Published in final edited form as:

Ann Assoc Am Geogr. 2014 January 1; 104(1): 40–59. doi:10.1080/00045608.2013.846152.

Modeling Social Ties and Household Mobility

Sara S. Metcalf

Department of Geography, The State University of New York at Buffalo, Buffalo, New York 14261
USA

Abstract

Underlying the aggregate phenomena of persistent problems such as urban sprawl and spatial socio-economic disparity is the individual choice of where to live. This study develops an agent-based model to simulate social and economic influences on neighborhood choice. With Danville, Illinois as an empirical context, a pattern-oriented approach is employed to examine the role of social ties in shaping intra-urban household mobility. In the model, household agents decide whether and where to relocate within the community based upon factors such as neighborhood attractiveness, affordability, and the density of a household's social network in the prospective block group. Social network and neighborhood choices are encoded with logit utility functions. The relative influence of factors affecting the formation of social ties in the simulated social network, such as geographic proximity, similarity of income, race, and presence of children, are adjusted using parameter variation to create alternative model settings. Simulated migration patterns resulting from different network and neighborhood choice coefficients are compared with observed migration patterns over a two-year period. Based upon 1000 simulation experiments, a regression of homeowner migration error (the difference between simulated and observed migration) relative to the parameter settings revealed components of social network choice such as income, race, and probability of local ties to be significant in matching observed migration patterns. A non-linear effect of simulated social networks on household mobility and thus migration error was exhibited in this study.

Keywords

social networks; neighborhood choice; agent-based modeling; pattern-oriented modeling

1. Introduction

The concept of spatial disparity is so embedded in urban America that residents may not recognize the problems implicit in referencing the “wrong side of town” in casual conversation. The problem of socio-economic disparity has been particularly acute in industrial cities of the U.S. Rust Belt, where population declined and poverty increased as middle-class jobs were outsourced in the late 20th century. Even so, many of these

Mailing Address: UB Geography Dept, 105 Wilkeson Quad, Ellicott Complex, Buffalo, NY 14261 smetcalf@buffalo.edu Telephone: (716) 645-0479 Fax: (716) 645-2329.

ⁱCode for the model developed in this study is provided at the following link: <http://www.acsu.buffalo.edu/~smetcalf/resources/ModelCode.htm>

communities continue to expand in area while shrinking or stagnating in population. Such a dynamic intensifies the dichotomy between new and old, or rich and poor, parts of town.

While fundamental economic forces ultimately drive the production of inequality, individual choices about where to live are proximate influences on the spatial manifestation of disparity in local urban geographies. This research explores the relationship between social network structure and these household-level choices, as expressed in intra-urban migration patterns. Because the relationship between social ties and location is complex and therefore difficult to untangle, an agent-based model is employed to simulate local community ties as they change over time. With this model, hypothesized effects of social networks on neighborhood choice are simulated dynamically. The model is constructed to allow systematic exploration of potential relationships between households' choices of where to live and the extent of their social networks.

A social network is an analytical abstraction of relationships between people that includes trusted channels for communication such as family, friends, and advisors. Homophily, or self-sorting according to similarity, has been shown to be a significant factor in shaping social network structure along dimensions of race, age, religion, education, occupation, and gender, with geographic proximity and family ties creating opportunities for such self-sorting connections to form (McPherson et al. 2001). Because humans are mobile and idiosyncratic, in this globalized digital age we are capable of forming and maintaining diverse social obligations and attachments that span significant distances (Rainie and Wellman 2012).

Similarities have been drawn between dynamics of social and residential mobility, such as the propensity of a 'vacancy chain' mechanism to operate. A home vacancy in a neighborhood, like a job vacancy in an organization, creates an opportunity for another to move up, into a more affluent position (White 1970). In the residential setting, the vacancy chain involves the out-migration of affluent households from a given neighborhood, which creates opportunities for lower-income households to purchase homes in a neighborhood previously unaffordable (Hoyt 1939). Although the vacancy chain is a logical mechanism, it does not operate in isolation. Ethnographic geographies of social networks reveal complex dynamics such as the dialectical interplay of social influence and life opportunities for the case of women who are further marginalized in terms of class and race (Rowe and Wolch 1990, Peake 1997).

Computational advances have rendered it feasible to include social networks in geospatial analysis and therefore to accommodate the increasing availability of networked data about human interactions. While distance continues to be relevant to human interaction, Kwan (2007) emphasizes that the nature of its relevance has changed with the widespread adoption of information and communication technologies. New methods of geographic analysis explicitly account for the possibilities and constraints that are created by the intersection of geographic space with social network structure. For example, structural analyses of empirical, georeferenced social network data have informed the study of gang behavior in Los Angeles (Radil et al. 2010) and diarrheal disease in Bangladesh (Emch et al. 2012). Similarly, simulation studies of socio-spatial network dynamics have been employed to

examine influenza diffusion (Bian et al. 2012) and hurricane evacuation (Widener et al. 2012). These analyses help to untangle social and spatial influences on the dynamics human behavior, and provide ways to better apprehend the transient dimensions of mobility and communication that are consistent with the hypertext model proposed by Kwan (2007).

Modeling methods that recognize autonomy, interactivity, and contextual constraints help reconcile macro and micro interpretations of persistent social dilemmas (Schelling 1971, 1978). Schelling's segregation model, played as a game or simulated using cellular automata, demonstrates how even slight preferences for similar neighbors can induce aggregate patterns of residential segregation. An insight from such an individual-based analysis is that preferences for diversity, rather than simply a tolerance of difference, must be cultivated to overcome systemic tendencies toward segregation.

Agent-based models are increasingly used to simulate human-environment interactions, test effects of heterogeneity, and simulate mobile objects in dynamic landscapes (Westervelt and Hopkins 1999, Batty 2005). In agent-based models, individual choices shape aggregate outcomes. Information generated by changes in aggregate characteristics, such as social norms and neighborhood affluence, can then feed back to influence other agents' choices as they re-evaluate their environment over time.

Bottom-up simulation of local interactions among diverse agents in a dynamic environment offers a way to explore complex social systems (Epstein and Axtell 1996). To simplify model interpretation and ease implementation, abstract representations of space are utilized in many agent-based models (Resnick 1994, Epstein and Axtell 1996). Simulation research has demonstrated ways to integrate agent-based models with empirical GIS data (Gimblett 2002, Parker et al. 2003). For example, recent work by Yin (2009) and Crooks (2010) demonstrates the continued utility of the Schelling (1971) model in formulating agent-based models of residential preferences in a dynamic urban landscape parameterized using GIS analysis.

Despite recent improvements in computational ease, the calibration and interpretation of agent- or individual-based models remains a challenge (Osgood 2009). To address this difficulty, a pattern-oriented approach was adopted for this research as a means of ground-truthing the simulation model (Grimm and Railsback 2005). The use of observed patterns to guide model development and analyze model output enables decoding of essential system information (Wiegand et al. 2003). Simulation studies enable indirect estimation of parameters for abstract models from patterns of spatial socio-economic data. For example, in analyzing unemployment, Conley and Topa (2003) use Census tract-based spatial proximity to structure a social network of local interactions, with the majority of interactions occurring within and adjacent to residents of a given tract. They demonstrate that this spatial algorithm anticipates shifts in unemployment better than an aggregate, non-spatial approach would. Similarly, the present study employs the technique of indirect estimation by simulating the influence of hypothetical social networks on migration patterns, using pattern-oriented modeling to compare the simulated results with observed patterns of household mobility.

