



## Mapping family connectedness across space and time

Caglar Koylu, Diansheng Guo, Alice Kasakoff & John W. Adams

To cite this article: Caglar Koylu, Diansheng Guo, Alice Kasakoff & John W. Adams (2014) Mapping family connectedness across space and time, *Cartography and Geographic Information Science*, 41:1, 14-26, DOI: [10.1080/15230406.2013.865303](https://doi.org/10.1080/15230406.2013.865303)

To link to this article: <https://doi.org/10.1080/15230406.2013.865303>



Published online: 04 Dec 2013.



Submit your article to this journal [↗](#)



Article views: 733



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 5 View citing articles [↗](#)

## Mapping family connectedness across space and time

Caglar Koylu\*, Diansheng Guo, Alice Kasakoff and John W. Adams

*Department of Geography, University of South Carolina, 709 Bull Street, Columbia, SC 29208, USA*

*(Received 25 July 2013; accepted 25 October 2013)*

Understanding the structure and evolution of family networks embedded in space and time is crucial for various fields such as disaster evacuation planning and provision of care to the elderly. Computation and visualization can potentially play a key role in analyzing and understanding such networks. Graph visualization methods are effective in discovering network patterns; however, they have inadequate capability in discovering spatial and temporal patterns of connections in a network especially when the network exists and changes across space and time. We introduce a measure of family connectedness that summarizes the dynamic relationships in a family network by taking into account the distance (how far individuals live apart), time (the duration of individuals' coexistence within a neighborhood), and the relationship (kinship or kin proximity) between each pair of individuals. By mapping the family connectedness over a series of time intervals, the method facilitates the discovery of hot spots (hubs) where family connectedness is strong and the changing patterns of such spots across space and time. We demonstrate our approach using a data set of nine families from the US North. Our results highlight that family connectedness reflects changing demographic processes such as migration and population growth.

**Keywords:** space-time visualization; family connectedness; network measure; social network; family tree

We dedicate this article to John W. Adams whose idea it was to make use of published genealogies for social science research, and who developed the data we use in our article. He worked on compiling and organizing the data throughout his lifetime, adding information from other sources and obtaining funding to expand it. He was one of the first to see the usefulness of genealogical material for the social sciences. He understood how important longitudinal data was for understanding social processes and documenting the lives of migrants. His wisdom made it possible for us to study the expansion of families over several hundred years.

### Introduction

The interaction between geography and social relationships has long been studied by researchers (e.g., Festinger, Schachter, and Back 1963; Hägerstrand 1976; Michelson 1970). Due to the wide use of social networking applications (e.g., Facebook and LinkedIn) and genealogy applications (e.g., Family Search and Ancestry), large social networks with geographic information have become increasingly available. Using such data, recent studies have proposed new ways of quantifying relationships, some of which make use of geography to infer social interactions (Backstrom, Sun, and Marlow 2010; Crandall et al. 2010), while others examine how geography and migration (or movement) influence relationships between individuals (Onnela et al. 2011; Phithakkitnukoon et al. 2011). Understanding of how

relationships (e.g., kinship, friendship) evolve across space and time is crucial for decision making in various fields such as disaster evacuation planning and provision of care to the elderly.

In a social network, each individual is represented by a node and each edge represents the relationship between two individuals. The weight of an edge can be quantified in a variety of ways such as the degree of kinship in a family tree; co-authorship in a scientific collaboration network; and the frequency of phone calls, text messages or e-mails exchanged in a communication network. A social network is dynamic because it evolves (changes) over space and time as individuals move (migrate), new individuals are added or removed, and relationships develop and change over time. In this article, we use the term “dynamic geo-social network” to refer to a dynamic social network embedded in space and time. Understanding the changing aspects of a dynamic geo-social network requires methods that can simultaneously account for the spatial, temporal and relational (network) dimensions of the network.

In order to understand the dynamics of social networks embedded in geographic space, a variety of computational and statistical methods such as graph theoretical measures (Scellato et al. 2011), random graph modeling (Schaefer 2012), factor analysis (Hipp, Faris, and Boessen 2012), simulation (Butts et al. 2012), and regression analysis (Viry 2012) have been introduced by studies in social networks. A similarity between these studies is that they

---

\*Corresponding author. Email: [koylu@email.sc.edu](mailto:koylu@email.sc.edu)

consider geography as a background variable to interpret the results of network analysis. However, the methodologies introduced by these studies have limited capability in analyzing the spatial, temporal and relational aspects of dynamic geo-social networks.

With the advancement of graph drawing algorithms, current methods of graph visualization (Lewis, Gonzalez, and Kaufman 2012; Patil 2011) are effective in discovering network (connection) patterns, e.g., clusters of connected members, or commonalities between friends who share interests and groups in a social networking application. However, existing graph visualization methods are inadequate for discovering the spatial and temporal patterns in social networks. On the other hand, spatiotemporal visualization methods (Andrienko et al. 2010; Fyfe, Holdsworth, and Weaver 2009) have been successfully applied to identify temporal variation of spatial patterns, which often do not adequately consider the network dimension (connections between individuals). Therefore, there is still a lack of methodology that can incorporate the relational aspect (sophisticated relations between individuals) of geographically embedded and time-varying social networks.

We introduce a measure and mapping approach to analyze connectedness in a dynamic family network and its changing patterns across space and time. Our approach differs from the current methods in that it takes into account the time that each pair of individuals spend together, the distance that they live apart, and the strength of their relationship (e.g., the degree of kinship). To demonstrate the approach, we use a data set of family trees derived from the published genealogies of nine families in the US North over a span of 300 years. The data also include information on migration of individuals. The remainder of the paper is organized as follows. First, we review the related work in the next section. We then introduce our data and describe our methodology in detail. Finally, we present the results and conclude with a summary and a discussion for the future research.

### Related work

This article introduces a methodology to understand the spatial, temporal and relational (network) aspects of a dynamic geo-social network. A dynamic geo-social network evolves (changes) over space and time as the actors of the network move (migrate), new actors are added or removed, and relationships between the actors develop and change over time. We demonstrate our approach using a dynamic family network embedded in space and time. Previous approaches to analyzing geo-social networks span a variety of themes and methodologies. In this section, we review the studies that aim at bridging social network analysis and spatial analysis in certain aspects.

### Computational and statistical methods

There are various studies on geo-social networks within the social network domain. For example, a number of studies (Daraganova et al. 2012; Lomi and Pallotti 2012; Sailer and McCulloh 2012; Schaefer 2012) used exponential random graph models to account for geographic embeddedness of individuals in modeling social networks and investigate the effects of social and spatial distance on the network structure. Doreian and Conti (2012) analyzed a set of empirical networks to understand how networks are shaped by social and spatial contexts using a variety of modeling strategies. Butts et al. (2012) conducted an exploratory simulation study to examine the influence of spatial variability of background population on the network structure and the social ties.

Viry (2012) examined the relationship between spatial dispersion of personal networks, residential mobility and network composition by conducting regression analyses. Cho, Myers, and Leskovec (2011) and Scellato et al. (2011) focused on online geo-social networks to describe the relationship between geography and social interaction using graph theoretical methods. Similarly, Radil, Flint, and Tita (2010) introduced a spatialized positional analysis to reveal spatial patterns of social relations. Hipp et al. (2012), and Mennis and Mason (2012) performed factor analyses to delineate neighborhood boundaries by taking into account the density of social ties and the physical distances between the members of a social network. A similarity between the studies that focus on geo-social networks in the social network domain is that they consider geography as a background variable to interpret the results of network analysis. However, the methodologies introduced by these studies have limited capability in analyzing the spatial, temporal, and relational aspects of dynamic social networks.

