Agent-based Simulation of Intra-urban Social Network Structures and Household Mobility

This paper assesses a number of influences on intra-urban neighborhood choice, using the community of Danville, IL as a real-world test case for the simulation work. Participant observation and several in-depth interviews among residents from a Danville neighborhood association were undertaken in late 2005 to develop a longitudinal sense of household mobility patterns. Although the interviewees were homeowners, many of their neighbors and prior experiences included the renter tenure category. This analysis revealed the significance of changes in family status (presence of children, marriage and divorce) as well as extended family ties. Geographic analysis of owner-occupied parcel data revealed that families sharing a last name lived statistically closer to each other than households with different last names. These sources of data were used to “ground-truth” the simulation model and identify its bounds, while Census data were combined with geographic parcel and tract boundaries to initialize the model.

With a starting distribution of renters, owners and other attributes specified according to the 2000 U.S. Census, parameters influencing household choice were adjusted to test consequences of more or less weight on a particular attribute (geographic proximity, similarity of income, race, family status). The probability of one household connecting to another was encoded with a logit choice structure in which utility depended upon household attributes. The influence of social network choice on neighborhood choice, as household agents considered whether and where to relocate within the community, was considered using a pattern-oriented calibration scheme in which the strength of social influence and other factors was varied and results were compared to observed migration patterns. This procedure minimized an error that combined total and directional (between Census blockgroups) moves, comparing simulated and observed patterns of owner-occupied household migration between 2001 and 2003.

Keywords: Social networks, agent-based modeling, neighborhood change

Contents:
Figure 1. Trends in Population and Land Area for Danville, IL, USA................................. 2
Figure 2. Map of Owner Occupancy in Danville Census Blocks................................. 3
Figure 3. Homeowner Migration Within Danville from 2001 to 2003.......................... 4
Figure 4. Home Ownership Changes Within Danville (A) and Overall (B) from 2001 to 2003... 5
Figure 5. Causal Hypothesis of Neighborhood System Dynamics................................. 7
Figure 6. Distance Between Households by Number of Households Sharing Last Names ....... 9
Figure 7. Distribution of Households, Names (L) and Matching Pairs (R) Compared.......... 10
Figure 8. Process for Creating Model Objects.............................................................. 11
Figure 9. Model Structure Screenshot in AnyLogic.......................................................... 12
Figure 10. Alternative Influences on Model Structure................................................... 14
Figure 11. Calibration Strategy......................................................................................... 15
Figure 12. View of Optimization Process.......................................................................... 16
Figure 13. Total and Directional Move Error for 1000 Experiments................................. 18
Figure 14. Spatial and Social Influences on Choice in Filtered Set of Solutions................. 18
Figure 15. Parameter Weight Ratios Among Most-Filtered Set of Solutions............... 19
Figure 16. Cluster Analysis of Parameter Weights in "Good" Solution Space............... 19

Table 1. Comparison of Households with Shared and Different Last Names............... 8
1. Background

This research links local ties with aggregate patterns of community cohesiveness to explore relationships between individual choices about where to live and social network structure, and between neighborhood and community networks. Underlying the aggregate phenomena of sprawl and spatial disparity is the individual choice of where to live. To systematically assess the role of social influence in neighborhood choice, this research simulates facets of neighborhood choice with dynamically simulated social networks. A social network is an abstraction of relationships that include trusted channels for communication - family, friends, and advisors. The potential effects of social networks on neighborhood choice over time are explored through virtual experimentation with an agent-based dynamic model. Simulated results are compared with local homeowner migration patterns recently observed within the community of Danville, Illinois, USA.

Over the past few decades, Danville has experienced population decline and stagnation, but still experiences sprawl in the form of new fringe development. The trends in population and land area over time in Figure 1 reveal that Danville’s land area has nearly doubled in the last half-century, though its population is the same as it was in 1920. As measured by the decade Danville’s population peaked in 1970, with well over 40,000 people. In the words of one city official, “we’re a town of 30,000 with an infrastructure for 60,000.” Even so, Danville remains the largest community in Vermilion county, with over 40% of the county’s population.

Until the 1980s Danville was a predominantly working class town hosting industries such as General Motors, General Electric, and ALCOA. The exodus of these industries created a
tension between attachment to hometown roots and the struggle to find work. Many middle-class workers left to find jobs elsewhere. Although much of Danville’s remaining population lives in poverty, the city maintains a powerful class of affluent professionals and doctors who serve the community’s hospitals and retirement homes. To stabilize local neighborhood dynamics, the city promotes the formation of community through resident-organized neighborhood associations.