A dynamic simulation of social networks may involve connections that change over time (as in Bian et al. 2012), or processes of information diffusion within network structures (as in Widener et al. 2012). This study emphasizes the former, rendering social networks dynamic by changing connections among simulated household agents according to homophily-based reassessments of friendship utility that depend in part upon geographic proximity. Model parameters for moving behavior, individual attributes, and network structure are estimated from different empirical sources. Household agents' choice algorithms are specified using utility functions to evaluate social connections and prospective residences. Simulated results are compared with observed homeowner migration patterns to explore the relative role of social influence on neighborhood choice. The next section describes the empirical context of Danville, Illinois that is used to ground-truth this simulation study.

2. Empirical Context: Danville, Illinois

Empirical observations for the case of Danville, Illinois provide a basis for specifying the household characteristics that shape social network and neighborhood choice in the agent-based model. Danville was selected as a reference case because its longitudinal patterns of population growth, decline, and stagnation, and its spatial pattern of areal expansion, echo the trajectories of other Rust Belt cities.

Located 215 km (134 miles) south of Chicago, Danville is a small city with an industrial legacy, whose population declined from its peak of 42,570 people in 1970 to 33,027 people in 2010. Figure 1 illustrates the population trend over time for Danville on the basis of estimates from the decennial U.S. Census. In the decades prior to World War II, Danville's land area was less than 8 square miles. Since then, Danville has grown to nearly 18 square miles, with a steadily declining population density associated with post-WWII growth in the automotive era. Although Danville's land area more than doubled since 1940, its population is now less than it was in 1920. Notably, Danville has continued to expand its geographic footprint despite population decline by annexing adjacent communities. As one Danville official explained, "we're a town of 30,000 with an infrastructure for 60,000."

The exodus of industry opportunities in the late 20th century created a tension between attachment to hometown roots and the struggle to find work. Many middle-class workers left to find jobs elsewhere. Nearly 30 percent of Danville's population presently lives in poverty, although the city maintains a powerful class of affluent professionals and doctors who serve the community's hospitals and retirement homes.

While this Rust Belt urban industrial dynamic was an important reason for selecting Danville for this study, another reason was its manageable size for simulation. As a smaller city, Danville's size enabled computation of each household as a separate agent, so that population sampling was not necessary. An additional, practical reason for selecting Danville was its geographic proximity to the author when fieldwork was conducted in 2005. Because the simulation model was developed alongside data collection, heuristics for modeling individual choices drew upon observations made during field interviews with residents of a Danville neighborhood.

2.1 Neighborhood Associations

With limited fiscal resources stretched further by an expanding areal footprint, the city of Danville actively encourages residents to form neighborhood associations. The geographic boundaries of these associations are based upon the judgment of the residents who organize them; there is no pre-determined size or scope. The city provides guidance and transitional support for newly formed neighborhood associations. There are 17 active and 5 inactive neighborhood associations in Danville. Most active associations hold monthly meetings during the Spring, Summer and Fall seasons.

This study is informed by observations of participants in the Kentucky-Tennessee-Delaware (KTD) neighborhood association, named for the three north-south streets spanning its width, and extending three blocks south of Main Street in the southeast part of Danville (highlighted in Figure 2 below). The original neighborhood organizers restricted the boundaries of the association to correspond with the street pattern and keep the size manageable. Participant observation and in-depth interviews with residents were undertaken in 2005 to develop an understanding of social factors affecting household mobility. All of the interviewees had participated in the neighborhood association at least once, and conveyed generally positive views about the role of the association in building trust and strengthening neighborhood connections.

Although the interviewees were invariably homeowners, many of their neighbors and prior experiences included renting. Characterizations of renters as transient and therefore disinclined to maintain their homes and neighborhood were revealed by some homeowners at neighborhood association meetings and in interviews, although others encouraged neighborhood renters to attend association meetings and neighborhood events. Danville city officials underscored the potential for the renter-owner ratio in a given neighborhood to exceed a tipping point of instability, inducing neighborhood outmigration. The map in Figure 2 illustrates the prevalence of owner-occupied households within Danville Census blocks. The darker shades indicate places where owner-occupied households are more prevalent than renter-occupied households. The percentage shown does not include vacant households, reflecting the fraction of homeowners out of all occupied households.

The KTD neighborhood is racially diverse. As one African American resident observed, the neighborhood is “mixed. It’s like black, white, black, white.” Household experiences were elicited from homeowners who are women and men, African Americans and whites, who span the age range and family stages that comprise the neighborhood. These interviews highlight the relevance of rent/own status and racial diversity among households in the neighborhood, as well as resident unease about changes in neighborhood composition and differences in family status.

The importance of children in neighborhood social networks was emphasized in the interviews. As shared by one homeowner with grown children, “My daughter knows more people around here [than me]. Like [my back-yard neighbor] – she met him because she would go out [to the back alley] and take out the garbage. They would be out there in the garage, because my grandchildren would ride their tricycles around the garage.” Although the friendships children form are significant social ties for a neighborhood, some youth

engage in activities – such as playing in the street – that concern neighbors accustomed to different norms of child safety and discipline. Residents shared different attitudes about whether they would confront their neighbors about a concern, or whether they would sooner call the authorities to resolve a disturbance, such as noise from a late night party.

These interviews with neighborhood residents provided a qualitative source of data to guide heuristics for agent behavior and reveal reasons for relocating that were beyond the scope of the simulation model, such as marriage and divorce dynamics. Another data source indicating homeowner records at the parcel scale for 2001, 2003, and 2005 was used to analyze homeowner migration, as described in the section below.

2.2 Homeowner Migration Patterns

While the qualitative insights from interviews helped to conceptualize the model and recognize its boundaries, homeowner migration data were used explicitly to test the model using the pattern-oriented approach. To generate migration patterns, parcel data provided by the city of Danville were matched to homeowner data for 2001 and 2003 and used to identify owners who moved during that two-year time period. Changes in the inclusion or exclusion of individual family members (e.g., Jane & John Smith in 2001; John Smith in 2003) in consecutive homeowner records suggested the frequency of marriages, divorces, and deaths as likely events shaping neighborhood tenure and homeowner mobility patterns. Although it was labor-intensive to individually distinguish between new owners, departures, and those who remained in their own home during the two-year interval, the process of matching names and addresses provided a novel way to capture intra-urban migration data at the scale of the individual homeowner household.

Analysis of these data revealed that during the 2001-2003 time period, Danville experienced a net gain of 43 owner households overall, with 1472 new owners, 1429 departures from the homeowner register, and 139 intra-urban moves (8.6 percent of all arrivals). From 2003 to 2005, Danville lost 22 owner households, corresponding with 1041 new owners, 1063 departures, and 119 intra-urban moves (10.2 percent of all arrivals). Figure 3 (A) shows the intra-urban migration patterns from 2001 to 2003 aggregated to the scale of the Census blockgroup. Blue shades signify areas with net gains in homeowners, such as the lake area in the relatively affluent northwest part of Danville, whereas red shades represent areas of outflow. However, as shown in Figure 3 (B), if all of the homeownership changes – new owners as well as owners who left Danville – are included, the pattern nearly inverts. An increase in the overall number of homeowners includes new arrivals to Danville as well as renters who become owners. So the overall increase in ownership in lower-income areas (such as the southeast section of Danville) likely reflects a significant number of transitions from renting to owning a home.