### Visualization

Alternative to modeling, graph theoretical and statistical approaches, network visualization methods have been developed to examine the dynamic nature of social networks. Dynamic network visualization methods allow the discovery of complex patterns in a network over time using animation (network movies) (Moody, McFarland, and Bender-deMoll 2005) and “small multiple displays” (Robertson et al. 2008). The layout of a graph is constructed by a graph drawing algorithm which often places nodes (individuals) that have strong relationships closer to each other. To enhance the perception of changes in a sequence of graph layouts, a collection of methods are developed by considering additional criteria such as minimizing edge crossings and ensuring repeatability and stability (Bender-deMoll and McFarland 2005). However, such graph layouts represent only the topological structure of the network while disregarding its geographic dimension.

To incorporate a geographic dimension into the network space, a number of studies (Faust et al. 2000; Nag 2009; Todo et al. 2011) mapped actors (people) based on their geographic location and drew edges between those actors using different width and color intensity to reflect relative strength of each relationship. However, a graph layout that positions nodes based on their geographic coordinates suffers from the visual cluttering problem. Moreover, with a relatively large network, it is difficult to perceive network structures that involve multiple dimensions (i.e., space, time, and social connections). Because social network data are highly dynamic, it is challenging to reveal how social relationships change across geographies and time by simply displaying a sequence of graphs.

Alternatively, some studies (Luo et al. 2011; Onnela et al. 2011) introduced integrated approaches that use dynamically linked views of network space and geographic space and allow user interactions to demonstrate the interplay of topological structure and geography. Discovering the interaction between geography and the network is useful in extracting microscale (individual level) patterns. However, there is also a need to summarize spatial, temporal, and relational aspects of such networks in order to provide a general overview of the data.

Hägerstrand (1976) introduced a space-time framework to conceptualize and represent human interactions over space and time. Adopting this framework, many spatiotemporal visualization approaches (e.g., space-time path, density surface, computational, and interactive approaches) have been introduced (Aigner et al. 2011). The space-time path approach (Chen et al. 2011; Lee and Kwan 2011) identifies human activity patterns in a social network by visualizing individuals' paths in a three-dimensional surface. Alternatively, the density surface approach summarizes the activity patterns by a density surface which is represented with either an animated sequence of continuous surfaces (Rana and Dykes 2003) or a three-dimensional surface of the space-time continuum (Demšar and Verrantaus 2010; Nakaya and Yano 2010). Additionally, some computational and interactive approaches such as self-organizing maps (Agarwal and Skupin 2008) have been used to identify temporal variation of spatial patterns.

The space-time approach by Shaw, Yu, and Bombom (2008), Fyfe, Holdsworth, and Weaver (2009) and Andrienko et al. (2010) examine geo-social interaction patterns across space and time, but it does not adequately consider the network dimension (connections between individuals). Therefore, there is still a lack of methodology that can incorporate the relational aspect (sophisticated relations between individuals) of geographically embedded and time-varying social networks. Another challenge in analyzing geographically embedded and time-varying social networks is the small area problem,

where a single node or connection is often too small (with insufficient data) for deriving stable statistical measures. Koylu and Guo (2013) introduced a smoothing approach to mapping graph measures in geographic space. In this research, we introduce a different space-time smoothing or interpolation method for visualizing both network measures and social relations in space and time.

## Data

To demonstrate our approach, we use family tree data derived from published genealogies of nine families from the US North over a span of 300 years. These books were compiled by family members with the help of professional genealogists. More information on migration has been added by linking the genealogies to the US censuses using data from Ancestry.com. A series of demographic events (e.g., births, deaths, migrations) were coded from the genealogy, including the places where events had occurred. From these event locations and dates, we can infer the migration paths of each individual in the families.

For the simplicity of methodology presentation and result explanation, in this article we only report the analysis results with the Chaffee family (Chaffee 1909), which was selected over eight other genealogies on the basis of better temporal resolution and information on migration. The Chaffee family includes 1225 males descended in the male line from the founder who came to Hull, Massachusetts from England in 1635. All men born into the family up to 1860 were included along with all siblings of men born through 1840. There were 2387 geocoded moves and 856 distinct locations where the family members lived in 296 years.

The family data involve only males because women changed their names at marriage; they were more difficult to follow. Although life expectancy changed over time, it was largely due to changes in infant and child mortality (Kasakoff and Adams 2000). This study included only men who survived to at least age 20. If we included those who had died young, we might have biased the study toward families with high infant and child mortality. Life expectancy at age 20 was remarkably moved westward, albeit at greater and greater distances (Egerbladh, Kasakoff, and Adams 2007).

Information on moves comes from records of vital events. If an event occurred in a place where a person had not previously lived, the move was assigned at a date close to the vital event. Most moves occurred before the vital event and thus the dates are approximate. The most accurate move dates come from the child bearing years because this population had children approximately every two years. Also only about 65% of the men had death dates recorded in the genealogy. For the rest, the last date on record was considered a death date. The animation of the migration of nine families including the Chaffee (CFE)

family in the US can be viewed at the link: <http://129.252.37.169:8400/flowvis/trajectories/index.html> (Koylu 2013a).

## Methodology

We introduce a measure and mapping approach to analyze the relationships in a family network embedded in space and time. Given a space-time window, the measure quantifies the family connectedness of each individual, considering his/her kinship to other family members coexisted in the window, their geographic distances, and the time duration of their coexistences. We then interpolate the measure values for all locations, map a series of space-time windows to examine the changing dynamics of the family relationships across space and time. Specifically, the approach consists of three steps. First, the time dimension is partitioned into a sequence of time intervals. Second, within each time interval, we calculate the measure of family connectedness for each individual at each location where he/she was present, considering the closeness (the degree of kinship) of his/her connections he/she has within a geographic distance threshold and the temporal duration of each connection. Third, given the family connectedness value for each individual at each unique location within a time window, a surface of family connectedness is produced using a smoothing and interpolation method based on inverse-distance weighting. In the following subsections, we introduce each of the steps.

### Time interval

To allow for a temporal analysis of connectedness in a family network, one can employ a data-driven approach such as sliding windows, top-down or bottom-up segmentation algorithms (Keogh et al. 2001; Warren Liao 2005) to obtain time intervals. For our case study, we employed a domain-specific approach to partition time series data into equal intervals and reflect meaningful stages of the family tree data. Because some patterns may fall between time windows and not appear, we use a sliding window approach.

In the family data set, the minimum period needed for a connection (coexistence) to occur is one year. On average a man is 35 years old when a son is born and 20 years is nearly the smallest generation, i.e., the youngest a man might be when he has a son. Also, a period of 20 years divides the life course into meaningful stages: age 1–20 would be before marriage; child bearing should stop by age 60 (Adams and Kasakoff 1984). So people in different 20-year windows should be in different life stages. Therefore, we partition the data into a time window (interval) length of 20 years. Theoretically, we can move this 20-year window one year at a time to obtain a smooth time series. To reduce the size of time series (and data

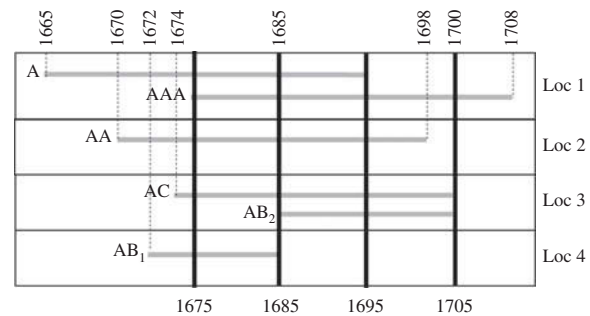


Figure 1. A sample subset of a family network. The horizontal axis illustrates time and the vertical axis represents unique locations (i.e., Loc 1, Loc 2, Loc 3, and Loc 4). An individual at a location is represented with a horizontal line with a beginning and an ending year. For example, AB1 refers to the period that AB lived at location 4 between 1672 and 1685, whereas AB2 refers to the period that AB lived at location 3 between 1685 and 1700.

redundancy), we move the window 10 years each step. In other words, there is a 10-year (i.e., 50%) overlap between neighboring time windows.