Research participants from a Danville neighborhood association who provided in-depth interviews for this study spanned race, age, and family dimensions, but were all homeowners. (Income was not elicited in the interview process, but was imputed from the 2000 Census for the simulation analysis.) Perspectives on renters versus owners were elicited informally through participant observation of neighborhood association meetings (which were dominated by homeowners) and through conversation with city officials, who openly noted that the renter-owner ratio can reach a tipping point that creates instability in terms of outmigration. Early on, therefore, the distinction between renters and owners was highlighted as one that might make a difference in the model. One participant commented that “a lot of the homes in this area are rentals, and rentals are going to attract more transient type of people.” Another noted: “The neighborhood became a lot of rental property, and I think that’s probably the biggest issue there. They’re not very selective maybe because many of them will go Section 8 housing. So they’re not taken care of in the way they might have been had they been homeowners.”

Figure 2. Map of Owner Occupancy in Danville Census Blocks

The map in Figure 2 illustrates the prevalence of owner-occupied households in census blocks. The light backdrop illustrates the larger blockgroup boundaries. The darker shades indicate that owner-occupied households dominate renter-occupied households. The percentage shown does not include vacant households, so that it is strictly a ratio of owner-occupied to all
occupied households. Danville city officials consider the renter-owner ratio to be a significant determinant of neighborhood stability, and resident interviews corroborate this hypothesis. Because of this importance, rent and own states were incorporated into the model at the parcel level.

1.1. Owner Migration Patterns

Parcel centroids provided by the city were used to create point feature sets, isolating only the owners who moved in each two-year time period. The from-to pairs of points were then used to create line features as vectors showing migration paths of individual households over two years. The set of overlapping arrows in Figure 3 represents vectors for homeowner moves between 2001 and 2003, derived from the name and address information provided by the city of Danville. The light dots represent the 2001 origin, and the darker dots represent the 2003 destination. While the entire set of arrows in Figure 3 makes directionality difficult to distinguish, selecting parcels in the center reveals a strong tendency to point to the Northwest direction, which is the affluent section of Danville. The desirability of the northwest neighborhood is confirmed by selecting the parcels in the northwest section by the lake — most of the moves are to the area, and not away from it. Selecting parcels in the lower East side (where the interviews were conducted) revealed a strong northwest outmigration tendency, though not reaching as far as the lake.

These ownership changes within Danville reveal internal migration preferences. Figure 4 (A) reveals the aggregate effect of the individual arrows shown in Figure 3. The dark areas, such as the lake area, have the most incoming owners, while the lighter areas represent areas of outflow. However, for the same time interval of 2001 to 2003, if all the ownership changes —
new owners as well as owners who left Danville – are included, the pattern nearly inverts, as in Figure 4 (B).

For this time period (2001-2003), Danville experienced a net outflow of 55 owner households, many of whom left the attractive northwest lake area. Even more confusing is the inflow of owners to the southeast side of town, which includes the less "attractive” but affordable low-income neighborhoods. The behavior exhibited in Figure 4 is explained by an increase overall includes both new arrivals to Danville as well as renters who become owners. It is not uncommon for landlords (once they've made enough money from renting) to sell houses to tenants on contract. In the neighborhood of study in the Southeast part of town, a housing investor is known to buy a house for refurbishing, find a good renter, and help them transition to ownership. Apparently this practice is common in low- to moderate-income areas (that may be implicitly ‘red-lined’ if banks won't lend readily there) and less common in higher-income areas where "normal" housing market dynamics are at play. So the overall increase in ownership in lower-income areas may reflect a significant number of renter to owner transitions.

In Figure 4 (B), the net exodus by the Northwest lake area (despite its appeal to movers within Danville) may be accounted for by the link between affluence and mobility. The North and Northwest areas are affluent in Danville, while the South and East areas of town are modest to very poor. Affluent professionals move in able to afford much more than Danville has to offer in terms of housing investments. Given its depressed housing market, the parts of Danville where housing investments stood the greatest chance of being recouped were the wealthiest ones, by the lake. Professionals move there but frequently have to move elsewhere for work. But in the lower income areas, housing values were perceived by some residents to have plateaued at the bottom of the market. Although no gains were in sight, the already-depressed market allowed homeowners to worry less about losing investments, as long as they stayed put in a sound structure. Hoyt’s (1939) vacancy chain hypothesis helps to reconcile the intra-Danville migration
patterns with the overall changes in homeownership (A and B of Figure 4). As affluent homeowners leave the lake area, vacancies become available to upwardly mobile households within Danville who aspire to that neighborhood. More moderate homes then become available to lower-income households, who then vacate homes that may be leased to renters or sold to new owners.

2. Social Network Dynamics

During the 1960s, Stanley Milgram (1967) conducted an experiment that came to undergird “small world” theory. Milgram gave research participants in Kansas and Nebraska a letter describing a target person in Massachusetts. If the participant knew the target on a personal basis, he/she was asked to send the letter directly to that person. Otherwise, the participants were to give the letter to a personal acquaintance who was more likely to know the person. Of the letters that were returned, the median number of intermediate links was 5.5, rounding up to the cliche “six degrees of separation.” Although Milgram’s research was inherently geographical in nature, measuring the social distance corresponding to geographic distance between individuals, advances in social network analysis have involved “relational” space existing apart from geography. Nonetheless, geographers have studied social networks to understand how individual networks shape and are shaped by life opportunities (Rowe and Wolch 1990, Peake 1995).

