1 Introduction

City residents face a variety of problems that depend upon coordinating with neighbors: informal social control of youth, maintaining or increasing property values, keeping the sidewalk free of ice in winter, electing officials, or getting permits for a block party.

Residents coordinate their actions through personal contacts with neighbors, formal organizations, or shared perceptions and beliefs about the neighborhood [Hunter, 1974, Taub et al., 1987, Bursik and Grasmick, 1993]. These different bases for action imply different sets of relevant neighbors. For keeping an eye out for truant school children, the relevant neighbors are mainly other residents on the block, known by and to the resident. For electing a city councilor, the relevant neighbors are all voters in an arbitrary political district. In reckoning whether to buy a home, the relevant neighbors are those whose recent fate seems especially informative about the economic fortunes
of an area.

In researching how neighbors develop capacity to act collectively, scholars have focused on the first two: investigating how residents come to have personal relations with their neighbors and the ways that local and city-wide organizations can solve local coordination problems.

This is not because researchers think that beliefs and perceptions are unimportant. The key results in the neighborhood effects literature are result about beliefs. Resident’s beliefs about neighbors, as reported in survey research, appear to influence a variety of measures of social order [Sampson et al., 1997]. From the place making literature, we know that local activists and developers also believe that the fate of an area can depend upon beliefs about what the area is or will be [Martin, 2003].

However, we know very little about a.) how residents come to perceive some common region of the city as their neighborhood, and b.) how they form beliefs about this area. In large part, this is because we have not had any means to systematically and reliably identify the parts of a city that residents will tend to see as their neighborhood. It’s hard to study how a thing emerges and changes if we can’t ever locate the thing.

Some researchers have attempted to identify perceived neighborhoods by surveying city residents, typically by asking residents to draw maps of their neighborhood. These studies persistently find that residents draw very idiosyncratic maps, often with little overlap.[Hunter, 1974, Campbell et al., 2009]. However, with great individual level variation, there can still be def-
inite districts of a city which residents will tend to perceive as their neighborhood.

In order to observe these average perceived neighborhoods, we need enough data. Historically, it has not been practical to collect these data at the volume and spatial resolution necessary.

In this paper, I discuss a novel Internet source for data on neighborhood perception, demonstrate that this data coheres together, and provide evidence that these differences in perception line up physical features and the spatial organization of social factors. Further, I show how we can use these data to predict what areas of a city are likely to be seen as distinct areas by residents even when we don’t have data on those resident’s perceptions.

2 Neighborhood Claims

Craigslist a large, online classified listing service, with about 60 million visitors each month from the United States [cra]. Many of the listings for sublets, roommates, and apartments on the service contain claims that the advertised location is part of a particular neighborhood. In aggregate, these claims can inform us about neighborhood perception.

2.0.1 Craigslist Listings

Craigslist is organized into more than 700 city specific sites. On a city specific site, like http://chicago.craigslist.org, the classified listings
are organized into categories like “Housing” and subcategories like “Sublets/Temporary.”

It is free to read the classified ads and free to post ads except for a few subcategories. It is free to post in all the “Housing” categories except for brokered apartment rentals in New York City.

Both private individuals and real estate companies use Craigslist to advertise housing listings. However, the listings in the categories for roommates, shared housing, sublets, are mainly made by individuals. The listings for apartments are mainly made by landlords both large and small.

When an advertiser posts a housing listing on Craigslist they have the option of putting in an address of the listing or a nearby street intersection. If the advertiser puts in the geographic information, she will be shown a map with a pin over the listing’s location. If the pin is not in the correct place, the advertiser can move the pin as necessary. The advertiser can also put a pin in the desired location without entering in address or street intersection information. When the listing is published, the geographic coordinates of the posting are embedded in the posting’s web page.

In addition, to location of the listing, the advertiser also has the option of filling out a field entitled “Specific location.” The advertiser can put whatever text she wants to in this field, but typically advertisers put in the name of a neighborhood. This information is also embedded in the published listing.

If a housing listing has a neighborhood name in the “Specific location”
and the geographic coordinates of the listing location, the listing represents a claim that a particular point in the city is part of a neighborhood.

I have written a computer program that checks for and downloads, every night, listings in the “apts/housing”, “rooms/shared”, and “sublets/temporary” subcategories for the following cities:

Another computer program parses the listing (a HTML document) and extracts the geographic information and the contents of the “Specific location” field. If the geographic coordinates are missing, but address or street intersection information is available, the program attempt to lookup the geographic coordinates associated with this information using a geocoding service [Shen, 2008].

For Chicago in 2014, there were about 230 new postings a day in the sublet, roommate, and apartment listings. Of these, about 130 listings include a claim that an identifiable, geographic location belongs to some “Specific location.”
2.1 Distribution of Claims

Using these claims about specific locations, we can estimate the areas that advertisers will claim to be a different neighborhoods. This will be a two step process. First, we will use claims about discrete locations to estimate how likely we are to see a particular neighborhood claim over an area of a city. Second, comparing these likelihoods, we will identify the areas where a particular neighborhood is more likely to be claimed than any other alternative. We will take these regions, where the plurality of neighborhood claims favor one neighborhood, to be measures of average perceived neighborhoods.