Tracking these owner-occupied household data over a two-year interval revealed a northwest migration pattern within Danville, as congruent with informal appraisals of Danville neighborhoods offered by city officials who were consulted during this study. This internal migration pattern contrasted with the overall flows of outmigration and new ownership. Figure 3 demonstrates that despite its appeal to residents of Danville, the attractive northwest lake area experienced net losses in ownership overall. Outmigration

from Danville was evident in the more affluent neighborhoods, ostensibly where property values could be recouped as mobile professionals relocated for better job opportunities. Balancing this outflow of relatively affluent owners was a net increase in homeownership in the southeast part of town, where homes are more affordable. An influx of ownership in lower income areas (such as the KTD neighborhood) revealed the significance of renter to owner transitions. Since the property values had reached bottom in lower income areas, some landlords were willing to sell houses to tenants on contract after they made enough money from renting.

Via a musical chair kind of mechanism, Hoyt's (1939) vacancy chain hypothesis helps to explain the intra-urban and inter-urban migration patterns observed for Danville in Figure 3. Hoyt's vacancy chain mechanism proposes that affluent homeowners who leave the lake area make vacancies available to upwardly mobile households within Danville who aspire to that neighborhood. These transitions create additional vacancies, as more modest homes become available to lower-income households, who then vacate deteriorating homes that may be leased to renters or sold to new owners.

For this study, both qualitative and quantitative data provided patterns that shaped the formulation and analysis of the model. In particular, the empirical homeowner migration patterns described above were employed as a direct basis of comparison for simulated migration outcomes.

3. Model Development

To simulate social influence on household mobility, an agent-based model is constructed to explore the role of social connections in shaping a household's decision to move to a new location in Danville. As household agents relocate over time, social ties adjust, shaping the nature of neighborhood networks. Embedded and replicated in the main class of the model, household agents make two dynamic decisions: social network choice and neighborhood choice. These decisions are described in the following sections. The full model code is available online.ⁱ

The process of identifying factors of household choice was informed by the empirical context of the Danville study. Because friendships among children were revealed through interviews to be a significant factor in forming neighborhood networks, the family status factor was included in the formulation for social network utility. According to the homophily heuristic, the agent-based model constructed in this study enables social connections to be formed and broken on the basis of proximity as well as similarity of demographic attributes. As specified in the sections below, income affects social network choice as well as neighborhood choice, through the affordability and attractiveness terms. The distance effect in choosing a social network, and the social network effect in choosing a neighborhood, together constitute a reflexive relationship between neighborhoods and social networks.

3.1 Social Network Choice

A discrete choice methodology is widely applied to urban and transportation problems (Ben-Akiva and Lerman 1985, McFadden 1991). Extensions demonstrate the use of logit models for discrete choice selection of social connections (Van de Bunt et al. 1999) and for choices made in the urban context (Waddell et al. 2003, Paez and Scott 2007). The social network choice in Equation 1 employs a binary logit expression, such that the probability of a household i connecting with household j is based upon the exponential of the utility of that connection, divided by one plus the same exponential term (Ben-Akiva and Lerman 1985, Train 2003).

$$P_{ij} = \frac{\exp(U_{ij})}{1 + \exp(U_{ij})} \quad (1)$$

where

P_{ij} = Probability of household i connecting to household j

U_{ij} = Utility of connecting household j to household i

The logit expression assumes that unobserved factors are independent over time in a repeated choice situation, and represents systematic variation of preferences (Train 2003). As is the case in this implementation, a normalization constant is often applied to the denominator of the utility term to account for the variance of the unobserved portion of utility and scale the elasticity of choice (Train 2003, p.44).

Household agents evaluate the utility of their social network at a stochastic and asynchronous frequency. At the moment of network re-evaluation, each household picks another household and tests whether the probability from the binary logit expression in Equation 1 passes a stochastic satisfaction threshold. Each simulated household determines the utility of a prospective social tie using the expression in Equation 2:

$$U_{ij} = C + \alpha_D \cdot D_{ij} + \alpha_I \cdot \frac{|I_i - I_j|}{I_{avg}} + \alpha_R \cdot R_{ij} + \alpha_F \cdot F_{ij} \quad (2)$$

where

C = Constant average utility

D_{ij} = Distance between centroids of Census blocks containing household i and j

I_i, I_j = Income of household i and household j respectively

I_{avg} = Average income across all households in the city

R_{ij} = Binary term for similarity of racial category

F_{ij} = Binary term for similarity of family status

α_D = Utility weight for the block distance effect, [0, -1]

α_I = Utility weight for the similarity of income effect, $[0, -1]$

α_R = Utility weight for the effect of racial similarity, $[0, 1]$

α_F = Utility weight for the effect of family status similarity, $[0, 1]$

Equation 2 describes multiple effects on social network choice. The utility of a social connection is expressed as a set of alpha (α) weights applied to various homophily effects. For the effects of block distance and income, a greater difference decreases the likelihood of connection, and thus the α_D and α_I weights range between zero and -1 . Household income is employed as a proxy for socio-economic status. Additional effects in Equation 2 include binary categories for racial similarity (set to one if both households identify as white, or if both households identify as nonwhite) and similarity of family status (set to one if children are present in both households). If the prospective household identifies with the same racial category, the value of the term R_{ij} is one. Likewise, if both households have children, the binary term F_{ij} is encoded as one. If either household does not have children, F_{ij} is zero.

3.1.2 Probability of Local Ties—The probability of connection as generated by the logit formula in Equation 1 and the utility expression in Equation 2 is part of an agent-level algorithm that also invokes a probability of local ties (*localProb*), where “local” is defined as within the same Census blockgroup. The *localProb* parameter establishes the odds of selecting a connection from within the same blockgroup. When the connection algorithm is executed, a random number between zero and one is compared to the *localProb* odds. For a prospective social tie, if the probability of local ties exceeds the random number and both households reside in the same blockgroup, or if *localProb* is less than the random number and they reside in different blockgroups, then the utility of the connection is evaluated in terms of the utility (Equation 2) and thus probability (Equation 1) of forming a particular connection. If the probability of connecting according to Equation 1 exceeds another random number between zero and one, then the connection is formed between the two households. Otherwise, additional prospective households are considered from within (or outside of) the home blockgroup until this probability threshold is exceeded.

Via this connection algorithm, the probability of local ties broadly determines what fraction of ties will be made within the same blockgroup. The probability of local (intra-blockgroup) ties is one of two geographic factors shaping whether a social connection forms between two households. While block-level centroids are used to establish a coarse measure of distance between households in Equation 2, the probability of local ties imposes a spatially variable, blockgroup-level “local” container upon the formation of social ties. The reason for including this probability in the connection algorithm is to impose a different geographic component on social network choice from the block centroid-based distance between households.

3.2 Neighborhood Choice

In considering where to live, the first step each household agent takes is to assess, via a heterogeneously assigned “happiness” threshold, whether it is satisfied in its current location. If the household agent is dissatisfied, it selects a set of vacant parcels at random

from the larger set of all available (vacant) parcels in Danville. If the utility of these alternative parcels exceeds the current location utility, the parcels become part of a consideration set. The limited size of the consideration set (up to 15 parcels under base case assumptions) is a proxy for a scan of the real estate section of the newspaper. Once the consideration set is defined, the location with the highest neighborhood utility is chosen.

