Measuring neighborhood change as the movement of emergent boundaries.

Jonathan Tannen

[Draft. Do not cite.]

Abstract

Cities' spatial patterns in ethnicity and race can be characterized by large clusters of blocks with similar composition and sharp boundaries between those clusters. While most neighborhood research uses fixed boundaries, such as Census tracts, I argue that boundaries are not predetermined, but emerge endogenously and can move over time. If we allow for moving boundaries, a city's spatial demographics can change in two distinct ways: within-cluster changes in demographic composition and boundary movements. In this paper, I develop a Bayesian algorithm, the Space-Time CRP, to identify spatial clusters of households' race and ethnicity through time from block-level Census data, and examine changes in those ethnoracial clusters in Philadelphia, PA from 2000 to 2010. I decompose Philadelphia's demographic change and find a previously unmeasured dynamic: Philadelphia's White and Asian clusters are growing spatially, even as all clusters are internally becoming proportionately more Black and more Hispanic. Gentrification particularly appears to occur by boundary moving–White clusters are "spreading", rather than boundaries remaining fixed and the ethnoracial mixtures within those boundaries changing. I replicate the analysis for the central cities of the 100 largest U.S. Census Metropolitan Statistical Areas.

1 Introduction

Do populations in cities "spread"? The answer, when posed to city-dwellers, is an obvious yes. Residents can quickly point to an unofficial boundary that has moved over time, as Population X now extends all the way up to Street Y. Methodologically, however, neighborhoods research often relies on the use of fixed boundaries, such as census tracts, and models all population change as changes in the demographic proportions within them. Yet population change would mean a very different thing for neighborhoods if it occured by the demographic group(s) of one region expanding block-by-block into space previously occupied by others, rather than the demographic mix within a tract uniformly and gradually shifting. How much of city population change occurs by the movement of boundaries, and how much occurs by actual changes in residential mixes? This paper provides an answer.

When we look at most American cities, we see substantial ethnoracial residential segregation (Massey and Denton, 1993). Cities are composed of vast regions of blocks with similar distributions in household race and ethnicity. In a clustering model, these spatial clusters can be defined by two attributes: their boundaries and each cluster's internal demographic proportions. Over time, then, there are two ways clusters can change: either (1) their internal demographic compositions change, as populations mix differently or (2) they keep the same demographic composition, but spatially grow or shrink to cover more or fewer blocks.[1] Is city population being driven by residents truly mixing differently, or by a simple displacement of one large group overtaking blocks previously inhabited by another?

To answer this question, we first need a method to identify these cluster boundaries and internal demographic proportions across time. I develop a Bayesian spatiotemporal clustering method for block-level Census household data. I use the cluster results to visually explore maps of Philadelphia's residential change from 2000 to 2010, and then to decompose that ethnoracial change into changes in ethnoracial mixing and movements in boundaries. I then replicate the analysis for the central cities of the largest 100 Metropolitan Statistical Areas (MSAs) in the United States. A substantial amount of the observed demographic change, including all of the observed increases in the non-Hispanic White population, was due to boundary shifting.

[1] There is actually a third way, which completes the accounting: a change in the number of households in a given block, which doesn't change the spatial clusters but does their relative populations. I discuss this later.

2 Emergent Boundaries

2.1 What are emergent boundaries?

I call the boundaries between demographic clusters "emergent". By this, I mean that they are created endogenously, as an outcome of residential decisions rather than being predefined. Emergent residential boundaries do not simply align with physical or political lines, though they can. Instead, boundaries between segregated spaces could emerge along streets that aren't obviously different from others. Most importantly, by not being predefined, emergent boundaries can *move*.

Emergent demographic boundaries have long been theorized and implicitly accepted, though perhaps without being named and without the necessary fine-scale data to measure them. Two common families of city model that yield emergent boundaries help illustrate the ubiquity of emergent boundaries in our understanding of cities. Schelling's (1971) foundational model of segregation allowed simulated agents with weak preferences for the type (e.g. race) of their neighbor to move around a grid; the result was surprisingly complete segregation with a sharp boundary dividing populations. This boundary was not predefined, but emerged as an outcome of the iterated sorting by the residents.

Alternatively, economic models of neighborhood formation also allow for sharp divisions of space to emerge from individual-level choices (e.g. Lucas and Rossi-Hansberg (2002)). Agents make residential decisions based on a suite of factors, including local amenities, distance to the urban center, and perhaps, endogenously, other agents' choices. Boundaries in such a model emerge, generally speaking, at points in space where the type of agent who prefers the location switches from one type to another. In these models, even smooth preferences create discrete boundaries at the point of the switch, and these boundaries do not preexist the agents.

These simplifying models have quite different sources of fine-scale boundaries–the first is only endogenous to residential decisions, the second can rely on both exogenous factors and endogenous preferences for neighbors–but both create boundaries that do not rely on pre-defined lines. In this paper, I don't rely on a specific model as a source of emergent boundaries, but use data to identify the spatial boundaries that both models are capable of yielding.

Let's examine some real-world emergent ethnoracial boundaries. Consider Figure 1, which presents maps of the race and ethnicity of households for a section of North Philadelphia in 2000 and 2010, aggregated to Census Blocks and Census Tracts (ignoring the third row for the moment). Census blocks are generally equivalent to a "city block", and are the smallest unit at which Census data is publicly available, serving as the unit from which tracts and my clusters are built. In the map, blocks and tracts are colored as a weighted average of their ethnoracial household composition on the Red-Green-Blue (RGB) color scale. A block that was

60% White households and 40% Black households, for example, would be shaded with a mixture of 60% blue ((0, 0, 1) in RGB) and 40% green ((0, 1, 0)), yielding the greenish-blue RGB of (0, 0.4, 0.6). In the map, the extreme segregation of households is immediate clear. A large region of Black households sits in the Northwest, a large region of Hispanic households in the Northeast, and regions of White households sit in the South and East. These clusters sometimes span multiple tracts and sometimes divide a single tract; they don't align perfectly with the scale and boundaries of tracts. The cluster boundaries are not particularly subtle; we could ask people completely unfamiliar with Philadelphia to try to draw boundaries using a pen and many of their lines would likely agree. Or we could use a clustering algorithm to group similar blocks together, creating those lines as the divisions around the clusters.

[Figure 1 about here.]

Many types of neighborhoods at many scales operate at the same time, there is no "correct" definition of neighborhood (Sampson et al., 2002; Hipp, 2007). Ethnoracial clusters are not a "correct" definition of neighborhood, but instead capture strong spatial demographic patterns which overlay other definitions of neighborhood, and define who lives in those neighborhoods. As such, these clusters are essential for understanding neighborhoods and their change. As an example, two particularly powerful forms of neighborhood are cognitive maps and physical boundaries, and our understanding of each is complemented by ethnoracial clusters. Cognitive map neighborhoods–a form of neighborhood studied by Hunter (1974) and Hwang (2007)–are formed by the boundaries perceived by residents. Residents may intuit the sharp ethnoracial boundaries and use them to inform their own cognitive maps, limiting their daily activities within those lines. Neighborhoods can also be defined by physical boundaries, such as a large street or a river. The importance of these boundaries may be mediated the existence of a demographic boundary; if a physical boundary isn't matched by an emergent one, that may be evidence that the physical boundary is less relevant in neighborhood structures. I do not test these ideas in this paper, but they are plausible and merit further research.

A clearly important type of neighborhood is the political neighborhood, such as school catchment area or police districtss. The movement of demographic boundaries will determine who lives within these other neighborhood boundaries; a school catchment area with an ethnoracial boundary moving through it will experience greater, though perhaps short-lived, diversity.

2.2 Moving boundaries

How do city populations change? Most analyses look at the changes within fixed boundaries, such as changes in the ethnoracial proportions within each Census tract.

However, if we allow for emergent boundaries, all of a sudden a new, unmeasured dynamic of population change becomes possible. Emergent boundaries are not fixed in space. As the inputs change–as spaces evolve, as residents' preferences shift, or as new populations arrive–the boundaries may move, causing the clusters they define to grow or shrink.

A cluster-aware theory of neighborhood change would posit that tract-level measures mask an important spatial reality. With fixed boundaries, we have only one means of population change: change in the internal composition of a unit (e.g. tract), which is treated as uniform within the unit. If boundaries are not predefined, however, there are two types of dynamics available: those internal mixture changes and, additionally, boundary movement. Figure 2 presents toy examples of a single tract experiencing the same tract-level changes by these two very different dynamics. The first column shows a boundary that remains still through time, with the internal composition of one cluster changing uniformly. The all-White cluster becomes internally and uniformly more Black, and the tract average changes accordingly. The second column now shows that demographic boundary moving while each cluster's internal demographic composition remains constant. The all-Black cluster takes over blocks that previously belonged to the all-White cluster, yielding the same green-ing of the tract average. Clusters can, of course, generally exhibit both of these dynamics at once.

There is also a third way the tract proportions can change: by differential changes in the number of households. If new household construction or a decline in vacancy rates occurred disproportionately in the green cluster, then the tract would overall be greener without either the boundary moving or the clusters' colors themselves changing, but simply because the blocks were weighted differently. This dynamic will complement the other two in the decomposition I develop later.

[Figure 2 about here.]

Tract-level measurements often display smooth transitions in population, as the averaged proportions change gradually. It is easy to read these models as the blocks within those tracts experiencing similarly-smooth transitions. If boundary-shifting is the active dynamic, this straightforward interpretation is wrong. Blocks don't experience smooth transitions, but instead sharp ones as an emergent boundary passes over them and they switch discretely from one type of block to another. What we interpret as a gradual change within a tract is really two distinct types of blocks that are being represented in different propotions as a boundary moves through.

Return to the maps in Figure 1. Consider the section of White households in the Southeast of the map. Between 2000 and 2010, the block-level map shows that the boundary has sharply shifted North and West; what had been a diverse mixture of blocks is now almost entirely White. In the tract-level map, the tracts through which the boundary moved are represented as having a smooth internal shift in ethnoracial

proportions; they internally became somewhat more White, without the additional–and readily visible at finer resolution–sense of the line moving within them. We perceive diversity at the tract level which is really just due to a moving boundary that is halfway through.

Just as emergent boundaries have long been present in neighborhoods literature, so too has their movement. In a Schelling-style model, a simple change in the population proportions would mean that the present clusters needed to expand or contract. Economic models of spatial equilibrium might identify shifting conditions or preferences, which would lead to different points of spatial equilibrium where boundaries form. Predating both, Park and Burgess (1925) wrote of populations "spilling over" their fixed boundaries. While their metaphor keeps boundaries in place, their identified dynamic is compatible with the boundary of an emergent cluster moving up to and beyond their drawn fixed boundaries. Much more recently, Guerrieri et al. (2013) showed that poor census tracts next to wealthier tracts were more likely to gentrify than those not neighboring wealthy tracts, a dynamic clearly compatible with a model of gentrified clusters spreading across tract boundaries. Of course, the real world is more complicated than these simplifying models; Massey and Denton (1993) discuss the Great Migration of Southern Blacks to Northern cities, in which landlords targetted blocks neighboring already all-Black blocks, and exploited racial fears and the fear of falling housing prices to acquire and then rent these properties, an active form of cluster expansion. I will not be able to distinguish among these storylines in this paper, but each suggests a specific kind of dynamics: blocks near a boundary exhibiting a discrete change from one type to another, clusters growing and shrinking; not a smooth, uniform change in a tract's composition.

Why would correctly identifying emergent moving boundaries be important for understanding neighborhood change? First, it would illuminate exactly where in a neighborhood the change happens. If boundary-shifting proves explanatory, then we would expect most demographic change to occur at the edges of clusters, as people move into or out of houses at the boundary, causing the observed clusters to grow or shrink. Second, emergent boundaries complicate our ideas of what gentrification means by disentangling two dynamics: how much of gentrification occurs by smooth shifts in the racial composition of blocks and how much by wholesale, discrete changes as a gentrified cluster spreads outward.[2]

[2]I use explicitly gentrification here both because it is an important topic in neighborhood change and because I show later that gentrification was predominantly defined by boundary shifting in 2000-2010, while increases in non-White households were not.

3 Analytic strategy

The analysis for this paper consists of two parts. The first identifies ethnoracial spatial clusters over time from block-level data. I develop a Bayesian model for this task below. Having identified the clusters, the second examines the story they tell us. It explores maps of moving boundaries, and then uses the given clusters to decompose a city's overall population change into differential construction, internal changes in composition, and boundary-movements.

The key limiting factors in studying emergent neighborhood boundaries have been methodological: (1) data availability and (2) computational power and techniques. First, we need data at a fine-enough scale to identify nuanced spatial boundaries. Such data is just now becoming easily available. This paper uses only block-level household data, which has been provided by the U.S. Census since 2000, but future research can and should exploit the large amounts of newly-available geocoded data on crime, property attributes, and even cell phone movements.