For a particular neighborhood name, the probability that the name will appear in a listings’s “Specific location” field can depends upon the location of the listing. In other words, the claims for a particular neighborhood name has a probability distribution across the space of the city.

For a particular neighborhood, we can imagine that the location of a particular claim is generated by a draw from this distribution. If so, we can estimate this probability distribution by looking at the empirical distribution of observed neighborhood claims.

2.1.1 Kernel Density Estimation

There are many ways to estimate a probability distribution from observed location data. For one dimensional space, the classic histogram is the most familiar method. We’ll use a simple method called Kernel Density Estimation (KDE). It behaves much like smoothed histogram (Figure 1).
With KDE, we associate every observed point with a two-dimensional normal distribution centered on the point. At any location in space, every observation’s associated normal distribution will have some probability density. At a particular location, if we sum up all these densities, we will get an estimate of the overall probability distribution, i.e. we will get an estimate of the probability density of observing a neighborhood claim at this location.

2.2 Neighborhood Areas

Using KDE, we estimate the probability distribution for every neighborhood name we observe. With this in hand, we can estimate the probability of seeing a neighborhood claim at any location in the city for any neighborhood: $\hat{Pr}(\text{Location}_j \mid \text{Hood}_i)$

Next, we transform these probabilities into the probabilities of observing
a particular neighborhood claim given a location, \( \hat{\Pr}(\text{Hood}_i \mid \text{Location}_j) \). This depends not just on the probability of seeing a claim for a particular neighborhood, but will also depend upon probabilities for of seeing a claim for each and every neighborhood.

\[
\Pr(\text{Hood}_i \mid \text{Location}_j) = \frac{\Pr(\text{Hood}_i) \Pr(\text{Location}_j \mid \text{Hood}_i)}{\sum_k \Pr(\text{Hood}_k) \Pr(\text{Location}_j \mid \text{Hood}_k)}
\]  

(1)

approximated by

\[
\hat{\Pr}(\text{Hood}_i \mid \text{Location}_j) = \frac{n_i \hat{\Pr}(\text{Location}_j \mid \text{Hood}_i)}{\sum_k n_k \hat{\Pr}(\text{Location}_j \mid \text{Hood}_k)}
\]  

(2)

Where \( n_i \) is the number of observations of neighborhood name \( i \).

In Figure 2, we map the estimated probabilities for neighborhood claims on the North Side of Chicago. Each color represents a different neighborhood claim distribution. The dark lines between mark the where the most probable neighborhood claim switches from one neighborhood to another.

3 Spatial and Social Organization

Visually, we can notice that each ‘neighborhood’ is contiguous. There is no neighborhood name that dominates in two separate, disconnected areas. There are definite districts of the city wherein Craigslist advertisers tend to claim that area is a neighborhood.
Figure 2: Estimated probabilities for neighborhood claims
I would like to treat this measure as a kind of ‘averaged neighborhood perception.’ Of course, it is a measure biased by both selection and strategy. As we are using listings for apartments, sublets, and roommates, we have much less data for parts of the city where most housing is owner occupied and single family. Moreover, listers are advertising, and they may be more likely to claim a desirable neighborhood than one with some stigma.

However, the distribution of claims of Craigslist advertisers may still generalize to the distribution of perceptions of residents. The spatial organization of ‘claimed neighborhoods’ is associated with the physical barriers and sociological similarities and differences that urban sociologists have long associated with neighborhood formation. We’ll make this case first visually and impressionistically, and then formally and mathematically.

In the Figure 2, the map also shows the large parks, the north branch of the Chicago river and railroad embankments (dashed line). Neighborhoods are generally separated by these large physical features. In light gray, we have also plotted the borders of Community Areas—areas defined by Ernest Burgess in the 1920s, which he believed had and would continue to have distinct social histories. The ‘claimed neighborhoods’ toe the lines of Burgess’s areas.

In order make these associations mathematically precise, we need a means of associating the spatial organization of a claimed neighborhood with the spatial organization of sociological factors and the locations of physical borders.
As a first step, we could split the city up into a large number of small areas, say city blocks. Then, for each block, we can attempt to predict the dominant neighborhood claim. Regardless of what else we put in the model, a prediction for a particular block should depend upon what we predict for the neighboring blocks. This is a problem of multinomial regression with spatial autocorrelation.

This type of problem is not tractable using classic techniques of maximum likelihood or MCMC methods. The inter-block coupling of discrete variables requires that we calculate terms for every possible configuration of assignments of neighborhood claims to blocks. With just a few blocks and possible neighborhood claims, the combinatorial explosion makes this computationally unfeasible.

