An Integrated Framework To Support Global and Local Pattern Assessment for Residential Movements

Yin Liu
Sichuan Normal University

Alan T. Murray
Drexel University

Abstract

Residential mobility is a defining characteristic of society in the United States. A 2003 U.S. Census Bureau migration report highlights that more than 22 million people were characterized as domestic migrants between 1995 and 2000. Understanding resulting patterns is important because it provides insights on rationale for movement and for housing, services, and supporting infrastructure implications. The method for facilitating pattern identification and exploration of movements unfortunately is lacking. It is often the case that migration and movement are considered in aggregate terms—between cities and counties in a state or region. Individual behavior reflective of a movement trajectory is therefore masked in various ways. Survey evidence also indicates that residential movements of short distance—for example, those occurring within a city or county—reflect the greatest proportion of total migrations. To address limitations, this research proposes a framework integrating spatial analytical methods to support pattern analysis for individual movements, relying on detailed information of origin and destination change. The framework can explore the patterns at both the global and local levels. The framework is designed using various visual analytic interfaces coupled with statistical evaluation and significance testing, representing both exploratory and confirmatory assessment. The integrated framework is applied to study residential movement involving 2,636 housing changes in Franklin County, Ohio, and effectively estimates some special global and local patterns from those events.

Introduction

Residential mobility is a defining characteristic of U.S. society today. Studies of residential changes can enable better understanding of how humans and the environment engage in mutual interactions in different places and at different scales. For the researchers, including geographers and urban planners whose major interest is to comprehend and design better urban environment, it is especially meaningful to examine the effect of a collection of residential movements, if such spatial data are available, as a way of studying the immediate changes in the composition of many urban neighborhoods. The information associated with change of residence can also provide insights into the mechanism under which urban structure constrains residential choices. Therefore, residential movement is an important issue.

Research on residential movement has been carried out for more than a century (Quigley and Weinberg, 1977; Ravenstein, 1885) and is still of broad interest to researchers today (Dieleman, 2001; Rae, 2009). A major part of those research attempts was rooted in cartographic technique and gradually has developed to spatial data visual analytics, in which computational tools have been extensively involved (Andrienko et al., 2008; Rae, 2009; Ravenstein, 1885; Tobler, 1987). The disadvantage of spatial data visual analytics is its incapability for confirmatory hypothesis testing in spatial pattern identification.

On the other hand, the confirmatory analyses actually have been applied into movement pattern since the 1960s. Depending on mathematical and statistical methods, those works essentially transformed the spatially complex form of movements into manageable geometry and so focused more attention on the movements' spatial pattern (Adams, 1969; Brunsdon and Corcoran, 2006; Fotheringham and Pitts, 1995; Morrill, 1963). The disadvantage of those works is that they were excessively focused on the "global map" or general "law." Meanwhile, such method is heavily based on aggregated data.

Compared with the study based on aggregated data at the macro level, it is actually more important for geographers and urban planners to understand the pattern of residential movement at the micro scale with individual records, because the pattern can directly and accurately illuminate the change of social composition in a county or city. The difficulty with such perspective shifting is the significant raising of data size. With hundreds and thousands of individual movement records, significant patterns can be masked (Andrienko et al., 2008). Taken with redesigned visual analytics, however, statistical methods might prove efficient and effective for exploring and evaluating patterns in mass movement data. Exploratory spatial data analysis (ESDA)-based studies recently have recognized this necessity, and researchers have suggested possibilities for combining visual analytics and statistical methods (Thomas and Cook, 2005). As yet, a systematic framework has not been developed that reflects this necessity, synthesizing both confirmatory and exploratory approaches.
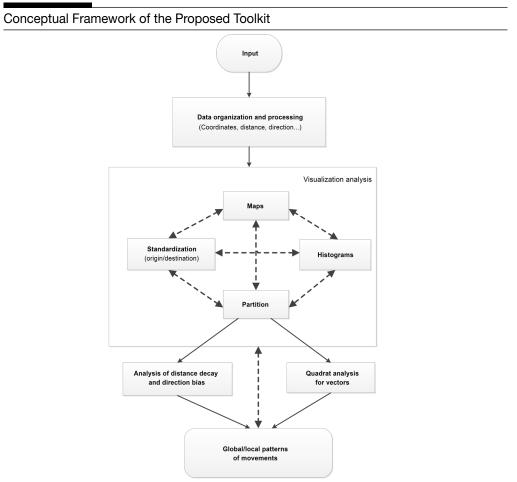
Method

