

IDENTIFYING AND BOUNDING ETHNIC NEIGHBORHOODS¹

*John R. Logan*²

**Department of Sociology
Brown University**

Seth Spielman

**Department of Geography
University of Colorado at Boulder**

Hongwei Xu

**Department of Sociology
Brown University**

Philip N. Klein

**Department of Computer Science
Brown University**

Abstract: This study presents three novel approaches to the question of how best to identify ethnic neighborhoods (or more generally, neighborhoods defined by any aspect of their population composition) and to define their boundaries. The authors use data on the residential locations of all residents of Newark, NJ, in 1880 to avoid having to accept arbitrary administrative units like census tracts as the building blocks of neighborhoods. For theoretical reasons the street segment is chosen as the basic unit of analysis. All three methods use information on the ethnic composition of buildings or street segments and the ethnicity of their neighbors. One approach is a variation of k-functions calculated for each adult resident, which are then subjected to a cluster analysis to detect discrete patterns. The second is an application of an energy minimization algorithm commonly used to enhance digital images. The third is a Bayesian approach previously used to study county-level disability data. Results of all three methods depend on decisions about technical procedures and criteria that are made by the investigator. Resulting maps are roughly similar, but there is no one best solution. We conclude that researchers should continue to seek alternative methods, and that the preferred method depends on how one's conceptualization of neighborhoods matches the empirical approach. [Key words: neighborhood, clustering, ethnicity, k-function, boundaries.]

North American cities have long been spatially segmented by the race, ethnicity, social class, and nativity of their residents. Consequently when social scientists write about a neighborhood, they inevitably note its population composition. Analysts think of middle-class or working-class or gentrifying neighborhoods, immigrant neighborhoods or more specifically Latino or Chinese neighborhoods, and white or black neighborhoods as though

¹This research was supported by research grants from National Science Foundation (0647584) and National Institutes of Health (1R01HD049493-01A2) and by the staff of the research initiative on Spatial Structures in the Social Sciences at Brown University.

²Correspondence concerning this article should be addressed to John R. Logan, Department of Sociology, Brown University, Providence, RI 02912, USA; telephone: 401-863-2267; fax: 401-863-3213; email: John_Logan@brown.edu

the neighborhood were defined by these characteristics. Even categories like *socially mixed neighborhoods* or *neighborhoods in transition* gain meaning from the implied reference to neighborhoods that are not mixed or not changing.

We review three approaches to identifying neighborhoods and their boundaries based on information about their social composition. Each of these methods is rooted in a distinct set of assumptions about the nature of neighborhoods. We exploit an unusual data file that includes the geocoded address of every household in Newark, NJ, in 1880, along with personal characteristics of residents as recorded in the 1880 Census of Population. Such information is available for other populations, such as the contemporary registration data for residents of several North European countries and sometimes census data.³ In this case the data are fully public, with no limitations on how it can be analyzed. We address the question of how methods appropriate for individual-level address data can be applied to the study of ethnic neighborhoods.

Three approaches are described and applied to the Newark case. The first is a variation of k-functions, calculating the density of same-group and other-group population at various distances from each resident. The second is a data mining technique commonly applied to the problem of sharpening photographic images, where decisions are made about whether a pixel of a given color is “really” that color or should be transformed to the color of its neighbors. The third is a Bayesian model that has been used to detect spatial clusters. While these approaches involve widely diverging procedures, they have two major elements in common. They all take into account information about people and their neighbors at varying distances, presuming that an ethnic neighborhood is one where members of a particular group have a disproportionate presence within some local area. In addition, they all result in spatially coherent zones of the city that are identified with a specific ethnic group or classified as “mixed.”

The problem of identifying neighborhoods is related to geographers’ longstanding interest in the definition of areal units for spatial analysis. Neighborhoods are geographic regions, but they are also collections of people, institutions, and infrastructure. Because we begin with data about individual people, our problem is much like the one that Hartshorne (1939, Chapter 11) enunciated decades ago: “We could study the geography of the area only from the study of the geography of the infinite number of points within it. This task, being infinite, is impossible. The problem of regional geography, as distinct from a geography of points, is how to study and present the geography of finite areas.” The task of combining points (or other units) into coherent finite areas is called regionalization. Early proponents of regionalization like Openshaw (1977) saw the approach as a way to solve the Modifiable Areal Unit Problem (MAUP), which is that the results of statistical analyses change as geographic units of analysis change in size and shape (Fotheringham and Wong, 1991). King (1997) argued that the MAUP arises only when the appropriate unit of analysis is unclear on theoretical grounds. The issues are both theoretical and empirical. The problem is not how to select from among a number of pre-formed alternatives (such as census blocks, tracts, precincts, cities, or counties) based on knowing which is theoretically “right.” Instead, we must recognize that each empirical approach incorporates a different conceptualization of the neighborhood.

³Such as the Israeli census data analyzed by Omer and Or (2005) or U.S. data at a confidential Census Research Data Center.

APPROACHES TO IDENTIFYING ETHNIC NEIGHBORHOODS

Recent ethnographic studies of cognitive maps show that social scientists and neighborhood residents draw on many of the same characteristics to identify neighborhood boundaries. Lacy (2007) described this process as “boundary work” and focused on how residents of middle-class black areas seek to signify that their neighborhood is distinct from neighborhoods of black working-class or poor people. In a Baltimore neighborhood, Rich (2009) found that white residents used both the racial composition and class background of specific blocks to mark the limits of their own neighborhood: “When speaking about the boundaries for themselves, many talked about race and, in the terms of one Villager, where the racial ‘Maginot Line’ fell. Homeowners deemed the more middle-class and white areas of the neighborhood ‘the heart of Village Heights’, while they questioned if the areas with a majority of black residents were included in the boundaries” (p. 838). But in another locale Campbell et al. (2009) reported that white residents preferred to draw wider boundaries in order to think of their neighborhood as more racially and occupationally diverse: They drew “large maps that extended beyond the census tract to include a northern area. Whereas the Broadmore block group was a white, middle- to upper-middle class neighborhood, North Broadmore was less affluent and included a higher percentage of African Americans” (p. 479).

Neighborhoods often have names. They sometimes have clear boundaries, and Suttles (1972, p. 4, cited in Stoneall, 1981) argued that residents tend to construct simplified images of the city in which differences between neighborhoods are magnified. Simplified images, he stated, “serve us well by reducing the complexity of the urban landscape to a range of discrete and contrastively defined ecological units despite the general continuity, gray areas, and constant changes in any section of the city. A cognitive map of our urban environs is useful for precisely the reason that it simplifies to the point of exaggerating the sharpness of boundaries, population, composition, and neighborhood identity.” Nonetheless, researchers have noticed that it is common for borders to be fluid. Hunter (1974), for example, reported that areas of Chicago that he studied had “rolling” boundaries—people might agree on the name of their neighborhood, but those living near its edge tended to perceive the neighborhood as extending farther in that direction. Ambiguous boundaries probably work fine for residents, but they complicate neighborhood research. To measure neighborhood characteristics and their changes over time analysts need ways of identifying where neighborhoods are.

If neighborhoods exist in the minds of residents, then the direct approach is to find out how residents understand their neighborhoods. In fact, as Logan et al. (2002, pp. 303–304) pointed out, “[E]thnic neighborhoods are most often identified and studied through fieldwork, where the researcher typically begins with the knowledge that the ethnic character of a given locale is socially recognized—certainly by group members and perhaps also by others. This ethnic character may be visible through observation of people in public places, the names of shops or the languages found on signs or spoken by clerks or patrons, or by community institutions such as churches, social clubs, and associations.” Yet many researchers are unable to do original field work, and many study designs require information on more locations than could feasibly be covered by field work. Administrative boundaries are convenient: If local officials have decided to divide their city into neighborhood planning areas and to provide data at that geographic level, those areas become

politically meaningful to some degree and practical for researchers. More routinely data only for census tracts or similar units are available, and the lack of alternatives makes it legitimate to treat those tracts as if they were “real” neighborhoods.

With census data one can easily identify tracts that are candidates for inclusion in an ethnic neighborhood. But how Chinese or Filipino must a Chinese or Filipino neighborhood be? There is no established criterion. In their study of Italians, Irish, and Germans in the New York region, Alba et al. (1997, p. 892) operationalized an ethnic neighborhood as “a set of contiguous tracts, which must contain at least one tract where a group is represented as 40% or more of the residents and whose other tracts each have a level of ethnic concentration among residents of at least 35%.”