However, computer scientists have developed techniques for just this kind of problem, called structured support vector machines. In brief, this technique puts together three ideas. First, a formalism that represents a distribution over coupled discrete random variables as an undirected graph. Second, a result that the best scoring configuration of that distribution is equivalent to a special division of the network called a minimum cut. Finally, a method of finding the minimum cut of a network in polynomial time. Let us now describe this method bit by bit.
3.1 Neighborhood Model Representation

Let’s imagine a very small city consisting of four blocks. We can represent this tiny town as a network where two blocks are connected if they face the same street. In this representation, blocks that are kitty corner are not directly connected. We’ll index the blocks as 1, 2, 3, and 4.

Suppose that, in our city, there are two neighborhoods. Each block belongs to either one or the other of these neighborhoods. Blocks that are similar are more apt to belong to the same neighborhood and blocks that are different are more apt to belong to different neighborhoods.

We want similar blocks to belong to the same neighborhood. One way to formalize this desire is to score every possible pattern of neighborhood labels in such a way that our preferred patterns have the best score.

First, we need some more precise terms. Let the two neighborhoods be called 0 and 1. A block belongs to either neighborhood 0 or 1, and we will denote this membership as $y_i$, so that $y_1 = 0$ is equivalent to saying that...
block 1 belongs to the 0 neighborhood. Let the similarity between blocks $i$ and $j$ be called $\phi_{i,j}$.

Let a particular pattern of assignment of blocks to neighborhoods be called $y$. The score of $y$ will be $E(y)^2$ which will take the following form:

$$E(y) = \sum_{<i,j>} N \epsilon_{i,j}(y_i, y_j, \phi_{i,j})$$  \hspace{1cm} (3)

Where $<i,j>$ indexes a pair of neighboring blocks, and where $i < j$. $N$ are all these indices of neighboring blocks. The $\epsilon_{i,j}$ function is

$$\epsilon_{i,j}(y_i, y_j, \phi_{i,j}) = \begin{cases} 
0 & y_i = y_j \\
\phi_{i,j} & y_i \neq y_j 
\end{cases}$$  \hspace{1cm} (4)

Suppose that we want our preferred neighborhood assignment to have the lowest score. If our similarity measures $\phi_{i,j}$ are positive when blocks are similar and negative when blocks are different, then we will encourage similar, neighboring blocks to belong to the same neighborhood.

Let our city have the following $\phi$’s.
Table 1: Preferred Assignment: If $y_i = 0$, the block is colored white. If $y_i = 1$, the block is black.

\[
\begin{align*}
\phi_{1,2} &= 1 \\
\phi_{1,3} &= -1 \\
\phi_{2,4} &= -1 \\
\phi_{3,4} &= 1
\end{align*}
\]

If we now score every possible pattern of neighborhood labels, we will find that the lowest scoring assignments are ones that put blocks 1 and 2 in one neighborhood and 3 and 4 in the other (Table 1, Table 2). By choosing the right $\phi$'s we can have made our preferred pattern have the lowest scores.

### 3.2 Learning similarities

Suppose we knew that we wanted some particular pattern of neighborhoods, but we didn’t know what similarity scores, $\phi$’s would give that pattern the lowest score. In order to find the $\phi$’s we might start with a likely looking set
Table 2: Scores of Neighborhood Assignments

For small networks, we could check the scores for each individual pattern, but this exhaustive strategy becomes quickly impractical for larger cities. With two possible neighborhoods, the number of possible assignments is $2^N$ where $N$ is the number of blocks. This permutational explosion means that
for even small cities, we cannot possibly check every possible neighborhood assignment in human-scale time.

### 3.2.1 Network Methods

Fortunately, researchers, largely in the field of computer vision, have developed methods to quickly find the lowest scoring assignment for problems like ours. This work traces back to a 1986 paper by Greig, Porteous, and Sehult, where they demonstrated that finding lowest scoring assignment for a two neighborhood case was equivalent to solving the problem of finding the minimum cut of a network.[Greig et al., 1989]

In that paper, the authors set up an scoring problem that is slightly different than ours. In addition to terms that depended upon pairs of neighbors, their score also included terms for individual nodes.

\[
E(y) = \sum_i \epsilon_i(y_i) + \sum_{<ij>}^N \epsilon_{i,j}(y_i, y_j, \phi_{i,j}) \tag{5}
\]

The node terms were of one of two forms, either

\[
\epsilon_i(y_i) = \begin{cases} 
0 & y_i = 0 \\
\phi_i \geq 0 & y_i = 1 
\end{cases} \tag{6}
\]

or
\[ \epsilon_i(y_i) = \begin{cases} \phi_i \geq 0 & y_i = 0 \\ 0 & y_i = 1 \end{cases} \] (7)

They had the same form of \( \epsilon_{i,j} \) as we do

\[ \epsilon_{i,j}(y_i, y_j, \phi_{i,j}) = \begin{cases} 0 & y_i = y_j \\ \phi_{i,j} & y_i \neq y_j \end{cases} \] (8)

but unlike us, \( \phi_{i,j} \geq 0 \) for all \( <i,j> \) in \( \mathcal{N} \).