Studies of social network structure have demonstrated the significance of homophily, or self-sorting according to similarity. Homophily has been demonstrated by similarity along race, age, religion, education, occupation, and gender lines, with geographic proximity and family ties creating opportunities for such self-sorting connections to form (McPherson et al. 2001). The Danville neighborhood studied was racially diverse. As one resident observed: “It’s mixed. It’s like black, white, black, white.” Amidst this diversity, some racial tension was noted by an older white male: “When you get right down to the nitty gritty, everybody has a prejudice and it may not show, but it can be felt. If we’re going to integrate we’re going to have to integrate right from intermarriage right on up. We’ll always have a separation until the youth, the preschoolers of today are old like I am and have outgrown all that.” Another resident, an African-American female homeowner, argued that such biases “take away from that total connection.”

The importance of children in social networks was emphasized but nuanced in its effects. Often, children serve as vectors for social ties between households. As shared by one homeowner with grown children, “My daughter knows more people around here. Like [back-yard/alley neighbor], she met him because she would go out there and take out the garbage and they would be out there in the garage, because my grandchildren would ride their tricycles around the garage.” Changes in neighborhood family structures were also noted: “Over the years the neighborhood has changed, and we have a younger group coming in now, that’s [got] kids and so forth. As we began to see that [older] age group leave the neighborhood, I kind of took a private stance, because I didn’t know if [the new people] were friendly or not.” Some residents voiced concerns about neighborhood children: “When the kids were little, the were cute little things and they’d come over and talk to you when they were doing things out in the yard; but as they got older, they got a little more malicious.” “Now the little kids – well, I’m sitting out in the summertime and they are running all over and it scares me to death. I don’t say anything but I think they let them run a little too loose and it’s not that safe on [street].”

Social network analysis enables a flexible treatment of scale, to assess the degree of connectivity within an individual (ego) network or clustered at various scales of aggregation. Scale free (Barabasi 2002) and small world (Watts 1999) network theories have drawn recent
attention in an era of abundant virtual networking. At their core, network representations are simply abstractions of human relationships defined by interactions between people. Social networks may be simulated dynamically as (1) structures whose connections change with the nature and frequency of communication, or as (2) a process of contagion (via word of mouth, as well as epidemiology) within relatively stable superstructures. To simulate social influence on household mobility within an agent-based model, a social network was encoded as a multiplicity of ego networks centered on household agents who have dynamically updated lists of contacts. From an aggregate, structural perspective, a local density of social networks was hypothesized to sustain household tenure via a reinforcing feedback mechanism.

The directionality and polarity of causal relationships in Figure 5 is indicated by arrows that create reinforcing (R) and balancing (B) feedback loops. Three major accumulations (stocks) are represented in the boxes: neighborhood residents, neighborhood social networks, and neighborhood affluence. The stock of neighborhood residents changes with rates (flows) of in and out-migration. As the number of residents increases relative to a fixed neighborhood housing capacity (in the simulated case of no new development), a balancing feedback loop (B1) is completed as the number of vacancies decreases.

An effect of increasing the number of residents, in conjunction with the duration of one’s tenure in the neighborhood, is to increase opportunities for local interactions. Such interactions may occur within deliberately formed neighborhood associations, or simply through casual contact. To mitigate the social costs of a dispersed, fragmented, and socio-economically polarized community such as Danville, the city has encouraged residents to form neighborhood associations. The research subjects were volunteers who participated in the neighborhood association, which was about two years old in 2005. “Before the neighborhood association came about, I never thought about the neighborhood. My neighbors, the person the left and to the right of me, the people directly across the street – past that, it's not like we had any really true contact with each other outside of saying ‘hi’ and ‘goodbye.’” Others agreed: “If we would get together more often, we would know each other more. Then we would trust each other more.” Via the neighborhood association, one homeowner remarked “I've become more familiar with who the
neighbors are. People have started to take more pride and connect more with one another.” Another reasoned that “if I show an interest in how you’re living over there and your yard, help you keep it clean, then you’ll do it, too.” This kind of reflexive role-modeling behavior might be labeled IMBY (In My BackYard), the process of local diffusion of neighborhood norms. While income is used in Figure 5 as a proxy for property value via the neighborhood affordability and attractiveness variables, one resident was quick to note that “keeping things picked up and respectable ... has nothing to do with income.”