Figure 1 shows a sample subset of a family tree data to illustrate the measure calculation. The horizontal axis represents the time periods of individuals (i.e., grandfather A, father AA, uncle AB, uncle AC, and son AAA), whereas the vertical axis represents the locations (i.e., Loc 1, Loc 2, Loc 3, Loc 4) of those individuals in those time periods. For example, AC lived in Location 3 between 1674 and 1700, whereas AB lived in Location 4 between 1672 and 1685, moved to Location 3 and lived there between 1685 and 1700. Additionally, the solid vertical lines represent the beginning and the end of time intervals: 1675–1695 and 1685–1705.

### Family connectedness

We argue that a potential spatial interaction between two individuals in a time period is often dependent on how close those individuals are to each other both in terms of their geographic and kin proximity. While we use geographic proximity to form a territory of potential spatial interaction for each individual, we conceptualize kin proximity by the closeness of the relationship (e.g., degree of kinship) between two individuals. Naturally, the potential for spatial interaction between individuals change across time as individuals move, new individuals are added or removed, and relationships develop and change over time. By taking into account the time-varying relationship between geographic and kin proximity between individuals, and the time duration of their coexistence, we introduce a measure of family connectedness as a proxy for potential spatial interaction.

For each time window, we derive the territory of each individual by using a geographic distance threshold

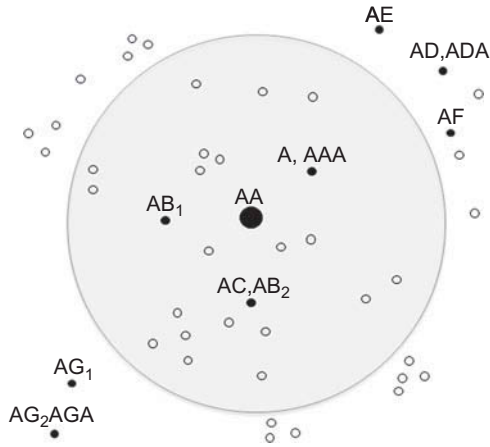


Figure 2. The potential connections of individual AA from the sample network given in Figure 1. The circular buffer illustrates the neighborhood of individual AA which is used to determine his/her potential connections. Nodes with labels within the neighborhood are potential connections of AA, whereas empty node symbols and labeled nodes outside the neighborhood are individuals that are not connected to AA. A subscript (e.g., AB<sub>1</sub>, AB<sub>2</sub>) for an individual indicates his/her existence at each unique location given the time interval.

around his location at the time and then calculate the family connectedness of an individual by considering his geographic closeness, temporal overlapping and family relationship to other individuals within the territory. Figure 2 illustrates the individual AA's family connections that are determined by his territory (gray circle). We provide a discussion on how to determine the territory of individuals using the distance threshold in the following paragraph. While the nodes with labels illustrate

connections of individual AA, empty nodes illustrate individuals that do not have any family relationship with the individual of interest. For the family tree data in this study, we define relationship as kinship and two individuals do not have a relationship if they are not members of the same family tree. For individual AA at location 2, he had five family connections, which are A, AAA, AC and AB (at two locations, noted as AB<sub>1</sub> and AB<sub>2</sub>) for the given time interval 1675–1695. Notice that, although individuals such as AD, AE, ADA, and AF were from the same family with AA, they are not considered as connections because they lived outside the neighborhood buffer of AA.

The choice of the distance threshold (bandwidth) and what constitutes a connection are two important decisions for determining potential connections of an individual at a location and time. To select an appropriate bandwidth, we evaluated the distribution of move distances over time. Figure 3 illustrates the box-plots of distances by time intervals. Migration was highly skewed toward shorter distances, as is always the case. Over time, the longest distances increased, but moves at such distances were relatively rare and overall median distance is approximately 60 km. Still these distances are much greater than they were in Europe (Pooley and Turnbull 1998), where population density was higher and people were more apt to remain in their local areas and reflect the Westward expansion of the US population.

Considering the temporal resolution of the data which is composed of recorded events from the late seventeenth century till the mid-twentieth century, increasing trend of migration distance could be attributed to what transportation medium was available for the given time period. Until the mid-nineteenth century when the first railway system

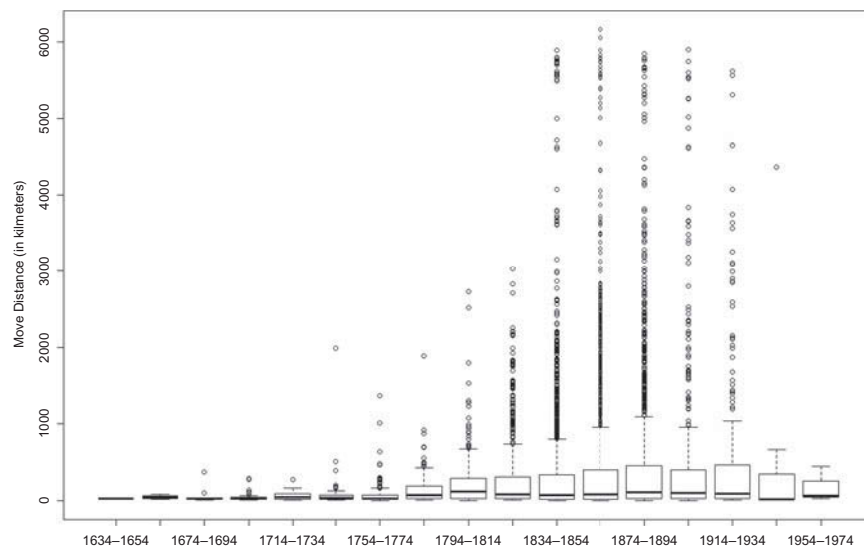


Figure 3. The distribution of move distances over time intervals. The median move distance is approximately 60 km and there is an increasing trend of individuals moving greater distances over time.

was built in the northeastern states, traveling was limited to the capability of horse carriages. Along the railway lines the ability to travel long distances greatly increased. However, horse carriage remained to be a major transportation medium. On average horse quality, terrain and weather conditions, a horse carriage was able to travel 32–64 km a day (Bogart 2005). Assuming that a potential for a consistent spatial interaction is possible without moving homes, we chose 60 km as a threshold distance to identify potential connections for each individual. The second important decision is to determine what constitutes a connection between individuals in a family network. In this study, we define connection as kinship and we assume that two individuals are connected if they are from the same family tree. Given the connections, we use Equation (1) to calculate each individual’s family connectedness at a specific time interval and a specific location

$$FC_{rt}(i) = \sum_{j \in N_{rt}} T_{rt}(i,j) * KP(i,j) \quad (1)$$

where  $FC_{rt}(i)$  is the family connectedness for individual  $i$  at location  $r$  and in time interval  $t$ . Individual  $i$  may have more than one locations (one at a time) in the time interval.  $N_{rt}$  are family members within the neighborhood of the individual  $i$ ’s location ( $r$ ) and the time interval  $t$ ;  $T_{rt}(i,j)$  is the duration of time that individuals  $i$  and  $j$  coexisted within the neighborhood of  $r$  and the time interval  $t$ ;  $KP(i,j)$  is the kin proximity which describes the degree of kinship between the family members  $i$  and  $j$ .

