Geographical Variability as a Determinant of Large-scale Network Structure*

Carter T. Butts†

It is a well-established result that the marginal probability of a social tie between two persons declines with geographical distance for a wide range of social relations (see, e.g., Bossard, 1932; Zipf, 1949; Festinger et al., 1950; Hägerstrand, 1967; Freeman et al., 1988; Latané et al., 1995; McPherson et al., 2001). While often regarded as a mere curiosity, others have argued that this relationship is a critical determinant of social structure (Mayhew, 1984). Indeed, Butts (2003) has shown that under fairly weak conditions, spatial structure is adequate to account for the vast majority of network structure (in terms of total entropy) at large geographical scales.

Spatial Network Models

The simplest family of network models to incorporate this notion is the family of spatial Bernoulli graphs, defined by pmfs of the form

\[
\Pr(Y = y | D) = \prod_{i,j} B(Y_{ij} = y_{ij} | \mathcal{F}(D_{ij}, \theta)),
\]

where \(Y\) is the graph adjacency matrix, \(D\) is a matrix of inter-vertex distances, \(B\) is the Bernoulli pmf, and \(\mathcal{F}\) is a function taking distances into the \([0, 1]\) interval (parameterized by real vector \(\theta\)). In this context, \(\mathcal{F}\) is referred to as a spatial interaction function, and can be interpreted directly as providing the marginal probability of a tie between two randomly selected individuals at some given distance. It can immediately be observed that this family is a special case of the inhomogeneous Bernoulli graphs (w/ pmf \(\Pr(Y = y | \Phi) = \prod_{i,j} B(Y_{ij} = y_{ij} | \Phi_{ij})\)), with parameter matrix \(\Phi\) given by \(\Phi_{ij} = \mathcal{F}(D_{ij}, \theta)\). Models of this form have been studied in the context of geographical distances by Butts (2002); Hipp and Perrin (2009); Butts and Acton (2011), and are closely related to the latent space models of Hoff et al. (2002); Handcock et al. (2007). They can also be viewed as special cases of the family of gravity models (Haynes and Fotheringham, 1984), which have been used for several decades in the geographical literature to model interaction between areal units. Butts (2006) has further shown that the spatial Bernoulli graphs can be written as a special case of a more general curved exponential family of graph distributions. By defining canonical parameters \(\eta(\theta, d) = \logit \mathcal{F}(d, \theta)\), we may write the pmf for adjacency matrix \(Y\) with support \(\mathcal{Y}\) as

\[
\Pr(Y = y | D, \theta, \psi) \propto \exp \left( \sum_{i,j} \eta(\theta, D_{ij}) y_{ij} + \psi^T t(y) \right),
\]

where \(\psi \in \mathbb{R}^p\) and \(t: \mathcal{Y} \to \mathbb{R}^p\) are respective vectors of parameters and sufficient statistics. The incorporation of additional statistics (via \(t\)) allows for the combination of both spatial and non-spatial effects (e.g., endogenous triangulation, as explored in recent work by Daraganova and Pattison (2007)).

Implications for Cross-Sectional Structure

Employing this model family with population data from the U.S. Census, we have explored the impact of geographical variability on the structure of large-scale interpersonal networks. A basic observation regarding the distribution of humans across geographical space is that this distribution is extremely heterogeneous. Even leaving aside the contrast between inhabited lands and uninhabited oceans (comprising the majority of Earth’s surface area), settlements are typically concentrated in a small set of regions having desirable geological, hydrological, and resource access properties. Within these regions, the resulting settlements are of extremely uneven size, distribution, and structure (Zipf, 1949; Brakman et al., 1999; White et al., 2008). Contrary to the intuition of an evenly inhabited Earth, then, humans are distributed unevenly across a wide range of geographical scales. This variability has important consequences for network structure.