The second limitation has been computational. A number of geographers have developed algorithms for identifying fine-scale contiguous neighborhoods in a single time period, called "Regionalization" in the literature (Duque et al. (2007) provides a comprehensive review, Spielman and Logan (2013) performed such an analysis for specifically high-resolution ethnic neighborhoods in a single time period). Here, I adapt the spatial distance-dependant Chinese Restaurant Process (ddCRP) (Blei and Frazier, 2011; Ghosh et al., 2011) to a bayesian spatiotemporal model, which I call the Space-Time CRP. Models using the ddCRP are very similar to the graph-based spanning tree algorithms of Regionalization first proposed by Maravalle and Simeone (1995), though the ddCRP was developed in the Machine Learning literature for Image Segmentation. I discuss this algorithm in the Methodology section, but an important point here is that while I prefer this method for a number of reasons, it does not behave fundamentally differently from any number of spatial clustering algorithms proposed in the literature; the clusters that we are trying to identify are not particularly subtle (as seen in Figure 1), and most reasonable clustering algorithms will agree on the broad strokes.

One strength of the Space-Time CRP is that one does not need to specify the types or number of types of neighborhoods, and the clusters identified are not expected to be racially homogenous or even defined by the high presence of a single race. The method clusters blocks that are close to each other with similar compositions, but those blocks can be similarly diverse as well as similarly homogenous. For example, in Philadelphia, the cluster including the University of Pennsylvania and Drexel University was identified as 64% non-Hispanic White, 14% non-Hispanic Black, 14% Asian, and 6% Hispanic. What's important is that these proportions are similar among the blocks, not that they are particularly high for one group or another.

For this paper, I measure only clusters of household ethnoracial distribution. I made this decision in an attempt to keep the analysis–a new form of boundary with a new method to identify them–relatively simple. Simplifying populations to only residential race and ethniticity is clearly flawed, and misses too many important aspects of neighborhoods to enumerate here. However, because of America's discouraging correlations among race, income, education, social status, etc., identifying sharp geographical boundaries in the patterns of household race and ethnicity will to a first-order capture patterns in most of these other variables. More importantly, this paper should be read as a purely descriptive analysis; I am simply observing *how* ethnoracial populations have spatially changed, without any ability to discuss *why*. That being said, I find that this overly-simple characterization still illuminates a stark process with important implications.

4 Methodology

4.1 Generative Model

The first challenge of the project, and a main contribution of this paper, is to develop an algorithm for identifying contiguous clusters of blocks across space and time. I develop a Bayesian method which samples cluster assignments using the Space-Time CRP, a spatiotemporal case of the ddCRP with some modifications.

The intuition in the full generative model is that blocks belong to clusters, and those clusters have an internal distribution of household race and ethnicity (p). Blocks can belong to different clusters in different times; in a given time, they draw their observed households from the distribution of the appropriate cluster. The clusters' traits themselves can vary over time, so that a cluster's proportions are correlated across time but not fixed. Thus, the model allows blocks' own proportions to change through time by either the traits of its cluster changing, or by the block switching clusters.

Let there be N blocks in T time periods. For block i in time t we observe an R-length vector of data X_{it}. In this paper, X_{it} is a vector of household counts in each of 8 Census ethnoracial groups (Non-Hispanic of each White, Black, Native American or Alaska Native, Asian, Native Hawaiian or Pacific Islander, Other, Two or More Races; and Hispanic of any race). Each block-time has a cluster membership, z_{it}, which determines the distribution from which X_{it} is drawn: a block in cluster z at time t draws its data from a multinomial distribution with probability vector p_{zt}, considering the number of households on the block, n_{it}, fixed. To sample p_{zt}, I model it as a logistic normal distribution, parametrized by γ_{ztr}; these γ are correlated over time, allowing for inertia in clusters' compositions. The full data model is:

- Hyperparameters:
  - $\rho \sim Beta(90, 10)$.
  - $\sigma_0, \sigma_1 \sim LogNormal(0, 1)$.
- For all blocks and times:
  - $z_{1:N,1:T} \sim$ Space-Time CRP(α, ρ, G).
- For each cluster z and race r:
  - $\gamma_{z,t=0,r} \sim N(0, \sigma_0^2)$.
  - $\gamma_{z,t>0,r} \sim N(\gamma_{z,t-1,r}, \sigma_1^2)$.
  - $p_{ztr} = exp\{\gamma_{ztr}\} / \sum_{r'} exp\{\gamma_{ztr'}\}$.
- For each block i and time t:
  - $X_{it,1:R} \sim Multinomial(p_{z_{it},t,1:R}, n_{it})$.

I'll consider the Space-Time CRP in a moment. This model provides the joint probability, dropping subscripts,

$$p(z, \gamma, X, \rho, \sigma_0, \sigma_1 | \alpha, G) = p(X|z, \gamma, n)\, p(z|\alpha, \rho, G)\, p(\gamma|\sigma_0, \sigma_1)\, p(\rho, \sigma_0, \sigma_1).$$

Notice that γ perfectly defines p.

Blocks in the same cluster have similar X because they were drawn from the same multinomial distribution; if a block's observed population is too unlikely to draw from that distribution, it will be assigned to a different cluster.

An important benefit of using the generative model is that it is straightforward to transparently incorporate new data. We might, for example, add a block-level observation in addition to X. This could be age or household size data from the Census, or even non-Census data geocoded to blocks, such crime, income, or business data. Adding in a level of observations here would require multiplying new terms to the full conditional probability, but could be relatively simple with well-chosen distributions. Adding in new data would allow for identification of new boundaries in areas where ethnoracial composition was the same but where there was, for example, a boundary between younger households and older households, or a boundary between regions of higher crime rates and lower crime rates.

4.2 The Space-Time CRP

Modeling cluster assignments z_{it} relies on the Space-Time CRP, which itself is a modified case of the ddCRP. The ddCRP is a distribution over partitions which

models clusters by maintaining a network among blocks; disconnected components of the network are labeled as separate clusters.

First, let's develop the ddCRP in a single time period. Figure 3 illustrates a sample realization as a guide. To generate the ddCRP, each block i chooses another block as an assignment, labeled c_i. These assignments are represented as arrows pointing from block i to block c_i in the figure (for example, $c_4 = 7$ would mean that an arrow points from Block 4 to Block 7, though indices are not labeled in the figure). For this paper, the probability of block i choosing c_i is simply the neighbor function: blocks can connect to only their neighbors or themselves. The full network of assignments form a directed graph. Disconnected components in the graph are the clusters, and we label the cluster assignment of block i as z_i. In the figure, the blocks shaded red form one component of the network, the blocks shaded blue another; these are our clusters, with perhaps all red blocks having $z_i = 1$ and all blue blocks $z_i = 2$ (the exact choice of labels for the clusters are interchangeable).

[Figure 3 about here.]

Though the model is typically conceived in a single time period, we can easily extend it to a multiple-time-period situation by treating time as a third spatial dimension. Now let there be T time periods, labeled 1 to T. We layer the time periods, as depicted in the second row of Figure 3 for $T = 2$. In addition to its neighbors in its own time period, each block in time period $t > 1$ can also connect to itself in the previous time period, $t - 1$. The two-dimensional single-time map thus becomes a three-dimensional multiple-time-period map, but the algorithm doesn't change; all the general ddCRP requires is a neighbor graph and a distance metric. The clusters are fit through space-time, eliminating the need to "match" clusters across time. Notice, in the figure, that two blocks have changed from the red cluster to the blue cluster between time periods.

We sample cluster assignments by reconnecting a given c_{it}. Let G be the neighbor graph of the block polygons (not the ddCRP graph, but the full graph with each block i connected to each of its neighbors). In this multiple time period model, I assign the probability of block i in time t connecting to block j in time t' (written as $c_{it} = (j, t')$) as

$$p(c_{it} = (j, t')|c_{-it}, \alpha, G, \rho) \propto p_\rho(z(c); \rho) \times p_{comp}(z(c); \alpha) \times \tag{1}$$

$$\begin{cases} 1, & i = j, t' = t \\ 1, & i = j, t' = t - 1 \\ 1, & i \sim j, t' = t \\ 0, & otherwise \end{cases}, \tag{2}$$

where c_{-it} is the set of c with c_{it} removed, $z(c)$ is the z assignments implied by the c network of c_{-it} with the potential c_{it} added. $i \sim j$ symbolizes that blocks i and j are neighbors in G. The Space-Time CRP departs from the original ddCRP,

besides the space-time layering, by the first two terms: the probability depends on the properties of the clusters produced, through an inertia component p_ρ, and a compactness component p_{comp}.

The inertia likelihood component, $p_\rho(z;\rho)$, contributes a probability ρ that blocks will keep the same cluster assignment in consecutive time periods. I treat blocks' staying in the same cluster as a binomial distribution:

$$p_\rho(z;\rho) = \prod_{t>0} \binom{N}{S_t} \rho^{S_t}(1-\rho)^{N-S_t}$$

with $S_t = \sum_i \delta_{z_{i,t-1},z_{i,t}}$ simply the number of blocks that have the same cluster assignment in time t as in time $t-1$. Here, $\delta_{z,z'}$ is the Kronecker delta, which equals 1 if $z = z'$ and 0 otherwise.

The compactness component, $p_{comp}(z;\alpha)$, models the compactness and number of the resulting clusters. It contributes an exponential likelihood penalty of

$$p_{comp}(z;\alpha) = \alpha \wedge \left\{ \frac{1}{4\pi} \sum_{z',t} Bound(z',t)^2 / Area(z',t) \right\},$$

where $Bound(z',t)$ is the boundary of cluster z' (in km) in time period t, and $Area(z,t)$ the area (in km^2). The normalization 4π means that the creation of a new circle cluster contributes a log-likelihood penalty of $-\ln(\alpha)$; less compact clusters will penalize more.[3] This compactness penalty is relatively weak, and its purpose is not so much to encourage circular clusters as to prevent "dumbbell" clusters that stretch across the city, which both aren't substantively meaningful and harm the model's mixing.

I will somewhat interchangeably use the language "change in blocks' cluster assignment z" and "boundary-shifting"; they are the same process perceived from different scales. What I explicitly estimate are changes in each block's cluster assignment, a block-level perspective. At a macro level, however, contiguous clusters of blocks define cluster boundaries as the lines dividing regions of different assignments (for example, between the red and blue blocks in Figure 3). A block switching clusters between time periods would appear on the map as the boundary between clusters moving through space, from one side of the block to the other.

4.3 Sampling from the Model

I identify cluster assignments and cluster ethnoracial proportions by sampling from the posterior, $p(c,p,\rho,\sigma_0,\sigma_1|X,\alpha,G)$, using a Gibbs Sampling step for c and Metropolis Hastings step for the other parameters, explained in Appendix A. I marginalize out γ for a given p. Remember that c perfectly determines z.

[3] I use this parametrization to keep results generally in-line with the original ddCRP, in which new clusters contributed $\ln(\alpha)$ to the log-likelihood.

4.4 Decomposing Spatial Population Change

Once we've fit the model above to sample z and p, we can use the results to decompose the city's popultion change. A powerful feature of the model above, and the core idea of this paper, is that a block's ethnoracial household distribution can change in three ways: (1) by the internal composition of its cluster ($p_{zt,1:R}$) changing over time, (2) by its own cluster membership (z_{it}) changing (boundary movement), and (3) by its number of households n_{it} changing. Substantively, I'm interested in the first two, but the third is necessary to complete the accounting decomposition. This section develops a decomposition of the total population change of a city into these components.

Consider the following equation for the proportion of households of each ethnoracial group, conditional on $z_{1:N,1:T}$:

$$\frac{X_{tot,tr}}{n_{tot,t}} = \sum_i \sum_z \frac{n_{it}}{n_{tot,t}} \delta_{z,z_{it}} p_{ztr},$$

in which the index i being replaced with tot symbolizes the sum over all blocks, and $\delta_{z,z_{it}}$ is the Kronecker delta function which is 1 if $z_{it} = z$ and 0 elsewhere. This equation simply multiplies a block's number of households by the correct cluster's ethnoracial proportions ($\delta_{z,z_{it}}$ will be zero for all of the clusters to which block i is not assigned). If we take p from the posterior of the clustering model, this is approximate (interpreted as the expected value of the number of households we would draw, conditional on p_{ztr}); if we use p as the observed proportions conditional on z, so that $p_{ztr} = \sum_{i|z_{it}=z} X_{itr} / \sum_{i|z_{it}=z} n_{it}$, then this is an accounting identity. I will use the latter; in practice the variance of the posterior of p conditional on z is tiny and the difference between the results of the two are negligible.