It is widely understood that the group does not necessarily have to be a majority in its identified ethnic neighborhood (a corollary is that some zones may contain “ethnic neighborhoods” of more than one group). Philpott (1978) has pointed out that the principal Swedish ghetto identified by Park and Burgess in Chicago in 1930 was only 24% Swedish; the German ghetto was only 32% German. Some places today have international reputations as ethnic neighborhoods despite having modest proportions of group members. For example, parts of Los Angeles “are so heavily identified with Armenians that when prospective emigrants in Armenia or Iran are asked about their destination, they may answer ‘Hollywood’ or ‘Glendale,’ respectively, instead of America” (Bozorgmehr et al., 1996, p. 368). Yet in 1990, Armenians made up only about 25% of residents of Hollywood and Glendale, reaching a maximum of 33% in their most “Armenian” tract, and only 10–15% in their peripheries. Among well-known contemporary Chinese neighborhoods, the core immigrant area of Flushing (in Queens, New York) studied by Zhou (1992) was only 14% Chinese in 1990. Monterey Park, California, was less than 25% Chinese in the mid-1980s when Horton (1995) began to study it. A recent study of minority groups in Los Angeles defined Asian residential enclaves as areas that were as little as 10% Asian (Bobo et al., 2000).

Should neighborhoods be built up from tracts or smaller units within tracts? Hipp (2007) showed that when neighborhoods are defined in terms of their *effects* on residents’ experience of crime, racial/ethnic composition is more significant at the scale of the census tract than the block. But to discover the effects of average income on experience of crime, research must be done at the more localized scale of the block. In some cases, the tract is too small a unit because the relevant social phenomena are organized at a larger scale. For this reason, recent developments in spatial statistics have encouraged the use of tract data in conjunction with information about surrounding tracts. As Sampson et al. (1999) have observed, small areas are “embedded” in a larger urban context. That is, real neighborhoods—areas whose characteristics can be consequential for residents—may be larger than a single tract. Logan et al. (2002) found that thematic maps of New York and Los Angeles in the 1990s showed visible concentrations of several major immigrant minority groups that typically extended across many tracts. They used a measure of local spatial clustering at the tract level (local Moran’s *I*) to identify statistically significant clusters and treated these larger areas as the groups’ ethnic neighborhoods (see also Logan and Zhang, 2004).

The advantage of building neighborhoods analytically from point data is that one need not know *a priori* what the appropriate spatial scale of the neighborhood should be. The building blocks of neighborhoods in our research are households (point locations) along

a street segment. In this respect we agree with Grannis (2009, pp. 8–9) who argued that neighbor relations develop when people’s “lifestyles cause them to casually and unintentionally encounter each other and thus to have the opportunity to learn about each other through observation and to acknowledge each other’s presence or choose not to.” The street network guides the paths of people’s daily lives, and it is reasonable to expect neighborhoods to build up from street segments. Neighborhoods extend beyond face blocks through a sort of ripple effect: people on one block may become acquainted with those on the next block, who are in turn tied to those on the next, and so forth. This process is not a necessary condition for any of the models that we present below, but it is our theoretical starting point.

A review of the literature suggests three main questions about ethnic neighborhoods. First, are their boundaries “rolling” (following Hunter, 1974) or discrete (as suggested by Suttles, 1972)? This may not be a question of residents’ experiences on the ground but rather a question of representation: Should ethnic neighborhoods be modeled as discrete territories or as areas with “fuzzy” boundaries? Second, how “ethnic” does a neighborhood have to be to be given an ethnic label? The literature suggests several approaches to this question and indicates that high-intensity ethnic settlement can warrant designation as an ethnic neighborhood. Third, the literature raises questions about the scale of ethnic neighborhoods: What are the atoms that constitute a neighborhood—streets, buildings, people? How geographically extensive do neighborhoods tend to be? There is no correct answer to these questions. In the following sections we present three methods for defining ethnic neighborhoods that allow considerable flexibility in operationalization.

DATA SOURCES

This study makes use of data assembled by the Urban Transition Historical GIS project (www.s4.brown.edu/utp). A critical component is the 100% digital transcription of records from the 1880 Census that was organized by the Church of Latter Day Saints and prepared for scholarly use by the Minnesota Population Center (MPC). This file includes approximately 50 million Americans, organized by household and with information on each resident’s name, age, race, gender, occupation, relation to head of household, state or country of birth, each parent’s state or country of birth, the enumeration district, ward, city, county, and state of residence. It is available from the North Atlantic Population Project (NAPP: <http://www.nappdata.org>).

We rely on the person’s and parents’ race and place of birth to create four exhaustive and mutually exclusive categories of race and ethnicity. “Yankees” are white, born in the United States, and have parents who are also native-born white. “Irish” and “Germans” are also white. For the foreign born, country of birth determines ethnicity. For those who were born in the United States but have at least one parent who was born abroad, ethnicity is primarily determined by mother’s country of birth. If only the father was foreign born (or if the mother was foreign born but her birthplace was not reported), the father’s country of birth is applied. People of other than the above three race/ethnic groups are categorized “all others.”

The 1880 census did not gather information on income or education, the most conventional indicators of socioeconomic standing. The available measure is the socioeconomic

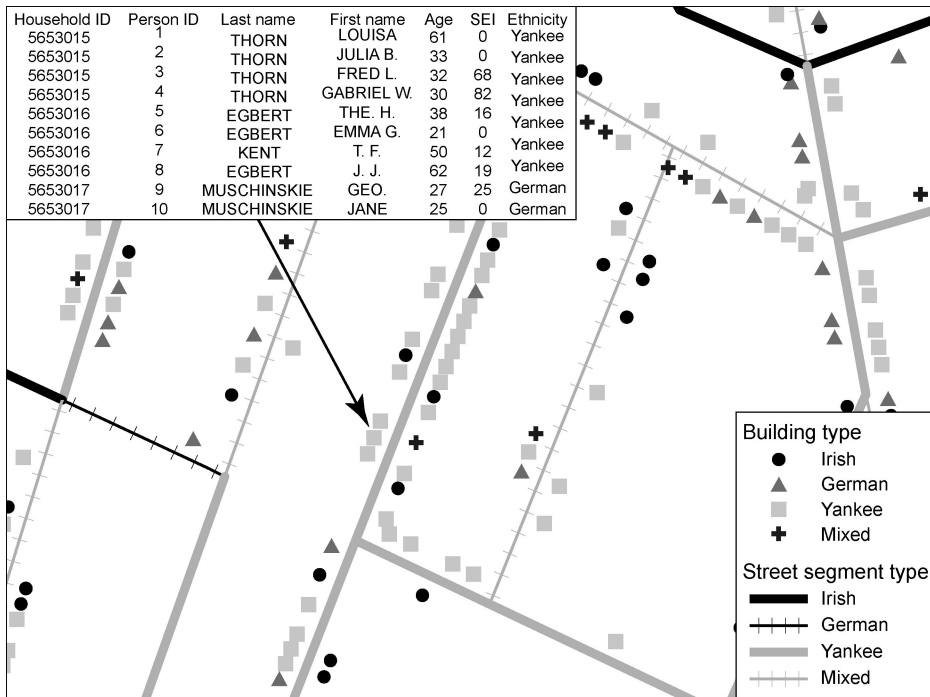


Fig. 1. Geocoded information for a section of Newark, NJ in 1880.

standing, measured on a scale of 0 to 100, of the person's occupation; this socioeconomic index (SEI), which was provided by MPC, is based on the average education and earnings of persons in 1950 in each occupation. An important question in using the SEI in a historical study is whether the relative standing of occupations is stable over time. Sobek (1996) provided some evidence that it is relatively stable by documenting a close relationship ($r = 0.93$) between the average income of men in each of 140 occupations in 1890 and the income of men in those occupations in 1950.

Addresses for residents of Newark (and 38 other cities not used here) were transcribed and merged with the information about residents in the NAPP files. We juxtaposed a map showing 1880 Essex County boundaries with U.S. Census Bureau TIGER street files (<http://www.esri.com/data/download/census2000-tigerline/index.html>). The street file was edited by hand to reconstruct the street layout of Newark in 1880. This task involved deleting streets or highways constructed after 1880, adding those that were demolished, changing names of others, and in some cases correcting the alignment of streets. The Newark city directory for 1880 included a listing of address ranges by block for most streets, and this information was used to geocode addresses. Additional checks were made to ascertain the boundaries of enumeration districts and ensure that individuals were placed in their recorded enumeration district.