Generally speaking, spatial patterns refer to the specific spatial configuration or arrangement of features of interest over space (Chou, 1995). Insights into the characteristics of spatial patterns result in knowledge of the dynamics of spatial processes (Getis and Boots, 1978).

Movement data are records of spatial trajectory for tracking behavior over space or, more abstractly, as a directional line with fixed length. In geometry, such form of a trajectory is essentially a vector of distance and direction. Based on this perspective, the spatial pattern in residential movements can refer to a spatial arrangement of the dual components: moving distance and moving direction. Distance is a basic means for measuring the space separating objects; it is also used to quantify the possible intensity of a relationship and interaction between geographical events. The distribution of moving directions is a reflection of urban evolution. Research in spatial arrangement of moving directions within a city can be used to examine whether a specific planning strategy has been effective in generating movements toward certain areas (Quigley and Weinberg, 1977). A research difficulty is figuring how to analyze the features of distance and direction simultaneously.

The proposed framework is to resolve this difficulty by both exploratory and confirmatory approaches.

Exhibit 1 illustrates a proposed framework with general structure for data analysis. First, the movement data records are processed and converted into a standard format. Using the manipulated

Exhibit 1

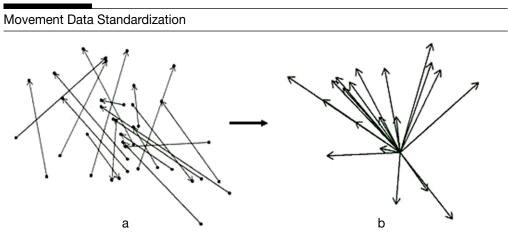Conceptual Framework of the Proposed Toolkit

data, visualization-based analysis is then performed, incorporating exploratory data analysis using multiple interfaces. Based on visual analytics, quantitative examination is included to test visually detected patterns at both the global and local scales.

Standardization of Movement Data

A unique method of data standardization is suggested here as the first step for pattern analysis: moving all the vectors' origin or destination to an identical center while maintaining distance and direction. The effect is shown in exhibit 2.

The standardized movements intuitively provide a more direct and clear illustration of the distribution of distances and directions. Distances are abreast with each other from the same start, and directions are arranged in a uniform circle to form something like a rose. In this standardized interface, potential patterns in the arrangement of distances and directions cannot be easily masked. Thus, patterns in the distances and directions can be simultaneously illustrated.

Exhibit 2

Movement Data Standardization

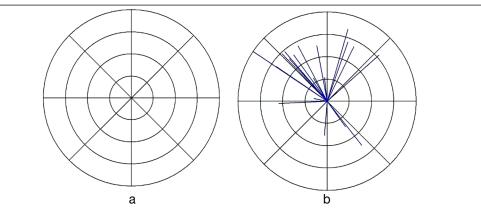a   b

Notes: (a) All the vectors are distributed at their original locations over the area before the standardization step. (b) All the vectors' origins are moved to an identical center while maintaining their distances and directions.

Exploratory Analysis Interface

Given the standardized movement data, one can imagine that they are distributed along a series of circular sections. The width for each circular section represents a unit range of length away from the established center. Further, each circular section can be separated into several equal fan sectors based on a directional range. Based on this idea, as shown in exhibit 3, a partition scheme is possible.

Based on the partition scheme over standardized vectors, a color gradation is applied to each section to enable visual understanding. The darkness of color in each section is in accordance with the counts of endpoints in this section. The more endpoints in a particular section, the darker the fill color.