With this equation, the change in the population proportions between time t_1 and t_2 can be written as

$$\frac{X_{tot,rt_2}}{n_{tot,t_2}} - \frac{X_{tot,rt_1}}{n_{tot,t_1}} = \sum_i \sum_z \Bigg(\underbrace{\left(\frac{\delta_{z,z_{it_2}} p_{zt_2r} + \delta_{z,z_{it_1}} p_{zt_1r}}{2} \right) \left(\frac{n_{it_2}}{n_{tot,t_2}} - \frac{n_{it_1}}{n_{tot,t_1}} \right)}_{D^{(n)}_{r,iz}} + \underbrace{\bar{n}_i \bar{\delta}_{zi} (p_{zt_2r} - p_{zt_1r})}_{D^{(p)}_{r,iz}} + \underbrace{\bar{n}_i \bar{p}_{zr} (\delta_{z,z_{it_2}} - \delta_{z,z_{it_1}})}_{D^{(z)}_{r,iz}} \Bigg) \tag{3}$$

where $\bar{n}_i = \frac{1}{2}(\frac{n_{it_1}}{n_{tot,t_1}} + \frac{n_{it_2}}{n_{tot,t_2}})$, $\bar{p}_{zr} = \frac{1}{2}(p_{zt_1r} + p_{zt_2r})$, and $\bar{\delta}_{zi} = \frac{1}{2}(\delta_{z,z_{it_1}} + \delta_{z,z_{it_2}})$. We can calculate the sum of each of the addends separately, yielding the decomposition

$$\frac{X_{tot,rt_2}}{n_{tot,t_2}} - \frac{X_{tot,rt_1}}{n_{tot,t_1}} = D^{(n)}_r + D^{(p)}_r + D^{(z)}_r.$$

$D_r^{(n)}$ is the change in the city proportion of households of race r due to differential changes in the blocks' numbers of households, at the average racial proportions of each block's cluster(s). Suppose we allowed block-level construction and vacancy rates to proceed without changing ethnoracial proportions at all (holding p and z fixed). Differential changes in numbers of households among blocks could by themselves cause the city-wide proportions to change. For example, in Philadelphia from 2000-2010, more construction and a larger drop in vacancy rates occured in predominantly-White blocks. Without any changes in block-level proportions, these relative changes would imply a more-White city.

$D_r^{(p)}$ is the change in city proportions due to clusters' changing ethnoracial compositions, evaluated at each block's time-averaged number of households and its "average" cluster membership (for blocks that change clusters, $\bar{\delta}_{zi} = 0.5$ for each of the z they assume). This supposes we fixed the number of households and didn't allow the cluster boundaries to move, but only allowed the internal ethnoracial proportions of clusters to change. This component is illustrated by the internal, uniform changes in colors in clusters or tracts.

$D_r^{(z)}$ is the change in city proportions due to changing cluster membership, evaluated at each block's average number of households and each cluster's time-averaged ethnoracial composition $\bar{p}$. Suppose we fixed the number of households in each block as well as the internal ethnoracial proportions of demographic clusters. The only way blocks could change their composition would be by changing cluster assignment, having the cluster boundaries move around. This component captures the extent of that dynamic, and is the new feature that I claim is important. Fixed tracts assume this to be zero.

Together, these three components completely describe the ways population can change in this model, with one entirely new component ($D_r^{(z)}$) that fixed-boundary analyses do not allow.

5 Results

I have divided discussion of the results into three sections. The first explores the clusters produced by the model to assess how the model and its clusters behave. The second uses those clustering assignments to understand substantively how populations changed, and the importance of boundary movement, in Philadelphia. The final section briefly replicates the analysis for the central cities of the largest 100 MSAs in the United States.

I fit all models using Java, using the network package JGraphtT (Naveh and Contributors, 2011). All data preparation used R, relying heavily on the rgeos (Bivand and Rundel, 2013), reshape2 (Wickham, 2007), spdep (Bivand, 2013), and

dplyr (Wickham and Francois, 2014) packages; and GRASS (GRASS Development Team, 2012), accessed via spgrass6 (Roger Bivand and Neteler, 2014). I created the plots and maps using ggplot2 (Wickham, 2009) and ggmap (Kahle and Wickham, 2013), with base maps provided by Stamen (Stamen Design, 2014) using OpenStreetMap (Haklay and Weber, 2008) data.

5.1 Model Settings

I fit the Space-Time CRP model for the city of Philadelphia using block-level U.S. Census data from 2000 and 2010 for household race and ethnicity (United States Census Bureau, 2012) and block polygons from TIGER/Line shapefiles (United States Census Bureau, 2013). I translated blocks in 2000 to blocks in 2010 using the Census crosswalk, distributing households proportionally to area (changes in block definitions during the data were not substantial). I measure only the City of Philadelphia (defined as the Census Place, which is the county) and not the full MSA. This is an attempt to better target emergent boundaries which occur within political entities, and to not simply have the measured boundaries identify fixed municipal boundaries (which are likely very strong, but are not in the scope of this paper). While I fit the clustering model using eight ethnoracial groups, in my discussion I'll use the five ethnoracial categories non-Hispanic White, non-Hispanic Black, non-Hispanic Asian, Hispanic, and total other (I drop "non-Hispanic" from the names of the first three for the rest of the paper).

I fit the model using six chains of the Gibbs sampler for each of eight values of the parameter α, with $\alpha = exp\{-50,-100,-200,-400,-1000,-2000,-4000,-8000\}$. Remember, α is intuitively a likelihood penalty on new clusters, with smaller values of α (more negative exponents) imposing a stronger penalty and thus yielding fewer clusters; I chose these values to provide a wide range of cluster sizes, from clearly-too-small to clearly-too-large. I present results for $\alpha = exp\{-1000\}$, with results for other values of α and a discussion of why $\alpha = exp\{-1000\}$ appears to be a characteristic scale left for Appendix B. This value of α corresponds to quite large clusters; Philadelphia's population change between 2000 and 2010 can be well-described as tectonic shifts in population.

The results present 300 samples from each of 6 chains of the Gibbs sampler, yielding 1800 sample cluster assignments from the posterior for each α. Every value presented has a Gelman-Rubin statistic (Gelman and Rubin, 1992) less than 1.2, supporting convergence.

5.2 Philadelphia's Clusters

Philadelphia is in many ways a typical large northern U.S. city, experiencing Black hypersegration amid extensive White flight up to the 1980's (Massey and Denton,

1993), then some small regions of gentrification beginning in the 1990's (Hackworth, 2007). While the public discourse has recently focused on gentrification, the city as a whole saw a decrease in the proportion of householders that were non-Hispanic White (from 47.9% in 2000 to 42.6% in 2010), and an increase in the proportions Hispanic (6.4% to 9.3%), Asian (3.6% to 5.2%), and Black (40.3% to 41.0%).[4] How did this change occur? Did every cluster become more Hispanic, Asian, and Black? Or did the clusters keep their initial compositions and simply grow and shrink spatially to accomodate the changing populations?

The third row of Figure 1 maps a single realization of cluster assignments. As for the maps before, clusters are shaded using a weighted average on the RGB scale.

There are two strong trends that the cluster reveal; one a boundary movement and one a change in ethnoracial composition. First, notice that the White cluster in the South expanded, taking over blocks that had been Black. In the western part of the map, the boundary moved north by a few blocks; in the easterern section, the boundary moved West. The second noticeable trend is that the Hispanic cluster in the Northeast became more Hispanic (more orange). In 2000, this cluster was 47% Hispanic, 25% White, 21% Black, and 5% Asian; in 2010 it was 59% Hispanic, 11% White, 24% Black, and 5% Asian.

Figure 4 presents a map of clustering results for the whole city. This sample identified 12 clusters; the full results at this scale produced a mean of 12.3 clusters with 95% of the samples in [10,15]. This map gives a sense of the sheer size of the clusters. First, notice how much *inertia* the clusters have; both the boundaries and the internal compositions seem to stay the same more than they change. However, tracing the lines carefully will find boundaries that did move nontrivially; the extension Westward of the White cluster in the center of the city (by the label "Philadelphia"), for example, is the much-publicized gentrification of University City.

[Figure 4 about here.]

5.2.1 Moving Boundaries

Let's quantify boundary movements. Again, I present results for $\alpha = exp\{-1000\}$; the results are generally robust across scales, though with stronger evidence of boundary movement at the larger scales. Appendix C presents decomposition results for all of the values of α in the previous section.

The population change was not gradual at the block level. It was extreme. Figure 5 shows the proportions of each race for blocks in 2000 and 2010; points below the

[4] It is worth noting that White households make up a larger proportion of Philadelphia's households than the White population does of the total population, as White households are on average smaller than non-White households.

45-degree line saw proportionate declines in the respective race, points above the line saw increases. What stands out is the enormous variance in the plots. White and Black proportions often changed dramatically, particularly for blocks that were diverse in 2000 (in the middle of the plots). It was not uncommon for a block that was 50% White in 2000 to be 20% White in 2010. These sharp leaps in blocks' proportions, while possible in a model with fixed boundaries if the internal compositions changed sharply enough, would be a hallmark of boundary-shifting. Blocks would change dichotomously when a boundary passed over them, differing from the more gradual tract- and city-level trends. Interestingly, Hispanic and Asian proportions seemed to increase more gradually and consistently across blocks, suggesting a systematic increase across the city rather than sharp, dichotomous switches.

[Figure 5 about here.]

The cluster results tell a similarly extreme story. From 2000 to 2010, 6.0% of Philadelphia blocks, representing 6.6% of all households, changed cluster assignments. Given that these are large-scale clusters which have strong boundaries, changing clusters means changing between two very different racial compositions.

Are certain types of clusters–for example, more-Hispanic clusters–growing and others shrinking, leading to the overall city trends? Figure 6A presents the clusters' proportionate growth (defined as the change in the number of blocks divided by the average number of blocks: $\frac{nb_{z2}-nb_{z1}}{\bar{nb}_z}$) versus each 2000 ethnoracial proportion for a single realization of the clusters. The fitted lines are from a regression weighted by the clusters' average number of households, $\bar{nb}_z$; each regression was run separately, so the White regression, for example, does not control for the proportion Black in the cluster. The White clusters grew, with a mean coefficient across all samples of 0.12 (the 2.5-97.5 credible interval is (0.04, 0.23)); this means that a cluster that had a proportion White 10 percentage points higher grew at a rate 1.2 percentage points higher. Black clusters shrank, with a mean coefficient of -0.10, (-0.22, -0.05). Asian clusters grew sharply in this sample, but there was large variance across samples due to their low representation, with a mean of 0.52 (-0.88, 1.94). Hispanic clusters didn't display a tendency to grow or shrink, with a mean of -0.07 (-0.55, 0.31); notice that the single highly-hispanic cluster exhibits strong leverage.

[Figure 6 about here.]

5.2.2 Changing Internal Composition

How could the Black population have grown and the White population shrunk from 2000 to 2010 if the clusters with more Black people are shrinking, and the clusters with more White people are expanding? And how could the Hispanic population

display such growth if clusters wth more Hispanic people are not themselves growing? The answer is that clusters are becoming internally more Black and Hispanic and less White. Figure 6B plots the ethnoracial proportions for clusters in 2000 and 2010 for the same cluster results; points below the 45-degree line experienced a drop in that race's proportion, points above an increase. Almost all of the clusters fall below the line in the White facet, and above the line in the Black, Asian, and Hispanic facets. Clusters became internally less White and more Black, Asian and Hispanic across the board.

5.3 Decomposing Spatial Population Change

The decomposition presented in Eq. 3 disentangles the competing boundary and internal composition effects in terms of their impact on the city population as a whole. Figure 7 displays decomposition results for the entire city for each race/ethnicity, comparing the clusters with tracts (these results are, again, for $\alpha = exp\{-1000\}$, the results for all values of α are presented in Appendix C). The y-axis is the value of the component, which represents the change in the city-wide proportion of a given race/ethnicity due to the given dynamic; for example, a value of -0.01 for $D^{(z)}_{wht}$ means that the city-wide proportion White decreased by 1 percentage point due to boundaries moving. The first column, dX_{tot}, shows the overall proportionate change in household races/ethnicities. These are, by the accounting definition, the same for every α and for tracts. The proportion of households that were White fell 5.4 percentage points, while Hispanic households led the other groups with an increase of 2.8 percentage points. The second column plots $D^{(n)}$, the component representing changes in blocks' numbers of households. Increases in households predominantly occured in clusters/tracts that were already White and decreases in units that were Black; if we were to only allow the number of households on each block to change as observed, we would have a 0.6 percentage point increase in the city's proportion White, and a 0.8 percentage point decrease in the proportion Black. Asian and Hispanic proportions would have increased slightly. These results are similar for clusters and tracts.

[Figure 7 about here.]

The city-wide trends were produced mostly by changes in clusters' internal compositions. The third column presents $D^{(p)}$, the change in population due to allowing only the clusters' (or tracts') internal ethnoracial compositions to change as observed. This produces an average 7.5 percentage point drop in the White population (greater that what we actually observed), and a 3.0 percentage point increase in Black population, 2.9 increase in Hispanic, and 1.4 increase in Asian. The values for cluster assignments are not different from the values for tract-level measures except for Black households; the tract measures assigned only a 1.6 percentage point growth to internal mixing.

The movement of boundaries worked in the exact opposite direction. The fourth column plots $D^{(z)}$, the change in population that would have occurred had only the cluster assignments changed (and the boundaries moved). Tract measures assume that boundaries don't move, fixing $D^{(z)}$ at zero. We saw before, however, that White clusters did tend to grow; here, $D^{(z)}$ contributes a city-wide 1.5 percentage point increase in the proportion White. Conversely, boundary movements caused a 1.5 percentage point decrease in the proportion Black. We don't see large effects for Hispanic or Asian; boundary movements across the city simply didn't impact these populations in a consistent way.

It appears that the boundary movements occured dominantly in the growth of White clusters. To further explore this, let's examine separately the blocks that increased in White proportion, and those that decreased. Figure 8 shows the decomposition when we limit our analysis to only the blocks that increased in proportion White (N = 3,829). To accomplish this, in Eq. 3, I change the sum over i to include only those i which experienced an increase in White population, and recalculate n_{tot} only among these blocks. I define dX_{tot} as the sum of the three components; the accounting identity no longer holds because some clusters will have blocks that both increased and decreased in proportion White, and are thus included in different summations. Furthermore dX_{tot} is no longer the observed changes among the blocks because p will tend to smooth the population changes in clusters where some blocks increased and some decreased. However, we can recover the full decomposition using a household-weighted sum of the increasingly-White and increasingly-non-White decompositions.