One of their networks could look like this

Based on this undirected network, Greig and his co-authors constructed a special, directed network where every edge has a cost associated with it. For every variable \( y_i \) in the undirected network, the directed network contains a node \( z_i \). For every edge \( (y_i, y_j) \) in the undirected network, they introduced a directed edge from \( z_i \) to \( z_j \) and another from \( z_j \) to \( z_i \) with associated costs of \( \phi_{i,j} \).

In addition to the original nodes, they also introduced two special nodes: a source node called \( s \) and a target, or sink, node called \( t \). If \( \epsilon_i(0) = 0 \) then there would be a directed edge from \( s \) to the \( i \)th node. If \( \epsilon_i(1) = 0 \) then there
would be directed edge from the $i$th node to $t$. Networks like this are called s-t networks.

Suppose that $\epsilon_1(1) = 2$, $\epsilon_2(0) = 1$, $\epsilon_3(1) = 3$, and $\epsilon_4(0) = 3$. The corresponding directed network is shown in Figure 3.

![Figure 3: S-T Network](image)

We must now introduce some terms. First, to ‘cut’ a network is to remove edges from the network so that the network is split into two unconnected, smaller networks. Second, an ‘s-t cut’ is a cut on a s-t network that causes the source node and target node to end up in separate, unconnected networks. Finally, a ‘minimum cut’ is an s-t cut where the sum of cost associated with
the removed edges is as small as any other possible s-t cut.

In our example, removing the edges \((z_2, t)\) and \((z_4, t)\) is an s-t cut, but not the minimum cut. The minimum cut would be removing edges \((z_1, z_2)\) and \((z_3, z_4)\) (Figure 4).

Figure 4: Minimum Cut

Every s-t cut can be mapped onto a pattern of neighborhood block labels by letting \(y_i = 1\) if node \(z_i\) ends up in the \(t\) subnetwork and \(y_i = 0\) otherwise. Greig and his coauthors demonstrated that the minimum cut of their constructed s-t network always corresponded to the minimum scoring pattern of block assignment.
Because computer scientists have known how to solve the minimum cut problem swiftly since the 1950s, this meant that a large class of network configurations problems were now tractable. Instead of checking every possible pattern out of exponentially possibilities, we can now find the best scoring pattern directly and quickly.[Ford and Fulkerson, 1956]

In particular, if we want to see if a certain set of similarities, $\phi$’s gives a preferred pattern of neighborhood labels a lower score than any other pattern, we can apply Greig’s approach. If the minimum cut of the associated s-t network corresponds to our preferred pattern, then we know our $\phi$’s gives us our preferred pattern the lowest score. If the minimum cut corresponds to some other pattern, then we know that these $\phi$’s do not give our preferred pattern the lowest score.

**Submodularity** Greig and his coauthors’ original result only applied to scoring functions where blocks could belong to only one of two classes and where the scoring function had an important property called submodularity. A submodular scoring function would be one where $\epsilon_{i,j}(0,0) + \epsilon_{i,j}(1,1) \leq \epsilon_{i,j}(0,1) + \epsilon_{i,j}(1,0)$ for all $i,j$ in $\mathcal{N}$. Our scoring function is decisively not submodular because we want it to cost more to put together very dissimilar neighboring blocks than to assign them to them to different neighborhoods. That is, for very dissimilar blocks we want it to hold that $\epsilon_{i,j}(0,0)+\epsilon_{i,j}(1,1) \geq \epsilon_{i,j}(0,1) + \epsilon_{i,j}(1,0)$.

Since we don’t have a submodular scoring functions and often will have
more than two neighborhood names, we can’t use the minimum cuts algorithm directly to find the lowest scoring neighborhood pattern. However, since the 1980s computers scientists have developed other graph cutting methods that allow us to find approximate solutions to multi-label non-submodular scoring problems like ours [Kolmogorov and Rother, 2007]

3.3 Learning similarities

These network methods mean that we can quickly check if a particular set of inter-block similarities gives a preferred neighborhood pattern the best score. We will now describe how to use this to efficiently find inter-block similarities that will give a preferred pattern the lowest score.

We will make our lives easier by only considering simple, linear similarity terms. As before, the scoring function will be

\[
E(y) = \sum_{<i,j>}^N \epsilon_{i,j}(y_i, y_j, \phi_{i,j})
\]

and

\[
\epsilon_{i,j}(y_i, y_j, \phi_{i,j}) = \begin{cases} 
0 & y_i = y_j \\
\phi_{i,j} & y_i \neq y_j 
\end{cases}
\]

but now, let the similarity term \(\phi_{i,j}\) be the weighted sum of observed similarities between block \(i\) and \(j\).
\[
\phi_{i,j} = w_0 + w_1 s_{1,i,j} + w_2 s_{2,i,j} + \ldots + w_n s_{n,i,j}
\] (11)

An observed inter-block similarity could be absolute difference in population, or dummy variable indicating whether the blocks are separated by a railroad, or similar.