We use consanguinity (Leutenegger et al. 2011) to quantify the degree of kinship (relation) between the members of a family, which is widely used in law and genetics. Figure 4 represents a family tree of four generations where A is the ancestor of all members in the family. The relation among two people is called lineal consanguinity if one is descendant from the other such as the son and the father (e.g., A-AA), or the grand-father (e.g., A-AAA), and so upward in a direct ascending line. The

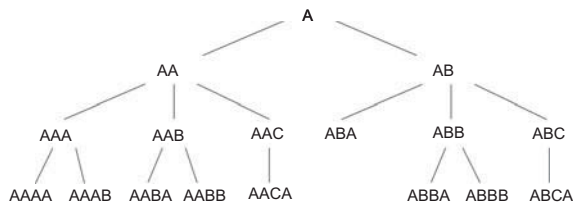


Figure 4. A sample family tree with four generations that descend from the ancestor, A. The relation among two people is called lineal consanguinity if one is descendant from the other such as the son and the father (e.g., A-AA), or the grandfather (e.g., A-AAA), and so upwards in a direct ascending line. For people who descend from the same ancestor, but not from each other (e.g., cousins or uncles-nephews), the relation is called collateral consanguinity.

degree of lineal consanguinity is directly measured by the number of lines (e.g., edges in Figure 4) between the two family members. For example, father-son relations (e.g., A-AA, AAB-AABA) are first degree; grand-father-grandson relations (e.g., A-AAA, AB-ABBB) are second degree, and great grandfathergreat grandson relations (A-AAAA, A-ABBA) are third degree.

The relation between individuals who descend from the same ancestor, but not from each other (e.g., cousins or uncles-nephews) is called collateral consanguinity. The degree for collateral relationship is calculated by finding the common ancestor then counting the number of steps downwards to reach the two individuals. If one of the individuals is more distant (remote) to the ancestor, the number of steps to the more remote person determines the degree of consanguinity. For example, a relation between brothers (e.g., AA-AB, AAB-AAC) is considered as a first degree consanguinity since there is only one step from the father to each of them, whereas an unclenephew relation (e.g., AA-ABA, AAB-AACA) is a second degree consanguinity because the nephew is two steps away from the common ancestor, and the rule of calculating the degree is extended to the more remote person of the collateral relationship. After determining the degree of relation (consanguinity) between two individuals, we assign a kin proximity value to each relation by simply taking the inverse of the degree. For example, the kin proximity of a first degree relationship (e.g., father-son, brothers) is  $1/1 = 1$ , whereas the kin proximity of a second degree relationship (e.g., grandparent-grandchildren, cousins) is  $1/2 = 0.5$ , and a third degree relationship (e.g., great uncle/grandnephew: AA and ABBA) is  $1/3 = 0.33$ , and so on.

**Spatial interpolation of family connectedness**

The components of the measure, which are cumulative kinship and time for an individual at a location, are highly correlated with the presence of individuals that live within a close distance to that location. We discuss that more people living close by increases the chance of potential interactions, thus the correlation between the presence of individuals and the measure components is appropriate and does not necessitate normalization. As we are not interested in family connectedness as a cumulative measured quantity, we produce a geographically weighted average surface of family connectedness by using a spatial smoothing and interpolation method rather than a cumulative density surface of family connectedness.

Given the family connectedness for each individual at each unique location within a time window, a surface of family connectedness is produced using a smoothing and interpolation method based on inverse-distance weighting (IDW). IDW assumes that each measured value has an influence on the prediction by applying weights that are

proportional to the inverse of the distance between the prediction location and the measured data point. The equation for smoothing and interpolation method is given below:

$$FC(x, t) = \sum_{i \in N_{xt}} \frac{w_i(x) FC_{r \in N_{xt}}(i, t)}{\sum_{j=0}^N w_j(x) + W_c} \quad (2)$$

$$w_i(x) = \frac{1}{d(x, r_i)^p} \quad (3)$$

where  $FC(x, t)$  is the interpolated value of family connectedness at location  $x$  in time interval  $t$ ;  $N_{xt}$  is the observations (existence of individuals at unique locations) within the neighborhood of  $x$  in time interval  $t$ ;  $FC_{r \in N_{xt}}(i, t)$  is the family connectedness value for individual  $i$  at location  $r$  in time interval  $t$ ;  $w_i(x)$  is a weighting function based on the distance  $d(x, r_i)$  from the location of the observation  $r_i$  to the unknown point  $x$ ;  $p$  is a positive real number called the power parameter;  $W_c$  is a constant penalty weight added to each estimation to remove the edge effect (Lawson et al. 1999).

We determine the neighborhood for each estimation point  $x$  by using the same distance threshold (60 km) we used in the previous step to identify connections between individuals. Because there are many observations (the coexistence of individuals) at the same or close by locations, the traditional IDW creates an interpolated surface which is greatly influenced by the edge effect (Figure 5(a)). After applying the penalty weight for locations with no or few observations the edge effect is removed (Figure 5(b)). A constant penalty weight does not have a significant effect on the estimation where there are many observed values by the estimation point; however, it does affect the estimation where there is a few or no observed points close by the estimation point. To balance between over-smoothing and under-smoothing, we empirically chose a value of 0.1 for the constant weight  $W_c$ .

## Results and discussion

We produced 29 surfaces of family connectedness each of which corresponds to a 20 year time window. Each surface was produced using a constant divergent classification scheme to enable comparison between each time window. While blue hue illustrates places with low family connectedness (i.e., low potential for spatial interactions), red hue illustrates places where family connectedness is higher.

The animations of family connectedness including all families and the Chaffee family can be viewed at the following link: <http://www.spatialdatamining.org/family-connectedness> (Koylu 2013b). For the simplicity of result explanation, in this article we only report the analysis results with the Chaffee family, which was selected over eight other genealogies on the basis of better temporal resolution and information on migration. Due to the limited space we only report a small subset of the time windows that we selected based on their relevance to historical events in chronological order.

The surfaces of family connectedness from the first time window (1634–1654) to the time window of 1854–1864 illustrate the demographic and spatial expansion of a colonizing population. Colonization proceeded in spurts with a family member moving out of the settled area and then most of his descendants remaining in the new location for three generations before spawning new settlements. It takes many years in a new location for connectedness to peak.

Before the American Revolution, the earlier window (Figure 6(a)), there is only a few individuals in the newly settled areas such as Scipio, Warren, Chittenden, Westminster and Berkshire. Sixty years later (Figure 6(b)), the core of the family stayed in the area between Woodstock and Becket while new hubs started to develop in Berkshire and Scipio as the family moved North and West after the Revolution. As compared to Chittenden and Westminster there were fewer individuals in Berkshire in 1824–1844 but Berkshire became a stronger hub than Chittenden and Westminster.