The increasingly-White blocks experienced a 4.5 percentage point increase in proportion White using tract proportions, which was smoothed out to an average 1.2 percentage point increase by the clusters. Notably, however, more than all of that increase in White population comes from boundary movement; the internal composition changes would suggest a *decrease* in the White population. The component $D^{(z)}$ reveals that boundary movements were responsible for a 4.4 percentage point increase in the proportion White, and a 5.0 percentage point decrease in the proportion Black. These blocks actually initially belonged to clusters that would experience drops in White proportion ($D^{(p)}$ is negative for White), but their White proportions increased because they switched to Whiter clusters. Tracts have the opposite sign for $D^{(p)}$ of White; the tract-level measure says that White blocks changed because the tracts they belonged to were internally becoming Whiter. It has completely misattributed the change because it didn't realize a boundary was moving through the tracts. Notice that boundary shifting explained very little of the Asian and Hispanic changes; for both tracts and clusters those proportions changed mostly due to $D^{(p)}$.

[Figure 8 about here.]

While moving boundaries defined the dynamic among blocks that became Whiter, the blocks that became less White were driven exclusively by their clusters' changing compositions. Figure 9 shows the decomposition for these blocks (N = 7,413). The cluster measures are almost indistinguishable from the tract measures; these blocks belonged to clusters that became more Black, Asian, and Hispanic, and boundary movements had a net zero effect on these blocks.

[Figure 9 about here.]

5.4 National Analysis

These results are forthcoming.

6 Discussion and Implications

Decomposing the spatial change in the emergent demographic clusters in Philadelphia–and then the 100 largest cities in America–reveals a dynamic that hasn't been measured before: the movement of boundaries represents a substantial piece of neighborhood change. Boundary-shifting is particularly responsible for changes among blocks that increased in proportions White. Gentrification specifically, then, is occuring by an existing population spreading into neighboring regions. Rather than gradual changes in mixing, boundary movement's component of change occurs by blocks along demographic boundaries switching sharply from one cluster to another. What does this distinction imply for theories of neighborhood change? A number of new questions are raised.

First, do other neighborhood definitions and neighborhood ecologies correlate to these emergent demographic boundaries? Those relationships, if they exist, could be causal or not: neighborhoods could be established with these demographic boundaries in mind (either explicitly or implicitly), or boundaries could be an outcome of residential sorting within other neighborhoods.

As a specific example, do these demographic boundaries structure the everyday experiences of residents? If residents are using these emergent boundaries to constrain their daily activities and social interactions, then when neighborhoods changed by boundaries moving, residents would expand or shrink the space within which they operate accordingly. In the extreme case where these boundaries completely bound movements, neighborhood change that occurred by moving boundaries would lead to exactly zero new mixing of populations; residents would simply reshape their activities to continue to operate within the same cluster. Neighborhood change via changes in cluster composition would not have this effect.

In fact, there is previous research that suggests that emergent boundaries do impact residents' cognitive maps. Hwang (2007) studied perceived neighborhood boundaries in a gentrifying neighborhood in South Philadelphia for a time in between these two maps. Hwang found that White residents' boundaries were based on housing values and perceived crime, while Black residents more often used physical landmarks or natural boundaries. It seems that the White residents were creating boundaries based on ecological feel of the neighborhood and their activity space. We don't have the exact household-level data in 2007 to corroborate, but a naive guess is that their perceived boundaries might align with the blocks that had already gentrified, and that these fine-scale ethnoracial boundaries were constraining their neighborhood defintions. This difference could also explain why gentrification is more dependent on boundary-shifting than other ethnoracial change.

How does boundary movement compare among American cities? Hackworth (2007) argues that neoliberalism and access to financial markets has changed the nature of gentrification, with Philadelphia relatively early in adopting strategies to attract large-scale financing. Particularly because boundary shifting appears to have its strongest effect in the process of gentrification, we might expect cities in different parts of the country, with different changes in demographics and different access to financial markets, to experience very different types of boundary shifting. Future research should explore the determinants of differences in boundary shifting among cities.

This paper simply measures *where* the boundaries moved; it leaves open why. What are the determinants of a boundary moving? What other spatial variables either precede or follow their movement? Bringing in other fine-scale variables–be they economic data, business data, crime data, or other variables that are structured at the "neighborhood"-level–as both an input into the identification of boundaries and as correlates of boundaries' moving would help us better understand why boundaries move.

Boundary moving exists and represents an important and distinct component of population change; once we've accepted this, many questions about its specific impact need to be explored.

References

Bivand, R. (2013). *spdep: Spatial dependence: weighting schemes, statistics and models*. R package version 0.5-65. http://CRAN.R-project.org/package=spdep.

Bivand, R. and Rundel, C. (2013). *rgeos: Interface to Geometry Engine - Open Source (GEOS)*. R package version 0.3-2. http://CRAN.R-project.org/package=rgeos.

Blei, D. M. and Frazier, P. I. (2011). Distance dependent chinese restaurant processes. *The Journal of Machine Learning Research*, 12:2461–2488.

Duque, J. C., Ramos, R., and Suriñach, J. (2007). Supervised regionalization methods: A survey. *International Regional Science Review*, 30(3):195–220.

Gelman, A. and Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472.

Ghosh, S., Ungureanu, A. B., Sudderth, E. B., and Blei, D. M. (2011). Spatial distance dependent chinese restaurant processes for image segmentation. In *Advances in Neural Information Processing Systems*, pages 1476–1484.

GRASS Development Team (2012). *Geographic Resources Analysis Support System (GRASS GIS) Software*. Open Source Geospatial Foundation.

Guerrieri, V., Hartley, D., and Hurst, E. (2013). Endogenous gentrification and housing price dynamics. *Journal of Public Economics*, 100:45–60.

Hackworth, J. (2007). *The neoliberal city: Governance, ideology, and development in American urbanism*. Cornell University Press.

Haklay, M. M. and Weber, P. (2008). Openstreetmap: User-generated street maps. *IEEE Pervasive Computing*, 7(4):12–18. http://dx.doi.org/10.1109/MPRV.2008.80.

Hipp, J. R. (2007). Block, tract, and levels of aggregation: Neighborhood structure and crime and disorder as a case in point. *American Sociological Review*, 72(5):659–680.

Hunter, A. (1974). *Symbolic Communities: The Persistence and Change of Chicago's Local Communities*. Midway Reprints. University of Chicago Press, Chicago, IL.

Hwang, J. (2007). The social construction of a gentrifying neighborhood: Redefining and reifying identity and boundaries in inequality. Stanford University undergraduate thesis.

Kahle, D. and Wickham, H. (2013). *ggmap: A package for spatial visualization with Google Maps and OpenStreetMap*. R package version 2.3. http://CRAN.R-project.org/package=ggmap.

Lucas, R. E. and Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica*, 70(4):1445–1476.

Maravalle, M. and Simeone, B. (1995). A spanning tree heuristic for regional clustering. *Communications in statistics-theory and methods*, 24(3):625–639.

Massey, D. S. and Denton, N. A. (1993). *American Apartheid: Segregation and the making of the underclass*. Harvard University Press, Cambridge, MA.

Naveh, B. and Contributors (2011). *JGraphT*. http://jgrapht.org.

Park, R. E. and Burgess, E. W. (1925). *The city*. University of Chicago Press, Chicago, IL.

Roger Bivand, R. K. and Neteler, M. (2014). *spgrass6: Interface between GRASS geographical information system and R.* R package version 0.8-6.http://CRAN.R-project.org/package=spgrass6.

Sampson, R. J., Morenoff, J. D., and Gannon-Rowley, T. (2002). Assessing "neighborhood effects": Social processes and new directions in research. *Annual review of sociology*, pages 443–478.

Schelling, T. C. (1971). Dynamic models of segregationï£¡. *Journal of mathematical sociology*, 1(2):143–186.

Spielman, S. E. and Logan, J. R. (2013). Using high-resolution population data to identify neighborhoods and establish their boundaries. *Annals of the Association of American Geographers*, 103(1):67–84.

Stamen Design (2014). *Toner Map Tiles*. Under Creative Commons by 3.0. http://maps.stamen.com/.

United States Census Bureau (2012). United States Census, 2000 & 2010. Accessed July 2, 2014. http://api.census.gov/data/.

United States Census Bureau (2013). United States Census, TIGER/Line shapefiles. Accessed July 2, 2014. https://www.census.gov/geo/maps-data/data/tiger-line.html.

Wickham, H. (2007). Reshaping data with the reshape package. *Journal of Statistical Software*, 21(12):1–20. http://www.jstatsoft.org/v21/i12/.

Wickham, H. (2009). *ggplot2: elegant graphics for data analysis*. Springer New York. http://had.co.nz/ggplot2/book.

Wickham, H. and Francois, R. (2014). *dplyr: a grammar of data manipulation*. R package version 0.2. http://CRAN.R-project.org/package=dplyr.

A The Gibbs Sampler

To fit the model in Section 4.1, I use a Gibbs Sampler to iteratively sample each c_i, with γ marginalized. I use a Metropolis Hastings step to sample σ_0, σ_1, and a Gibbs step for ρ. Once we've sampled c, which defines z, it is straightforward to calculate the posterior $p(\gamma|X, z(c), \Sigma)$.

A.1 Sampling c_i

First, let's calculate $p(c|X, \alpha, \rho, G, \sigma_0^2, \sigma_1^2)$, in which I've suppressed the subscripts for readability and have marginalized out γ. Note that c perfectly defines cluster membership z and γ perfectly defines p (I will denote the z defined by a given c as $z(c)$ and the p defined by γ as p_γ). For each set of c, we can then calculate $p(p|c, X, \sigma_0^2, \sigma_1^2)$. The conditional distribution is

$$\begin{aligned} p(c|X, \alpha, \rho, G, \sigma_0^2, \sigma_1^2) &= \int_\gamma \frac{p(c, X, \gamma|\alpha, G, \sigma_0^2, \sigma_1^2, \rho)}{p(X|\alpha, G, \sigma_0^2, \sigma_1^2, \rho)} d\gamma \\ &= \frac{p(c|\alpha, G, \rho)}{p(X|\alpha, G, \rho, \sigma_0^2, \sigma_1^2)} \int_\gamma p(X|p_\gamma, z(c)) p(\gamma|\sigma_0^2, \sigma_1^2) d\gamma. \quad (4) \end{aligned}$$

The numerator is given by Eq. 2. The denominator $p(X|...)$ is intractable–it requires summing over all possible combinations of c–so we turn to Gibbs Sampling. This sampler iterates through each individual c_{it}, sampling from it's current distribution conditional on the current values of all other c, which I denote as c_{-it}.

Let's simplify the integral. Notice that the full $\gamma_{z,1:T,r}$ is distributed as a multivariate normal with mean 0. The variance of a given γ_{ztr} is $\sigma_0^2 + t\sigma_1^2$ and the covariance between γ_{zt_1r} and γ_{zt_2r} is $\sigma_0^2 + min(t_1 - 1, t_2 - 1)\sigma_1^2$, yielding the $T \times T$ covariance matrix Σ_T for $\gamma_{z,1:T,r}$:

$$\Sigma_T = \begin{pmatrix} \sigma_0^2 & \sigma_0^2 & \sigma_0^2 & ... \\ \sigma_0^2 & \sigma_0^2 + \sigma_1^2 & \sigma_0^2 + \sigma_1^2 & ... \\ \sigma_0^2 & \sigma_0^2 + \sigma_1^2 & \sigma_0^2 + 2\sigma_1^2 & ... \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

Writing γ_z as a TR-length stacked vector with r varying fastest, $\gamma_z = (\gamma_{z,1,1}, \gamma_{z,1,2}, ..., \gamma_{z,1,R}, \gamma_{z,2,1}, ..., \gamma_{z,T,R})$, we can write the covariance matrix as $\Sigma_{TR} = \Sigma_T \otimes I_R$, with $\otimes$ the Kronecker product and I_R the $R \times R$ identity matrix. We can similarly stack the γ_z to form the ZTR-length vector γ, which is similarly distributed with $p(\gamma|z, \Sigma_{ZTR})$ a multivariate normal with mean 0 and covariance $\Sigma_{ZTR} = I_Z \otimes \Sigma_T \otimes I_R$.

Turning to the first multiplicand in the integral of Eq. 4, X is multinomial, which means that

$$\begin{aligned} p(X|p_\gamma, z, \Sigma_T) &= \prod_{t,z} \prod_{i|z_{it}=z} \frac{n_{it}!}{X_{it1}! \cdots X_{itR}!} p_{zt1}^{X_{it1}} \cdots p_{ztR}^{X_{itR}} \\ &= \prod_{t,z} \prod_{i|z_{it}=z} \frac{n_{it}!}{X_{it1}! \cdots X_{itR}!} exp\left\{ \gamma_{zt1} X_{it1} + \cdots + \gamma_{ztR} X_{itR} - n_{it} \log \sum_r exp\{\gamma_{ztr}\} \right\} \\ &= \left(\prod_{it} \frac{n_{it}!}{X_{it1}! \cdots X_{itR}!} \right) exp\left\{ \gamma' X^{(z)} - \sum_{zt} n_{zt} \log \sum_r exp\{\gamma_{ztr}\} \right\}, \end{aligned}$$

where $X^{(z)}$ is a ZTR-length vector constructed similarly to γ, and $X^{(z)}_{ztr} = \sum_{i|z_{it}=z} X_{itr}$ is the blocks' X aggregated to their respective clusters.