With this form, learning a scoring function means finding weights that give a better score to our preferred pattern of neighborhood assignments than any other pattern. In other words, we want to find some weights, \(w\), that solves this system of linear equations:

\[
\begin{align*}
E(y_1, s, w) &\geq E(y^*, s, w) \\
E(y_2, s, w) &\geq E(y^*, s, w) \\
&\ldots \\
E(y_{M-1}, s, w) &\geq E(y^*, s, w) \\
E(y_M, s, w) &\geq E(y^*, s, w)
\end{align*}
\]

Where \(y^*\) is our preferred pattern, \(M\) is the number of possible neighborhood patterns, and \(s\) are the similarities between blocks.
3.3.1 Specifying a unique solution

In order to use some convenient machinery, we will modify the learning problem somewhat. These modifications will only act to specify a particular solution to the system of inequalities. Any solution to the modified problems is also a solution to the system of inequalities and if there is solution to the system of inequalities there will be a solution to the modified problems.

First, observe that if a set of weights gives the preferred assignment a lower score than any other assignment, there is some difference between the score of the preferred assignment and the score of the next best scoring assignment. Call this difference the ‘margin.’ Our first modification to the learning problem is that we now seek the weights that give our preferred assignment the largest margin.

\[
\arg\max_w \gamma \\
\text{such that } \quad \tag{12}
\]

\[
E(y, s, w) - E(y^*, s, w) \geq \gamma \\
\text{for all } y \text{ where } y \text{ is in the set of possible neighborhood assignments } \quad \tag{14}
\]

\[
\text{and } y \neq y^* \quad \tag{16}
\]

This still does not specify a unique set of weights. To do that, we constrain the sizes of the weights. We could do this by may ways, for example by
limiting the sum of the absolute values of the weights. As will be momentarily clear, more convenient choice is to require that \( \sqrt{\sum_i^M w_i^2} = 1 \). The learning problem is now:

\[
\begin{align*}
\arg\max_{w: \sqrt{\sum_i^M w_i^2} = 1} \quad & \gamma \\
\text{such that} \quad & E(y, s, w) - E(y^*, s, w) \geq \gamma \\
\text{for all} \quad & y \text{ where } y \text{ is in the set of possible neighborhood assignments} \\
\text{and} \quad & y \neq y^* \\
\end{align*}
\]

With these two modifications, the learning problem is now a quadratic program, a well known class of constrained optimization problems.\(^3\)

Unfortunately, we cannot directly use an off-the-shelf quadratic program solver, because the number of constraints will typically be too large. We require every possible pattern to have a higher score than our target pattern, and the number of possible patterns grows exponentially with the number of blocks.

Instead, we will solve a similar problem that takes advantage of the network cutting algorithm but which can be solved quickly. The algorithm for solving this problem is typically called the ‘structured support vector machine.’\(^{[Szummer et al., 2008]}\)
3.3.2 Structured Support Vector Machine

First, we initialize the weights to some starting value, create an empty set of constraints $\mathcal{K}$, and set a counter $i$ to 0.

1. Use a minimum cuts algorithm to find the neighborhood pattern that has a lower score than any pattern given the current weights. Call this pattern $y_i$ and add it to the constraint set $\mathcal{K}$.

2. Update the weights by solving the quadratic program:

$$\arg\max_{w : \sqrt{\sum w_i^2} = 1} \gamma$$

such that

$$E(y, s, w) - E(y^*, s, w) \geq \gamma$$

for all $y$ where $y$ is in $\mathcal{K}$

and $y \neq y^*$

3. If the weights changed in the previous step, set $i = i + 1$ and go to step 1. If the weights did not change, stop the routine.

If we find a set of weights that gives our target pattern a lower energy then any pattern in the constraint set $\mathcal{K}$, then we are assured that our target pattern has a lower score than any other possible pattern. If another pattern had a lower score given the final set of weights, then it would have been added to $\mathcal{K}$ in step 1, and our optimization routine would have continued.
With this structured support vector machine, we have the means of efficiently learning weights that give a targeted pattern a better score than any other pattern, but only if those weights exist.

There is a similar, but more general form that can find weights that either give the target pattern the lowest score or minimize the difference between the score of the target pattern and the another, lowest scoring pattern. There are also a tweaks that penalize the scores of patterns that empirically diverge from the target pattern, and which can help the algorithm learn better weights for predicting new patterns.

The details of these extensions are best found in x, y or z. For our current purposes, we use one the extensions, where the slack variables are weighted by block-by-block Hamming loss.

4 Modeling

Now that we have the machinery, we can now measure the association between the spatial organization of our estimated neighborhoods and the spatial organization of physical and social factors in a city. First, we’ll describe the data.

4.1 Data

Our basic units will the Census block, a geographic unit which, in Chicago typically covers half a city block. Using census blocks gives us a fine geo-
Figure 5: Training Data

graphic resolution and also allows for convenient use of Census data. For each census block, we will calculate the most likely neighborhood from our estimated neighborhood distributions at the center of each block. The most likely neighborhood will be our block label $y_i$ (Figure 4.1).

We will count two blocks as neighbors if they both share a bordering street or alley. In the training data, there are 5,857 blocks and 13,887 block-neighbors.
4.1.1 Physical Barriers

For the inter-block similarities, we will use physical, demographic, and administrative features (Table 3).