The map in Figure 1 illustrates the resulting data. For any single building, we know the location and the characteristics of every resident, including what household they lived

TABLE 1. ETHNIC COMPOSITION OF POULATION AND STREET SEGMENTS, NEWARK 1880

	Population		Population age 18+		Classified street segments	
	No.	Percent	No.	Percent	No.	Percent
Yankee	37,180	27.8	22,608	28.2	659	37.0
German	40,042	30.0	22,561	28.2	522	29.3
Irish	30,158	22.6	18,740	23.4	395	22.2
Other/mixed	26,174	19.6	16,207	20.2	204	11.5
Total	133,554	100.0	80,116	100.0	1,780	100.0

in (clearly at least three households lived at the selected address in the figure, 41 Walnut Street). Key characteristics used here are age, ethnicity, and SEI. In this figure we have labeled buildings and street segments according to the dominant ethnicity of residents. At the selected address 8 of 10 adult residents are Yankee. Four residents, all women, have no occupation (SEI = 0). Some others (Fred and Gabriel Thorn) have very high status occupations, like doctors or judges.

In Table 1 we summarize information on the 1880 population of Newark. The total city population was 133,554, of whom 80,116 were adults (age 18 and above). Residents were fairly evenly divided among the three main ethnicities: Germans were the largest group, followed closely by Yankees (native-born whites with native parents), and Irish. About 20% of the population was classified into other races and ethnicities, of which the largest (about 10%) was British. For the sake of simplicity, we disregard this other population in the remainder of this paper. The analysis is based on the adult population (age 18 and above), whose ethnic composition is very similar to that of the total population.

In parts of the analysis, we examine the ethnic composition of the population by street segment in Newark. Newark had 1,780 street segments with at least one Yankee, German, or Irish adult resident. These segments have been classified in two ways. In the energy minimization model, each street segment begins with a label, which denotes the dominant resident group. Street segments whose population has 50% or more of one ethnic group have been labeled as Yankee, German or Irish; all other street segments have been labeled as mixed. For the Bayesian model there are three independent categorizations of street segments, and these reflect the clustering of members of each ethnic group. The approach begins with arbitrary cutoff points to classify neighborhoods as high, medium, or low for each group, and the final cutting points are derived in the analysis.

The disaggregated nature of the Urban Transition HGIS data obviates the need to think about neighborhoods as territorial units defined by geographic boundaries. Instead we can define neighborhoods based upon the ethnic composition of an area, potentially on a building-by-building basis. The neighborhood-definition problem becomes how to combine buildings or streets into some larger unit while maintaining flexibility in terms of scale, the nature of neighborhood boundaries, and the number and composition of neighborhoods.

STUDYING NEIGHBORHOODS THROUGH K-FUNCTIONS

A single building or street does not by most definitions constitute a neighborhood. Our first approach to the creation of neighborhoods is rooted in “local k-functions” (Getis and Franklin, 1987). A k-function measures the expected number of “events” around an “arbitrary event” as a function of distance. The term *event* refers to occurrences, such as diseases, accidents, or crimes, that have a probability of occurring at a geographic location. The k-function is estimated by counting the number of events around each event and dividing by an intensity parameter λ , which measures the density of events in the study area (Bailey and Gatrell, 1995). K-functions are typically used to answer questions about the geographic distribution of events, such as whether they are clustered or occur randomly in space.

In this case events are buildings, and their location is determined by exogenous factors like the layout of streets. We use the k-function in a somewhat atypical way: instead of looking for places where there are more or fewer events than expected, we use the k-function to describe the character of the neighborhood around each building. We compute thousands of *local* k-functions to provide a statistical summary of the prevalence of the Irish, German, and Yankee ethnic groups around every building in the study area (in contrast to a global k-function which would summarize an entire study area). More concretely we draw a series of concentric rings around each of the 15,000+ residential buildings in Newark and then count the number of people by ethnicity within 50, 75, 100, 200, 300, 400, and 500 meter radial buffers. The result is a series of three graphs (one for each ethnicity) for each building. To create realistic graphs we had to consider the scale of the environment experienced by a resident of Newark in 1880. Hershberg et al. (1981, p. 136) estimated the average commute in Philadelphia in 1880 to be between 0.2 km (for blacksmiths) to 1 km (for lawyers), and we adopted the 1 km diameter for our largest distance band. Hence the buffer around each building includes the area that a typical working adult might have experienced in a given day. Within this catchment area we measure ethnic prevalence at regular intervals, oversampling at the sub-block (< 100 m) scale.

Our k-functions include a constant intensity parameter (λ). This means that in densely settled parts of the city the observed values of the k-function will tend to be higher simply because the total population density is higher. In addition, the observed values for points near the edge of the study area will be lower because their concentric rings include areas that lie outside the study area. We control for both of these distortions by calculating two different k-functions: an ethnic-specific one and another for the total population, as shown in Equations (1) and (2) below.

$$K_{ij}(h) = \frac{1}{N_i} \lambda^{-1} (\# \text{ of people of ethnicity } i) \quad (1)$$

$$K_j(h) = \frac{1}{N} \lambda^{-1} (\# \text{ of people}), \quad (2)$$

where K_{ij} is the k-function for ethnicity i at location j , K_j is the k-function for the total population at location j , N is the total population, and N_i is the total population of ethnicity i . The analysis focuses on the difference between these k-functions. Conceptually, this approach is similar to the $L(h)$, an adaptation of the k-function whereby the $E\{k(h)\}$ under

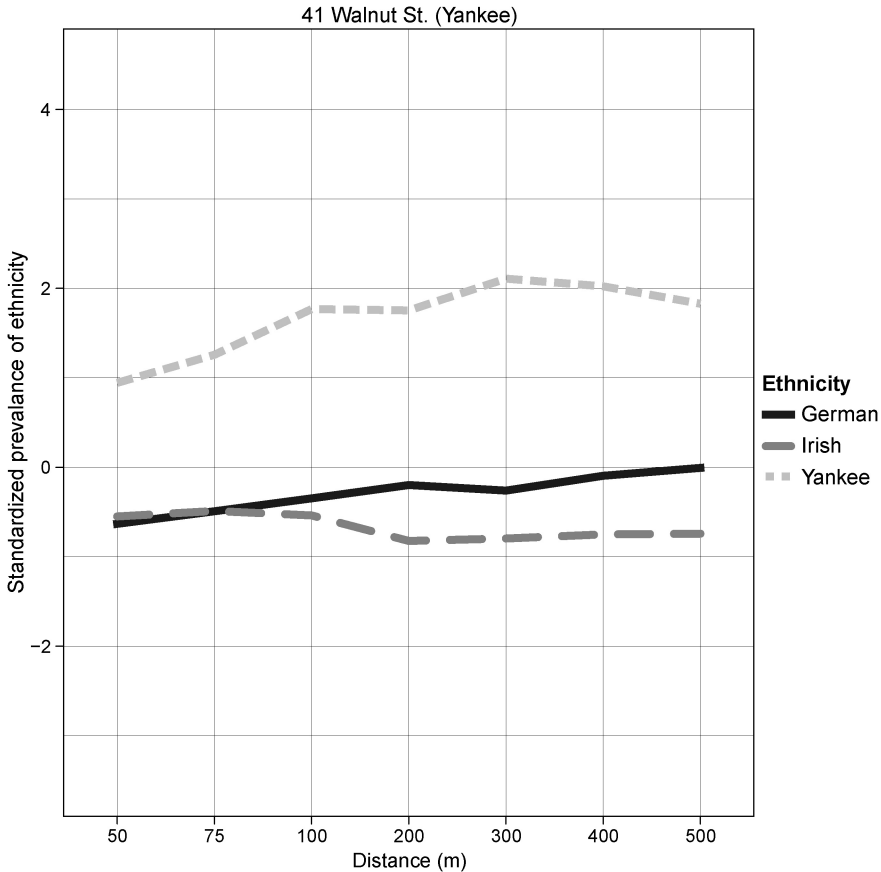


Fig. 2. K-functions for a Yankee building in Newark.

complete spatial randomness is subtracted from the observed value of $k(h)$ (Bailey and Gatrell, 1995). In addition, we normalize the densities so that each distance band has a zero mean and unit variance.