Including $p(\gamma)$, the full integral in Eq. 4 becomes

$$\int_\gamma p(X|\gamma, z, \Sigma_T)p(\gamma|z, \Sigma_T)d\gamma = \left(\prod_{it} \frac{n_{it}!}{X_{it1}! \cdots X_{itR}!}\right)(2\pi)^{-ZRT/2}|\Sigma_T|^{-ZR/2} \times \int_\gamma exp\left\{-\frac{1}{2}\gamma'\Sigma^{-1}_{ZTR}\gamma + \gamma' X^{(z)} - \sum_{zt} n_{zt} \log \sum_r exp\{\gamma_{ztr}\}\right\} d\gamma. \tag{5}$$

This is again not tractable, but can be approximated with its own multivariate normal distribution. Approximate the term in the exponent with a quadratic expansion:

$$\begin{aligned} f(\gamma) &= -\frac{1}{2}\gamma'\Sigma^{-1}_{ZTR}\gamma + \gamma' X^{(z)} - \sum_{zt} n_{zt} \log \sum_r exp\{\gamma_{ztr}\} \\ &\approx f(\hat{\gamma}) + J'_{\hat{\gamma}}(\gamma - \hat{\gamma}) + \frac{1}{2}(\gamma - \hat{\gamma})' H_{\hat{\gamma}}(\gamma - \hat{\gamma}) \\ &= f(\hat{\gamma}) - \frac{1}{2} J'_{\hat{\gamma}} H^{-1}_{\hat{\gamma}} J_{\hat{\gamma}} + \frac{1}{2}(\gamma - \hat{\gamma} + H^{-1}_{\hat{\gamma}} J_{\hat{\gamma}})' H_{\hat{\gamma}} (\gamma - \hat{\gamma} + H^{-1}_{\hat{\gamma}} J_{\hat{\gamma}}). \end{aligned} \tag{6}$$

where $\hat{\gamma}$ is the point about which we are approximating and $J_{\hat{\gamma}}$ and $H_{\hat{\gamma}}$ are the Jacobian and Hessian of f evaluated at $\hat{\gamma}$, respectively. The last step illustrates that this exponent is simply a constant term plus an unnormalized multivariate normal term with covariance $-H^{-1}_{\hat{\gamma}}$ and mean $\hat{\gamma} - H^{-1}_{\hat{\gamma}} J_{\hat{\gamma}}$. We can optimize $\hat{\gamma}$ to maximize f, strengthening this approximation.

With Eq. 6, we can calculate the normalization factor of the implied multivariate normal, and the integral is thus tractable:

$$\int exp\left\{f(\gamma)\right\} d\gamma \approx exp\left\{f(\hat{\gamma}) - \frac{1}{2} J'_{\hat{\gamma}} H^{-1}_{\hat{\gamma}} J_{\hat{\gamma}}\right\} \times (2\pi)^{ZRT/2} | - H_{\hat{\gamma}_T}|^{-1/2}$$

which yields the full integral in Eq. 5 as

$$\int_\gamma p(X|\gamma, z, \Sigma)p(\gamma|z, \Sigma)d\gamma \approx \left(\prod_{it} \frac{n_{it}!}{X_{it1}! \cdots X_{itR}!}\right) \times exp\left\{f(\hat{\gamma}(z, \Sigma_T)) - \frac{1}{2} J'_{\hat{\gamma}(z,\Sigma_T)} H^{-1}_{\hat{\gamma}(z,\Sigma_T)} J_{\hat{\gamma}(z,\Sigma_T)}\right\} \times | - H_{\hat{\gamma}(z,\Sigma_T)}|^{-1/2} |\Sigma_T|^{-ZR/2}$$

where I've added the parentheses to $\hat{\gamma}$ to emphasize that it depends on z and Σ_T. Combining these results and dropping terms that are constant with respect to z, we get the convenient approximation:

$$p(c|X, \alpha, G, \rho, \Sigma) \propto p(c|\alpha, G, \rho) \cdot exp\left\{f(\hat{\gamma}(z(c), \Sigma_T)) - \frac{1}{2} J'_{\hat{\gamma}} H^{-1}_{\hat{\gamma}} J_{\hat{\gamma}}\right\} \times | - H_{\hat{\gamma}(z(c),\Sigma_T)}|^{-1/2} |\Sigma_T|^{-ZR/2}.$$

The complete conditional for c_{it} is, finally,

$$p(c_{it}|c_{-it}, X, \alpha, G, \rho, \sigma_0^2, \sigma_1^2) \propto p(c_{it}|c_{-it}, \alpha, G, \rho) \cdot exp\left\{ f(\hat{\gamma}(z(c), \Sigma_T)) - \frac{1}{2} J'_{\hat{\gamma}} H^{-1}_{\hat{\gamma}} J_{\hat{\gamma}} \right\} \times \\ | - H_{\hat{\gamma}(z(c),\Sigma_T)}|^{-1/2} |\Sigma_T|^{-ZR/2}, \quad (7)$$

with the first term the ddCRP distribution given in Eq. 2. At each step, you can optimize for $\hat{\gamma}$.

The Gibbs Sampler proceeds by deleting a given c_{it}, calculating the clusters that would be produced by every possible connection (all of the block's neighbors, and itself in this time period and the previous one), calculating the approximation in Eq. 7 for every possible connection, and sampling c_{it} with those probabilities.

A.2 The posterior of γ

Once we have sampled a set of c with γ marginalized out, we can calculate the posterior distribution of γ (which itself yields the more intuitive p) as

$$\begin{aligned} p(\gamma|X, z(c), \Sigma) =& \frac{p(X|\gamma, z(c))p(\gamma|z(c))}{p(X|z(c))} \\ \propto& exp\{f(\gamma; X, z(c))\}, \end{aligned}$$

in which f is the function defined in Eq. 6, with dependence on X and z made explicit. In this paper, the clusters are broad enough that X is large, meaning this distribution sharply peaks at the family $\gamma_{zrt} = \log(X^{(z)}_{zrt}) - \log(n_{zt}) + C$, so that the distribution of p_γ collapses to the observed proportions. For that reason, I simply use the observed proportions of each cluster; this also makes the decomposition in Section 4.4 an accounting rule rather than the clunkier "expected value of the posterior distribution of the decomposition".

B Clustering Results at Other Scales

The results presented in this paper focus on the scale of $\alpha = exp\{-1000\}$. Here, I discuss the choice of that scale and what the clusters look like at other scales, each of $\alpha = exp\{-50, -100, -200, -400, -1000, -2000, -4000, -8000\}$.

First, let's consider the size of these clusters. Figure A1 presents the number of clusters identified for each scale. With 18,872 blocks, the cluster sizes range from 18,872/5.5 = 3,431 blocks for $\alpha = exp\{-8000\}$ to 18,872/104 = 181 blocks for $\alpha = exp\{-50\}$.

[Figure A1 about here.]

Given the different cluster sizes, the results vary in the types of boundaries they identify and how those clusters smooth the differences among blocks. Figure A2 presents the map of North Philadelphia clustered at different scales. Notice that for $\alpha = \{-200\}$, the gentrified region of White blocks appears in the center of the map (at Temple University). This region was too small to be represented at larger scale. The boundaries at the small scales are somewhat weaker, in that they divide clusters that are less different; the single Black region at larger scales was divided into two regions with slightly different compositions.

Most importantly, though, the model at $\alpha = \{-200\}$ identified a new, diverse cluster at the boundary of the Hispanic and White cluster in 2010. Rather than stating that the White cluster spread northward, this model created a new cluster which changed internally. This is the danger of using a scale that creates clusters of the same size as the typical boundary movement: rather than realizing that change occurred due to a larger cluster shifting over time, the model will create small new diverse clusters that change internally.

[Figure A2 about here.]

Let's quantify the notion that the boundaries are stronger at the larger scales. I measure the strength of a boundary as the euclidean distance between the 8-dimensional proportion vector p_z (one dimension for each ethnoracial group) for two neighboring clusters z, z', normalized: $strength = ||\vec{p}_z - \vec{p}_{z'}||/\sqrt{2}$, averaged over all pairs of neighboring blocks with different cluster assignments. For two clusters that were each 100% of a different race, the magnitude of the difference in p_z–the numerator of the strength calculation–would be $\sqrt{2}$, so the normalization places strength on a 0-1 scale: 0 if all neighboring clusters have the same p_z, 1 if neighboring clusters are each homogenous of a different race. Figure A3 presents the boundary strengths for different α; we see a steady increase in strength as the α become smaller and thus the clusters larger.

[Figure A3 about here.]

The ddCRP model assumes that neighboring clusters have independent ethnoracial compositions, $\vec{p}_z$. It doesn't model, for example, White clusters tending to neighbor other White clusters. This might still happen in the observed data, and be produced in the results even though the model assumes it to not exist. In Figure A2, this in fact seems to happen at small scales, as a number of the small clusters have similar ethnoracial compositions to their neighbors. In the style of a residual-analysis, we can evaluate the assumption by measuring if the p_z of identified neighboring clusters are, in fact, more correlated than would occur randomly. I test this by taking the observed clusters' proportions and randomly reassigning them among the clusters,

then calculating the simulated boundary strength of the random clusters. If the observed neighboring clusters were truly independent, their boundary strength would fall in this simulated distribution.[5] If the observed neighboring clusters instead had correlated p_z, the observed boundary strength would be weaker, as neighboring clusters would have smaller differences in p_z than clusters sampled at random. The rectangles in Figure A3 show the results from reshuffling the $\vec{p_z}$ for each of the 1600 samples for each α; the rectangles reach from the 2.5th to the 97.5th percentile of the simulated boundary strengths. Notice that the larger values of α (with smaller clusters) exhibit much weaker boundaries than the random simulations, indicating that at this level neighboring clusters are more similar to each other than the model assumes. Only for $\alpha = exp\{-400\}$ and smaller–especially smaller–do the clusters exhibit neighboring independence. The assumptions of the model I am fitting here are most representative of the dynamics at large cluster sizes.

Figure A4 presents the proportion of blocks that change cluster assignments between 2000 and 2010, weighted by average household population $\bar{n}_i$ (and thus presents the proportion of households that change clusters). Many more households change clusters at smaller scales; mostly because with more clusters there are many more boundaries that can move. At the larger scales, fewer households–though still 6.6% of households for $\alpha = exp\{-1000\}$–change cluster. However, the boundaries are much stronger, so the households that do change clusters are experiencing more drastic change.

[Figure A4 about here.]

With all of this evidence, I argue that the results for larger scales such as $\alpha = exp\{-1000\}$ are both conservative and best representative of the true dynamic. Larger scales are conservative because the boundaries are stronger, less noisy, and fewer. Because of this, a block changing clusters is a significant event. Larger scales are best representative of the spatial change dynamic because they do not overfit diverse clusters at boundaries of change, and the results best satisfy the assumptions of the model. Smaller scales may be desired for substantive reasons for a given study, but the population change of Philadelphia can be well-described by tectonic shifts.

C Decomposition for Philadelphia at multiple scales

The results for the decomposition across values of α are generally consistent, though with stronger evidence for the importance of boundary shifting at larger scales (and

[5] Actually, the boundaries that we identify should be stronger than this simulation, as the algorithm would tend to group two neighboring clusters with randomly similar $\vec{p_z}$ into the same cluster, thus undercounting weak boundaries. We do notice that at the larger scales the results are above the center of the random values.

smaller α). This corroborates the 'characteristic scale' of the model, as the boundaries become more representative at larger scales (see Appendix B). Figures A5, A6, and A7 present the results for all, increasingly-White, and increasingly-non-White blocks, respectively.

[Figure A5 about here.]

[Figure A6 about here.]

[Figure A7 about here.]