Highways, rail lines, rivers, and major streets often act as neighborhood boundaries. For each of these types of features, we code an similarity feature, \( s \), as 1 if the two adjacent blocks are separated by the feature or code and 0 otherwise.

We also measure the difference in orientation between blocks. Blocks are nearly all longer than they are wide, and we calculate the angle of the longest side. For each pair of blocks, we calculate the difference between the orientations of the blocks. We normalize the difference to fall in \([0, 1]\). This measure was inspired by Vivien Palmer’s work on the persistent effect of primary settlements.

In the late 1920s, Palmer argued that when a subdivision of a city was originally opened for residential settlement, the area would start out largely homogenous in its building stock and population. This homogeneity would lead to the subdivision maintaining a common fate through the history of the city. The residents would tend to respond similarly to larger urban processes and the building stock would tend to have the same patterns of depreciation and reevaluation. The primary settlement should act as atomic, spatial unit for urban processes.[Palmer, 1932]

When subdivision were originally laid out, the developers tended to align the streets within their subdivision. Misalignment of blocks are a rough
measure of the where different subdivision meet.

4.2 School Boundaries

If two adjacent blocks are in different elementary school attendance boundary then we code an similarity feature with 1, 0 otherwise. Similar for high schools.

4.3 Demographic Features

From the U.S. Census, we have block level information on race, age, family structure, and housing ownership patterns. We can define measures between blocks for these data.

We will use two measures for the demographic factors. The first is the absolute difference. The second is the \( \chi \) distance \( \sqrt{\frac{1}{2} \sum_i \frac{(x_i-y_i)^2}{x_i+y_i}} \), where \( x_i \) is the frequency of bin \( i \) of the histogram \( x \) and similar for \( y_i \). This is just the square root of the familiar \( \chi^2 \) statistic. By taking the square root, \( \chi \) becomes a metric.[Pele, 2011]

For race, the distribution is the number people coded by the Census as “Hispanic or Latino”, “Not Hispanic or Latino : White alone”, “Not Hispanic or Latino : Black Alone”, and “Not Hispanic or Latino : Asian alone.”

For age, we’ll use the absolute difference between the log median age of neighboring blocks.

For family structure, the distribution is of households with children versus
households without.

For housing structure, the distribution is of housing units in the rental market, housing market, or otherwise disposed. We also use the absolute difference between the log median density of housing units.

Many blocks have no one living in them or only a handful. In such cases, it makes little sense to compare demographic distributions. If either adjacent blocks has fewer than 30 persons living in it, we do not calculate the demographic differences. There are 7,092 edges where we calculate these demographic similarities.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>Min</th>
<th>25th Quant.</th>
<th>Median</th>
<th>75th Quant.</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sufficient Population</td>
<td>0.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rail</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highway</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>0.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elementary School</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grid Street</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Block Angle</td>
<td>0.20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Log Housing Density</td>
<td>0.51</td>
<td>0.00</td>
<td>0.16</td>
<td>0.36</td>
<td>0.70</td>
<td>4.00</td>
</tr>
<tr>
<td>Log Median Age</td>
<td>0.14</td>
<td>0.00</td>
<td>0.04</td>
<td>0.09</td>
<td>0.17</td>
<td>1.55</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.24</td>
<td>0.01</td>
<td>0.16</td>
<td>0.22</td>
<td>0.30</td>
<td>0.94</td>
</tr>
<tr>
<td>Housing</td>
<td>0.23</td>
<td>0.00</td>
<td>0.12</td>
<td>0.19</td>
<td>0.31</td>
<td>0.93</td>
</tr>
<tr>
<td>Family</td>
<td>0.13</td>
<td>0.00</td>
<td>0.05</td>
<td>0.11</td>
<td>0.19</td>
<td>0.81</td>
</tr>
</tbody>
</table>

Table 3: Distribution of Barriers and Distances
Figure 6: Rail Lines
Figure 7: Highways
Figure 8: Major Streets
Figure 9: River
Figure 10: Elementary School Attendance Boundaries
Figure 11: High School Attendance Boundaries
Figure 12: Sufficient Population
Figure 13: Difference in Block Orientation
Figure 14: Difference in Distribution of Race and Ethnicity
Figure 15: Difference in Log Median Age
Figure 16: Difference in Log Density of Housing Units
Figure 17: Difference in Distribution of Family Type
Figure 18: Difference in Distribution of Housing Type
5 Modeling

Since many census blocks are empty, the demographic similarity between these blocks is not well defined. We might be tempted to ignore these cases, but deleting these cases would change the topology of the city. Removing unpopulated blocks would turn populated blocks into islands separated by commercial corridors. This would not allow for a neighborhood to span a highway or industrial corridor, a possibility we do not want to preclude.

We deal with this variation by using dummy variables to effectively learn two model at once: a nonpopulated and populated block model.