These interactions among households, integrated over time, form neighborhood social networks. As these networks are strengthened, increased duration of tenure in the neighborhood is likely to result, reinforcing (R1) the network effect further. A “co-flow” representation in Figure 5 enables the flow of incomes associated with migration patterns to contribute to neighborhood affluence, which feeds back to in-migration through affordability (B2) and attractiveness (R2) effects on neighborhood choice. These feedback loops operate across scale as characteristics such as neighborhood affluence contribute to the household’s choice of where to live. As to the decision of whether to stay or leave, the homeowners interviewed were by definition ‘stayers’, one of whom asserted that “deterioration of the neighborhood has never been a force to make me want to leave.” In the simulation model, homeowners have a baseline move frequency of 10 years, while renters move annually on average. As such, renters would be less likely to develop local ties than homeowners.

Determining what constitutes a social connection is no trivial task, arguably easier to simulate than to estimate from observed interactions between people. Likewise, neighborhood definition can vary. The link between networks and neighborhoods was made clear by one resident: “A neighborhood is where I could come to you, you could come to me, we can help each other out instead of just going past, ‘hi,’ and that’s it. I don’t even know your name.” While some residents wanted to know each other, others (including renters who chose not to attend neighborhood association meetings) preferred privacy.

2.2. Family Ties: What’s in a Name?

Homeowner last names were included in the Danville city records obtained for analysis, and were geocoded using parcel data. This information was used to compute distances between all owner-occupied households in the most recent (2005) record. Distances between households sharing the same last name were compared with households having different last names.

Table 1. Comparison of Households with Shared and Different Last Names

<table>
<thead>
<tr>
<th></th>
<th>Shared Last Name</th>
<th>Different Last Name</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Households</td>
<td>5627 (72%)</td>
<td>2117</td>
<td>7745</td>
</tr>
<tr>
<td>Number of Pairs Compared</td>
<td>32,961 (0.11%)</td>
<td>29,955,679 (99.9%)</td>
<td>29,988,640</td>
</tr>
<tr>
<td>Mean Distance</td>
<td>9796 feet</td>
<td>10,237 feet</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 compares households with shared and different last names. The total number of pairs among the 5627 households sharing last names with at least one other household is specified by Equation 1.
Equation 1. Pairs Compared for Sets of K Households Sharing a Last Name

\[
P_k = \frac{K \cdot (K - 1)}{2}
\]

\[
P_{tot} = \sum_{K=1}^{96} N_k \cdot P_k = 32,961
\]

where

\[P_k\] = Matching pairs among \(K\) households sharing a last name
\[N_k\] = Number of unique names shared in a set of \(K\) households

Although the formula in Equation 1 applies to the total set of possible links in Table 1, the constraint of sharing a last name restricts the number of matching pairs among the set of 5627 households to 32,961 pairs among households sharing last names in pairs, triads, quadruples, etc. For the full set of 7745 owner-occupied households, this number of pairs approaches 30 million. Although the majority of households (72%) share their last name with at least one other household, a very small fraction (0.11%) of household pairs were among those sharing a family name.

The comparison in Table 1 reveals over 400 feet closer proximity on average for households with shared last names, as compared to households with different last names. Household distances were computed using a Python script to convert coordinate pairs into Euclidean distance, and were then imported into R for statistical analysis. The Welch two-sample t-test was performed on the shared versus different last name distances, and revealed a significant difference of 440.6 feet closer proximity for households sharing the same last name. The 95% confidence interval for this difference in distance was 382.3 to 499.0 feet.

![Figure 6. Distance Between Households by Number of Households Sharing Last Names](image)

The horizontal axis in Figure 6 reveals the commonness of a name as the number of households sharing the same last name (e.g., “2” for two households with the last name of “Metcalf”). The vertical axis shows how the average distance varies with such commonality of
name. The solid line in Figure 6 shows the average distance between households with different last names, which is over 400 feet farther than the distance for those sharing the same last name. However, this average is significantly diluted by the number of households with very common last names (i.e., the “Smith” and “Johnson” effect). For example, the average for households sharing a last name with only one other household is approximately 8500 feet, which is substantially lower than the overall average.

The dilution of average distance between households by common names stems from the number of pairs among high-frequency names. If a name such as “Johnson” is shared by 90 households, the number of distances to be measured (according to Equation 1) is $90 \times 89 / 2 = 4005$. In contrast, a name shared by only two households has just one measurable distance. In this set, approximately 600 names are shared by only two households, less than 300 names are shared in triads, and just over 100 names are shared by quadruples, etc. The frequency of unique names declines according to a power law with the number of households sharing the same last name (Figure 7). This distribution has a long tail of single occurrences for very popular sets, up to nearly 100 households sharing the same last name.

The analysis of distance between households sharing last names provides a means of testing an implicit social network effect. Two limitations of this analysis are that: 1) not all households sharing last names are related, and 2) not all sets of relatives in Danville are reflected due to their different last names as well as the exclusion of renters from the source data. These limitations render the analysis difficult to embed in a simulation model, although a consideration of both last name and uniqueness could strengthen the social network effect on migration choices that result in spatial patterns of socioeconomic disparity. This option is reserved for future extensions of the model.