List of Figures

1 South Philadelphia household race and ethnicity aggregated to the block, tract, and cluster level, 2000 & 2010. Units are shaded using a weighted average of the legend colors on the RGB scale based on their Census household composition; gray blocks have no Census households. 32

2 A toy example of a single rectangular tract exhibiting two types of spatial change, and the tract-level averaged population. Column A presents a boundary that remains stationary, but with the internal composition of the White cluster becoming more Black. Column B presents a boundary between an entirely Black and an entirely White region which moves over time; each cluster remains internally homogenous. Both tracts move from 75% White-25% Black to 50-50 to 25-75. 33

3 A sample cluster assignment of the ddCRP for one time period and two time periods. Arrows represent block assignments c, while the color represents the cluster assignments z. 34

4 A sample cluster assignment produced by $\alpha = exp\{-1000\}$ for all of Philadelphia. 35

5 The relative proportions of each race and ethnicity in Philadelphia households for blocks in 2000 and 2010 (N = 18,872). The 45° line represents the points where racial proportions would remain the same. 36

6 The changes in clusters' sizes and ethnoracial composition for a single cluster realization. Sizes of the dots are the average number of households in the cluster, $(n_{z,t=1} + n_{z,t=2})/2$. N(clusters) = 12. (A) The proportionate change in clusters' sizes, measured in blocks, as a function of the proportion of each ethnoracial group in 2000. Fitted lines are from univariate regressions weighted by the average number of households in each cluster, $\bar{nb}_z$. (B) Clusters' ethnoracial proportions in 2000 vs. 2010. Clusters along the 45-degree line experienced no change in proportions. 37

7 The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results with $\alpha = exp\{-1000\}$ and tracts. Columns are components of the decomposition, and error bars are 2.5th and 97.5th percentiles of the posterior samples (2.5-97.5 credible intervals). $D^{(n)}$ represents the change due to within-block household numbers, $D^{(z)}$ due to cluster changes, and $D^{(p)}$ due to within-cluster level changes. Tracts fix $D^{(z)}$ at zero. N = 18,872 blocks, 1800 samples from posterior cluster assignments. . 38

8 The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for clusters with $\alpha = exp\{-1000\}$ and tracts, limited to only blocks that experienced an increase in the proportion White. Columns are components of the decomposition. Tracts fix $D^{(z)}$ at zero. N = 3,829 blocks. 39

9 The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for clusters with $\alpha = exp\{-1000\}$ and tracts, limited to only blocks that experience an increase in the proportion non-White. Columns are components of the decomposition. Tracts fix $D^{(z)}$ at zero. N = 7,413 blocks. 40

A1 The number of clusters produced for samples from six chains of cluster assignments for each value of α, jittered. N(blocks) = 18,872. 41

A2 Cluster results for a single sample of the chain for $\alpha = exp\{-1000, -400, -200\}$, for the same section of North Philadelphia pictured in Figure 1. Clusters are shaded using a weighted average of the legend colors on the RGB scale based on their Census household composition. . . 42

A3 Mean boundary strength for realized cluster assignments, measured as the mean valwue of $||\vec{p}_z - \vec{p}_{z'}||/\sqrt{2}$, where z, z' are cluster assignments of neighboring block pairs with different cluster assignments. Shaded rectangles represent 2.5th and 97.5th percentiles of the boundary strength achieved by randomly reassigning observed $\vec{p}_z$ among clusters. Observations outside of the rectangles are evidence that neighboring clusters are not independent, as the model assumes. 43

A4 The proportion of households that change cluster assignments between 2000 and 2010, jittered, for each α. Because number of households for each block can change, this is a weighted proportion of blocks that change, weighted by blocks' average numbers of households, $\bar{n}_i$. 44

A5 The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results for each α and tracts. Columns are components of the decomposition, and error bars are 2.5-97.5 credible intervals. $D^{(n)}$ represents the change due to within-block household numbers, $D^{(z)}$ due to cluster changes, and $D^{(p)}$ due to within-cluster level changes. Tracts fix $D^{(z)}$ at zero. N = 18,872 blocks. 45

A6 The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results for each α and tracts, limited to only blocks that experienced an increase in the proportion White. Columns are components of the decomposition, and error bars are 2.5-97.5 credible intervals. Tracts fix $D^{(z)}$ at zero. N = 3,829 blocks. 46

A7 The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results for each α and tracts, limited to only blocks that experience an increase in the proportion non-White. Columns are components of the decomposition. Columns are components of the decomposition, and error bars are 2.5-97.5 credible intervals. Tracts fix $D^{(z)}$ at zero. N = 7,413 blocks. 47

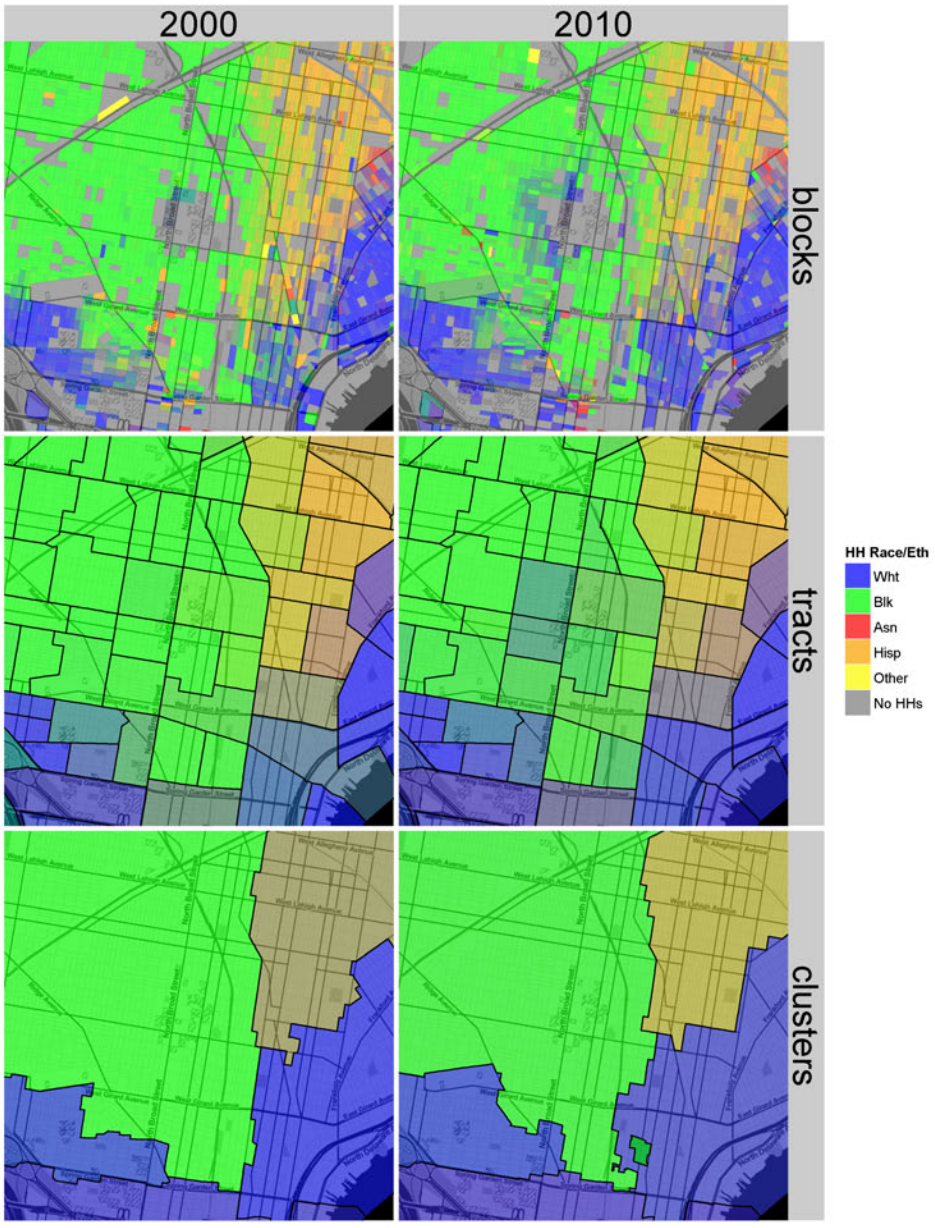

Figure 1: South Philadelphia household race and ethnicity aggregated to the block, tract, and cluster level, 2000 & 2010. Units are shaded using a weighted average of the legend colors on the RGB scale based on their Census household composition; gray blocks have no Census households.

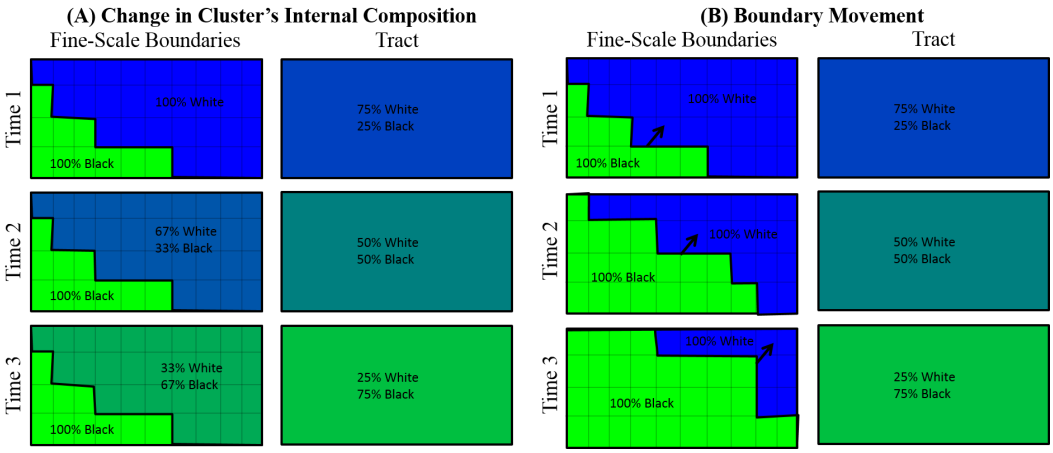

Figure 2: A toy example of a single rectangular tract exhibiting two types of spatial change, and the tract-level averaged population. Column A presents a boundary that remains stationary, but with the internal composition of the White cluster becoming more Black. Column B presents a boundary between an entirely Black and an entirely White region which moves over time; each cluster remains internally homogenous. Both tracts move from 75% White-25% Black to 50-50 to 25-75.

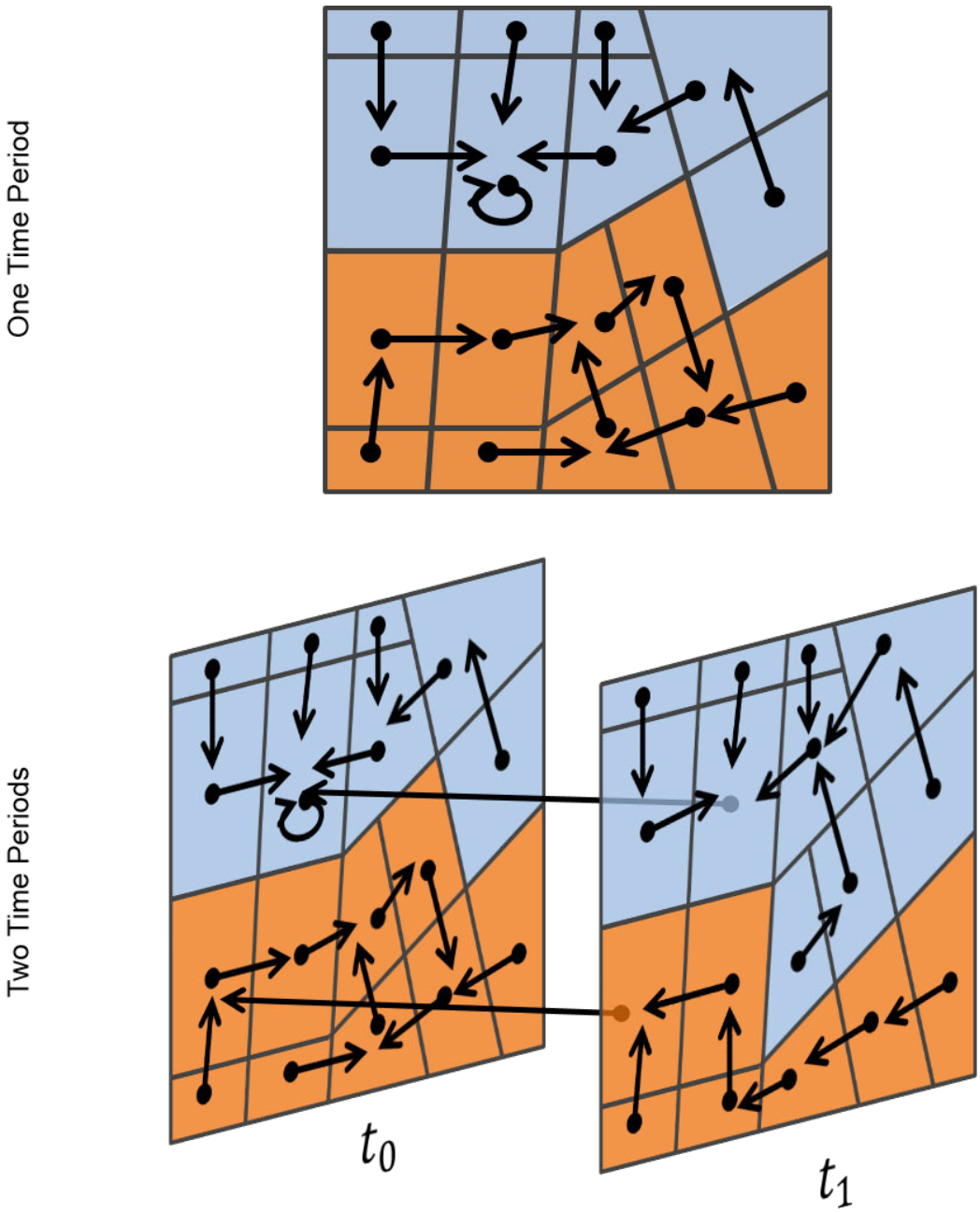

Figure 3: A sample cluster assignment of the ddCRP for one time period and two time periods. Arrows represent block assignments c, while the color represents the cluster assignments z.

