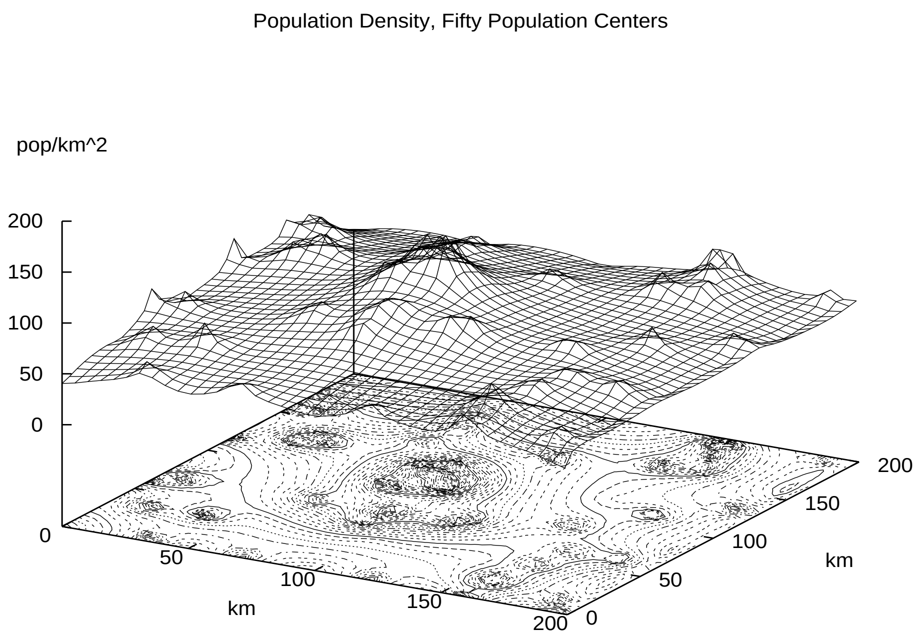

Figure 5: Simulated Population Density, 200km x 200km Region, 50 Population Centers

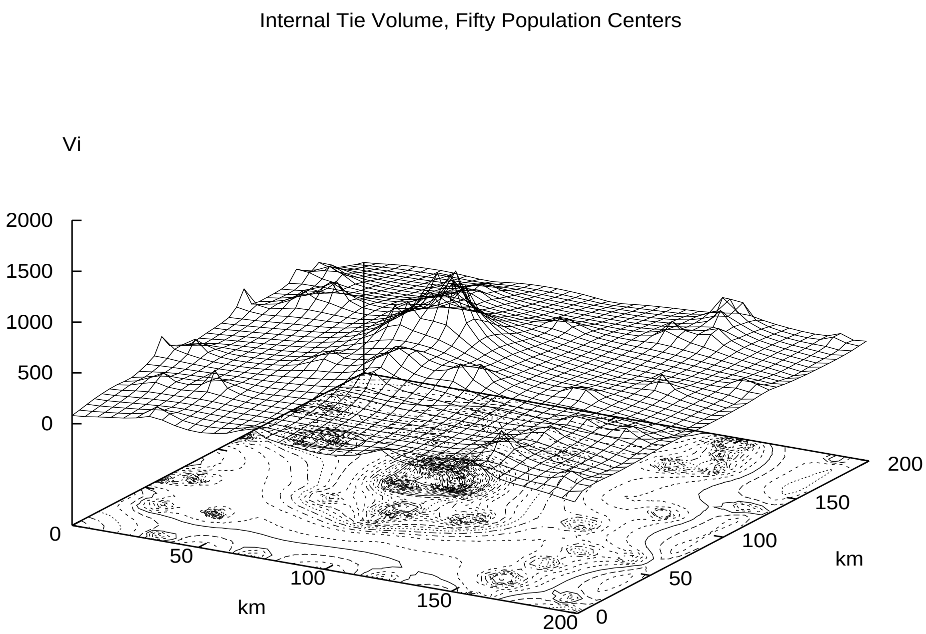

Figure 6: Estimated Internal Tie Volume (50 Pop Centers, 1km x 1km Cells)