![Figure 7. Distribution of Households, Names (L) and Matching Pairs (R) Compared](image-url)
3. Model Formulation

This section describes the construction of an agent-based simulation model that enables testing, or virtual experimentation, of the role friends and family might play in influencing a household’s location choice. In the model, individual agents interact over space and time, and neighborhood networks emerge from these simulated interactions. Heuristics for modeling individual choices draw upon field interviews with residents of a rapidly changing Danville neighborhood. Model parameters for moving behavior, individual attributes, and network structure are estimated from a variety of empirical sources. Alternative model structures were calibrated to observed migration patterns to discern the potential role of social influence on neighborhood choice.

For the 764 selected blocks within Danville, Census data were extracted for tenure (housing units owned, rented and vacant) and race by tenure (white householders who owned or rented). Because the owner-occupied parcels were already established from city data, the renter-occupied parcels were imputed using the ratio of rental to vacant housing units in the block. Racial status was imputed on the basis of tenure at the Census block level. To do so, the probability of a household being white was based upon the fraction of white households within the owner-occupied and renter-occupied households. The Census’ racial category “Black or African American” constitutes the majority of non-white households in Danville. For simplicity of representation, the racial category was reduced to “White” or “Non-white.”

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. Census racial categories were used to infer children and income attributes because cross-tabulations were available at the blockgroup level, and race had already been imputed at the finer-grain block level. By accounting for the cross-distributions of these attributes by race, this approach maximized the use of available information in matching initial conditions to Danville's demographic distribution.

Figure 8. Process for Creating Model Objects

Figure 9 illustrates a screenshot of the model structure in AnyLogic. At the upper left are the state objects Parcel, Block, Blockgroup, Household, and ModelObjects. The Household state object is a container for state information used to create the dynamically active Household agent object. The ModelObjects class is a container for the sets of objects (encoded as java ArrayLists) blockList, householdList, etc. For full model documentation, see Metcalf (2007).
Household agent objects are embedded and replicated (as indicated by the asterisk in Figure 9) in the main view of the model, and make two dynamic decisions: social network choice and neighborhood choice. The social network choice in Equation 2 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 1 plus the same exponential term (Ben-Akiva and Lerman 1985). In turn, the utility of the connection is expressed as a set of alpha ($\alpha$) parameters multiplied by effects.

**Equation 2. Social Network Choice**

$$P_{ij} = \frac{\exp \left( \frac{U_{ij}}{b} \right)}{1 + \exp \left( \frac{U_{ij}}{b} \right)}$$

$$U_{ij} = C + \alpha_D \cdot D_{ij} + \alpha_I \cdot \frac{|I_i - I_j|}{I_{avg}} + \ldots$$

where

- $P_{ij}$ = Probability of household $i$ to connect to $j$
- $U_{ij}$ = Utility of connecting household $j$ to household $i$
- $D_{ij}$ = Distance between household $i$ and household $j$
- $I_i, I_j$ = Income of household $i$ and household $j$ respectively
- $I_{avg}$ = Average income of all households in the Danville model
- $C$ = Constant average utility
- $b$ = Scaling constant
- $\alpha_D, \alpha_I$ = Parametric weights for the distance and income effect respectively
The two effects shown in Equation 2 are the distance effect and the income effect. In both cases, a greater difference decreases the likelihood of connection. The weights for these effects, represented by the alpha terms, would be negative when making a connection and positive when breaking a connection. The $C$ term represents the constant average utility. This term, as well as the alpha parameters, will be adjusted as part of the model calibration. Additional terms may also be included to create alternative model structures. These terms include binary (dummy variable) representation of race (“white” or “non-white”) and children (present or not). If the prospective household has the same race, the value of the term is 1. Similarly, if both households have children, the binary term is encoded as unity. If one or both households does not have children, the effect is zero.

Household agents evaluate their social network at a stochastic frequency designed to create asynchronous decisions – which enables heterogeneous “human” behavior, and is more efficient computationally. At this point in time, a household picks another household at random and tests whether the probability from the binary logit passes a “satisfaction threshold,” which is heterogeneously defined as a proxy for human idiosyncrasy.

The frequency with which households evaluate prospective parcels is informed by their tenure as renters or owners. The agent-based decision dynamics were implemented using asynchronous evaluation at a frequency of times derived from an exponential distribution of owner or renter move rates. The exponential distribution of rate invokes a Poisson process and approximates a first-order delay in an aggregate stock and flow representation of system dynamics.

The first step a household takes is to assess, via a pre-determined “happiness” threshold, whether they are satisfied in their current location. If they are not satisfied, then they search the available (vacant) parcels up to 15 times (a proxy for a scan of the real estate section) to assess location alternatives. If the utility of the alternative parcels exceeds the current location utility, they become part of a consideration set. After the consideration set is complete, the location with the highest neighborhood utility is chosen.