Exhibit 3

A Partition Scheme for Standardized Vectors

a   b

Notes: (a) The partition scheme is generated by identical distance and direction intervals; concentric circles surrounding an established center partition the vectors by equal length interval (the radius of the largest circle equals the largest vector length), and homocentric rays from the center partition the vectors by equal angle interval. (b) The partition scheme is set over standardized vectors.

Statistical Analysis of Movement Patterns

Because the partitioned scheme parallels the quadrat analysis interface, a type of statistical testing, the goodness-of-fit metrics, can be applied to examine whether the actual distribution of distance and direction follows a theoretical pattern. Such a method is also meaningful for exploring the pattern for a specific subset of movements and for testing whether this subset possesses a pattern that is similar to or different from the global pattern. Using this kind of comparison, it is possible to explore whether community dissimilation or mixture is occurring in a city or region.

In practice, the number of endpoints from the entire data (global) and the subset (local) in each section is counted respectively. It follows that the global counts for each section can be compared statistically with the local number, by linear regression analysis. The counts for the local data are treated as the dependent variable and the counts for the entire data are treated as the independent variable. Then, the regression function is—

$$L_i = (a \times G_i) + b + \varepsilon_i, \tag{1}$$

where L_i equals the counts of the endpoints for local data in section i, G_i equals the counts of the endpoints for entire data in the same section i, a and b are the regression parameters, and ε_i is the regression error for the section i. The similarity between global and local patterns can be tested by the coefficient of determination, R-square, of the linear regression model. Here, R-square indicates the proportion of the local pattern "explained" by the global pattern. The lower the R-square value, the less consistency exists between local and global patterns.

Distance Decay and Directional Bias

That locations separated by shorter distances are more related has been globally examined and has become a fundamental theory for spatial science, the theory of "*Distance Decay*" (Fotheringham and

O'Kelly, 1989; Taylor, 1975). Given this global theory, within a study area, more residential movements will reasonably be expected with short distance than with longer movements. The question is how to quantitatively measure the frequency of movements with specific moving distances. Taylor (1975) has given two model categories and five specific expressions for describing the quantitative relationship between the moving distance and the movement intensity.

1. Single-log models $f(d) = d^m$.

Square root exponential: $\log I = a - (b \times \sqrt{d}\,)$ (2)

Exponential: $\log I = a - (b \times d)$ (3)

2. Double-log models $f(d) = \log^m d$.

Normal: $\log I = a - (b \times d^2)$ (4)

Pareto: $\log I = a - (b \times \log d)$ (5)

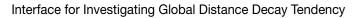Log-normal: $\log I = a - [b \times (\log d)^2]$,   (6)

where I is the intensity of movements, d is the distance, and a and b are regression parameters to be estimated. Among these five specific forms, the issue to resolve is which one, if any, would best describe the distance decay tendency in observed behavior. Different estimates and values in parameters may help reveal information associated with moving behaviors. The estimation relies on a statistical test, such as R-square or P-value. The test can be formalized as—
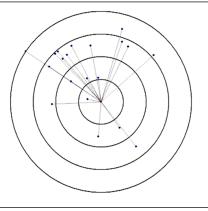
$$\lambda_x \leftarrow g_x (I,d), \tag{7}$$

where λ_x is a test index (that is, R-square or P-value) used for evaluating the regression equation $g_x (I,d)$.

In the standardized interface, intensity of movements is equivalent to intensity of endpoints; thus, movement intensity is derived based on endpoints within different distance ranges. Exhibit 4 illustrates this consideration.

Exhibit 4

Interface for Investigating Global Distance Decay Tendency

Intensity I_i is calculated based on the counts of endpoints and the area of each circular section. The series of intensities, then, is examined based on their relationship to the radius of the corresponding circular section d_i.
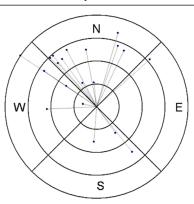
To detect the potential directional variation in distance decay, each circular range is separated into different directional sections. Exhibit 5 is showing an example in which each circular range is directionally partitioned into north, south, east, and west sections.