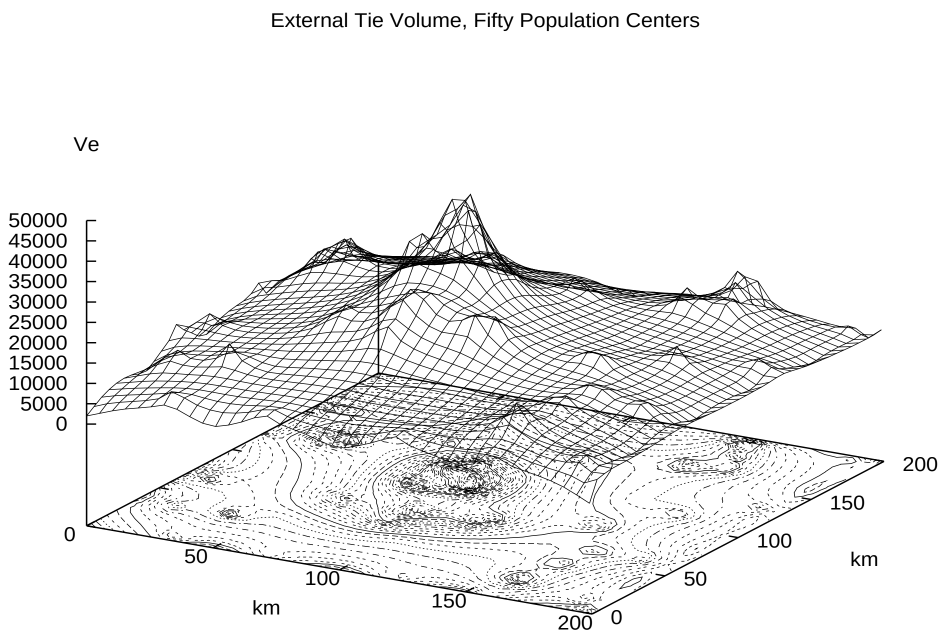

Figure 7: Estimated External Tie Volume (50 Pop Centers, 1km x 1km Cells)