For neighborhood choice, the algorithm is similar to that of social network choice. The utility of a parcel $p$ to household $h$ invokes a constant $C$ term and a set of weighted effects, where the alpha parameters are the adjustable weights.

**Equation 3. Neighborhood Choice**

$$U_{p,h} = C + \alpha_A \cdot A_{p,h} + \alpha_I \frac{I_{b(p)} - I_{b(p)\_avg}}{I_{b(p)}} + \alpha_{SN} \frac{SN_{h,bg(p)}}{SN_h} + ...$$

where

- $U_{p,h}$ = Utility of parcel $p$ to household $h$
- $A_{p,h}$ = Affordability of parcel $p$ to household $h$
- $SN_h$ = Social network of household $h$
- $b(p), bg(p)$ = Census block, blockgroup containing prospective parcel

The first effect on a prospective parcel in Equation 3 is an affordability constraint, such that the effect is negative if the household’s income is less than the average block income of the parcel. If the household’s income is sufficient, affordability is not a constraint.
excluded for evaluating the utility of staying in the current location). The next effect assesses attractiveness as the ratio of the average income of the parcel’s block, normalized to the average income of the entire community. The third term is the social network effect, expressed as the fraction of a household’s social network (i.e., number of connections) that reside in the destination blockgroup, divided by the household’s complete social network. Additional effects may be added to create alternative structures.

When a household moves to a new location, it induces a state change in the old and new parcels. First, the old parcel becomes vacant. The new parcel will be occupied, but whether it will be a rental or owner-occupied unit is not yet clear. To enable transitions between the rent and own state, both parcels and households keep track of their prior rent or own state. If both the parcel and the household were formerly of “rent” status, that status was retained as the most likely state. The same logic was applied to the “own” status. But if the prior parcel state was “rent” and the household was formerly “own”, or vice versa, the new state is uncertain. A simple way to handle this uncertainty is to flip a coin for determining the new state under these conditions. Tracking ownership status and enabling transitions between “own” and “rent” status were central to formulating the model in sync with homeowner migration patterns.

4. Model Testing

Identifying which parameters are worth testing involves model boundary choices informed by the empirical context of the Danville study. For example, friendships among children were revealed by interviews to be a significant factor in forming neighborhood networks. As illustrated in Figure 10 and specified in the equations above, income affects social network choice as well as neighborhood choice, through the affordability and attractiveness terms. In addition to distance and income, several other effects were included in the model. The rent or own status, informed by the Census, is used to vary the frequency of evaluating a move. Census-informed ethnicity (in the form of a “white” or “nonwhite” racial dichotomy), and presence of children were also included as factors in social network choice.

Effects which were revealed to be statistically significant in both linear and non-linear regression of simulated outcomes against observed residential mobility patterns are noted with an asterisk (*) in Figure 10. Feasible model extensions could include shared last name as a proxy for family connections, and neighborhood housing composition by vacancy or rent/own tenure. Housing composition could impact neighborhood attractiveness through the presence or
persistence of vacant units in the neighborhood. The real-life effect of excess vacancy in some Danville neighborhoods is to enable spaces for illegal activities such as drug dealing, meth labs, and vandalism. This, in addition to aesthetic signs of deterioration and neglect, renders vacancy a negative effect on a neighborhood. Other neighborhood composition effects could include ownership fractions and ethnicity.

Grimm and Railsback (2005, pp. 343-345) outline the following calibration methodology for pattern-oriented modeling:

1. Specify parameters using empirical analysis (e.g., location of parcels/owners, income from Census), noting which ones remain uncertain.
2. Identify ranges of values within which to vary uncertain parameters.
3. Create a set of permutations of parameter values across their possible ranges.
4. Define patterns (e.g., observed intra-urban homeowner mobility) to filter out unacceptable parameterizations.
5. Design a scenario for the circumstances under which the filter patterns were observed. Here, 2001 owner data were employed for model initialization, temporally adjacent to the 2000 census data used for attribute specification.
6. Run simulations for all permutations of model parameterizations. Save output variables to match with patterns, as well as major model predictions.
7. Determine the parameter combinations (model structures) that reproduce the filter pattern at a coarse level.
8. Perform sensitivity testing on selected parameters within combinations, assessing influence in simulated output.
9. Assess structural robustness of model through testing of effects at the model boundaries.

The first two steps involve setting practical model boundaries. As with the overall modeling process, the calibration process is often iterative. Figure 11 illustrates the flow of the calibration steps described above. Beginning at the top of Figure 11, parameters are set where possible using data analysis – for income, race, parcel location, etc. Then broad ranges are initially chosen for the remaining uncertain parameters – whether they are negative or positive, for example. These bounds are readily adjusted using the OptQuest optimizer of the AnyLogic software. Alternative model structures are specified as combinations of parameters for the effects just described.