Then, the five specific forms of decay models are examined with respect to each directional section.

The similarity or difference between models for global decay and directionally partitioned decay can be evaluated by comparing the confidence intervals of regression parameters. Because the interest is decay tendency, the evaluation needs to focus on only the slope parameter b in a regression model. After a regression model for global data has been derived, $y = \alpha + \beta x + \varepsilon$, a confidence interval with a specific significance level for the slope parameter β can be derived as well (β_{min}, β_{max}). Then, the slope parameter b_i (i represents the four directions respectively) in the regression model for each directional partition is tested for whether $b_i \in (\beta_{min}, \beta_{max})$. If the test derives a positive result, it indicates that the two regression models possess a significantly similar slope. In practical terms, $b_i \in (\beta_{min}, \beta_{max})$ means that the data located within the i directionally partitioned data zone possess a similar decay tendency as the entire dataset. The global regression model for distance decay can be appropriately used to describe the decay tendency for movements along a specific direction.
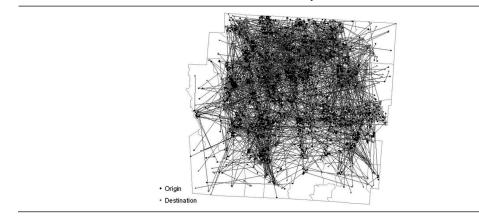
Exhibit 5

Interface for Investigating Distance Decay Over Different Directions

Application

To test those functions in the framework, 2,363 residential changes, which were derived from 39,232 residential transactions of home sales and purchases in Franklin County, Ohio, from October 2004 to April 2006, are geocoded as matched pairs in a Geographic Information System, or GIS, database. The 2,363 records are regarded as 2,363 movement vectors, as shown in exhibit 6.

Exhibit 6

The 2,363 Residential Movements in Franklin County, Ohio
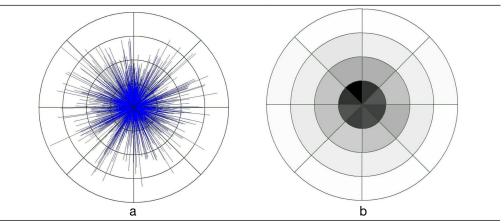

• Origin
• Destination

According to a series of reports by the Mid-Ohio Regional Planning Commission, from 2000 to 2030 the northwest region of Franklin County is planning for more employment opportunities and increased economic growth. Thus, cities in the regions surrounding U.S. Highway 33, including Dublin, Upper Arlington, and townships in the northwest part of Franklin County, will likely attract more residents. Because the research data are derived from residential changes from 2004 to 2006, it is reasonable for the planners from the Mid-Ohio Regional Planning Commission who made the regional planning strategies for Franklin County to assume that the patterns of those residential movements will reflect aspects of planned behavior. To assess observed behavior, this application will therefore use the framework to thoroughly explore patterns in the 2,363 residential movements.

Exploratory Analysis of Distance and Direction

To get more meaningful insights about spatial pattern for these residential movements, visual analytics are used. Following the procedures of vector standardization and the partition scheme introduced previously, a visual display for the distribution of the dual features is created. This scheme with standardized data offers an integrative version for the two spatial features, as illustrated in exhibit 7.

The counts of endpoints in each distance-direction section of the partition scheme reflect the number of movements for a corresponding distance and direction range. The counts of endpoints, or the frequencies of movements, can be visually illustrated by color gradation. By reading the color gradient tendency, the general pattern for the 2,363 movements becomes more evident. The eight sections of the innermost circle are much darker than those for any peripheral section, suggesting that distance decay is a dominant pattern in residential changes. For the distribution of directions, an imbalanced pattern is clear, which suggests a tendency for more residential changes to be oriented toward the northwest.

Exhibit 7

Partition Scheme Over the Standardized Vectors (a) and the Color Gradient Showing
Endpoint Density (b)

a   b

Directional Bias in Distance Decay Tendency

To further explore the directional bias in distance decay for the 2,363 residential movements, the
relationship between intensity of movements and moving distance is explored. Exhibit 8 illustrates
the interface for such exploration.