For neighborhood choice, the utility of a parcel p to household h invokes a constant C term and a set of weighted effects, as expressed in Equation 3.

$$U_{p,i} = C + \alpha_{SN} \left(\frac{SN_{i,bg(p)}}{SN_i} \right) + \alpha_{Attract} \left(\frac{I_{b(p)} - I_{avg}}{I_{avg}} \right) + \alpha_{Afford} \cdot A_{p,i} \quad (3)$$

where

$U_{p,i}$ = Utility of parcel p to household i

α_{SN} = Utility weight for the social network effect, [0,1]

SN_i = Size of social network for household i , as number of connections

$SN_{i,bg(p)}$ = Size of social network for household i within blockgroup containing parcel p

$\alpha_{Attract}$ = Utility weight for the effect of neighborhood attractiveness, [0,1]

$I_{b(p)}$ = Average income of the block containing the prospective parcel p

α_{Afford} = Utility weight for the affordability effect, [0,1]

$A_{p,i}$ = Neighborhood affordability for parcel p to household i

After the constant C term on the right side of Equation 3, the first effect on the utility of a prospective parcel is that of the social network, expressed as the proportion of a household's social ties that reside in the destination blockgroup. The composition of the social network is determined by the utility (Equation 2) and consequent probability (Equation 1) of forming a social tie with another household. The social network effect in Equation 3 completes a reinforcing, or positive, feedback mechanism in which the longer a household resides in a neighborhood, the more likely it is to have social ties in the neighborhood, and the utility of staying in the neighborhood therefore increases.

The second effect on neighborhood choice in Equation 3 assesses neighborhood attractiveness as the ratio of the average income of the parcel's block relative to the average income of the entire community (I_{avg}). As with the social network effect, the attractiveness effect also creates a positive feedback mechanism: as more affluent households move to a neighborhood, the neighborhood becomes more affluent, and therefore more attractive to other households. This attractiveness is counterbalanced by the third term in Equation 3, which represents the affordability constraint, $A_{p,i}$, as defined in Equation 4.

$$A_{p,i} = \max \left(\frac{I_t - I_{b(p)}}{I_{b(p)}}, 0 \right) \quad (4)$$

The affordability term is negative if the household's income (I_i) is less than the average income of block containing the prospective parcel. If the household's income is sufficient, affordability is not a constraint. The affordability constraint is also excluded when evaluating the utility of staying in the current location. Therefore, the relative affordability of staying in place reflects the inconvenience or “cost” of relocating. This induces a balancing, or negative, feedback mechanism on the propensity of households to move to an affluent neighborhood.

Although income is part of both attractiveness and affordability terms, the former is via aggregate assessment of block income relative to the city, whereas the latter compares household income to prospective block income. An alternative indicator of neighborhood affluence would be a measure such as property value. However, the use of household income enabled a computationally efficient model design, since Census-derived income was encoded as a factor in both social network formation (Equation 2) and neighborhood choice (Equations 3 and 4). Associating household income with mobile agents facilitates the completion of both balancing and reinforcing feedback mechanisms across the individual and aggregate spatial scales of the model.

3.3 Decision Dynamics

The simulated rate at which households decide whether to move and connect with each other is probabilistic, sampling from an exponential probability distribution so as to facilitate asynchronous computation of household agent algorithms. The exponential form of probability distribution corresponds to a Poisson process in which events recur at a constant average rate. When aggregated, these individual events reflect a first-order time delay. While the exact timing of household decisions is probabilistic, the frequency of considering a move is normalized to once per decade for owners and annually for renters. Therefore, the frequency with which households revisit the decision of whether and where to move depends directly upon their current status as renters or owners.

The logic for simulating owner migration at a slower rate than that of renters is a heuristic drawn from interviews. For example, one homeowner noted that “a lot of the homes in this area are rentals, and rentals are going to attract more transient type of people.”

In contrast to the frequency of the household's relocation decision, social network behavior is not directly affected by renter or owner status, as the model simulates the frequency with which all households evaluate their social networks on an annual basis or when relocating. Indirectly, however, the initial distribution of renters and owners influences network connections through the proximity effect and the probability of local, intra-blockgroup ties.

3.4 Data Integration

The first phase of model development involved the construction of an abstract but spatially explicit prototype, non-GIS, two-neighborhood model of household network and neighborhood choice (Metcalf and Paich 2005). As consistent with the mixed-methods research design, GIS data processing was undertaken alongside model development and ethnographic fieldwork so that the final form of the agent-based model was empirically grounded. The process of data integration is illustrated in Figure 4, in which model objects were created from Danville data for model initialization. This process began at left with unstructured, or raw, data files. Raw data are considered unstructured because they lack semantic meaning. Geographic data from the city of Danville and the U.S. Census were processed to create four unstructured text files whose rows correspond to the number of objects.

As displayed at the left of Figure 4, the unstructured data were organized by Census blockgroup, Census block, parcels, and households. Household income, presence of children, and migration patterns were contained by 28 blockgroups. Race and housing status (rent, own, or vacant) were contained by 764 blocks; location coordinates were specified for the centroids of 13,166 parcels. Initial homeowner assignments were contained by the subset of 7,576 owner-occupied parcels.

The unstructured data were imported into the Eclipse Java development platform for processing. In this stage, shown in the center panel of Figure 4, the first task was to properly link the data so that blockgroups contain blocks, which in turn contain parcels. Parcels were linked to initial owners from the 2001 owner-occupied household list. Once the structured data were properly linked, the remaining objects and attributes were identified.

For the 764 selected blocks within Danville, attributes from the 2000 U.S. Census data were extracted for tenure (number of housing units owned, rented, and vacant) and for racial identity by tenure status (owners and renters). After owner-occupied parcels were specified from city data, the number of renter-occupied parcels was imputed from the ratio of rental to vacant housing units in each Census block. A racial status for each household was imputed on the basis of tenure at the Census block level. To do so, the probability of a household being classified as white was set to the fraction of white households within the owner-occupied and renter-occupied households. To represent race in binary form, simulated household racial categories were reduced to “white” or “nonwhite.” The Census racial category “Black or African American” constitutes the majority of nonwhite households in Danville. Because cross-tabulations were available at the Census blockgroup level by race, and race was first assigned at the finer-grain block level, household racial status was also used to impute whether children were present and to assign household income. Specifically, Census data at the blockgroup level were used to determine the prevalence of children among white and nonwhite households, as well as the distribution of income among white and nonwhite households.

The right panel of Figure 4 distinguishes between agent objects and state objects. In this nomenclature, state objects reflect attributes alone, whereas agents include a capacity for choice. This distinction streamlines the dynamic computational requirements of the model.

Spatially fixed state objects such as parcel, block, blockgroup, and household were serialized and structured in AnyLogic software. During model initialization, each dynamically active household agent was created from attribute information contained in the corresponding household state object.

4. Analysis of Simulation Results

Pattern-oriented modeling involves the use of empirical patterns to guide the process of both building and testing individual-based models (Grimm and Railsback 2005). The pattern-oriented strategy for model calibration used in this study is illustrated in Figure 5, as adapted from Grimm and Railsback (2005). Beginning at the top left in Figure 5, model parameters are set where possible using available data for attributes such as income, race, and parcel location (as described in Section 3.4 above). Then broad ranges are initially chosen for the remaining uncertain parameters, such as whether they are negative or positive. In this study, the distance and income parameter weights shaping social network choice (a_D and a_I in Equation 2) were bounded between zero and -1, so that a larger difference between households would reduce the perceived utility (and thus likelihood) of forming a social connection. All other parameter weights and linear constants in the logit functions specified in Equations 2 and 3 were bounded between zero and one. To provide a range of elasticity for the choice mechanism, normalizing constants applied to the denominator of the utility equations were bounded between 0.1 and 2. In addition, the probability of evaluating social connections within the home blockgroup, *localProb*, was varied between a 20 and 70 percent chance of selecting a household from within an agent's blockgroup.

Another step in the pattern-oriented process, as indicated in the top right of Figure 5, is to specify alternative model structures. In this study, such structures are specified using different parameter settings for the utility weights of social network and neighborhood choice. After specifying initial conditions and alternative model structures, observed patterns are selected to serve as filters for reasonable model settings. The intra-urban homeowner migration data described in Section 2.2 and illustrated previously in Figure 3 (A) are used to define an objective function that minimizes the difference between simulated and observed migration (see Section 4.1 below).