For example, Figure 2 shows the k-functions for 41 Walnut Street (the selected location in Fig. 1). As previously noted, this building itself has a high prevalence of Yankees at all distance bands, and Figure 1 seemed to show many Yankee buildings around it. Figure 2 quantifies this observation with three k-functions, one for density of Yankees, another for density of Irish, and a third for density of Germans. It shows that the line representing the Yankee k-function has values above 0, and the density rises rapidly to 100 meters. In contrast the lines are below 0 for Irish and German ethnicity. With only this information, one might guess that this building is in a “Yankee neighborhood.”

Figure 2 displays only 3 of the roughly 45,000 local k-functions we computed to describe Newark. Graphs for three other buildings are shown in Figures 3, 4, and 5. Figure 3 shows a house at 39 Belmont St. that has a high prevalence of Germans at all distance bands, especially at 300 meters. Figure 4 (144 Newark Street) shows a case where the highest prevalence is Irish, especially at above 75 meters’ distance. And Figure 5 (83 Chambers Street)

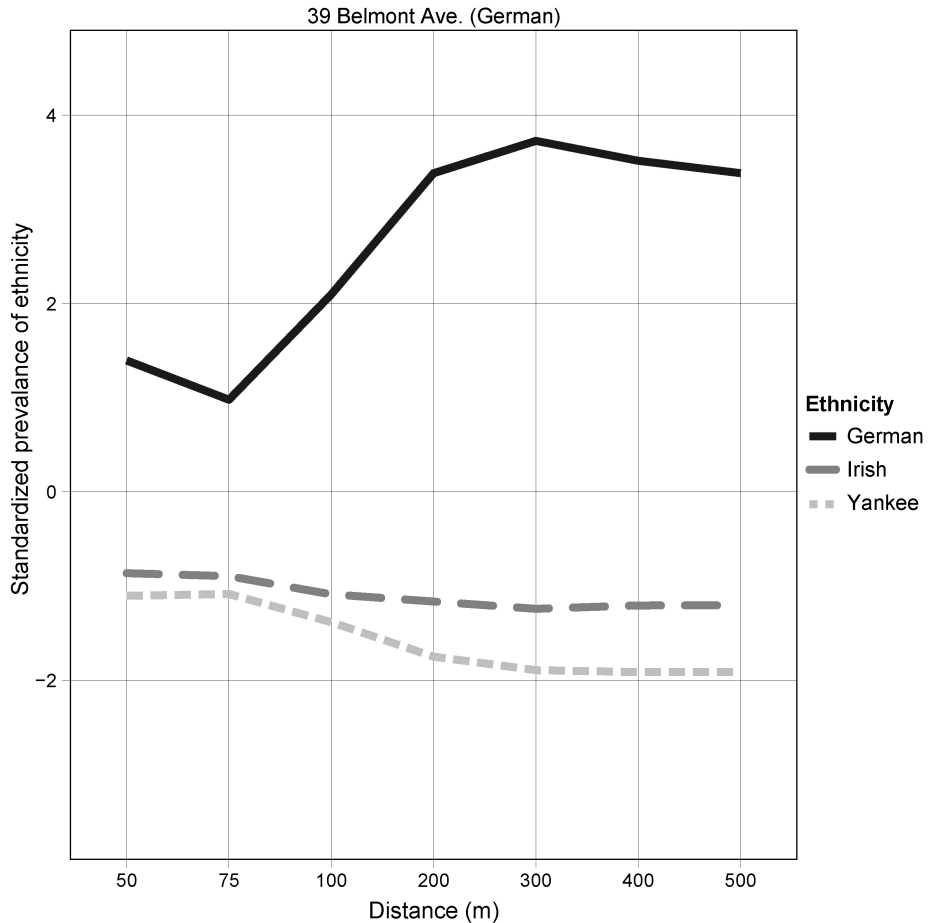


Fig. 3. K-functions for a Yankee building in Newark.

show a more mixed case where the highest density of neighbors is German up to about 300 meters, but beyond that distance the densities of Germans and Yankees are equal.

By design, local k-functions have a high degree of spatial autocorrelation; that is, those at locations near each other tend to be similar. This property is desirable for our purpose because it means we can expect to find spatial clusters of buildings with similar k-functions and use these clusters to identify types of neighborhoods. If a city has ethnic districts, we would expect the k-functions to reveal regions where the curve for one ethnicity is dominant. If, however, the city is spatially undifferentiated, we would not expect clearly defined regions where one ethnic curve is measurably higher than others: cases like Figure 5 would be more common than cases like Figures 2–4. Burstein (1981) notes that 1850–1880 was a time when American cities were in transition from a 19th century form that was not residentially differentiated by ethnicity or class toward a twentieth century pattern with clearly defined class and ethnic zones. Were ethnic neighborhoods clearly defined in Newark, NJ in 1880?

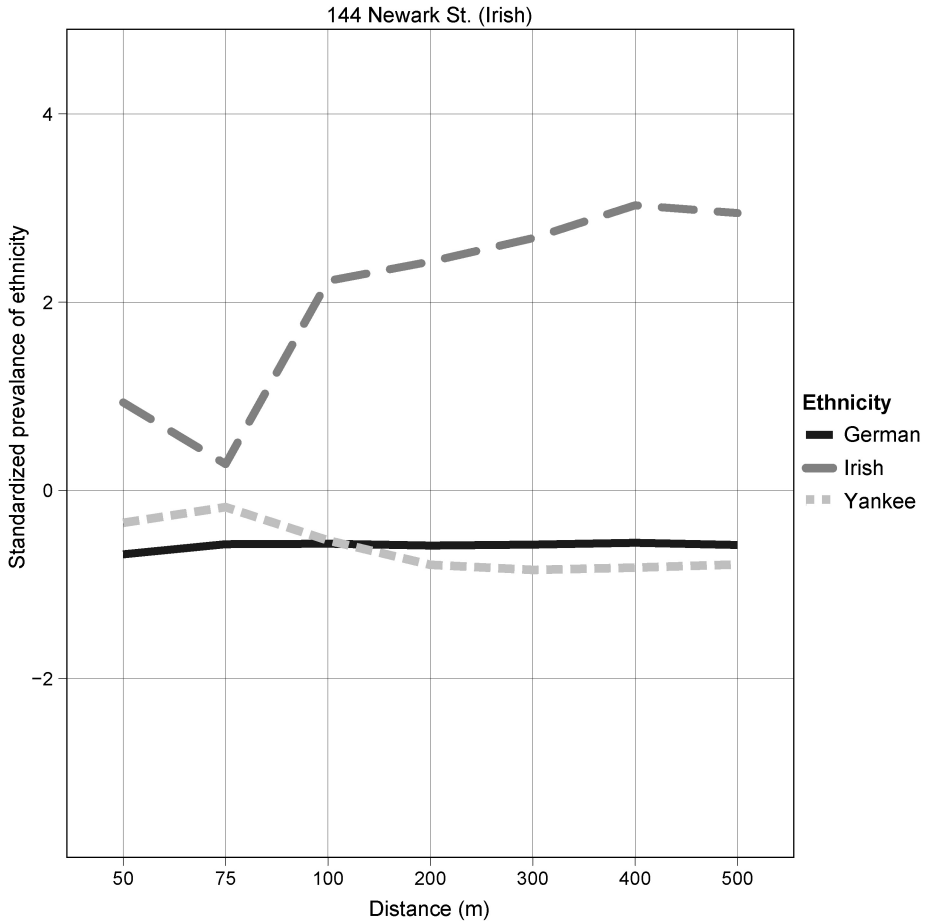


Fig. 4. K-functions for a German building in Newark.

Think of ethnic neighborhoods as zones of the city where a specific pattern of k-functions tends to cluster. We make no *a priori* assumptions about what are the types of patterns, but we do set the number of patterns that should be distinguished. In this example the number is four, and we are looking for Yankee, German, Irish, and mixed neighborhoods without defining them in advance. A cluster is defined simply as buildings with similar k-functions. The solution is provided by a simple k-means cluster analysis on the 45,000 local k-functions.

K-means is a way to partition a data set into a user-specified number of groups, where the groups are created through an algorithm that assigns an observation to a group based on how similar it is to the mean of each group (Romesburg, 2004). The procedure is iterative; after the initial assignment of observations to classes, the mean of each group is recalculated and observations are re-assigned. The procedure terminates once the classes become stable (i.e., there is little or no reassignment of observations between iterations).