When the new hubs were created, it took several generations to achieve the degree of family connectedness

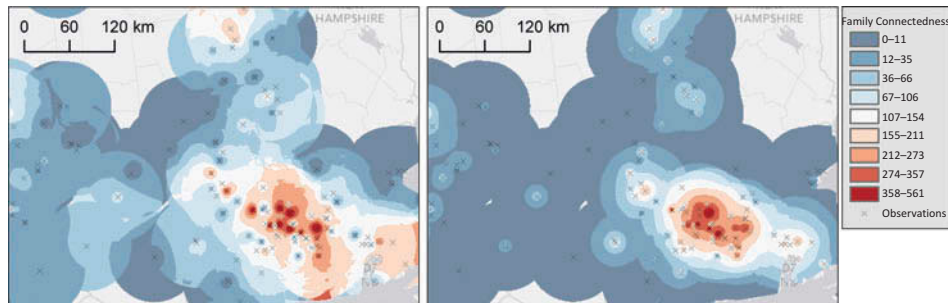


Figure 5. The comparison of the traditional IDW (a) with the modified IDW (b). The edge effect is noticeable throughout the traditional IDW surface (a). By applying additional weight that penalizes locations with no observations or few observations, the edge effect is removed in the modified IDW surface (b).



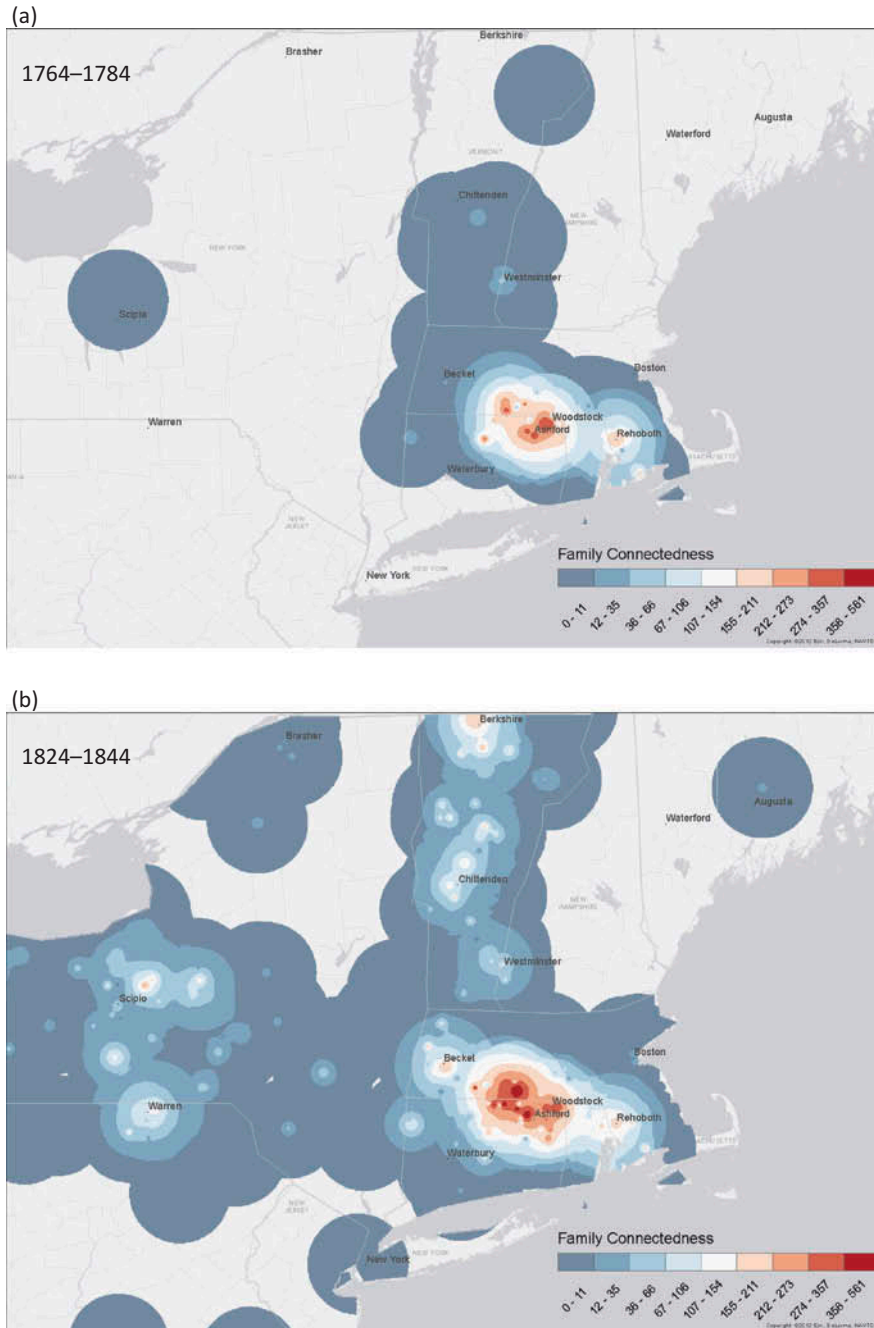


Figure 6. Family connectedness after the American revolution: 1764–1784 (a), 1824–1844 (b).

of the places that had been settled by the earliest generations. Due to the new births and new migrations of close kin into the area, Berkshire was able to increase its strength as a family hub in the later periods (Figure 7).

Family connectedness is a composite measure of shared time (coexistence of individuals in a neighborhood) and kin proximity (e.g., the closeness of their kinship). To better understand the relationship between shared time and kin proximity, one could decompose the family connectedness of an individual at a location and

time interval (Equation (1)) into its components of total shared time (Equation (4)) and total kin proximity (Equation (5)). We plot these components for two distinct time intervals (i.e., 1764–1784 and 1844–1864) to capture the temporal variation of the relationship between time and kin proximity (Figure 8). Both components are correlated with and influenced by the presence of individuals at close by locations (i.e., 60 km); thus time and kin proximity were highly correlated in both time intervals. The difference between the intervals

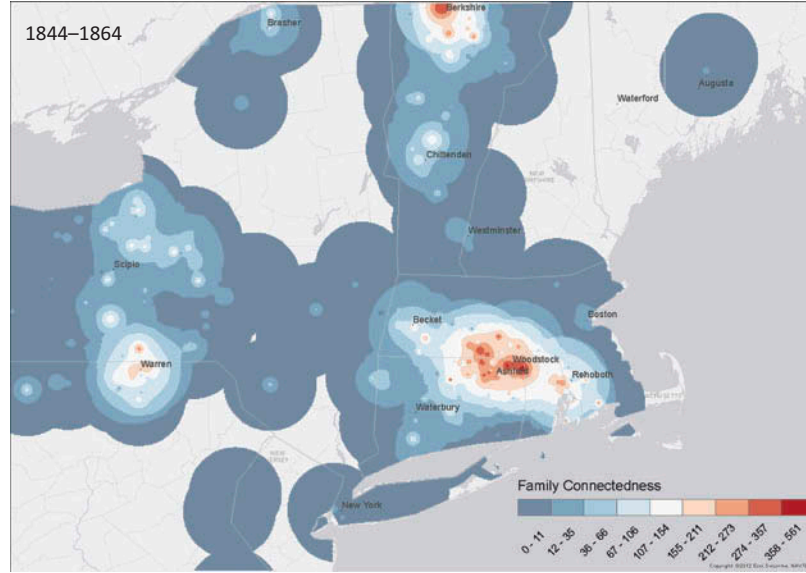


Figure 7. Family connectedness throughout the process of urbanization (1844–1864).

of 1764–1784 (Figure 8(a)) and 1844–1864 (Figure 8(b)) suggests that over time individuals spend more time in close by locations, whereas the availability of kin in their territory stayed the same. This trend could partially be explained by increased coexistence of individuals with distant kin.

$$\text{Total kin proximity}_{rt}(i) = \sum_{j \in N_{rt}} KP(i, j) \quad (4)$$

$$\text{Total shared time}_{rt}(i) = \sum_{j \in N_{rt}} T_{rt}(i, j) \quad (5)$$

The contribution of kin proximity and the shared time to the measure result vary across space and time. For example, an area with high family connectedness might

be a result of high shared time but low kinship due to the coexistence of a large group of distant relatives (e.g., cousins, 2nd level cousins). In an opposite case, an area with high family connectedness might be a result of low shared time but high kinship because of the coexistence of close relatives (e.g., parent-children, siblings) in shorter periods of time.