The model testing phase involved simulating outcomes from different model settings to evaluate which parameter combinations best matched the observed migration patterns. To do so, a Monte Carlo method of optimization was undertaken in which agent choice parameters were varied systematically using repeated random sampling within the specified bounds. This process was performed using AnyLogic's OptQuest algorithm for 1000 consecutive, unique parameter sets to facilitate simulation experiments under uncertainty. Each experiment was comprised of a batch of 25 stochastic simulation runs from a single initial seed and set of parameter values. Therefore, for the 1000 experiments conducted, a total of 25,000 individual simulations were run. Average statistics for each batch were computed to account for stochastic variations between individual simulation runs. Sources of stochasticity include the timing of household choice and thresholds for forming connections.

4.1 Measuring Migration Error

Each simulation was run for a two-year period and evaluated relative to the observed homeowner migration patterns from 2001 to 2003. Directional migration patterns were assessed using a matrix to represent moves from and to each of Danville's 28 blockgroups. A directional move error was computed as the sum of the move error for each cell in the 28×28 blockgroup matrix. In addition, aggregate move error was computed as the difference between overall simulated and observed homeowner migration patterns in Danville.

$$dirErr = \sum_{i=1}^n \sum_{j=1}^n |simM_{ij} - obsM_{ij}| \quad (5)$$

where

$dirErr$ = Directional homeowner migration error

$simM_{ij}$ = Simulated migration from blockgroup i to blockgroup j

$obsM_{ij}$ = Observed migration from blockgroup i to blockgroup j

The aggregate migration error provides a check on the overall accuracy of simulated intra-urban migration volume. As expressed in Equation 6, the cumulative simulated moves are compared with cumulative observed moves to determine a total, aggregate measure of move error. For the period from 2001 to 2003, there were a total of 139 homeowners who moved within the city. Therefore, the right-most expression in Equation 6 sums to 139 moves.

$$totErr = \left| \sum_{i=1}^n \sum_{j=1}^n simM_{ij} - \sum_{i=1}^n \sum_{j=1}^n obsM_{ij} \right| \quad (6)$$

where

$totErr$ = Total, aggregate homeowner migration error

The average move error was calculated as a summary statistic for each batch of 25 simulation outcomes by combining the directional and overall moves. The combined, average migration error provides the objective function expressed in Equation 7.

$$\min \quad avgErr = \frac{\sum_{s=1}^r dirErr_s + \sum_{s=1}^r totErr_s}{r} \quad (7)$$

where

$avgErr$ = Average homeowner migration error

r = Batch size (25), number of replications for a single parameter combination

s = Index for individual simulation run in an experiment

This objective function was designed to minimize the average homeowner migration error by adjusting the parameter settings to create 1000 unique simulation experiments. The optimization algorithm used in this study specifically tests the boundaries of different parameters in combination, 'learning' to avoid parameter spaces that result in the most mismatches relative to the observed migration data (April et al 2004).

4.2 Regression Analysis

As an indicator of how well the simulated migration patterns matched the Danville homeowner moves, the average homeowner migration error (*avgErr* in Equation 7) was used as a dependent variable in a regression analysis of the simulation results. In this analysis, the independent variables are household choice factors (α terms in Equations 2 and 3) that were adjusted to create the 1000 experiments simulated. Because the agent-based model is inherently dynamic, the regression analysis is exploratory in nature. Therefore, both non-linear and linear regressions were estimated. With the simulation experiment as the unit of analysis, Table 1 presents the regression results, reporting the coefficients and t -statistics for the two specifications. The non-linear regression included squared and linear terms for the independent variables. While the non-linear form of regression fit the data better (R^2 of 0.701 versus 0.478), it also carried the explanatory burden of having twice as many independent variables.

Lower average migration error suggests that the agent-based model is better accounting for the Danville migration decisions. So, for the linear regression, the choice factors that best account for the observed migration are those that have a statistically significant negative effect. For the non-linear regression, though, the interpretation is more subtle. Factors with a significant positive coefficient for the linear term and a significant negative coefficient for the non-linear term could either be irrelevant or highly important since mid-range values (across the experiments) produce the highest migration error. In contrast, logit choice utility weights with a significant negative coefficient for the linear term and a significant positive coefficient for the non-linear term are factors that clearly matter in explaining the Danville migration choices, but whose impact is bounded since setting the value too high may result in an increase in migration error.

Considering the geographic factors first, Table 1 shows that the probability of local (intra-blockgroup) ties reduced migration error, with large negative, statistically significant coefficients in both the linear and non-linear regressions (even though the coefficient of the squared term is positive, the net effect is negative over most of the parameter's range from 20 to 70 percent across the experiments). The influence of distance on network formation (α_D) also reduced migration error, but its coefficients are weak substantively and statistically. Rather than centroid-based block distance, therefore, the most significant geographic influence was the probability of forming social ties within the same blockgroup.

As with the geographic effects, utility weights for other components of social network choice also had negative coefficients, indicating that they reduce simulated migration error. Among these, utility weights for racial similarity (α_R) and income difference (α_I) reduced migration error with statistically significant negative coefficients in both the linear and nonlinear specifications. Utility weights for family status (α_F) and neighborhood

affordability (α_{Afford}) also help to account for homeowner migration patterns. The effect for homophily of family status, inferred by the presence of children, consistently reduced migration error in both regressions (with an inflection point outside the parameter range in the non-linear specification), even though its t -statistic does not reach the standard level of statistical significance. Similarly, the neighborhood affordability constraint has a negative but insignificant coefficient in the linear regression, while lowering migration error over most of its range in the non-linear regression (only increasing error for values above 0.67).

The utility weight for the social network effect (α_{SN}) on neighborhood choice resulted in a positive coefficient in the linear regression, indicating that increases in the social network influence on neighborhood choice would decrease the ability of the simulation to match observed homeowner migration. Although it is not significant in the linear regression, the influence of the social network on location choice is highly significant in the non-linear regression. The coefficients signify that the non-linear social network effect mitigates migration error when α_{SN} exceeds 0.69. This non-linearity indicates that some parameter combinations reduce simulated error under a strong social network effect, while other settings reduce error with a weak network effect.

Finally, Table 1 indicates that neighborhood attractiveness ($\alpha_{Attract}$) has a statistically significant positive coefficient in the linear regression, and the positive coefficient for its squared term in the non-linear regression is large enough that its net effect is positive over most of the parameter range across experiments. More specifically, above the value of 0.25, increases in the utility weight on neighborhood attractiveness exacerbate simulated migration error. For this statistically significant influence, smaller weights produce better matches with observed migration patterns.

4.3 Analysis of Filtered Monte Carlo Experiments

The relationship between aggregate and directional migration error is shown in Figure 6 for the set of 1000 Monte Carlo simulation experiments. This relationship reveals that if directional error were minimized completely, the trivial solution of no migration would be preferred over more meaningful scenarios. As shown in the left-most results of Figure 6, if no intra-urban migration occurs, the comparison to each of the 139 observed homeowner moves for the 2001-2003 time period produces equivalent errors in both the aggregate and directional dimensions, resulting in a combined error of 278 moves (200 percent of observed migration). If aggregate moves were exactly matched to the observed moves (zero aggregate error), but all of the moves were from and to the wrong blockgroups, directional error would be 278 moves.