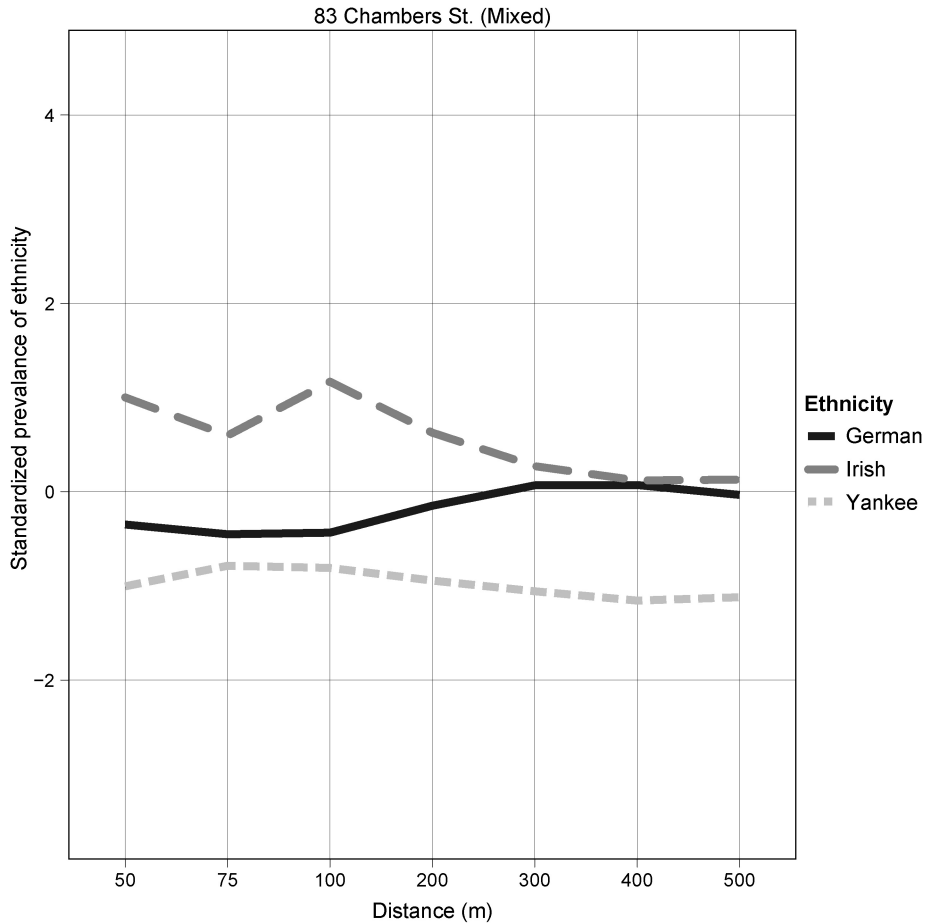


Fig. 5. K-functions for an Irish building in Newark.

The input to the k-means cluster analysis was the observed value of the local k-functions for each of three ethnicities at each of seven distance bands for all 15,000 buildings in Newark. The groups that emerged from the k-means cluster represent observations that have similar characteristics. Mapping the four-cluster result (Fig. 6) shows four clearly defined regions; each of these regions corresponds to groups of buildings with similar k-functions.

This map paints a picture of the ethnic spatial structure of the city. A section of the city running north-south and not far from the river on the east end is identified as Yankee. This identification is based on inspecting the typical k-function curves of buildings in this zone such as 41 Walnut Street. For example, we could easily plot the average densities at every distance band for all buildings in this zone. There is one large concentrated area in the western part of the city identified as German (this is where 39 Belmont Avenue is found). Four or five areas are identified as Irish, including the area where 144 Newark Street is located. Finally, other areas are designated as mixed, including the vicinity of 83 Chambers Street.



Fig. 6. Ethnic neighborhoods identified through local k-functions.

This use of local k-functions offers one answer to the question of how Newark neighborhoods were structured by ethnic composition in 1880. It identifies areas that can be treated as ethnic neighborhoods and provides a guide to their boundaries. The extensive territories that are categorized as mixed suggest that there are intermediate areas, or perhaps zones of transition, around the borders of these neighborhoods, which is consistent with urban theory. One could search for more complicated patterns; for example, would allowing a larger number of neighborhood types produce clusters that reflect Irish-Yankee or Irish-German combinations? It is also possible to extend the analysis to include not only ethnicity but also other dimensions such as the distinction between first and second generation, or gradations of occupational SEI.

The k-functions approach is superior to a simple mapping of the ethnicity of buildings, street segments, or larger areas on the street grid because it takes into account the full array

of other street segments at varying distances. Because spatial dependence is built into the k -functions for street segments that are near to each other, the method smoothes out local variations and yield the simpler picture that social scientists have in mind when they speak of ethnic neighborhoods.

ENERGY MINIMIZATION

A second approach uses the same data to produce a somewhat different map of ethnic neighborhoods. Energy minimization refers to an optimization technique that is most commonly used to process digital images such as medical images (Boykov et al., 2001). Its assumption is that a pixel that is unlike adjacent pixels may represent the edge of an object (like a tumor) or could just be noise (measurement error). Information from surrounding pixels is used to refine images and pixels are reclassified, i.e., given a new label that describes the ethnicity of the neighborhood.

As an indicator of whether a street segment falls within a largely Yankee, Irish, German, or mixed neighborhood, data on the ethnic composition of the segment is incomplete. An all-Irish street segment has some probability of signifying an Irish neighborhood, but the characteristics of adjacent street segments also count. The question is how to balance these two criteria.

In formal terms, we determine if a street segment is part of a larger neighborhood by minimizing the following quantity:

$$E(f) = E_{smooth}(f) + E_{data}(f) \quad (3)$$

Here $E(f)$ is the net energy cost of the neighborhood label that is assigned to a street segment. This cost has two components. The first, $E_{smooth}(f)$, controls the smoothness of the resulting image (or map). A smooth map consists of large homogenous regions whereas a rough map has a patchwork of small neighborhoods. The smoothness term applies a penalty for adjacent street segments that have different ethnicity labels, but no energy (cost) is added if they have the same label. For a given street, the smoothness term considers all connecting street segments, and the costs associated with every one are part of the net energy. We are assuming that cities consist of ethnically homogeneous neighborhoods and so we assign a “cost” to calling a street “Irish” if it is surrounded by streets where German is the dominant ethnicity.

The second component, $E_{data}(f)$, is referred to as the data cost. It assesses a penalty for assigning to a street a label that is different from the observed value. That is, there is a cost of labeling a predominantly Irish street segment as being in a German neighborhood. The notion is that the observed values have significance. The sum of the data and smoothing costs for each street is the quantity that we wish to minimize, and that controls the final map that the method produces. Like k -functions, it is up to the researcher to decide which street segments are near enough to consider as relevant neighbors; in this example we count only street segments that are connected at an intersection. The researcher also controls the parameters that are used to calculate each cost. The final map depends on how one chooses to weigh the smoothing costs vs. the data costs.

The approach is illustrated in Figure 7. Here Street A is observed from census data to be predominantly Irish (i.e., more than 50% of adult residents are Irish), and three of

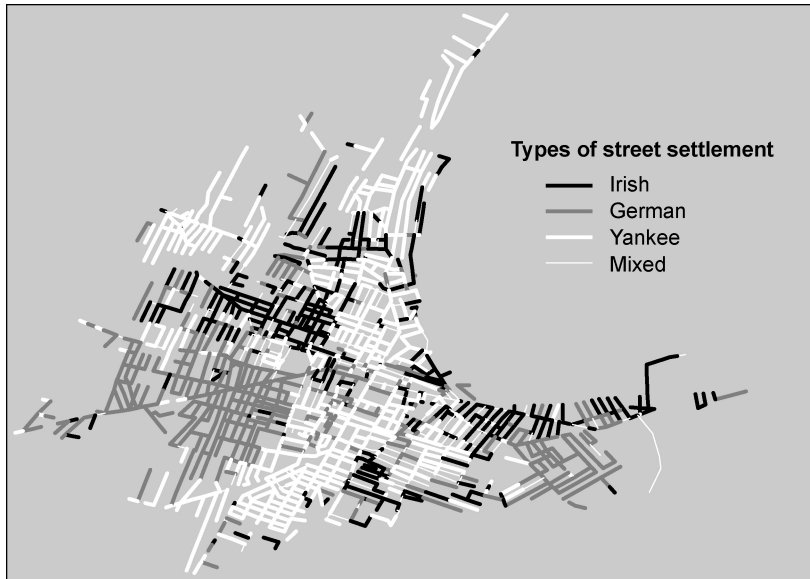


Fig. 8. Map of input data for energy minimization.

the four connecting streets are German. We ask, “Is Street A part of an Irish or a German neighborhood?” The figure shows two scenarios in which the relative weights of smoothing cost and data cost are different. Under Scenario 1 the cost of changing the label for Street A (from the observed Irish to a potential German) is set at 20 and the cost of being connected to each street that is a different ethnicity is 4. The total energy cost is easily calculated to be 12 if Street A is labeled Irish, but 24 if it is labeled German. Using these parameters, Street A remains Irish. In Scenario 2 there is a higher cost for having different types of streets connected to each other (12) but the same data cost (20). In this scenario the total energy cost would be minimized by labeling the street German.