To examine the spatial variation of the relationship between shared time and kin proximity, we performed bivariate local indicators of spatial association (LISA) (Anselin 1995). Bivariate LISA examines whether local correlations between values of a variable (e.g., time) at a location and those of its neighboring values of another variable (e.g., kinship) are significantly different from what you would observe under conditions of spatial randomness. For example, a significant low-high cluster means that low values of a variable such as shared time

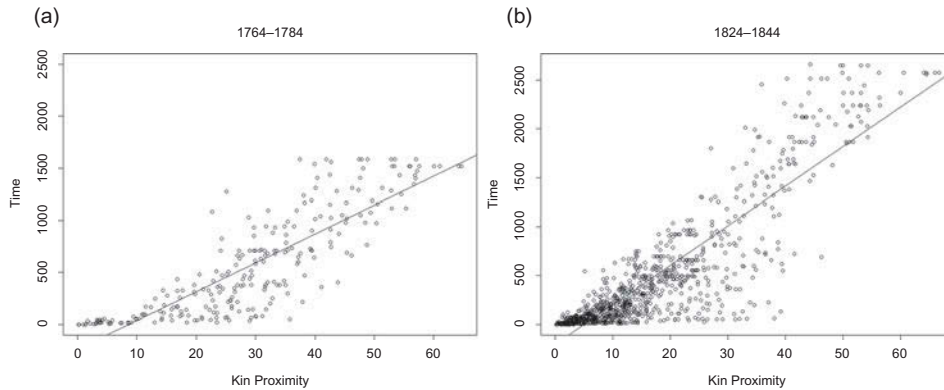


Figure 8. Relationship between the total shared time and the total kin proximity (kinship) of each individual's connections within a neighborhood in time intervals 1764–1784 (a) and 1824–1844 (b). The vertical axis represents the total shared time, whereas the horizontal axis represents the total kinship.

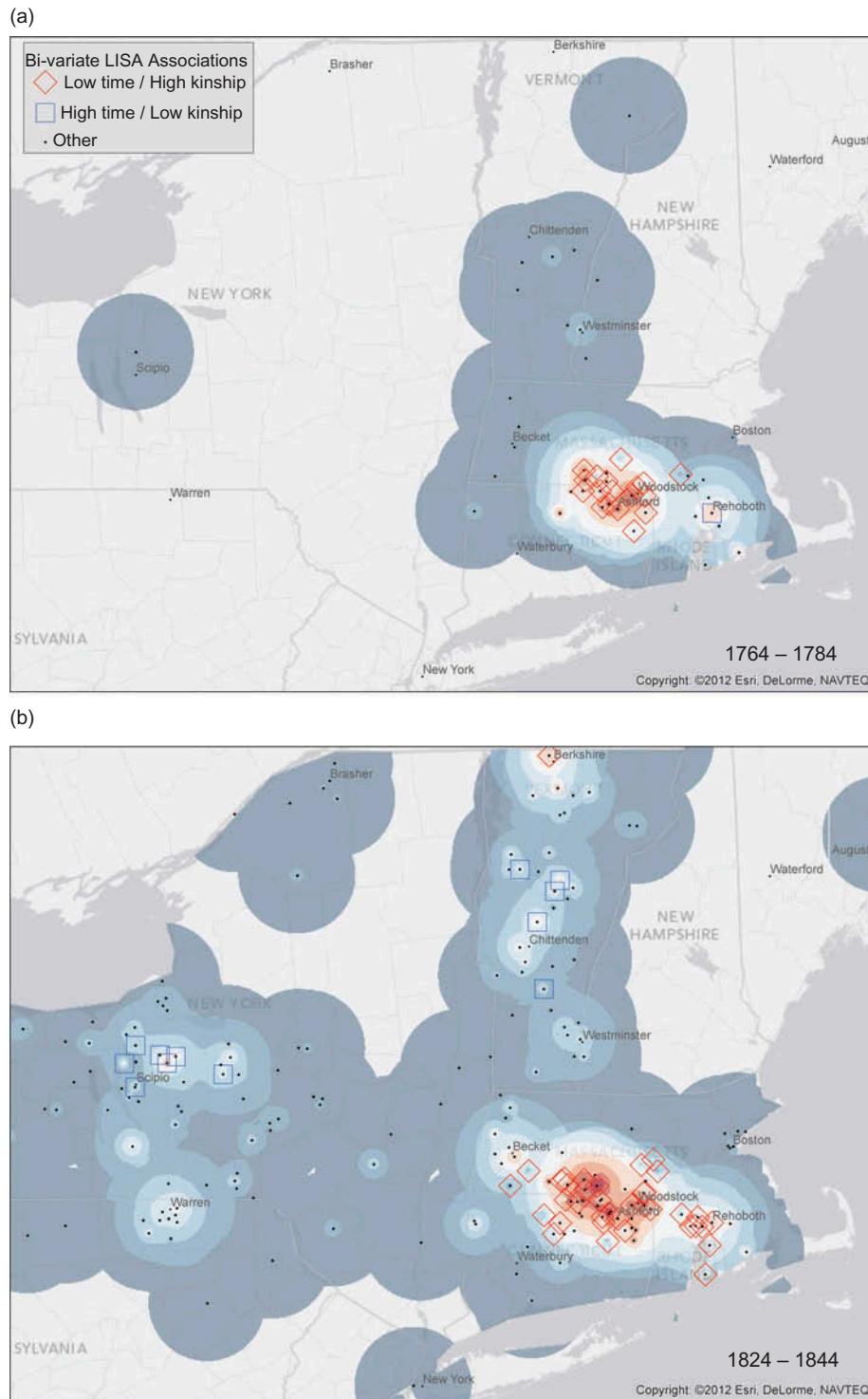


Figure 9. High-low and low-high associations of shared time and kinship in time intervals 1764–1784 (a) and 1824–1844 (b). While red diamonds represent low shared time and high kinship, blue squares represent high shared time and low kinship. The contrasting associations of cumulative shared time and kinship vary across space and time throughout the spatial and demographic expansion of the population.

are significantly correlated with high neighboring values of another variable such as kin proximity.

We are particularly interested in understanding contrasting patterns of kin proximity (kinship) and shared time, thus, in this article, we only report statistically significant associations with high kinship-low time, and low kinship-high time for the time intervals 1764–1784 (Figure 9(a)) and 1824–1844 (Figure 9(b)). In the early stages of the expansion (1764–1784) we observe a cluster of low time–high kinship values especially around Ashford, whereas Rehoboth continued to be a location with high time and low kinship. In the later period after the American Revolution (1824–1844) the family hubs located around Ashford still had low time but high kinship values, but the spatial extent of the hubs became more dispersed.

Moreover, we start to see the formation of new hubs around Scipio and Chittenden which have high time but low kinship values. Low kinship and high time associations occurred in especially Scipio and Chittenden because younger individuals which were distant relatives moved to these new hubs, whereas patriarchs stayed around the old established hubs. On the contrary, we observe a contrasting pattern, high kinship–low time (red diagonals), around Berkshire in the later time period of the expansion (1824–1844). This is because Berkshire became an established hub as a result of in-migration of close relatives and high presence of patriarchs.