Our overall scoring function is

$$E(y, s, w) = \sum_{<ij>} \epsilon_{i,j} (y_i, y_j, s_{i,j}, w)$$  \hspace{1cm} (17)

where

$$\epsilon_{i,j} (y_i, y_j, s_{i,j}, w) = \begin{cases} 
0 & y_i = y_j \\
\phi(s_{i,j}, w) & y_i \neq y_j
\end{cases}$$  \hspace{1cm} (18)

and the unpopulated block model is

$$\phi(s_{i,j}, w) = w_0 + w_1 \text{Rail}_{i,j} + w_2 \text{Water}_{i,j} + w_3 \text{Highway}_{i,j}$$  \hspace{1cm} (19)

$$+ w_4 \text{Major Street}_{i,j} + w_5 \text{Elementary School}_{i,j}$$  \hspace{1cm} (20)

$$+ w_6 \text{High School}_{i,j} + w_7 \text{Block Angle}_{i,j}$$  \hspace{1cm} (21)

$$+ w_8 \text{High School}_{i,j} + w_9 \text{Block Angle}_{i,j}$$  \hspace{1cm} (22)
While the model for the populated neighboring blocks will be

\[
\phi(s_{i,j}, w) = w_8 + w_9 \text{Rail}_{i,j} + w_{10} \text{Water}_{i,j} + w_{11} \text{Highway}_{i,j} \\
+ w_{12} \text{Major Street}_{i,j} + w_{13} \text{Elementary School}_{i,j} \\
+ w_{14} \text{High School}_{i,j} + w_{15} \text{Block Angle}_{i,j} \\
+ w_{16} \text{Family Structure}_{i,j} + w_{17} \text{Race and Ethnicity}_{i,j} \\
+ w_{18} \text{Age Structure}_{i,j} + w_{19} \text{Housing Structure}_{i,j} \\
+ w_{20} \text{Housing Density}_{i,j}
\] (23)

We can combine these models by creating a dummy variable that takes a value of 1 if neighboring blocks have sufficient population to support the demographic distance measures, and 0 if the neighboring blocks do not. We’ll interact this dummy with the populated model and add the variables to our unpopulated model.\(^4\)
\[ \phi(s_{i,j}, w) = w_0 + w_1 \text{Rail}_{i,j} + w_2 \text{Water}_{i,j} + w_3 \text{Highway}_{i,j} \]
\[ + w_4 \text{Major Street}_{i,j} + w_5 \text{Elementary School}_{i,j} \]
\[ + w_6 \text{High School}_{i,j} + w_7 \text{Block Angle}_{i,j} \]
\[ + \text{Populated Blocks}_{i,j} \cdot \]
\[ (w_8 + w_9 \text{Rail}_{i,j} + w_{10} \text{Water}_{i,j} + w_{11} \text{Highway}_{i,j} + w_{12} \text{Major Street}_{i,j} + w_{13} \text{Elementary School}_{i,j} + w_{14} \text{High School}_{i,j} + w_{15} \text{Block Angle}_{i,j} + w_{16} \text{Family Structure}_{i,j} + w_{17} \text{Race and Ethnicity}_{i,j} + w_{18} \text{Age Structure}_{i,j} + w_{19} \text{Housing Structure}_{i,j} + w_{20} \text{Housing Density}_{i,j}) \]

6 Results

After estimating the model using the PyStruct structure learning framework,[Miller and Behnke, 2014] we see that the average, perceived neighborhoods are associated with the location of physical barriers and the spatial organization of social factors.

We’ll focus on the parameters for the populated blocks (Table 4). A positive parameter means that, all else equal, the total score will be lower if ad-
jacent blocks are allocated to separate neighborhoods. Negative parameters, means that all else equal, adjacent blocks should be allocated to the same neighborhood. So, for example the negative intercepts mean that adjacent blocks that are identical will prefer to be assigned to the same neighborhood.

Nearly all parameters for barriers and administrative zones have a positive sign, which is what we expect: dissimilar blocks and blocks separated by barriers are more apt to be placed in separate neighborhoods. The negative ‘attractive’ parameter for major street is likely due to fact that every one of our training neighborhoods contains multiple major grid streets.

The demographic distances are more complicated. Differences in ethnic and racial composition is strongly positive. Such a ‘divisive’ parameter is required to have face validity in Chicago. Differences in the density of housing units also has the expected sign. Differences in median age and housing structure are unexpectedly weak, but also very close to 0. The relatively strong ‘attractive’ parameter for differences in family structure is very surprising. According the this model, a block that is all families and a block that has no families are very apt to belong the same neighborhood.

**Model Fit**

In order to evaluate our model fit, we will compare the neighborhoods our model predicts with the training data (Figure 19).