Exhibit 8

Visual Analytics Exploring Movement Intensity (a) and Moving Distance (b)
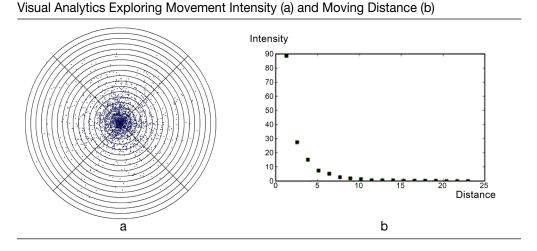

a   b

Exhibit 8 also illustrates the global relationship between intensity of endpoints in each distance
rings (without considering the directional partitions) and the radius of each distance circle.

Based on such an interface, the framework further estimates that, for all the movements, the most appropriate regression function is the *exponential* equation, because it has the highest R-square value, 0.9757. Thus, the function used for describing the global distance decay tendency is established as—

$$I = \exp [4.273 - (0.384 \text{ x } d)]. \tag{8}$$

Such exploration is conducted for data that have been categorized into different directional zones as well. Exhibit 9 summarizes the quantitative descriptions of distance decay for the four directional zones. It is interesting to researchers to notice that *square root exponential* function is most appropriate for each of the four directions, but none of them follow the same trend as the global tendency. The *square root exponential* function mathematically results in a more gradual rate of decay than the basic *exponential function*. This result actually indicates that the decay tendency is faster globally than it is for any of the directional zone subsets considered.

The framework further compares the global decay to the directionally partitioned decay statistically. To implement this comparison, the global pattern must be modeled by the *square root exponential* function in accordance with the model for each partitioned subset. A *square root exponential* regression equation for the global pattern is—

$$I = \exp [7.503 - (2.377 \text{ x } \sqrt{d})], \tag{9}$$

with an R-square value of 0.964, which means this model can also appropriately summarize the global distance decay tendency. The test is to evaluate whether the value for the global slope, -2.377, is significantly different from the slope for each of the directionally partitioned subsets. Exhibit 10 summarizes the confidence interval (with 95-percent significance level) of the global slope as well as the slope of each partitioned subset.

Exhibit 9

Estimated Distance Decay Functions for Movements in Each of the Four Directional Zones

Directional Zones	Function	R-Square
North	$I = \exp [5.576 - (2.166 \text{ x } \sqrt{d})]$	0.987
South	$I = \exp [5.395 - (2.162 \text{ x } \sqrt{d})]$	0.974
East	$I = \exp [5.390 - (2.123 \text{ x } \sqrt{d})]$	0.963
West	$I = \exp [5.293 - (2.060 \text{ x } \sqrt{d})]$	0.983

Exhibit 10

Slope Values of Regression Models for Global and Local Datasets

Global	95% Confidence Interval		East	West	North	South
	Lower	Upper				
− 2.377	− 2.621	− 2.134	− 2.123	− 2.06	− 2.116	− 2.162

Exhibit 10 clearly shows that the slope values for the north, east, and west subsets are not within the 95-percent confidence interval (-2.621, -2.134) of the global slope. This result indicates that the decay tendency for the north, east, or west partition is significantly (with 95-percent significance level) different from the globally estimated decay tendency. While the globally estimated tendency indicates a sharper decay slope, the decay tendency for the north, east, and west partitions is significantly smoother. So, if the globally based model is used to describe the pattern of distance decay for the entire set of movements, such a directional effect cannot be appropriately captured.

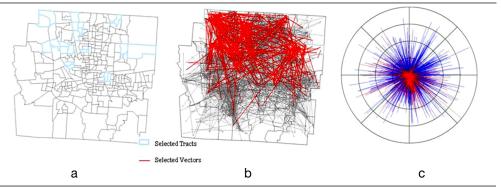## Pattern Comparison Between Global and Local

Exploration of whether the pattern for a subset of the data is consistent with the global pattern can help reveal whether residential changes lead to social differentiation in a city or region. As an example for such exploration, the 20 richest census tracts are considered locally interesting. Then, the residential movements with their destinations located inside these 20 census tracts are selected to form a sample subset. A research question is: "Is the distribution of distances and directions for those movements with destinations in the richest neighborhoods significantly similar to the respective distributions for all movements in Franklin County?"