As illustrated in Figure 6, aggregate migration error ($totErr$ in Equation 6) approaches zero when directional error ($dirErr$ in Equation 5) reaches approximately 220 moves. A threshold of 230 moves for the combined (directional and aggregate) migration error as averaged for the 25 realizations of each experiment ($avgErr$ in Equation 7) was used to filter 486 of the 1000 experiments for further analysis. Figure 6 highlights the aggregate and directional migration error of these filtered experiments.

A cluster analysis of parameter weights within this set of filtered experiments reveals the relative performance of simulated combinations of effects. This analysis was performed in SPSS as a two-step cluster classification using the Bayesian Information Criterion with a log-likelihood measure of distance in parameter space. Table 2 shows the average error and parameter settings for each cluster, along with the best and worst cases produced in the set of 1000 Monte Carlo experiments.

Three clusters of similar size (N=148, 163, 175) emerged from analysis of the 486 filtered solutions. Cluster 3 resulted in a slightly lower average migration error than the other two clusters. Cluster 3 is similar to cluster 2 in terms of parameters such as the affordability (α_{Afford}) and social network (α_{SN}) effects on location choice. Examination of these clusters shows that the probability of selecting social ties within one's blockgroup (*localProb*) approximated or exceeded 50 percent for all three clusters. In contrast, the distance effect on social network choice varied by cluster, which is consistent with the weak effect of distance in the regression analysis described above.

As indicated by the best case scenario in Table 2, the smallest average migration error achieved in a simulation experiment was 219 moves over two years. While this simulated migration error is 158 percent of the 139 observed moves, it produces a much better match than the worst case scenario of 978 moves. The large error of the worst case is induced by excessive intra-urban homeowner migration, creating 408 extraneous moves during the two-year time span. As shown in Table 2, the worst case effectively turns off the affordability constraint by setting α_{Afford} to zero while maximizing the attractiveness and social network effects on neighborhood choice. The dynamic behavioral consequences of these differences in parameter settings are explored in the following section.

4.4 Simulation of Spatial Inequality

To reveal how simulated household behavior impacts spatial inequality, the Gini coefficient was applied to incomes averaged at the blockgroup level. The Gini coefficient is evaluated as a fraction between zero and one, with higher values representing greater inequality (Gini 1921). The spatial dimension of this inequality was provided by comparing the 28 blockgroups containing households. Equation 8 articulates the formula used to assess the Gini coefficient across the 28 blockgroups in the Danville model.

$$G = \frac{1}{n} \cdot \left(n+1 - 2 \cdot \frac{\sum_{k=1}^n (n+1-k) \cdot I_k}{\sum_{k=1}^n I_k} \right) \quad (8)$$

where

G = Gini coefficient

n = number of blockgroups (28)

k = index of blockgroup in non-decreasing rank order

I_k = average income of k^{th} blockgroup

Figure 7 illustrates the dynamics of this spatially-defined Gini coefficient over a period of 20 years under the average parameter settings for Clusters 1, 2, and 3 (see Table 2). These three scenarios exhibit a qualitative similarity in that the inequality between blockgroups increases from a starting Gini coefficient of approximately 0.12 in the first few years of the simulation, then declines to a level similar to the starting point. Because it reflects the differences in average income by Census blockgroup, the blockgroup Gini coefficient is lower (more equal) than it would be if household incomes were directly compared to each other. For example, the 2010 American Community Survey of the U.S. Census indicates that the Gini coefficient for income variation between households in Danville is 0.43, slightly less than the Gini coefficient for households nationwide and statewide, which is over 0.46 and has been trending toward increased inequality in recent decades.

The model's simulated return to a dynamic equilibrium near the starting level of blockgroup inequality in Figure 7 reflects the impact of inter-urban migration. Under the base case assumption of 500 annual household arrivals and departures, random (and thus equalizing) effects were induced from the stochastic parcel assignments for immigrants and stochastic selection of emigrants. In simulations for which there were no arrivals to or departures from Danville, blockgroup inequality for the filtered solutions increased to a new steady state more rapidly and with less noise than for the base case of inter-urban migration.

Inequality dynamics for the three clusters shown in Figure 7 are qualitatively similar, revealing an initial transient to a state of heightened inequality that eventually returns to a dynamic equilibrium closer to the initial condition. The transitional period illustrated in Figure 7 for the dynamic realization of inequality under Cluster 2 settings lasts approximately 10 years, whereas the other two cluster settings produce a steadier trend of reduced inequality. In contrast to the cluster comparison, more dramatic differences in dynamic behavior are evident in Figure 8, which shows trajectories of inequality resulting from the parameter settings that generated the best and worst matches to observed data in the Monte Carlo optimization experiments (Table 2 lists the corresponding parameter settings).

The dynamic behavior of the best-matched scenario in Figure 8 continues the qualitative pattern of the cluster dynamics in Figure 7, in which blockgroup inequality increases at first but eventually returns to a level similar to the starting point. In contrast, the worst-matched scenario in Figure 8 shows a sharp drop in blockgroup inequality during the first two years, followed by a more gradual decline over the remainder of the time horizon. In addition to contrasting best and worst cases, Figure 8 illustrates the dynamics associated with a scenario of no migration into or out of Danville. In contrast to the base case assumption of 500 households per year arriving to and departing from Danville, the absence of inter-urban migration creates a steadier transition to a higher equilibrium value for the Gini coefficient.

Simulated inequality dynamics reveal that increases in rates of both intra-urban migration (as in the worst-matched scenario) and inter-urban migration (to and from Danville) have the effect of lowering the inequality between blockgroups. While households leaving Danville break social ties, arriving households become integrated into the simulated social network. Transient model dynamics thus reflect the movement of households and adjustment of their social connections.

4.5 Limitations and Extensions

In developing the model, simplifying assumptions were made so that computation would be efficient during the iterative Monte Carlo optimization process. For example, the use of 764 census block centroids as a basis of distance between households enabled a much smaller distance matrix to be referenced during computation than the full set of 13,166 parcels. This simplification improved computational ease and eliminated the need for distance to be computed dynamically, but the resolution lost from using block centroids limited the significance of the distance effect on simulated migration error. While the probability of local ties was a significant influence on migration outcomes, the use of blockgroups to bound local ties introduced limitations from variability of the geographic area occupied by different blockgroups.

Home ownership was the most accurately initialized attribute in the model, since it was linked to the parcel data used to deduce household migration patterns. Initialization of other household agent attributes (race, income, presence of children) from Census data produced limitations of disaggregating from frequencies reported at the block or blockgroup level. Although homeowner migration data were available at the parcel level, migration patterns were aggregated to the blockgroup level in part to compensate for potential errors induced by model initialization of the other household attributes.

The difficulty of matching directional migration patterns is largely due to the sparseness of the observed homeowner migration matrix aggregated to the level of the blockgroup. Just 11 percent (86 of 784) of the cells in the matrix were non-zero, containing the 139 total intra-urban homeowner moves from 2001 to 2003. The sparseness of this matrix reflects the relatively few homeowner moves in the two years used to assess homeowner migration patterns. Inclusion of renter moves would provide a richer pattern to match, but such data were not available for this study.

While the Monte Carlo parameter variation undertaken herein is in keeping with the pattern-oriented approach for exploring an underdetermined system, further experiments could be designed to vary, test, and assess: the transition between renting and owning, move evaluation frequency for renters and owners, rates of Danville in- and out-migration (assumed to be in equilibrium for this study), size of the vacant parcel consideration set, and network size (a minimum number of connections was encoded at three ties for this study, although the algorithms produce a range of household network sizes). Because the simulated inequality dynamics exhibit initial transients, extensions of this work would benefit from allowing a transition prior to the calibration period.