Third, our model seems to do an adequate job in capturing neighborhoods. We have a Rand Index score of 0.93 and an Adjusted Rand of 0.33.
Figure 19: Predicted Neighborhoods
Table 4: Parameter Estimates for Populated Blocks

<table>
<thead>
<tr>
<th>Variable</th>
<th>Parameter Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>-0.0026</td>
</tr>
<tr>
<td>Railroad</td>
<td>0.0104</td>
</tr>
<tr>
<td>Water</td>
<td>0.0029</td>
</tr>
<tr>
<td>Highway</td>
<td>0.0064</td>
</tr>
<tr>
<td>Major Street</td>
<td>-0.0021</td>
</tr>
<tr>
<td>Elementary School</td>
<td>0.0005</td>
</tr>
<tr>
<td>High School</td>
<td>0.0043</td>
</tr>
<tr>
<td>Block Angle</td>
<td>0.0021</td>
</tr>
<tr>
<td>Family Structure</td>
<td>-0.0021</td>
</tr>
<tr>
<td>Race and Ethnicity</td>
<td>0.0040</td>
</tr>
<tr>
<td>Age Structure</td>
<td>-0.0008</td>
</tr>
<tr>
<td>Density of Housing Units</td>
<td>0.0037</td>
</tr>
<tr>
<td>Housing Structure</td>
<td>-0.0004</td>
</tr>
</tbody>
</table>

The Rand Index is a measure of the overlap between two ways of grouping elements. With a set of blocks $B = \{b_1, \ldots, b_n\}$, let $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_p\}$ be a partition, or grouping, of the blocks into $p$ neighborhoods. Each block is in some neighborhood $\mathcal{X}_i$ and no block is in more than one neighborhood. Let $\mathcal{Y} = \{\mathcal{Y}_1, \ldots, \mathcal{Y}_s\}$ be another grouping of the same blocks into $s$ neighborhoods.

Let $a$ be the number of blocks in $B$ that grouped into same neighborhoods in both $\mathcal{X}$ and $\mathcal{Y}$. Let $b$ be the number of blocks in $B$ that are placed in different neighborhoods in both $\mathcal{X}$ and $\mathcal{Y}$. $a$ and $b$ are the number of agreements between the two neighborhood groupings.

Let $c$ be the number of blocks that are grouped in the same neighborhood in $\mathcal{X}$ but put in separate neighborhoods in $\mathcal{Y}$. Finally, let $d$ be the number of blocks that are in separate neighborhoods in $\mathcal{X}$ but put in the
same neighborhoods in $\mathcal{Y}$. $c$ and $d$ are the number of disagreements between the neighborhood groupings.

The Rand index is

$$R = \frac{a + b}{a + b + c + d}$$

(40)

The index has a value of 0 if the neighborhood groupings completely disagree and 1 if they completely agree.

However, for random groupings, the value of the Rand Index is not expected to be 0. Indeed, as the number of groupings increase, the expected value of the Rand Index tends toward 1.

The Adjusted Rand Index corrects for this.

$$AR = \frac{\text{Rand index} - \text{Expected index}}{\text{Maximum Rand index} - \text{Expected index}}$$

(41)

The maximum Rand index is the highest Rand index possible given the number of groupings in $\mathcal{X}$ and $\mathcal{Y}$. It will only be 1 if the number of groupings are the same. For details see on the Rand index see [Yeung and Ruzzo, 2001]. Both the Rand Index are difficult to interpret as isolated numbers. They are most useful in comparing multiple grouping techniques against another. I report these numbers mainly so that others researchers can compare their models and technique against mine.
Discussion

In the social sciences, we have a long tradition of learning about society by asking people questions. However, we usually don’t do this to try to identify what social groupings are meaningful. In part this is because people, like social scientists, have a very difficult time coming up with credible ways of identifying the boundaries of a social movement, class, or status group.

However, people often reveal patterns of affiliation, either directly when we ask what kind of person they are, or indirectly when we ask about their friends and things they don’t like. We haven’t had good ways of aggregating those patterns or of thinking about the logic of those patterns. Traditional, network community detection methods can help us see patterns of affiliation. They don’t usually help us test our ideas about what kind of processes produce those pattern.

The case of the neighborhood is easier than many social groupings. But, if the approaches we have discussed here help us a grasp on what parts of the social world people think are real, and techniques for thinking about why, then we can extend this work to approaches for understanding the core distinctions of our society and our science.
Notes

1In order to converge on the mode of the distribution of scores, we have to calculate a normalizing constant, which is the sum of the scores of all possible assignments. This is computationally too expensive. There have been a number of attempts to find an acceptable substitute for the normalizing constant, but the empirical results have disappointed.[Li, 2009]

2E as in energy not expectation

3The problem is easier to recognize as a quadratic program in its dual, canonical form:

$$\text{argmin}_w \frac{1}{2}||w||^2$$

such that

$$E(y, s, w) - E(y^*, s, w) \geq 1$$

for all $y$ where $y$ is in the set of possible neighborhood assignments

and $y \neq y^*$

4Actually, our model is a little more complicated because we are not just comparing groups of people, but also groups of households and groups of housing units. Sometimes there is sufficient population to compare race, but not sufficient households to compare family structure. We use an additional
dummy variables to handle these cases
References

craigslist | about > factsheet. URL http://www.craigslist.org/about/factsheet.


Figure 20: Predicted Neighborhoods in Chicago