In exhibit 11a, highlighted polygons are the 20 richest census tracts in Franklin County. In exhibit 11b, the movements whose destinations fall within the 20 census tracts are highlighted. This sample contains 538 movements. Exhibit 11c is similar to exhibit 7 but the vectors are standardized by destinations. Those highlighted vectors correspond to the selections in exhibit 11b.

Researchers in this study noticed that most of the highlighted vectors are oriented toward the north, because 18 of the 20 selected tracts are in the northern portion of Franklin County.

Further, based on the comparison metrics introduced previously, the framework tests the significant similarity or difference between the local pattern of the 538 selected movements and the global pattern of all the data by *linear regression* model. Exhibit 12 summarizes the counts of endpoints with respect to the entire (global) and subset (local) movements in each partition section.

Exhibit 11

Selected Movements With Destinations in the 20 Richest Census Tracts

Selected Tracts
Selected Vectors

a   b   c

Notes: Exhibit 11a highlights the 20 richest census tracts in Franklin County, Ohio. Exhibit 11b highlights the movement vectors whose destinations fall within the 20 census tracts. Exhibit 11c standardizes the vectors in exhibit 11b.

Exhibit 12

Counts of Endpoints in the Partitioned Sections for Entire Data and Subset Data

Circle Order	Global or Local Counts	NNE	ENE	ESE	SSE	SSW	WSW	WNW	NNW
First inside circle	Global	201	179	212	265	214	177	199	186
	Local	35	30	55	92	57	43	47	28
Second inside circle	Global	51	64	62	84	83	63	81	61
	Local	7	5	22	34	23	14	12	6
Third inside circle	Global	17	21	18	27	17	16	20	17
	Local	0	1	4	5	3	3	7	1
Outside circle	Global	3	4	7	2	2	6	2	2
	Local	0	0	2	0	1	1	0	0

ENE = east-northeast. ESE = east-southeast. NNE = north-northeast. NNW = north-northwest. SSE = south-southeast. SSW = south-southwest. WNW = west-northwest. WSW = west-southwest.

Note: Each row of the table records the counts in each distance circle, starting from the NNE corner and progressing in clockwise order.

Then, to estimate the relationship between the global and local patterns, the framework derives a linear regression function as follows.

Local = -2.02 + (0.255 x *Global*), (10)

with a coefficient of determination (R-square value) of 0.87. This R-square value indicates that the two patterns, in general, are statistically similar. In short, the local pattern, which is for the subset of movements with their destinations in the richest regions of the study area, is similar to the global pattern for the complete set of 2,363 residential changes.

Summary

Through a series of analyses, the framework facilitates pattern exploration for the 2,363 residential movements in Franklin County, Ohio. At the global level, distance decay has been confirmed as a significant tendency, but the framework also quantitatively detects that a single equation derived from the entire dataset is not able to fully describe characteristics in the distance decay tendency, because the directional bias is also established as a significant effect. For these northwest-oriented movements, the slope for their distance decay tendency is not as steep as the general trend.

The framework also facilitates investigation of local characteristics associated with the identified movement patterns as well. The research examines movements into the richest 20 census tracts in Franklin County. Their movement pattern is established and compared with the global pattern. The framework confirms that the local pattern is actually similar to the global. This finding indicates that these "rich" residential movements did not exhibit a significant difference in spatial behavior from the complete set of movements in this research.