For the sake of parsimony, the model simulates social networks using a logit choice formulation that enables indirect parameter estimation but does not reflect the ways that social ties between households actually form and influence migration patterns. As with dynamics of marriage and divorce, the model excludes social institutions, yet affiliations with schools, churches, and places of work clearly facilitate connections between households. The model also omits institutional constraints such as discriminatory lending in the housing market that would exacerbate the inequalities simulated. While the U.S. foreclosure crisis was beyond the scope of the model as designed, aspects of the crisis could

be proxied by adjusting migration rates, household income, and odds of owner-renter transitions.

Although institutional constraints are not explicitly represented in the model, household agent choices are constrained significantly by the particular parameter settings governing the simulation. These reflect tendencies toward homophily in terms of the attributes simulated (proximity, race, income, presence of children). The nature and intensity of a homophily effect was tested by varying logit choice utility weights to create alternative model structures reflecting distinct social norms. Thus, while agents are assigned heterogeneous attributes and are free from institutional constraints, they are subjected to universal social norms in the form of their choice functions. Because agent choices are also probabilistic, an important element of chance is embedded in the simulated household decisions.

5. Conclusion

This research develops an agent-based model for examining the dynamics of spatial and social influences on neighborhood and network choice. This study demonstrates the use of pattern-oriented modeling to explore simulated relative to observed migration patterns. Results show how a simulation model can be used to assess the importance of different socio-economic factors in shaping migration patterns. Inclusion of choice at the scale of the household in a spatial dynamic simulation model is an important step in representing multidimensional complex systems (Agarwal et al. 2002). Drawing upon patterns from qualitative and quantitative data sources, the model functions as a computational laboratory to test effects of alternative assumptions for household heuristics and social structures.

In the pattern-oriented calibration strategy, a Monte Carlo optimization algorithm was designed to test unique combinations of utility weights for each experiment, or batch of 25 simulation runs. The objective function minimized homeowner migration error as averaged for each experimental setting. With an average migration error of 219 moves in the best-matched scenario (Table 2), an error threshold of 230 moves was used to filter 486 reasonable solutions from the 1000 simulation experiments, indicating that the Monte Carlo algorithm produced many alternative parameter settings with error in the range of 220-230 moves. While effectively halving the parameter space precludes confidence about a singular extrapolation into the future, the filtering process enabled a finer analysis of utility weights and inequality dynamics among relatively fit solutions.

Regression of simulated results relative to factors shaping household choice sheds insight on hypothesized socio-economic dynamics. While increasing the influence of social networks on neighborhood choice exacerbated simulated migration error, the effect was not statistically significant in a linear regression model of the simulation results. However, the social network effect was significant in a non-linear form of the regression. Both regression and cluster analyses reveal the role of local (intra-blockgroup) ties in matching observed migration patterns.

Social network factors of racial similarity, income difference, and probability of local ties had significant mitigating effects on the homeowner migration error in both forms of

regression. The finer block-level resolution of the household race assignment from the Census data, along with the use of racial category in imputing other household attributes such as tenure (rent or own) status, presence of children, and income at the blockgroup scale, may have contributed to the significance of race in the regression of simulation results against observed migration patterns. These results, along with the significance of income-based homophily, suggest that such social sorting can contribute to intra-urban inequalities.

The model's emphasis on individual choice, exclusion of institutional constraints, and application of preference settings to all agents, are consistent with micro-level, bottom-up explorations such as the classic Schelling (1971) segregation model and the suite of models developed by Epstein and Axtell (1996). These models demonstrate the potential for simple rules to explain complex patterns of aggregate behavior. Specifically, Schelling (1971) considered a situation in which location preferences are shaped by an adjustable parameter for tolerance of diversity, finding that a preference for homophily with just a minority of one's neighbors nevertheless resulted in segregation. Continuing this line of inquiry, the model described herein examines linkages between social network structure and household migration using an experimental design that adjusts parameters for the strength of network homophily and the relative role of social networks, neighborhood attractiveness, and affordability in neighborhood choice.

By explicitly linking social network structure with neighborhood choice and integrating GIS data into a dynamic model, this study contributes to a growing body of research that simulates social networks to illuminate a wide range of issues in geographic context, from health (e.g., Bian et al. 2012) to hurricane evacuation (e.g., Widener et al. 2012). Using both quantitative and qualitative patterns to guide model design, this research demonstrates how simulation models can be empirically grounded while used to explore hypothetical scenarios that deal with persistent social problems. The pattern-oriented modeling approach and agent-based framework employed for this work can be adapted to other urban areas facing issues associated with economic decline, such as community fragmentation, polarization, and disparity.

Acknowledgments

As a reflection of original research conducted for my dissertation (Metcalf 2007), this study benefitted from the guidance of my advisor, Bruce Hannon, and committee members Mark Paich, Sara McLafferty, Mu Lan, and Jim Westervelt. Lyle Wallis assisted with technical aspects of model implementation. Peter Rogerson, Teri Metcalf, Harvey Palmer, and anonymous reviewers provided valuable input into the presentation and interpretation of the results in this paper. This work was supported in part by NSF award HSD-0433165 (model development and analysis activities) and NIH (NIDCR/OBSSR) awards 1R21DE021187-01 and 1R01DE023072-01 (synthesis and documentation activities).

References

- Agarwal, C.; Green, GM.; Grove, JM.; Evans, TP.; Schweik, CM. General Technical Report NE-297. U.S.Department of Agriculture, Forest Service, Northeastern Research Station; Newton Square, PA: 2002. A Review and Assessment of Land-Use Change Models: Dynamics of Space, Time, and Human Choice..
- AnyLogic software, version 5.5. XJ Technologies. St. Petersburg, Russia: <http://www.anylogic.com/>
- April J, Glover F, Kelly JP, Laguna M. The exploding domain of simulation optimization. Newsletter of the INFORMS Computing Society. 2004; 24(2):1–14.