We use Boykov et al.’s (2001) α -expansion algorithm to assign ethnic labels to streets that minimize the “energy” (Eq. 3) of the resulting map. The input data for this analysis is shown in Figure 8, and the result is mapped in Figure 9. The result is a product of some assumptions. Our initial classifications were based on the observed proportion of each ethnic group on the 1,780 populated street segments of Newark, classifying each street segment as Yankee, German, Irish (50% or more of adults were of that ethnicity), or mixed. We experimented with many combinations of smoothing and data costs. The outcome shown here represents an appropriate balance between a rough and a smooth map.

Note that compared to the input map, the result of energy minimization is a simplified depiction of the ethnicity of different sections of Newark. Many street segments whose actual composition was unlike adjacent segments have been relabeled, and a large zone has been labeled as mixed, even though only a few street segments began in the mixed category given how we defined that category. Very few street segments appear as isolated enclaves, and these are found only on the edges of the city.

Boycov et al.’s α -expansion algorithm produced many maps that also minimized Equation 3 (using different sets of parameters) but did not correspond to our understanding of



Fig. 9. Map of results from energy minimization.

the city based on detailed inspection of the input data. The connectivity of the street network in a given city, the number and definition of categories used for the input data, and the relative difference between the data and smoothing costs all affect the resulting map. Whereas in this sense Figure 9 is arbitrary, it is not any more arbitrary than other methods, and the iterative process of varying and evaluating the result offers a systematic way of probing the spatial pattern of Newark. The figure shows clearly defined Irish, Yankee, German, and mixed neighborhoods rather similar to the local k -function result, which is also reassuring.

Each approach (local k -functions and energy minimization) gives different results depending on how it is applied. It would be desirable to find historical, empirical, or theoretical justification for a given set of parameters and assumptions. In image processing it is often possible to objectively compare the result of an algorithm to “ground truth” because one knows more or less what the picture should look like. A better grounded approach to identifying ethnic neighborhoods with these methods would be possible if additional information were available to validate the results, such as archival sources giving the location of ethnic institutions like churches or stores that social scientists understand to be additional markers of ethnic turf. We return to this possibility in the conclusion.

BAYESIAN MODEL OF NEIGHBORHOOD CLUSTERS

The third approach is based in Bayesian analysis, which has been widely applied in spatial statistics (Congdon, 2003). Bayesian analysis is about making statistical inference from combining researchers’ prior beliefs and sample data (Gelman et al., 2004). The crucial feature is that any unknown parameter in a Bayesian model is considered to be a

random variable. Any information about the parameters before collecting the data such as previous findings and expert knowledge is specified as a “prior probability distribution.” The process of inference consists of using the sample data to adjust prior beliefs and deriving the “posterior distribution” of the parameters. We draw here on an application of Bayesian methods by Xu and Bao (2004), who used this approach to detect spatial clusters and identify disability patterns at the county level in Mississippi and Alabama.

The basic unit of analysis here is the street segment. As in the energy minimization example, we define the neighboring street segments as those that are directly connected to the focal street segment, and the ethnic classification of a street segment in the observed data is based on the ethnicity of the adult residents.

The input data are of the following form: among n_i people living along street segment i , there are y_i people of a given ethnic group. We posit that the probability for observing this composition is governed by the underlying spatial pattern of ethnic neighborhoods. That is, we posit a “real” ethnicity of neighborhoods and treat the observed values on a given street segment as its realization with some random variation. Specifically the approach here is to think of neighborhoods’ ethnicity in three dimensions, degree of Yankee-ness, German-ness, and Irish-ness. On each dimension neighborhoods are assumed to be high, medium or low (i.e., ethnic, mixed, or non-ethnic). This means, for example, that in theory a neighborhood could be high ethnic Yankee and high ethnic German at the same time. For each group let p_0 , p_1 , and p_2 denote the probabilities of observing y_i group members living among a total number of n_i persons in each of these three types of neighborhoods.

These probabilities must be estimated separately for each group. Under a Bayesian framework, any unknown quantity is assumed to be random.⁴ Our task is to estimate the distribution of this random quantity and draw summary inferences from this distribution for each street segment. The procedure is as follows:

First, following Xu and Bao (2004), we consider the following binomial distribution for the number of people from an ethnic group given the underlying k types of neighborhoods:

$$P(Y_i = y_i | X_i = k) = \binom{n_i}{y_i} p_k^{y_i} (1 - p_k)^{n_i - y_i} \quad (4)$$

X denotes the random set of ethnic neighborhood configurations. Given the underlying process for neighborhoods $X = x$ in one realization, we assume that X takes on values 0, 1, and 2 for the three types of neighborhood (k). Equation (4) shows that we estimate a binomial model for the probability of observing a given number of ethnic individuals in each street segment. If we ignored the spatial structure of the data, we would simply estimate a conventional binomial model as for other dichotomous outcomes in social science research. This approach would use aggregate data at the street segment level, similar to the way one models binary outcomes from a contingency table (Agresti, 2002).

⁴In our application, we make another assumption, that for a given group the probabilities are constant throughout the city.

Second, we incorporate the spatial structure by imposing a neighboring effect in the model. A Potts model is specified as the prior distribution for the underlying neighborhoods X . The Potts model has the property of a Markov random field (MRF), which simply means that proximate streets are more likely to be assigned the same ethnicity than are distant streets (François et al., 2006). The MRF property makes the Potts model a good choice for modeling spatial data (Cressie, 1991; Xu and Bao, 2004; François et al., 2006). If x is one possible realization of the random set X with each component being one of the k three types of neighborhood, then the prior distribution is given by

$$P(X = x) = Z(\rho)e^{\rho \sum_{k=0}^2 \sum_{i \sim j} I_k(i)I_k(j)} \quad (5)$$

Here ρ is a parameter that specifies strength of the spatial autocorrelation between neighboring areas. $Z(\rho)$ is a normalizing constant that can be dropped in our Bayesian model specification and estimation (Gelman et al., 2004). Neighborhood relationships are captured through a contiguity-based indicator function $I_k(i)$, which is 1 if $x_i = k$ and 0 otherwise. Here, $i \sim j$ means that in this case, street segments i and j are neighbors, provided that they are directly connected to each other through an intersection. We consider global, instead of local, spatial autocorrelation for each ethnicity in the study area (Liu, 2001) and use Moran's I computed using the first-order contiguity street segments as an estimate for ρ (Xu and Bao, 2004).

Third, a useful feature of this model is that it facilitates incorporating covariates into the model by linking $\text{logit}(p_i)$ or $\text{probit}(p_i)$ to a linear combination of predictors Z . Model estimation is done by using Gibbs sampling, a particular useful Markov chain Monte Carlo (MCMC) algorithm for solving multidimensional problems (Geman and Geman, 1984). We illustrate the introduction of covariates by controlling for SEI of the street segment (the mean for employed adults) in the final step of the analysis.

Figure 10 depicts the results for an area of Newark in 1880. Map A shows the initial Yankee categories for these street sections, divided into three equal terciles based on the number of Yankees (low, medium, and high). The same procedure is performed for Germans and Irish separately.⁵ Note that street segments with a high number of Yankees may not have a high proportion of Yankees if the density of Irish or Germans is also high. Map B shows the Bayesian solution (the posterior median estimates of classifications of street segments for Yankees) without controlling for SEI. It is notable that quite a few street segments initially classified as having high proportion of Yankees in Map A have been relabeled in this solution as having medium or low proportions, and some street segments with a medium proportion of Yankees have been relabeled as low. This suggests that some street segments may have been "mistakenly" classified in the input data, provided that a binomial model is appropriate.