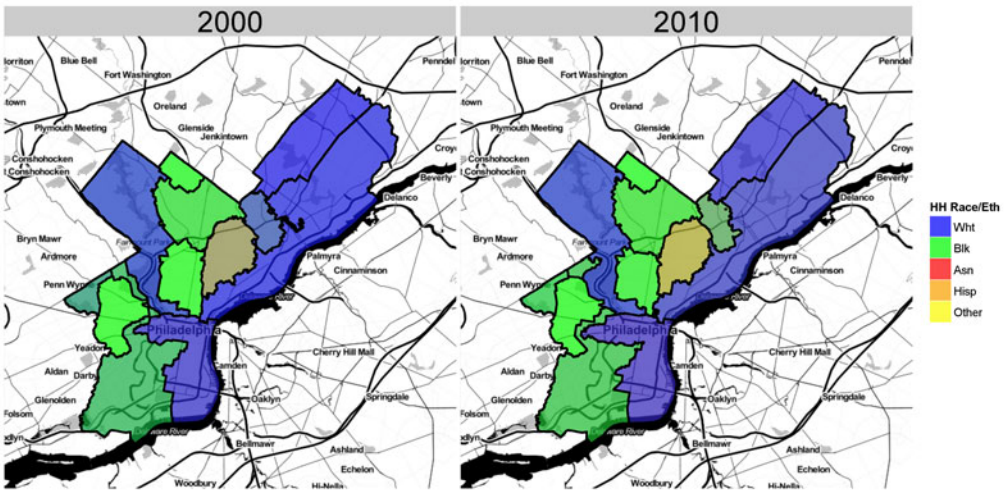

Figure 4: A sample cluster assignment produced by $\alpha = exp\{-1000\}$ for all of Philadelphia.

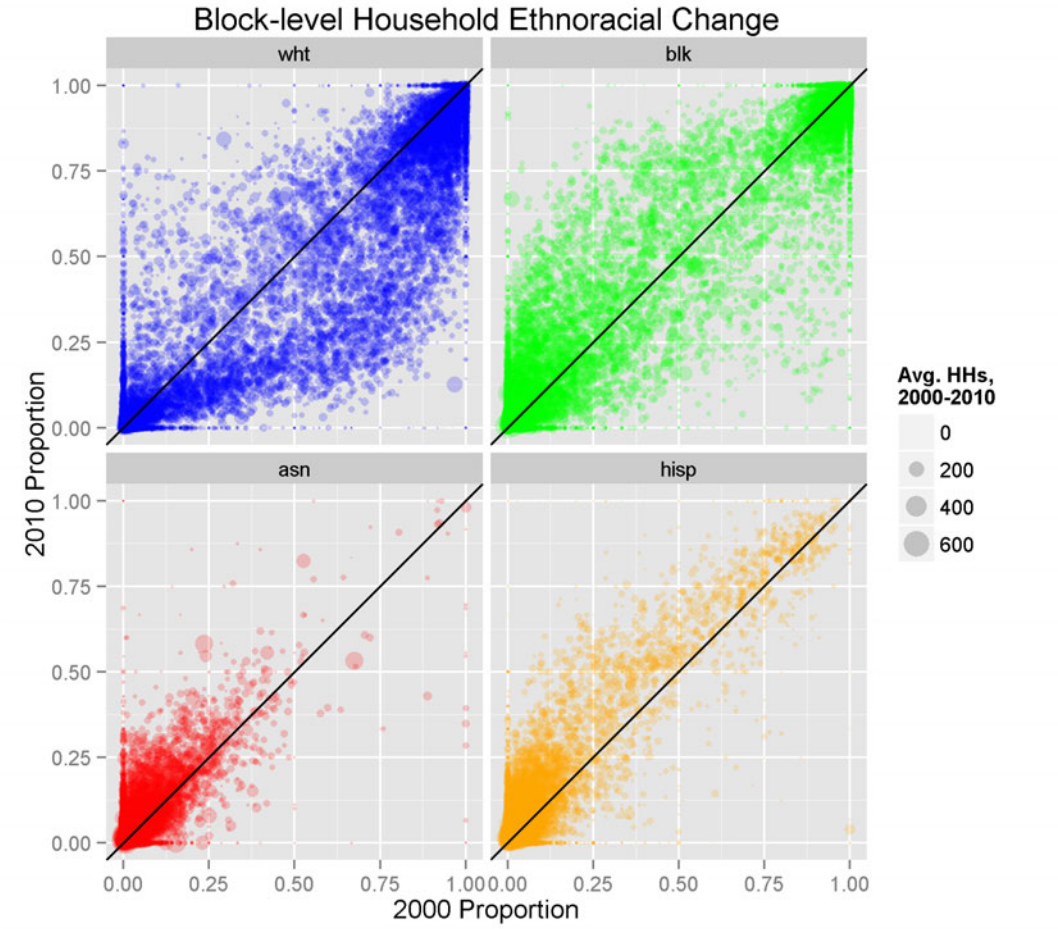

Figure 5: The relative proportions of each race and ethnicity in Philadelphia households for blocks in 2000 and 2010 (N = 18,872). The 45° line represents the points where racial proportions would remain the same.

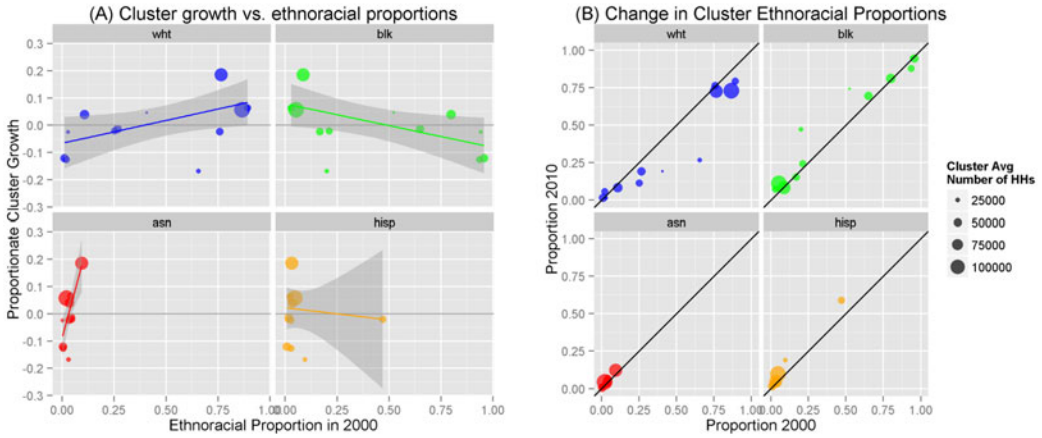

Figure 6: The changes in clusters' sizes and ethnoracial composition for a single cluster realization. Sizes of the dots are the average number of households in the cluster, $(n_{z,t=1} + n_{z,t=2})/2$. N(clusters) = 12. (A) The proportionate change in clusters' sizes, measured in blocks, as a function of the proportion of each ethnoracial group in 2000. Fitted lines are from univariate regressions weighted by the average number of households in each cluster, $\bar{nb}_z$. (B) Clusters' ethnoracial proportions in 2000 vs. 2010. Clusters along the 45-degree line experienced no change in proportions.

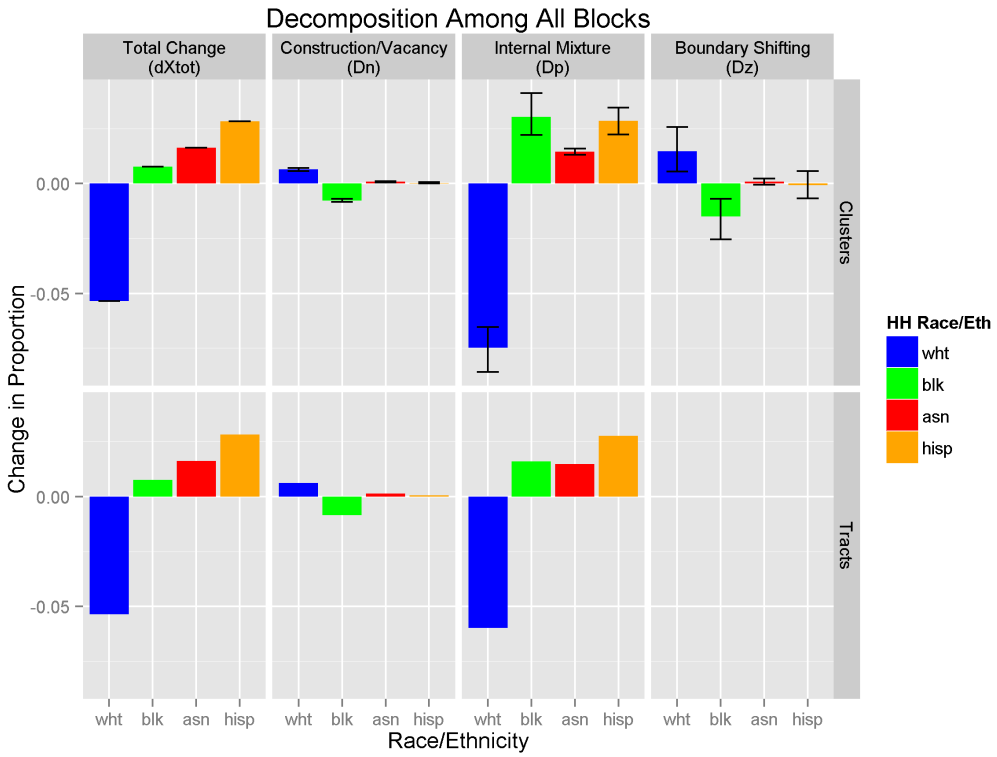

Figure 7: The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results with $\alpha = exp\{-1000\}$ and tracts. Columns are components of the decomposition, and error bars are 2.5th and 97.5th percentiles of the posterior samples (2.5-97.5 credible intervals). $D^{(n)}$ represents the change due to within-block household numbers, $D^{(z)}$ due to cluster changes, and $D^{(p)}$ due to within-cluster level changes. Tracts fix $D^{(z)}$ at zero. N = 18,872 blocks, 1800 samples from posterior cluster assignments.

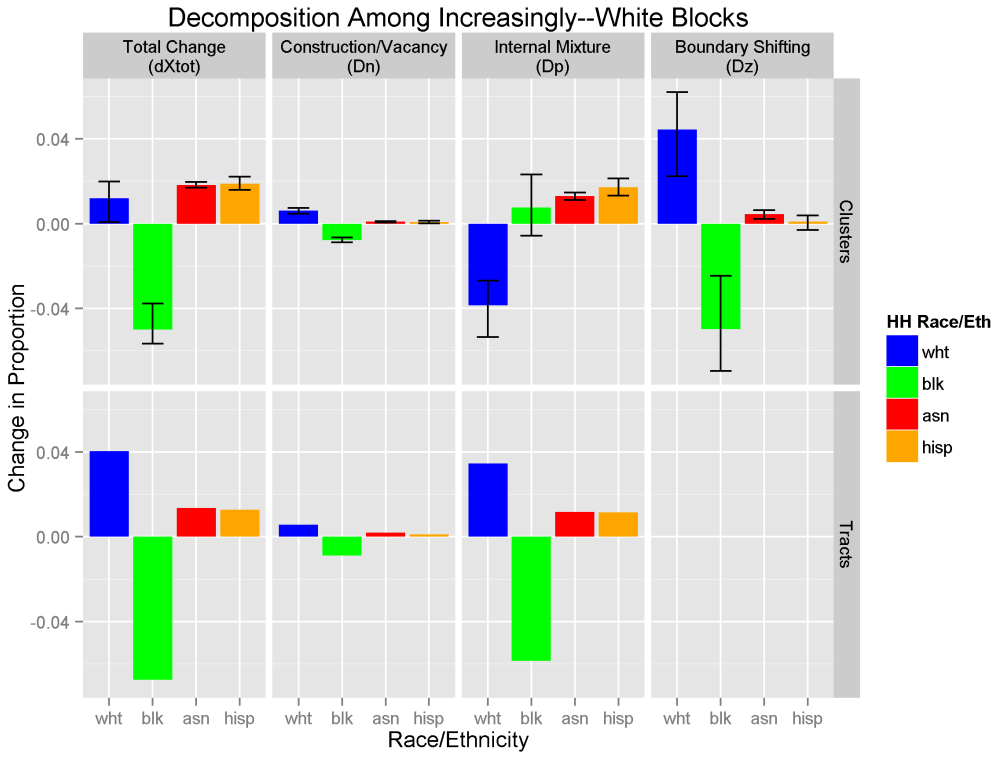

Figure 8: The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for clusters with $\alpha = exp\{-1000\}$ and tracts, limited to only blocks that experienced an increase in the proportion White. Columns are components of the decomposition. Tracts fix $D^{(z)}$ at zero. N = 3,829 blocks.