Conclusion

This research introduces a framework consisting of a series of functions and methods for exploratory and confirmatory examination of spatial patterns from mass residential movements at a micro scale. The spatial pattern of movement in this research is treated as the arrangement of distances and directions over space. Then, to efficiently detect patterns from mass movements, the framework provides ways to visually standardize movements. Based on standardized movements, a partition scheme is introduced. The distribution of movement distance and direction can then be qualitatively, as well as quantitatively, investigated. By using such an analytical process, movement patterns for 2,363 residential changes within Franklin County, Ohio, have been evaluated. Further, local deviations associated with subsets of movements can be identified from global movements. This technology has been used to examine patterns for 538 movements from the complete dataset. The successful application has confirmed the effectiveness of the proposed framework.

Some future developments can be made for improvement. First, some functions of the framework should be more objective for achieving statistically robust insights; for example, the size of the distance-directional interval in the partition scheme. Second, temporal dimensions should be included for analysis as well, because the distribution of residential movements over time can also suggest insights into how humans and the environment interact. Such developments indicate that this framework can exhibit additional strength for effectively and efficiently detecting information from mass residential movements.

Acknowledgments

Authors

Yin Liu is an assistant professor in the Key Laboratory of Land Resources Evaluation and Monitoring in Southwest China at the School of Geographical and Resource Science, Sichuan Normal University, Chengdu, China.

Alan T. Murray is a professor in the Center for Spatial Analytics and GeoComputation at the College of Computing & Informatics and in the School of Public Health, Drexel University, Philadelphia, Pennsylvania, United States.

References

Adams, John. 1969. "Directional Bias in Intra-Urban Migration," *Economic Geography* 45 (4): 302–323.

Andrienko, Gennady, Natalia Andrienko, Ioannis Kopanakis, Arend Ligtenberg, and Stefan Wrobel. 2008. "Visual Analytics Methods for Movement Data." In *Mobility, Data Mining and Privacy: Geographic Knowledge Discovery*, edited by Fosca Gionetti and Dino Pedreschi. Berlin, Germany: Springer: 375–410.

Brunsdon, Chris, and Jon Corcoran. 2006. "Using Circular Statistics To Analyse Time Patterns in Crime Incidence," *Computers, Environment and Urban Systems* 30 (3): 300–319.

Chou, Yue-Hong. 1995. "Spatial Pattern and Spatial Autocorrelation," *Spatial Information Theory Lecture Notes in Computer Science* 988: 365–376.

Cook, Kristin, and James Thomas. 2005. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. Los Alamitos, CA: IEEE Computer Society Press.

Dieleman, Frans. 2001. "Modelling Residential Mobility: A Review of Recent Trends in Research," *Journal of Housing and the Built Environment* 16 (3): 249–265.

Fotheringham, Stewart, and Morton O'Kelly. 1989. *Spatial Interaction Models: Formulations and Applications*. Boston: Kluwer Academic Publishers.

Fotheringham, Stewart, and Timothy Pitts. 1995. "Directional Variation in Distance Decay," *Environment and Planning A* 27: 715–729.

Getis, Arthur, and Barry Boots. 1978. *Models of Spatial Processes: An Approach to the Study of Point, Line and Area Patterns*. Cambridge Geographical Studies. New York: Cambridge University Press.

Morrill, Richard. 1963. "The Distribution of Migration Distances," *Papers in Regional Science* 11 (1): 73–84.

Quigley, John, and Daniel Weinberg. 1977. "Intra-Urban Residential Mobility: A Review and Synthesis," *International Regional Science Review* 2 (1): 41.

Rae, Alasdair. 2009. "From Spatial Interaction Data to Spatial Interaction Information? Geovisualisation and Spatial Structures of Migration From the 2001 UK Census," *Computers, Environment and Urban Systems* 33 (3): 161–178.

Ravenstein, Ernest. 1885. "The Laws of Migration," *Journal of the Statistical Society of London* 48 (2): 167–235.

Taylor, Peter. 1975. *Distance Decay in Spatial Interactions*. Concepts and Techniques in Modern Geography No. 2. Norwich, United Kingdom: Geo Abstracts.

Tobler, Waldo. 1987. "Experiments in Migration Mapping by Computer," *Cartography and Geographic Information Science* 14 (2): 155–163.