⁵For Irish, a street segment is initially classified as low if it has 2 or fewer Irish, medium if it has 3–9, and high if it has 10 or more. For Germans, a street segment is initially classified as low if it has 2 or fewer Germans, medium if it has 3–11, and high if it has 12 or more. For Yankees, a street segment is initially classified as low if it has 2 or fewer Yankees, medium if it has 3–13, and high if it has 14 or more.

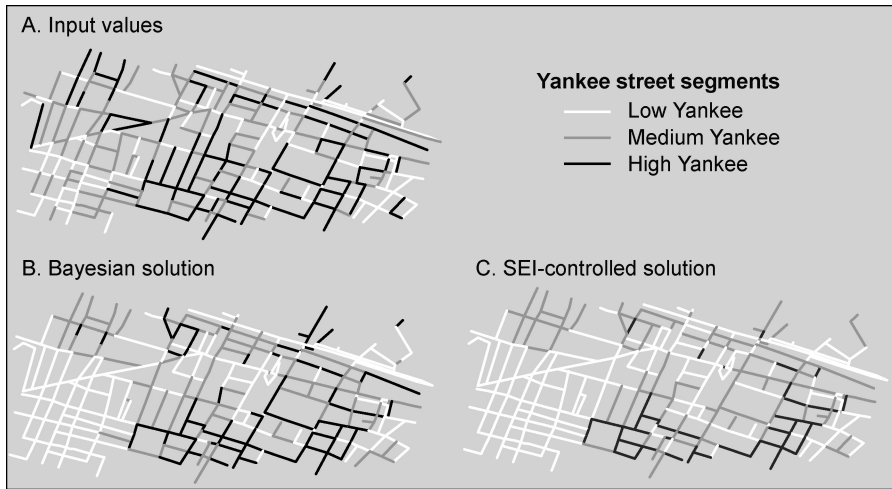


Fig. 10. Initial values and posterior medians of types of street segments for Yankees in a section of Newark, 1880.

Map C shows the posterior median of the types of street segments for Yankees after controlling for the average SEI of street segments in the same area. It is notable that some street segments that are classified as having a high (or medium) proportion of Yankees in Map B now are labeled as medium (or low) proportions. This suggests that for some street segments, the spatial concentration of Yankees can be attributed to the concentration of individuals of similar socioeconomic status (in particular, high status). In other words, Yankees may live close to each other not because of their preference for the same ethnicity but because of their preference for the same socioeconomic status group. Given a certain level of SEI (i.e., controlling for SEI), the proportion of Yankees is no longer great enough for the street segments to be classified as high.

Table 2 summarizes the classifications of street segments in the initial Bayesian solution and in the solution controlling for SEI. It shows that about 4% of street segments have been relabeled from having medium to low, and 6.6% from having high to medium proportion of Yankees after controlling for SEI. On the other hand, about 8.7% of street segments for Irish have been reclassified from low to medium, and 4.7% from medium to high after controlling for SEI. This suggests that for some street segments, the spatial concentration of Irish can be confounded by the concentration of people of similar socioeconomic status. Some street segments actually have relatively high proportions of Irish, but they could be mistakenly classified as having low proportions because they appear to be clusters of low-SEI individuals who are also Irish. Such flips (only about 3%) rarely happen to Germans. This is consistent with the fact that Yankees were on average of the highest socioeconomic status, Irish had the lowest status, and Germans ranked in the middle in 1880 Newark. That only a relatively few street segments become reclassified after controlling for SEI suggests that SEI matters much less than ethnicity itself.

When we create city-wide maps of Irish-ness, German-ness, Yankee-ness, we can then overlay these maps to identify whether an area is “mono-ethnic” (high on only one

TABLE 2. PERCENTAGES OF TYPES OF STREET SEGMENTS BEFORE AND AFTER CONTROLLING FOR SOCIOECONOMIC STATUS

Control for socioeconomic status		No control for socioeconomic status			Row sum
		Low	Medium	High	
Irish	Low	42.2	1.1	0.2	43.5
	Medium	8.7	23.6	2.4	34.7
	High	0.1	4.7	17	21.8
	Column sum	51	29.4	19.6	100.0
German	Low	49.1	1.1	0.0	50.2
	Medium	0.7	26.3	0.7	27.8
	High	0.0	0.5	21.6	22.1
	Column sum	49.8	27.9	22.3	100.0
Yankee	Low	36.4	4.0	0.3	40.7
	Medium	3.4	18.4	6.6	28.4
	High	0.0	3.1	27.7	30.8
	Column sum	39.8	25.6	34.6	100.0

dimension) or mixed. We finally overlay the three maps of posterior median of ethnic proportions for Irish, Germans, and Yankees after controlling for SEI. We can classify the street segments that are labeled as having a high proportion of Irish and low proportion of both Germans and Yankees as “Irish” neighborhoods, and so on for “German” and “Yankee” neighborhoods, resulting in a map like that shown in Figure 11. “Other” neighborhoods are those in which the density of no group is high. This specific solution could yield different maps. For example, we could add additional categories such as high on two or three ethnicities to highlight particular combinations of groups. The overall pattern is similar to those obtained from k-functions and energy minimization, although the resulting ethnic neighborhoods seem to be more fragmented, especially for some Irish and Yankee street segments.

DISCUSSION AND CONCLUSION

Geographic information technologies have led to an explosion of new data sources, allowing analysts to revisit old questions about the organization of cities. The study of ethnic differentiation within cities has depended largely on geographic aggregations of people such as census tracts or wards, which become the lens through which the city is viewed. Spatially aggregated data constrain what we can see because we cannot look inside the aggregations. They also limit how we can think about neighborhoods. Many social scientists begin with the concept of neighboring—the connections and daily interactions between households living in the same building, or the buildings next door or across the street, or maybe in the next block. But our data impose a different notion that is well captured in the nature of the exposure indices that are a standard measure of segregation. A higher exposure index means that members of one group tend to live in areas with larger



Fig. 11. Combined Bayesian classification of street segments for Irish, Germans, and Yankees in Newark, 1880.

shares of another group. But such exposure does not necessarily mean more social interaction between these groups, especially because the geographic scale at which the index is calculated may not be related to the scale of neighboring.

Census data will never directly measure social interaction. But the increasing availability of fine-grained information on individual people and households offers the possibility of constructing maps of neighborhoods that begin at a much finer scale than that of the block group or tract. Yet an abundance of disaggregated data poses potential challenges for analysts. Dealing with 100 or 500 census tracts in a city is much simpler than dealing with 100,000 locations. This challenge is the motivation for our experimentation with alternative methods of neighborhood identification, to learn how to study spatial patterns at a much finer spatial scale while linking operational decisions to theoretical conceptions of what a neighborhood is.

The three methods used to reveal the underlying social structure of the city differ in their assumptions. The Bayesian and energy minimization approaches assume that the types of neighborhoods that exist within a city are known *a priori*. The question these methods seek to answer is, “Where are the ethnic neighborhoods?” The k-function approach, on the other hand, does not make assumptions about the number and/or types of neighborhoods—it asks, “What are the neighborhoods and where are they located?” It is not our intention to argue for the superiority of any one of these approaches. They all begin with the same information about Newark and it is not surprising that the resulting maps are generally similar.

As there is no conventional template for assessing these models, the right result could be seen as being indeterminate. There are potential empirical guides for making choices. For example, as in many cluster analyses one might seek to maximize variation between neighborhoods while also maximizing homogeneity within them. This criterion presumes, however, that each neighborhood has one dominant ethnic group, while people from many ethnic groups can live in a neighborhood. Another criterion would employ external validation. Archival research could determine the locations of churches whose denomination or name denotes ethnic affiliation; archival research could also reveal the addresses of members of ethnic voluntary associations. The best mapping of neighborhoods could seek to locate these institutions most centrally in zones represented by their ethnic constituents. Historical documents that give ethnic labels to areas of the city would also be of use for validation.

Ultimately, empirical indicators (like the labels given to buildings or street segments) are best judged by their consistency with a theoretical concept. The three methods presented in this paper differ in their theoretical underpinnings.