Figure 11. Calibration Strategy
Table 2 describes model parameters that were selected for variation. Defining model alternatives helps to discern which combinations of parameter effects best match the observed migration patterns. Each parameter was varied Monte Carlo style such that random combinations were tested across the range of possible parameter values. The bounds were defined as in the table above and tested in this optimization process. Parameter weights and constants were bounded between zero (0) and unity (1), since each effect was normalized. The scaling constants (in the denominator of the utility equations) were bounded between 0.1 and 2 to reflect a range of reasonable elasticity. The probability of evaluating social connections within the home blockgroup, \( \text{localProb} \), varied between 20% and 70%.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Lower Bound</th>
<th>Upper Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>locNorm</td>
<td>Normalizing Constant, Denominator of Parcel Location Utility</td>
<td>0.1</td>
<td>2</td>
</tr>
<tr>
<td>locConst</td>
<td>Linear Constant for Parcel Location Utility</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>locAfford</td>
<td>Weight on Household’s Affordability Constraint</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>locAttract</td>
<td>Weight on Income-based Attractiveness of Block</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>locSN</td>
<td>Weight on Fraction of Social Network in Blockgroup</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>localProb</td>
<td>Probability of Forming Local Ties within Blockgroup</td>
<td>0.2</td>
<td>0.7</td>
</tr>
<tr>
<td>nwNorm</td>
<td>Normalizing Constant, Denominator of Social Network Utility</td>
<td>0.1</td>
<td>2</td>
</tr>
<tr>
<td>nwConst</td>
<td>Linear Constant for Household’s Social Network Utility</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>nwChild</td>
<td>Weight on Boolean Effect of Children on Social Network Utility</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>nwDist</td>
<td>Weight on Effect of Block Distance on Social Network Utility</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>nwInc</td>
<td>Weight on Effect of Income Difference on Social Network Utility</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>nwRace</td>
<td>Weight on Boolean Effect of Racial Similarity on Network Utility</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

5. Simulation Results

Figure 12 shows a screenshot of the optimization window in AnyLogic, which is powered by OptQuest software to enable a variety of simulation experiments under uncertainty. The left panel reveals the parameter names that have been selected to vary, along with the best value attained so far.
Each experiment was run for the first two years and evaluated relative to the observed migration patterns. The objective function combined directional error (relative to the blockgroup matrix of observed migration patterns) and cumulative error to ensure that approximately the right number of intra-urban owner moves are made. If a new test failed to produce better results than the previous test, the previous “best” would carry over to the right-most plot of Figure 12 (which is illustrative only, reflecting a single simulation outcome rather than a batch average of error across runs).

In the regression analysis, the independent variables were the parameter values for each of 1000 experiments performed. Each experiment consisted of 25 simulation runs at the same parameter values to compensate for random (stochastic) variations between runs. The dependent variable was the average move error, combining both directional and cumulative moves across each batch of simulation outcomes.

Table 3. Regression Summary of Parameter Effects on Average Move Error

<table>
<thead>
<tr>
<th></th>
<th>Non-Linear Model</th>
<th>Linear Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>R-squared</td>
<td>.701</td>
<td>.478</td>
</tr>
<tr>
<td>Standard Error of the Estimate</td>
<td>50.6</td>
<td>66.3</td>
</tr>
<tr>
<td>Neighborhood Constant (locConst)</td>
<td>significant</td>
<td>significant</td>
</tr>
<tr>
<td>Neighborhood Affordability Constraint (locAfford)</td>
<td>significant</td>
<td>NOT significant</td>
</tr>
<tr>
<td>Neighborhood Attractiveness (locAttract)</td>
<td>significant</td>
<td>significant</td>
</tr>
<tr>
<td>Social Network Effect on Neighborhood (locSN)</td>
<td>significant</td>
<td>NOT significant</td>
</tr>
<tr>
<td>Network Constant (nwConst)</td>
<td>significant</td>
<td>NOT significant</td>
</tr>
<tr>
<td>Effect of Income on Network (nwInc)</td>
<td>significant</td>
<td>significant</td>
</tr>
<tr>
<td>Effect of Race on Network (nwRace)</td>
<td>significant</td>
<td>significant</td>
</tr>
<tr>
<td>Effect of Children on Network (nwChild)</td>
<td>NOT significant</td>
<td>almost significant</td>
</tr>
<tr>
<td>Effect of Distance on Network (nwDist)</td>
<td>NOT significant</td>
<td>NOT significant</td>
</tr>
<tr>
<td>Probability of Local Ties (localProb)</td>
<td>significant</td>
<td>significant</td>
</tr>
</tbody>
</table>

Table 3 provides an overview of the regression analyses performed on the simulation data. Two models were created for comparison of parameter significance. The full model includes nonlinear (squared) as well as linear forms of the 12 independent variables. While the non-linear model has a better fit (i.e., a higher R-squared term) than the linear model, it carries the burden twice as many independent variables.