## Conclusion

We introduced a measure of family connectedness that summarizes the dynamic relationships in a family network embedded in space and time. The new measure is unique because it takes into account the duration of time that each pair of individuals spend together, the distance that they live apart, and the strength of their relationship (e.g., the degree of kinship). By mapping the relational aspects of a family network across space and time, our method facilitates the discovery of hot spots (hubs) where potential for spatial interaction between individuals is relatively higher across space and time.

One aspect of social context that is often not considered by the studies that incorporate social network analysis and spatial analysis is the addition or removal of members (e.g., birth/death, entry/exit) of the network. Our work has shown that especially deaths of individuals (e.g., patriarchs) who link many others together can greatly affect family connectedness particularly in the locations where those individuals live. The family data set we analyzed had a very high rate of demographic increase as was characteristic of the US North at the time. If death rates were higher, presumably there would be fewer hubs or hot spots. In other words, hot spots

would disappear much more quickly with those in Europe where death rates were much higher.

We are studying potential, not actual, interaction and this is one limitation of our work. We do not have a measure of actual interaction that we could compare with the potential interaction we have described. But there are other historical data sets which do have such measures. One example would be data on witnesses to marriages or other events, which exist for several European countries in the past (Bras 2011). Our measure could also be computed using current kinship networks and then compared with the actual interactions of family members obtained from questionnaires or by other means.

We demonstrated our approach using a family tree data set from a population that was growing and colonizing the Northern part of the US. This measure has demonstrated how important migration, birth, and death of individuals to family connectedness. In this study, we define relationship as kinship and assume that two individuals have a relationship if they are from the same family. Our methodology can readily be extended to develop a measure of social connectedness using other forms of relationships, such as friendship and coworkers.

Depending upon the context of the social network, one can define the connections in any form of interaction and quantify those interactions in a variety of ways such as using the frequency of the shared content, common friends in an online social network; the number of email exchanges or meetings held together in a business network. In this regard, the voluminous data collected from social networking platforms such as Twitter, Flickr and Foursquare and genealogy applications such as Family Search and Ancestry provide an excellent opportunity to study online social networks and family trees using our approach.

## Acknowledgments

This work was supported in part by The National Science Foundation under Grant No. 0748813. We greatly appreciate the constructive comments and suggestions from anonymous reviewers. The data set for the US North has been compiled with funds from NICHD Center for Population Research for the grants “Migration and Intergenerational Processes,” and “Spatial and Family Processes in the Transmission of Wealth”; a fellowship from the National Endowment for the Humanities awarded by the Newberry Library; an EPSCOR grant and grants from the National Science Foundation (Geography Regional Science Program) and the Anthropology Program; a Newberry Library Fellowship (Spring 1987); and a research award from the College of Liberal Arts, University of South Carolina.

## References

- Adams, J. W., and A. B. Kasakoff. 1984. “Migration and the Family in Colonial New England: The View from

- Genealogies." *Journal of Family History* 9 (1): 24–43. doi:10.1177/036319908400900102.
- Agarwal, P., and A. Skupin. 2008. *Self-Organising Maps: Applications in Geographic Information Science*. West Sussex: Wiley.
- Aigner, W., S. Miksch, H. Schumann, and C. Tominski. 2011. *Visualization of Time-Oriented Data*. New York: Springer.
- Andrienko, G., N. Andrienko, U. Demsar, D. Dransch, J. Dykes, S. I. Fabrikant, M. Jern, M. Kraak, H. Schimann, and C. Tominski. 2010. "Space, Time and Visual Analytics." *International Journal of Geographical Information Science* 24 (10): 1577–1600. doi:10.1080/13658816.2010.508043.
- Anselin, L. 1995. "Local Indicators of Spatial Association – LISA." *Geographical Analysis* 27 (2): 93–115. doi:10.1111/j.1538-4632.1995.tb00338.x.
- Backstrom, L., E. Sun, and C. Marlow. 2010. "Find Me if You Can: Improving Geographical Prediction with Social and Spatial Proximity." Paper presented at the Proceedings of the 19th international conference on World Wide Web, Raleigh, NC, April 26–30.
- Bender-deMoll, S., and D. McFarland. 2005. "The Art and Science of Dynamic Network Visualization." *Journal of Social Structure* 7 (2): 1–38.
- Bogart, D. 2005. "Turnpike Trusts and the Transportation Revolution in 18th Century England." *Explorations in Economic History* 42 (4): 479–508. doi:10.1016/j.eeh.2005.02.001.
- Bras, H. 2011. "Intensification of Family Relations? Changes in the Choice of Marriage Witnesses in the Netherlands, 1830–1950." *Tijdschrift voor Sociale en Economische Geschiedenis* 8 (4): 102–135.
- Butts, C. T., R. M. Acton, J. R. Hipp, and N. N. Nagle. 2012. "Geographical Variability and Network Structure." *Social Networks* 34 (1): 82–100. doi:10.1016/j.socnet.2011.08.003.
- Chaffee, W. 1909. *The Chaffee Genealogy*. New York: Privately Printed.
- Chen, J., S.-L. Shaw, H. Yu, F. Lu, Y. Chai, and Q. Jia. 2011. "Exploratory Data Analysis of Activity Diary Data: A Space–Time GIS Approach." *Journal of Transport Geography* 19 (3): 394–404. doi:10.1016/j.jtrangeo.2010.11.002.
- Cho, E., S. A. Myers, and J. Leskovec. 2011. "Friendship and Mobility: User Movement in Location-Based Social Networks." Paper presented at the Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, CA, August 21–24.
- Crandall, D. J., L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. 2010. "Inferring Social Ties from Geographic Coincidences." *Proceedings of the National Academy of Sciences* 107 (52): 22436–22441. doi:10.1073/pnas.1006155107.
- Daraganova, G., P. Pattison, J. Koskinen, B. Mitchell, A. Bill, M. Watts, and S. Baum. 2012. "Networks and Geography: Modelling Community Network Structures as the Outcome of Both Spatial and Network Processes." *Social Networks* 34 (1): 6–17. doi:10.1016/j.socnet.2010.12.001.
- Demšar, U., and K. Verrantaus. 2010. "Space–Time Density of Trajectories: Exploring Spatio-Temporal Patterns in Movement Data." *International Journal of Geographical Information Science* 24 (10): 1527–1542. doi:10.1080/13658816.2010.511223.
- Doreian, P., and N. Conti. 2012. "Social Context, Spatial Structure and Social Network Structure." *Social networks* 34 (1): 32–46. doi:10.1016/j.socnet.2010.09.002.
- Egerbladh, I., A. B. Kasakoff, and J. W. Adams. 2007. "Gender Differences in the Dispersal of Children in Northern Sweden and the Northern USA in 1850." *The History of the Family* 12 (1): 2–18. doi:10.1016/j.hisfam.2007.05.001.
- Faust, K., B. Entwisle, R. R. Rindfuss, S. J. Walsh, and Y. Sawangdee. 2000. "Spatial Arrangement of Social and Economic Networks Among Villages in Nang Rong District, Thailand." *Social Networks* 21 (4): 311–337. doi:10.1016/S0378-8733(99)00014-3.
- Festinger, L., S. Schachter, and K. W. Back. 1963. *Social Pressures in Informal Groups: A Study of Human Factors in Housing*. Stanford, CA: Stanford University Press.
- Fyfe, D. A., D. W. Holdsworth, and C. Weaver. 2009. "Historical GIS and Visualization: Insights from Three Hotel Guest Registers in Central Pennsylvania, 1888–1897." *Social Science Computer Review* 27 (3): 348–362. doi:10.1177/0894439308329762.
- Hägerstrand, T. 1976. "Geography and the Study of Interaction between Nature and Society." *Geoforum* 7 (5–6): 329–334. doi:10.1016/0016-7185(76)90063-4.
- Hipp, J. R., R. W. Faris, and A. Boessen. 2012. "Measuring 'Neighborhood': Constructing Network Neighborhoods." *Social Networks* 34 (1): 128–140. doi:10.1016/j.socnet.2011.05.002.
- Kasakoff, A. B., and J. W. Adams. 2000. "The Effects of Migration, Place, and Occupation on Adult Mortality in the American North, 1740–1880." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 33 (2): 115–130. doi:10.1080/01615440009598954.
- Keogh, E., S. Chu, D. Hart, and M. Pazzani. 2001. "An Online Algorithm for Segmenting Time Series." Paper presented at the Data Mining, 2001. Proceedings of IEEE International Conference on ICDM 2001, San Jose, CA, November 29–December 2.
- Koylu, C. (Producer). 2013a. *Animation of Family Migration*. Accessed May 5. <http://129.252.37.169:8400/flowvis/trajectories/index.html>
- Koylu, C. 2013b. *Family Connectedness*. Accessed October 22. <http://www.spatialdatamining.org/familyconnectedness>
- Koylu, C., and D. Guo. 2013. "Smoothing Locational Measures in Spatial Interaction Networks." *Computers, Environment and Urban Systems* 41: 12–25. doi:10.1016/j.compenvurbsys.2013.03.001.
- Lawson, A., A. Biggeri, D. Böhning, E. Lesaffre, J.-F. Viel, and R. Bertollini. 1999. *Disease Mapping and Risk Assessment for Public Health*. Chichester: John Wiley & Sons.
- Lee, J. Y., and M. Kwan. 2011. "Visualisation of Socio-Spatial Isolation Based on Human Activity Patterns and Social Networks in Space-Time." *Tijdschrift voor economische en sociale geografie* 102 (4): 468–485.
- Leutenegger, A. L., M. Sahbatou, S. Gazal, H. Cann, and E. Génin. 2011. "Consanguinity Around the World: What Do the Genomic Data of the HGDP-CEPH Diversity Panel Tell Us?" *European Journal of Human Genetics* 19 (5): 583–587. doi:10.1038/ejhg.2010.205.
- Lewis, K., M. Gonzalez, and J. Kaufman. 2012. "Social Selection and Peer Influence in an Online Social Network." *Proceedings of the National Academy of Sciences* 109 (1): 68–72. doi:10.1073/pnas.1109739109.
- Lomi, A., and F. Pallotti. 2012. "Relational Collaboration Among Spatial Multipoint Competitors." *Social Networks* 34 (1): 101–111. doi:10.1016/j.socnet.2010.10.005.
- Luo, W., A. M. MacEachren, P. Yin, and F. Hardisty. 2011. "Spatial-Social Network Visualization for Exploratory Data Analysis." Paper presented at the Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks, Chicago, IL, November 1–4.