- Batty, M. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals*. MIT Press; Cambridge, MA: 2005.
- Ben-Akiva, M.; Lerman, SR. *Discrete Choice Analysis*. MIT Press; Cambridge, MA: 1985.
- Bian L, Huang Y, Mao L, Lim E, Lee G, Yang Y, Cohen M, Wilson D. Modeling individual vulnerability to communicable diseases: a framework and design. *Annals of the Association of American Geographers*. 2012; 102(5):1016–1025.
- Conley TG, Topa G. Identification of local interaction models with imperfect location data. *Journal of Applied Econometrics*. 2003; 18(5):605–618.
- Crooks A. Constructing and implementing an agent-based model of residential segregation through vector GIS. *International Journal of Geographical Information Science*. 2010; 24(5):661–675.
- Emch M, Root ED, Giebultowicz S, Ali M, Perez-Heydrich C, Yunus M. Integration of spatial and social network analysis in disease transmission studies. *Annals of the Association of American Geographers*. 2012; 102(5):1004–1015.
- Epstein, JM.; Axtell, R. *Growing Artificial Societies: Social Science from the Bottom Up*. Brookings Institution Press; Washington, DC: 1996.
- Gimblett, HR., editor. *Integrating Geographic Information Systems and Agent-Based Modeling Techniques for Simulating Social and Ecological Processes*. Oxford University Press; New York, NY: 2002.
- Gini C. Measurement of inequality and incomes. *The Economic Journal*. 1921; 31:124–126.
- Grimm, Volker; Railsback, Steven F. *Individual-Based Modeling: Lessons from Ecology*. Princeton University Press; Princeton, NJ: 2005.
- Hoyt, H. *The Structure and Growth of Residential Neighbourhoods in American Cities*. Federal Housing Administration; Washington, DC: 1939.
- Kwan M-P. Mobile communications, social networks, and urban travel: hypertext as a new metaphor for conceptualizing spatial interaction. *The Professional Geographer*. 2007; 59(4):434–446.
- McFadden D. Advances in computation, statistical methods, and testing of discrete choice models. *Marketing Letters*. 1991; 2(3):215–229.
- McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. *Annual Review of Sociology*. 2001; 27:415–444.
- Metcalf, S. *Simulating the Social Dynamics of Spatial Disparity through Neighborhood Network Evolution*. Doctoral dissertation. University of Illinois at Urbana-Champaign; Urbana, IL: 2007.
- Metcalf, S.; Paich, M. Spatial dynamics of social network evolution.. *Proceedings of the 23rd International Conference of the System Dynamics Society*.; Boston, MA. System Dynamics Society; 2005. <http://www.systemdynamics.org/conferences/2005/proceed/papers/METCA487.pdf>
- Osgood N. Lightening the performance burden of individual-based models through dimensional analysis and scale modeling. *System Dynamics Review*. 2009; 25(2):101–134.
- Paez A, Scott DM. Social influence on travel behavior: a simulation example of the decision to telecommute. *Environment and Planning A*. 2007; 39(3):647–665.
- Parker DC, Manson SM, Janssen MA, Hoffman MJ, Deadman P. Multi-agent systems for the simulation of land-use and land-cover change: A Review. *Annals of the Association of American Geographers*. 2003; 93(2):314–337.
- Peake LJ. Toward a social geography of the city: race and dimensions of urban poverty in women's lives. *Journal of Urban Affairs*. 1995; 19(3):335–361.
- Radil SM, Flint C, Tita GE. 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*. 2010; 100(2):307–326.
- Rainie, L.; Wellman, B. *Networked: The New Social Operating System*. MIT Press; Cambridge, MA: 2012.
- Resnick, M. *Turtles, Termites, and Traffic Jams: Explorations in Massively Parallel Microworlds*. MIT Press; Cambridge, MA: 1994.
- Rowe S, Wolch J. Social networks in time and space: homeless women in Skid Row, Los Angeles. *Annals of the Association of American Geographers*. 1990; 80(2):184–204.

- Schelling TC. Dynamic Models of Segregation. *Journal of Mathematical Sociology*. 1971; 1(2):143–186.
- Schelling, TC. *Micromotives and Macrobehavior*. W.W. Norton and Company; New York, NY: 1978.
- Van de Bunt G, Duijin MV, Snijders T. Friendship networks through time: an actor-oriented dynamic statistical network model. *Computational and Mathematical Organization Theory*. 1999; 5(2):167–192.
- Waddell P, Boring A, Noth M, Freier N, Becke M, Ulfarsson G. Microsimulation of urban development and location choices: design and implementation of UrbanSim. *Networks and Spatial Economics*. 2003; 3(1):43–67.
- Westervelt JD, Hopkins LD. Modeling mobile individuals in dynamic landscapes. *International Journal of Geographical Information Science*. 1999; 13(3):191–208.
- Widener M, Horner M, Metcalf S. Simulating the effects of social networks on a population's hurricane evacuation behavior. *Journal of Geographical Systems*. 2012 doi: 10.1007/s10109-012-0170-3.
- Wiegand T, Jeltsch F, Hanski I, Grimm V. Using pattern-oriented modeling for revealing hidden information: a key for reconciling ecological theory and application. *Oikos*. 2003; 100(2):209–415.
- Yin L. The dynamics of residential segregation in Buffalo: an agent-based simulation. *Urban Studies*. 2009; 46(13):2749–2770.

Figure 1.
Danville Population Trend.

Figure 2.
Owner Occupancy in Danville Census Blocks.

Figure 3.
Home Ownership Changes (A) Within Danville and (B) Overall from 2001 to 2003.

Figure 4.
Process for Data Integration.

Figure 5.
Pattern-Oriented Calibration Strategy.

Figure 6.
Aggregate and Directional Homeowner Migration Error.

Figure 7.
Inequality Dynamics under Average Parameter Settings for Three Clusters.

Figure 8.
Blockgroup Inequality for Best and Worst Cases with Inter-Urban Migration Effect.

Table 1

Coefficients from Regression of Neighborhood and Network Effects on Migration Error.

	Non-Linear Regression		Linear Regression
	Coefficient for linear term	Coefficient for squared term	Coefficient
Utility Weights in Neighborhood Choice (Equation 3):			
Affordability (<i>locAfford</i> ; a_{Afford})	-2.007 (-4.94)	0.015 (4.44)	-0.129 (-0.90)
Attractiveness (<i>locAttract</i> ; a_{Afford})	-1.955 (-5.40)	0.039 (10.65)	1.526 (9.81)
Social Network Density (<i>locSN</i> ; a_{SN})	1.934 (3.96)	-0.014 (-3.00)	0.237 (1.27)
Utility Weights in Social Network Choice (Equation 2):			
Distance (<i>nwDist</i> ; a_D)	-0.265 (-0.55)	0.483 (0.54)	-0.196 (-0.85)
Income (<i>nwInc</i> ; a_I)	2.828 (5.82)	-0.034 (-7.63)	-0.766 (-4.85)
Race (<i>nwRace</i> ; a_R)	1.149 (2.92)	-0.014 (-4.30)	-0.629 (-4.05)
Family Status (<i>nwChild</i> ; a_F)	-0.339 (-0.62)	-0.003 (-0.63)	-0.305 (-1.69)
Probability of Local Ties (<i>localProb</i>)	-5.297 (-3.73)	0.045 (3.01)	-0.762 (-2.83)

For each of the regression coefficients listed above, the t-statistic is indicated in parentheses (). Coefficients with statistical significance ($|t| > 2$; p -value < 0.05) are highlighted in **bold**. To avoid squaring fractions, all independent variables were multiplied by 100 before conducting the regression. The dependent variable was the average migration error, *avgErr*, as defined in Equation 7.

Table 2

Average Error and Parameter Settings for Clusters and Extreme Cases.

	Cluster 1 (N=148)	Cluster 2 (N=163)	Cluster 3 (N=175)	Best Case	Worst Case
Components of Migration Error (Equations 5-7):					
Average, Combined Migration Error (<i>avgErr</i>)	226.63	226.90	225.43	219.20	977.76
Directional Migration Error (<i>dirErr</i>)	217.53	216.56	215.97	210.80	569.64
Total, Aggregate Migration Error (<i>totErr</i>)	9.10	10.34	9.46	8.40	408.12
Utility Weights in Neighborhood Choice (Equation 3):					
Affordability (<i>locAfford</i> ; α_{Afford})	0.9713	0.6922	0.9677	0.9754	0
Attractiveness (<i>locAttract</i> ; $\alpha_{Attract}$)	0.4621	0.3138	0.1957	0.2847	1
Social Network Density (<i>locSN</i> ; α_{SN})	0.6831	0.4167	0.6826	0.7373	1
Utility Weights in Social Network Choice (Equation 2):					
Distance (<i>nwDist</i> ; α_D)	-0.6721	-0.3468	-0.2116	-0.3566	-1
Income (<i>nwInc</i> ; α_I)	-0.7150	-0.4797	-0.4089	-0.4648	0
Race (<i>nwRace</i> ; α_R)	0.9324	0.5918	0.3262	0.5318	1
Family Status (<i>nwChild</i> ; α_F)	0.9238	0.6114	0.8720	0.9035	0
Probability of Local Ties (<i>localProb</i>)	0.5343	0.4659	0.6305	0.5912	0.2