As expected, the wildly unequal distribution of population across space leads to dramatic differences in local connectivity and tie volume. This is graphically illustrated in Figure 1, which shows simulated ties among individuals in blocks near the center of Cookeville, TN based on a model calibrated to data on friendship ties. While activity is present throughout the region, the intense clustering of persons in blocks like that near the center of the figure creates a corresponding social cluster whose members have both higher mean degree and who are on average more cohesively connected than those in nearby blocks. Even at scales on the order of 1km, we thus expect to see substantial heterogeneity in structural characteristics that are driven in part by geographical variation.

The unequal concentration of tie volume can have subtle implications for cohesion. For instance, Figure 2 shows the convex hulls covered by members of cohesively connected subsets of the 2-core of the Cookeville, TN network. We have shown that such groups develop relatively suddenly when a sufficiently large area exceeds a characteristic threshold density; the location of large cores “covering” the high-density regions of the figure is emblematic of this behavior. Such spatially large cohesive sets are of potential interest for theories such as those of Sampson et al. (1997), which relate to the

---

*This research was supported by NSF award BCS-0827027 and ONR award N00014-08-1-1015.

†Department of Sociology and Institute for Mathematical Behavioral Sciences; University of California, Irvine; SSPA 2145; Irvine, CA 92697-5100; buttes@uci.edu

‡This talk reflects work jointly authored with Zack Almquist (UC Irvine), John R. Hipp (UC Irvine), Ryan M. Acton (UMass Amherst), and Nicholas N. Nagle (UT Knoxville).
ability of social groups to monitor and control activities within a given area. Models of the kind studied here suggest a relatively sharp boundary between the conditions under which such cohesion is feasible, and those under which it is not.

Implications for Network Dynamics

It should be emphasized that the effects of geographical variability are in no way limited to the static case. For instance, we have also investigated the role of geography in shaping the emergent multi-organizational network (EMON) of collaborative relationships that formed in response to the 2005 Hurricane Katrina disaster. Figure 3 depicts the headquarter locations of the 1,577 organizations mobilized within the first 13 days following storm formation, with edges connecting those organizations who were observed collaborating on response-related tasks during the period. As shown by Butts and Acton (2011), pre-disaster headquarter location is a strong influence on tie formation, even given the dynamic nature of the network.

This marginal relationship does not tell the whole story, however. Modeling of the dynamics of the Katrina EMON reveals that factors such as proximity to the evolving storm track (Figure 3, blue curve) were important predictors of mobilization in the disaster, with immediate effects on tie formation. Thus, not only was the distribution of organizational headquarter locations important as a general factor encouraging or inhibiting collaboration (in the sense of a global propinquity effect), but this distribution was also consequential in determining which particular organizations were mobilized at any given time (and, hence, which pairs of organizations were available for collaboration). Where networks emerge in response to events that are localized in time and space, the geographical properties of the events themselves become significant influences on network structure. These influences may manifest themselves both in effects on the propensity of actors to form or dissolve ties, and on the likelihood that particular actors will be active in the first place (a powerful and generally underappreciated determinant of network structure).

Summary

Our experiments with extrapolative network simulation using detailed population data have shown that spatial variability exerts substantial influence on network structure at the settlement level. The highly uneven density of population within typical settlements results in “lumpy” networks that are characterized by regions of differential local connectivity, spatially correlated gradients of expected degree and core number, and other such properties. At small spatial scales, then, we predict that the character of the local structural environment will – for many types of relations – depend heavily on local population distribution.

While spatial heterogeneity does induce substantial within-network heterogeneity, we also observe that geography drives many aggregate network properties in a predictable way. For the relatively proximate relations we have examined in our work, properties such as aggregate mean degree, edge length, and local clustering can be well-predicted by the mean nearest-neighbor distance, together with SIF-specific factors. This implies that, for these sorts of relations, it should be possible to predict differences in a number of aggregate structural properties from fairly basic features of the underlying social geography.

The study of geographical effects on network dynamics is still in its infancy, owing in large part to a lack of available data. However, we have found in studying cases such as the Katrina EMON that the spatial distribution of both actors and external stimuli (e.g., an evolving hazard) can shape tie formation and the dynamic composition of the vertex set. It is clear that both types of effects will need closer study before their impact on network evolution can be well-understood.
1 References