- Mennis, J., and M. J. Mason. 2012. "Social and Geographic Contexts of Adolescent Substance Use: The Moderating Effects of Age and Gender." *Social Networks* 34 (1): 150–157.
- Michelson, W. M. 1970. *Man and His Urban Environment: A Sociological Approach*. Reading, MA: Addison-Wesley.
- Moody, J., D. McFarland, and S. Bender-deMoll. 2005. "Dynamic Network Visualization." *American Journal of Sociology* 110 (4): 1206–1241. doi:10.1086/421509.
- Nag, M. 2009. "Mapping Networks: A New Method for Integrating Spatial and Network Data." Unpublished manuscript, Department of Sociology, Princeton University.
- Nakaya, T., and K. Yano. 2010. "Visualising Crime Clusters in a Space-Time Cube: An Exploratory Data-Analysis Approach Using Space-Time Kernel Density Estimation and Scan Statistics." *Transactions in GIS* 14 (3): 223–239. doi:10.1111/j.1467-9671.2010.01194.x.
- Onnela, J.-P., S. Arbesman, M. C. González, A.-L. Barabási, and N. A. Christakis. 2011. "Geographic Constraints on Social Network Groups." *PLoS ONE* 6 (4): e16939. doi:10.1371/journal.pone.0016939.
- Patil, D. J. (Producer). 2011. *Visualize Your LinkedIn Network with InMaps*. Accessed 19 August, 2012. [http://www.youtube.com/watch?v=PC99Nw2JX8w&feature=player\\_embedded](http://www.youtube.com/watch?v=PC99Nw2JX8w&feature=player_embedded)
- Phithakkitnukoon, S., F. Calabrese, Z. Smoreda, and C. Ratti. 2011. "Out of Sight Out of Mind – How Our Mobile Social Network Changes During Migration." Paper presented at the 2011 IEEE third international conference on Privacy, security, risk and trust (passat) and 2011 IEEE third international conference on social computing (socialcom), Boston, MA, October 9–11.
- Pooley, C. G., and J. Turnbull. 1998. *Migration and Mobility in Britain Since the Eighteenth Century*. London: UCL Press.
- Radil, S. M., C. Flint, and G. E. Tita. 2010. "Spatializing Social Networks: Using Social Network Analysis to Investigate Geographies of Gang Rivalry, Territoriality, and Violence in Los Angeles." *Annals of the Association of American Geographers* 100 (2): 307–326. doi:10.1080/00045600903550428.
- Rana, S., and J. Dykes. 2003. "A Framework for Augmenting the Visualization of Dynamic Raster Surfaces." *Information Visualization* 2 (2): 126–139. doi:10.1057/palgrave.ivs.9500043.
- Robertson, G., R. Fernandez, D. Fisher, B. Lee, and J. Stasko. 2008. "Effectiveness of Animation in Trend Visualization." *IEEE Transactions on Visualization and Computer Graphics* 14 (6): 1325–1332. doi:10.1109/tvcg.2008.125.
- Sailer, K., and I. McCulloh. 2012. "Social Networks and Spatial Configuration – How Office Layouts Drive Social Interaction." *Social Networks* 34 (1): 47–58. doi:10.1016/j.socnet.2011.05.005.
- Scellato, S., A. Noulas, R. Lambiotte, and C. Mascolo. 2011. "Socio-Spatial Properties of Online Location-Based Social Networks." *Proceedings of ICWSM* 11: 329–336.
- Schaefer, D. R. 2012. "Youth Co-Offending Networks: An Investigation of Social and Spatial Effects." *Social Networks* 34 (1): 141–149.
- Shaw, S. L., H. Yu, and L. S. Bombom. 2008. "A Space-Time GIS Approach to Exploring Large Individual-Based Spatiotemporal Datasets." *Transactions in GIS* 12 (4): 425–441. doi:10.1111/j.1467-9671.2008.01114.x.
- Todo, Y., D. M. Yadate, P. Matous, and R. Takahashi. 2011. "Effects of Geography and Social Networks on Diffusion and Adoption of Agricultural Technology: Evidence from Rural Ethiopia." Paper presented at the CSAE 25th Anniversary Conference, Oxford, March 20–22.
- Viry, G. 2012. "Residential Mobility and the Spatial Dispersion of Personal Networks: Effects on Social Support." *Social Networks* 34 (1): 59–72. doi:10.1016/j.socnet.2011.07.003.
- Warren Liao, T. 2005. "Clustering of Time Series Data – A Survey." *Pattern Recognition* 38 (11): 1857–1874. doi:10.1016/j.patcog.2005.01.025.