The first question raised in the introduction is whether neighborhood boundaries are rolling or discrete. The Bayesian method views neighborhoods probabilistically: a street segment is not in or out of a neighborhood but rather has a probability of belonging to one of nine neighborhood types (high, medium, or low intensity of each ethnicity). This method suggests that a neighborhood is not a discrete territory, but rather a field of varying membership probabilities. Zones where these probabilities are low could be understood as borders or zones of transition. By contrast, in the energy minimization and k-function approaches neighborhood assignment is discrete not probabilistic.

The second question concerns degree of ethnicity needed for a neighborhood to be given an ethnic label. An advantage of all of the methods presented here is that they do not require setting an *a priori* threshold of ethnicity. Instead, the question is framed in terms of some form of spatial autocorrelation: From knowledge about “who are your neighbors” we infer neighborhood ethnicity and boundaries.

The third question concerns the scale and components of the neighborhood. An advantage of all three approaches and of the underlying dataset is that geographic scale is not predetermined, and any of the approaches could in principle be applied from information about buildings or street segments, or any other geographic unit. In one case we began with data about individual buildings and asked about their relation to other buildings at a variety of spatial scales. In two other cases we presumed that nearby buildings should be aggregated into street segments, i.e., that people are likely to have the most face-to-face interaction with people living along the two sides of the street on the same block and therefore share the same sense of neighborhood. We treat intersections of streets as effective delimiters of shared social space, but we also treat them as connectors. In both the energy minimization and Bayesian models we choose to treat only connected street segments as meaningful neighbors of the focal street segment. We could have stretched to greater distances, taking into account the ethnic composition of the next connected street segment, and the next. In the k-function analysis we created distance bands as concentric rings, but measuring those distances along the street network might better represent our expectations about how streets shape interactions. With any of these approaches we could have adopted the view that “neighbors” at all distances count but that nearer ones count more.

We present these alternatives and describe the choices we made in operationalizing them as a step toward better understanding of neighborhoods, emphasizing that we do not believe there is only one answer. The principal challenge for researchers interested in using large high resolution geographic data bases to study cities is connecting theory to methods and data. Selecting an approach and applying it to data is a thoughtful process, and the right one is the one that most clearly reflects the way the analyst thinks about underlying social processes.

REFERENCES

- Agresti, A., 2002, *Categorical Data Analysis*, 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc.
- Alba, R. D., Logan, J. R., and Crowder, K., 1997, White neighborhoods and assimilation: The greater New York region, 1980–1990. *Social Forces*, Vol. 75, 883–909.
- Bailey, T. and Gatrell, A., 1995, *Interactive Spatial Data Analysis*. Englewood Cliffs, NJ: Prentice Hall.
- Bobo, L. D., Oliver, M. L., Johnson, J. H., Jr., and Valenzuela, A., 2000, *Prismatic Metropolis: Inequality in Los Angeles*. New York, NY: Russell Sage Foundation.
- Boykov, Y., Veksler, O., and Zabih, R., 2001, Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, 1222–1239.
- Bozorgmehr, M., Der-Martirosian, C., and Sabagh, G., 1996, Middle Easterners: A new kind of immigrant. In R. Waldinger and M. Bozorgmehr, editors, *Ethnic Los Angeles*. New York, NY: Russell Sage Foundation, 345–378.
- Burstein, A. N., 1981, Immigrants and residential mobility: The Irish and Germans in Philadelphia 1850–1880. In T. Hershberg, editor, *Philadelphia: Work, Space, Family, and Group Experience in the Nineteenth Century*. New York, NY: Oxford University Press, 174–203.
- Campbell, E., Henly, J. R., Elliott, D. S., and Irwin, K., 2009, Subjective constructions of neighborhood boundaries: Lessons from a qualitative study of four neighborhoods. *Journal of Urban Affairs*, Vol. 31, 461–490.
- Congdon, P., 2003, *Applied Bayesian Modelling*. New York, NY: Wiley.
- Cressie, N., 1991, *Statistics for Spatial Data*. New York, NY: Wiley.
- Fotheringham, A. S. and Wong, D. W. S., 1991, The modifiable areal unit problem in multivariate statistical analysis. *Environment and Planning A*, Vol. 23, 1025–1044.
- François, O., Ancelet, S., and Guillot, G., 2006, Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics*, Vol. 174, 805–816.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B., 2004, *Bayesian Data Analysis*, 2nd ed. New York, NY: Chapman Hall.
- Geman, S. and Geman, D., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 6, 721–741.
- Getis, A. and Franklin, J., 1987, Second-order neighborhood analysis of mapped point patterns. *Ecology*, Vol. 68, 473–477.

- Grannis, R., 2009, *From the Ground Up: How the Layered Stages of Neighbor Networks Translate Geography into Neighborhood Effects*. Princeton, NJ: Princeton University Press.
- Hartshorne, R., 1939, *The Nature of Geography: A Critical Survey of Current Thought in the Light of the Past*. Lancaster, PA: Association of American Geographers.
- Hershberg, T., Light, D., Cox, H. E., and Greenfield, R., 1981, The "Journey-to-Work": An empirical investigation of work, residence, and transportation, Philadelphia, 1850 and 1880. In T. Hershberg, editor, *Philadelphia: Work, Space, Family, and Group Experience in the Nineteenth Century*. New York: Oxford University Press, 128–173.
- Hipp, J. R., 2007, Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review*, Vol. 72, 659–680.
- Horton, J., 1995, *The Politics of Diversity: Immigration, Resistance, and Change in Monterey Park, California*. Philadelphia, PA: Temple University Press.
- Hunter, A., 1974, *Symbolic Communities*. Chicago, IL: University of Chicago Press.
- King, G., 1997, *A Solution to the Ecological Inference Problem*. Princeton, NJ: Princeton University Press.
- Lacy, K. R., 2007, *Blue-Chip Black: Race, Class, and Status in the New Black Middle Class*. Berkeley, CA: University of California Press.
- Liu, J. S., 2001, *Monte Carlo Strategies in Scientific Computing*. New York, NY: Springer.
- Logan, J. R., Alba, R. D., and Zhang, W., 2002, Immigrant enclaves and ethnic communities in New York and Los Angeles. *American Sociological Review*, Vol. 67, 299–322.
- Logan, J. R. and Zhang, W., 2004, Identifying ethnic neighborhoods with census data: Group concentration and spatial clustering. In M. Goodchild and D. Janelle, editors, *Spatially Integrated Social Science*. New York, NY: Oxford University Press, 113–126.
- Martin, D., 1998, Automatic neighbourhood identification from population surfaces. *Computers, Environment and Urban Systems*, Vol. 22, 107–120.
- Omer, I. and Or, U., 2005, Distributive environmental justice in the city: Differential access in two mixed Israeli cities. *Tijdschrift voor Economische en Sociale Geografie*, Vol. 96, 433–443.
- Openshaw, S., 1977, A geographical solution to scale and aggregation problems in region-building, partitioning and spatial modeling. *Transactions of the Institute of British Geographers*, Vol. 2, 459–472.
- Philpott, T. L., 1978, *The Slum and the Ghetto: Neighborhood Deterioration and Middle-Class Reform, Chicago, 1880–1930*. New York, NY: Oxford University Press.
- Rich, M. A., 2009, It depends on how you define integrated: Neighborhood boundaries and racial integration in a Baltimore neighborhood. *Sociological Forum*, Vol. 24, 828–853.
- Romesburg, H. C., 2004, *Cluster Analysis for Researchers*. Raleigh, NC: Lulu Press.
- Sampson, R., Morenoff, J., and Earls, F., 1999, Beyond social capital: Spatial dynamics of collective efficacy for children. *American Sociological Review*, Vol. 64, 633–660.
- Sobek, M., 1996, Work, status, and income: Men in the American occupational structure since the late nineteenth century. *Social Science History*, Vol. 20, 169–207.
- Stonecall, L., 1981, Cognitive mapping: Gender differences in the perception of community. *Sociological Inquiry*, Vol. 51, 121–127.
- Suttles, G., 1972, *The Social Construction of Communities*. Chicago, IL: University of Chicago Press.

- Xu, L. and Bao, S., 2004, Detecting spatial clusters: A Bayesian approach with application to identifying disability patterns in Mississippi and Alabama. In A. Getis, J. Mur, and H. G. Zoller, editors, *Spatial Econometrics and Spatial Statistics*. New York, NY: Palgrave Macmillan, 265–295.
- Zhou, M., 1992, *Chinatown: The Socioeconomic Potential of an Urban Enclave*. Philadelphia, PA: Temple University Press.