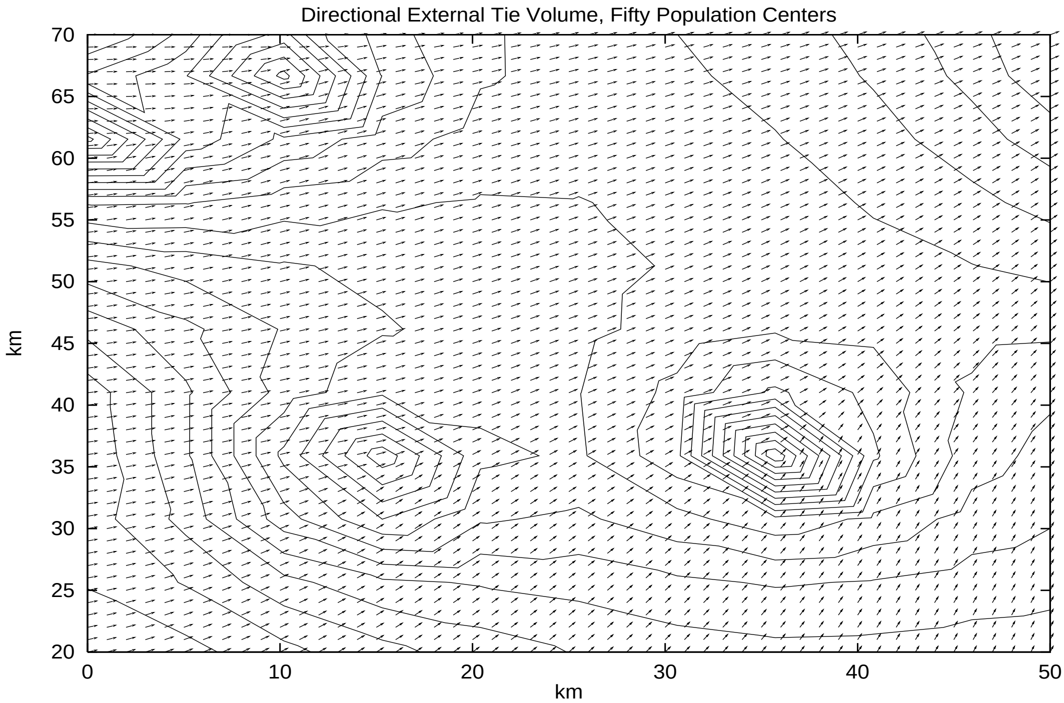

Figure 8: Estimated Directed External Tie Volume (50 Pop Centers, 1km x 1km Cells)

Notes

[1]Indeed, even the memory devoted to keeping track of old friends and the time taken to hunt for phone numbers or email addresses have costs; the degree to which such trivialities can be sufficient to affect our interactions should be obvious to anyone who has ever had to apologize to a friend for not keeping in touch.

[2]The exchange of genetic (or, perhaps, mimetic) material can also be constrained in a similar fashion. See, for instance, Boorman and Levitt (1980).

[3]Allowing four hours of sleep per night and reducing the speed to a friendlier 5mph reduces the radius covered to 10 miles, for those who are interested.

[4]Obviously, modern communication technology lessens this requirement somewhat (though see Carley and Wendt (1991)), but even in virtual environments participants must be "close" in some communication space to engage in interaction. Whether or not these communication spaces are in fact "smaller" than their physical counterparts for purposes of interaction is an open (and interesting) question.

[5]Radios and the like being excepted. This, however, falls under the category of communication space interactions mentioned above.

[6]Or, at least, are not included; given that many if not most empirical analyses of relational networks consider individuals who are physically proximate, this may be less of an omission than has been suggested. (But see Festinger et al. (1950).)

[7]It is interesting to note, however, that the influence structures of Latané's Dynamic Social Impact Theory (DSIT) (Latané, 1996) can be represented in terms of valued graphs; hence, the dichotomy is clearly a somewhat exaggerated one.

[8]It is worth noting that the relations described accounted for upwards of 90% of the variance in number of interactions within their three samples, making this a finding of truly remarkable strength.

[9]Clearly, it is quite possible that this assumption does not hold in all cases; it seems, however, to be a reasonable point from which to start. We thus present this model (and the gravity model) as *baseline models* in the spirit of Mayhew (1984), with an eye towards beginning with "obvious" physical constraints and adding additional complexity only as required.

[10]Since this may take any value in [0,1], this does not imply loss of generality.

[11]Or the minimum arc distance, in the case of spherical embeddings.

[12]Generally, we shall represent vertices with lowercase v's, and the vectors which designate their positions by boldface (e.g., $\mathbf{v}$).

[13]Interestingly, this implies that fitting a gravity model on data with physical and social spatial dimensions will yield social distances in physical equivalents; thus, in principle one could ascertain the "extra distance" in km between persons of differing race or gender. The relative "size" of these Blau dimensions across cultures could be an interesting tool for comparative analysis, particularly if obtained at multiple points in time.

[14]Specifically: "Given a stochastic 'multigraph'. represented by the collection of random matrices $\mathbf{X}$, actors i and i' are *stochastically equivalent* if and only if the probability of any event concerning $\mathbf{X}$ is unchanged by an interchanging of actors i and i'." (Wasserman and Faust, 1994).

[15]E.g., under relaxation of independence.

[16]Obviously, these are also real cities. Here, however, we shall consider them only hypothetically.

[17]In order for such multi-level clusterings to be "clean", human populations would have to be laid out in an extremely sparse fashion. Indeed, the sort of self-similar fractal which would be generated by a perfectly clean pattern of this type would be (in form) somewhat like the Cantor set, which is often referred to as an example of "fractal dust" due to its low dimensionality.

[18]The same procedures could be applied to other distance models, but may or may not yield closed-form solutions.