In addition to the quantitative comparison of R-squared and standard error terms, Table 3 provides a qualitative comparison of the two models, highlighting which parameters appeared as significant. The significance of some effects depends upon whether the regression of parameters was in linear or non-linear form. For example, the social network effect on location choice was deemed significant when squared terms were included, but not in the linear model. The distance effect on network formation appeared insignificant in both model forms, but the probability of local ties (a means of incorporating spatial effects within a discrete blockgroup area rather than along a spectrum of block distance) was significant in both. The distance effect in choosing a social network, and the social network effect in choosing a neighborhood, constitute the recursive relationship between neighborhoods and networks as hypothesized.

The effect of children on social network formation was not significant in the full model, but was more significant in the linear model. The significance level of the role that presence of children played in social networks was indicated by a p-value of 0.092 in the linear regression, greater than the 0.05 cutoff generally used for significance. However, the role of social networks on location choice was not significant (p-value 0.205) in the linear form of regression.
Conversely, in the non-linear form of regression, presence of children was not significant (p-value > 0.5 for both linear and squared terms), but the influence of the social network on location choice was highly significant.

A range of outcomes is revealed in Figure 13, illustrating the relationship between total move error and directional error. If directional error were minimized completely, the trivial solution of no moves would be preferred over more meaningful solutions. This is because the observed 139 intra-community owner moves would produce equivalent errors in both the cumulative and directional dimensions. Total move error was minimized when directional error missed by approximately 220 moves. The difficulty of matching the directional dimension was largely due to the predominance (698 of 784) of blank cells\(^1\) in the observed migration matrix.

\[278 \text{ Directional Error} = 139 \text{ Total Moves from and to the Wrong Blockgroups}\]

\[139 \text{ Intra-Danville Owner Moves between 2001 & 2003}\]

**Figure 13. Total and Directional Move Error for 1000 Experiments**

**Figure 14. Spatial and Social Influences on Choice in Filtered Set of Solutions**

\[^1\] The 28 x 28 owner move matrix (solely for internal migration patterns) contained 784 total cells, corresponding to all possible from-to combinations. Only 86 of these were nonzero, containing the 139 total moves. The remaining 698 cells were blank, containing zero (0) moves.
The recursive relationship between spatial and social effects on household choice is revealed in Figure 14, which shows the parameter weights of the 486 solutions filtered as ‘good’ from the set of 1000 experiments. The vertical axis displays the weights for the social network effect on location choice, and the horizontal axis displays the weights for the distance effect on social network choice. Among the filtered solutions, a positive correlation exists between these effects. As distance increasingly constrains social network choice, the network effect on location choice decreases. But when the influence of distance on social ties is lowered, more variation is observed in the social network effect. A finer filtering of solutions revealed thresholds and boundaries in the ratio of the Attractiveness to Network effect on Location Choice, as mapped relative to the ratio of the Distance to Income effect on Network Choice (Figure 15).

![Figure 15. Parameter Weight Ratios Among Most-Filtered Set of Solutions](image)

The parameter weights of analyzed in Figures 14-16 reveal tendencies among relationships in filtered solutions, such as \( \text{locAttract} < \text{locSN} \) and \( \text{nwDist} < \text{nwInc} \). However, a low parameter weight does not imply a less significant contribution to simulated migration patterns.

![Figure 16. Cluster Analysis of Parameter Weights in "Good" Solution Space](image)
A cluster analysis of parameter weights in the set of 486 “good” solutions was performed to delve deeper into the simulated patterns of effects. Figure 16 reveals the results of this analysis as performed in SPSS as a two-step cluster classification using the Bayesian Information Criterion (BIC) with a log-likelihood measure of distance. In addition to the parameter weights shown, both directional and total move errors were included in the cluster analysis. Three clusters of similar size (N=148, 163, 175) resulted from this analysis. The y-axis in Figure 16 shows the average for each parameter weight in the cluster, multiplied by 100 for scaling purposes. Cluster 3 (triangles connected with a bold line) resulted in a slightly lower overall migration error in this analysis, with one less move error on average than the other two clusters. Cluster 3 overlaps with cluster 2 for several parameter weights, such as the affordability (locAfford) and social network (locSN) effects on location choice, as well as the constant term (locConst) for that choice. The probability of selecting social ties within one’s blockgroup (localProb) is above 50% for all three clusters. Although the distance effect on social network choice is variable, this determination of local ties within blockgroup boundaries is consistently substantial.

6. Conclusion

Sensitivity testing of simulated results relative to the observed migration patterns reduced the 1000 simulation experiments to 486 reasonable solutions. Although this halving of parameter space precludes a singular extrapolation into the future, analysis of component weights within the narrowed solution set reveals insights regarding the relative importance of household attributes in neighborhood and network choice. While social network influence on neighborhood choice appeared insignificant in a linear regression model of the simulation results, it became significant in a non-linear form of the regression. Extensions of this work would benefit from including intra-household dynamics of age and life stage, and multi-scalar networks associated with urban governance.

7. References