[19]This assumption allows us to consider a lower bound on the conditions for approximate equivalence, and aids in clarity of presentation. The same basic logic may be applied to derive exact bounds in the general case.

[20]This is because $d_s(v_j, v_k) = \Delta$ and $d_s(v_i, v_k) = \delta + \Delta$.

[21]This is, essentially, a likelihood ratio approach (in the sense that require there to be minimal difference

between the edge likelihoods); it is employed here (rather than a raw difference measure of some sort) both because it provides a reasonable representation of our notion of "difference" in this case, and because it is an effective means of comparing arbitrarily small values independent of scaling effects.

[22]Note that units here are arbitrary, so long as α, Δ, and δ are in the same scale.

[23]Since the net effect of α is bounded by $\sqrt{\alpha(\tau - \tau^2)}/\tau$, α would have to be in the vicinity of 10 to potentially have a unit impact on δ_τ (assuming a fairly lax $\tau = 0.9$). In practice, however, the effect would be reduced for large Δ; in general, the message appears to be that this aspect of the model becomes increasingly well-behaved as Δ grows relative to the other parameters.

[24]Again, it should be repeated that this is an upper bound on the minimum radius. Obviously, then, it is true that the equivalence will exist for any actor beyond this minimum radius, though it may be the case that some closer actors will share the relationship as well.

[25]Or summations, in some cases. We shall be rather cavalier throughout the text in treating space as continuous and our associated functions as (abstractly) integrable, but these quantities can be interpreted as summations if required. (See Grimmett and Stirzaker (1992).)

[26]Thus, we are in fact assuming that the associated spatial position is a *sufficient statistic* for tie probability, not that it is in fact the only *actual* position ever occupied by the actor.

[27]For instance, a stochastic model could be constructed which treated persons as "events" in space in something of an analogy to the Poisson process. A rigorous treatment of such a model, however, is beyond the scope of the present work.

[28]Of course, it could be argued that quantities which are poorly suited to empirical measurement are necessarily of limited theoretical value; insofar as the purpose of theory is to identify useful relationships between observables, this is certainly the case.

[29]And that it is not uniformly zero! Technically, of course, if we truly allow our area of interest to go to zero, it must be the case that (in the real world) our area cannot possibly include more than one person. That said, the instantaneous measure should be interpreted as a description of the way in which local tie volume changes across space, rather than as a true estimate of volume at a given point in space.

[30]The entry and exit of ties, for instance, need not be "conserved"; nor are tie flows incompressible (which amounts to the same thing). Perhaps most importantly, the endpoints of ties need not be spatially local, which is clearly not the case for the movement of molecules.

[31]Note that the *distance* covered by these ties is intentionally removed. We shall consider the question of expected tie distance presently.

[32]The number and size of the population centers was controlled; the contribution of each center to a given point was an inverse linear function of the distance from the center in question.

[33]All plots for this document were produced using gnuplot. Surface plots were created using a 40x40 sample grid, smoothed using city block distances (norm of 2).

[34]All vectors have been normalized to facilitate display.

[35]In addition, it is obviously assumed here that any determinants of social ties *other* than those accounted for by position in socio-physical space are ignorable with respect to the spatial model. This assumption may limit the applicability of the model in certain contexts.

[36]Fortunately, excellent data sets exist mapping population to physical space; with the advent of the GPS, this data can only be expected to improve.

[37]Indeed, one can imagine a dynamic model in which actors, rather like electrons, are portrayed as "clouds" of possible positions with varying probability, and in which these position distributions evolve over time. Unfortunately, the complexity of such a model would likely be prohibitive.

[38]The authors have performed several preliminary simulation experiments using the gravity model on various layouts containing several thousand actors; thusfar, early results seem to confirm the importance of population density (and configuration), and point to scaling effects as having important consequences in many cases.

[39]The presence of stochastic equivalence classes in spatial networks, for instance, implies that such graphs should have lower levels of algorithmic complexity than their random counterparts (Butts 2000a; 2000b); this may in turn have implications for other properties of spatial networks.