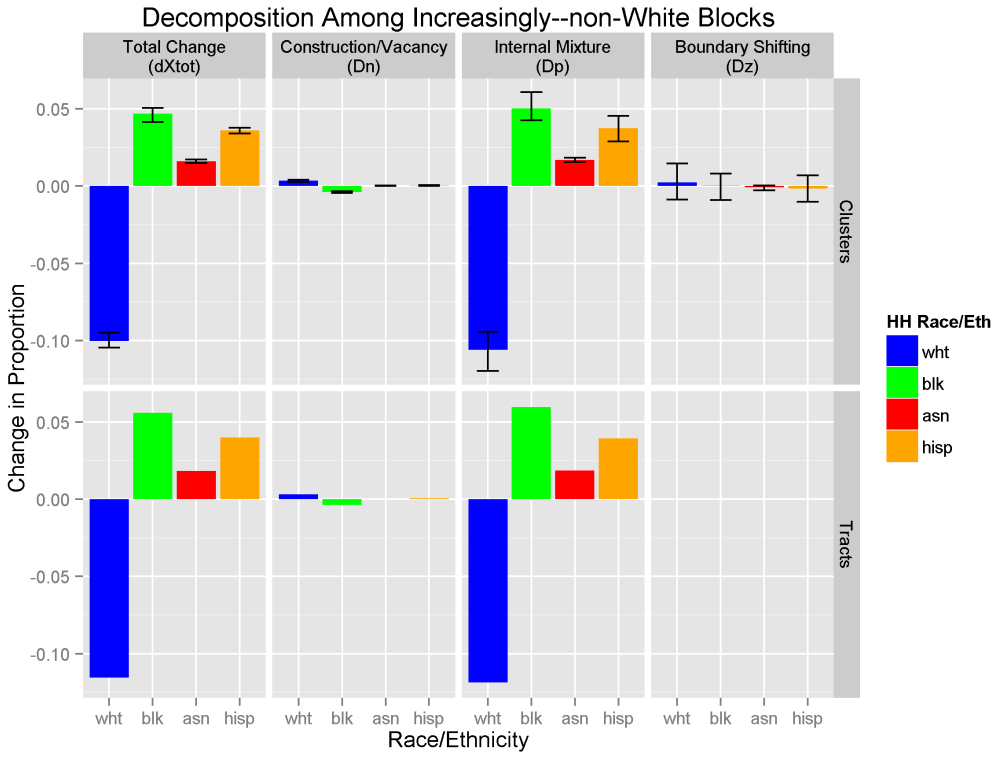

Figure 9: The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for clusters with $\alpha = exp\{-1000\}$ and tracts, limited to only blocks that experience an increase in the proportion non-White. Columns are components of the decomposition. Tracts fix $D^{(z)}$ at zero. N = 7,413 blocks.

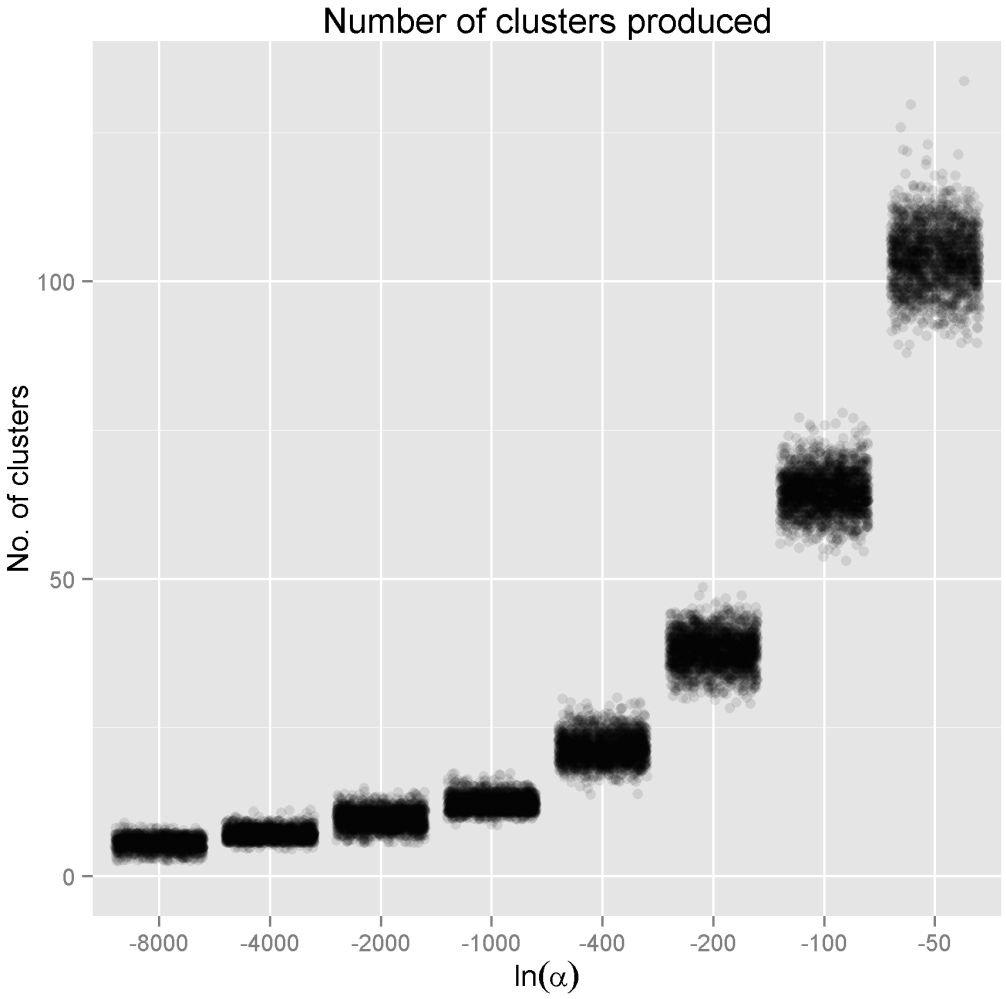

Figure A1: The number of clusters produced for samples from six chains of cluster assignments for each value of α, jittered. N(blocks) = 18,872.

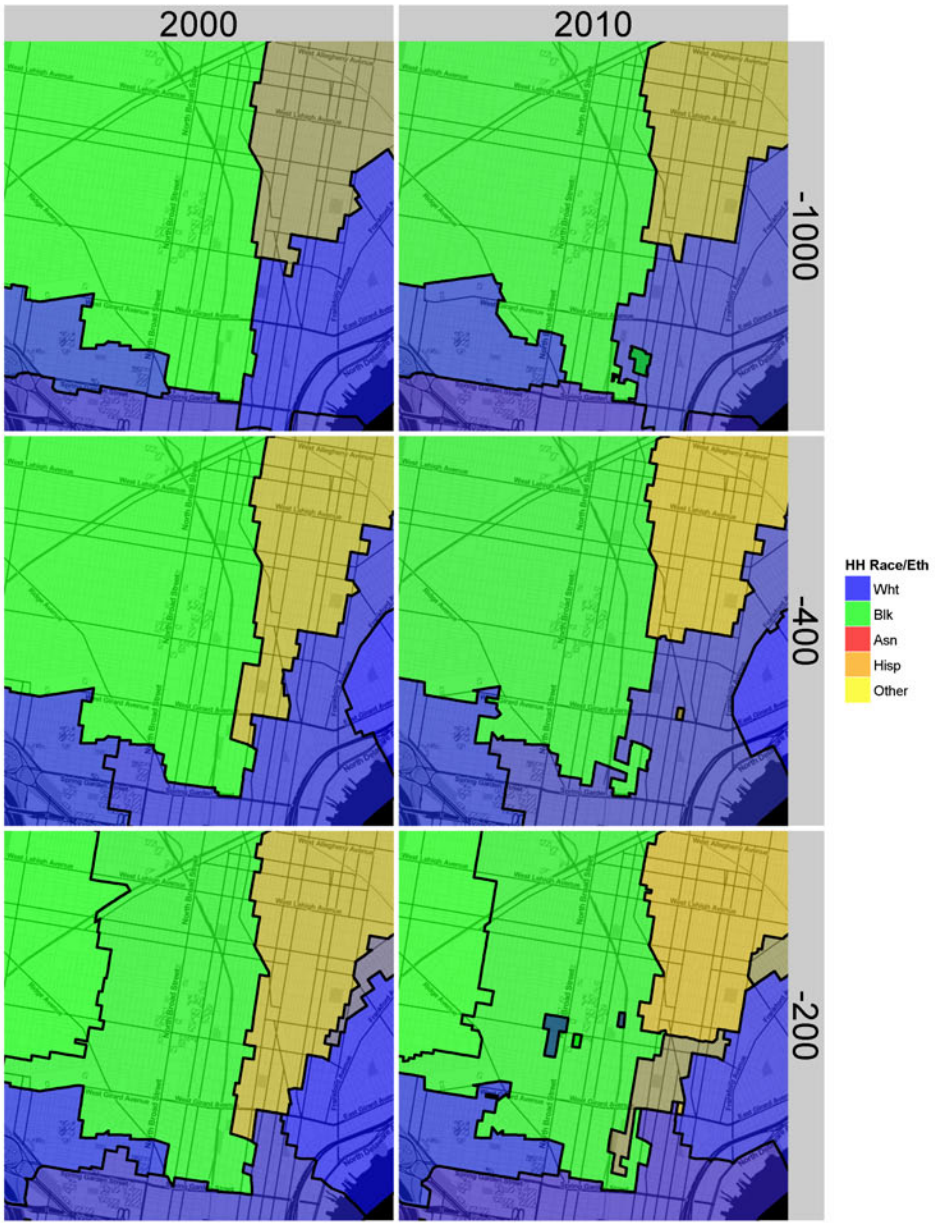

Figure A2: Cluster results for a single sample of the chain for $\alpha = exp\{-1000, -400, -200\}$, for the same section of North Philadelphia pictured in Figure 1. Clusters are shaded using a weighted average of the legend colors on the RGB scale based on their Census household composition.

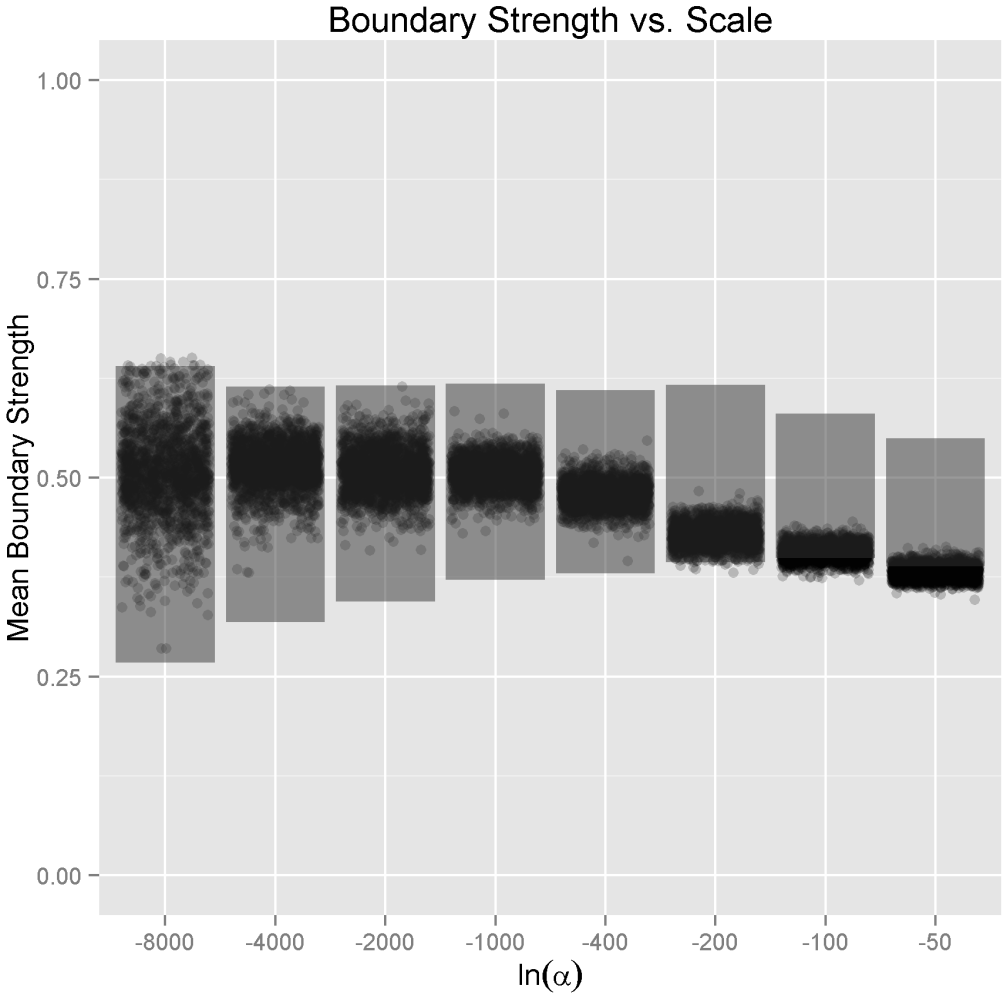

Figure A3: Mean boundary strength for realized cluster assignments, measured as the mean valwue of $||\vec{p}_z - \vec{p}_{z'}||/\sqrt{2}$, where z, z' are cluster assignments of neighboring block pairs with different cluster assignments. Shaded rectangles represent 2.5th and 97.5th percentiles of the boundary strength achieved by randomly reassigning observed $\vec{p}_z$ among clusters. Observations outside of the rectangles are evidence that neighboring clusters are not independent, as the model assumes.

Figure A4: The proportion of households that change cluster assignments between 2000 and 2010, jittered, for each α. Because number of households for each block can change, this is a weighted proportion of blocks that change, weighted by blocks' average numbers of households, $\bar{n}_i$.

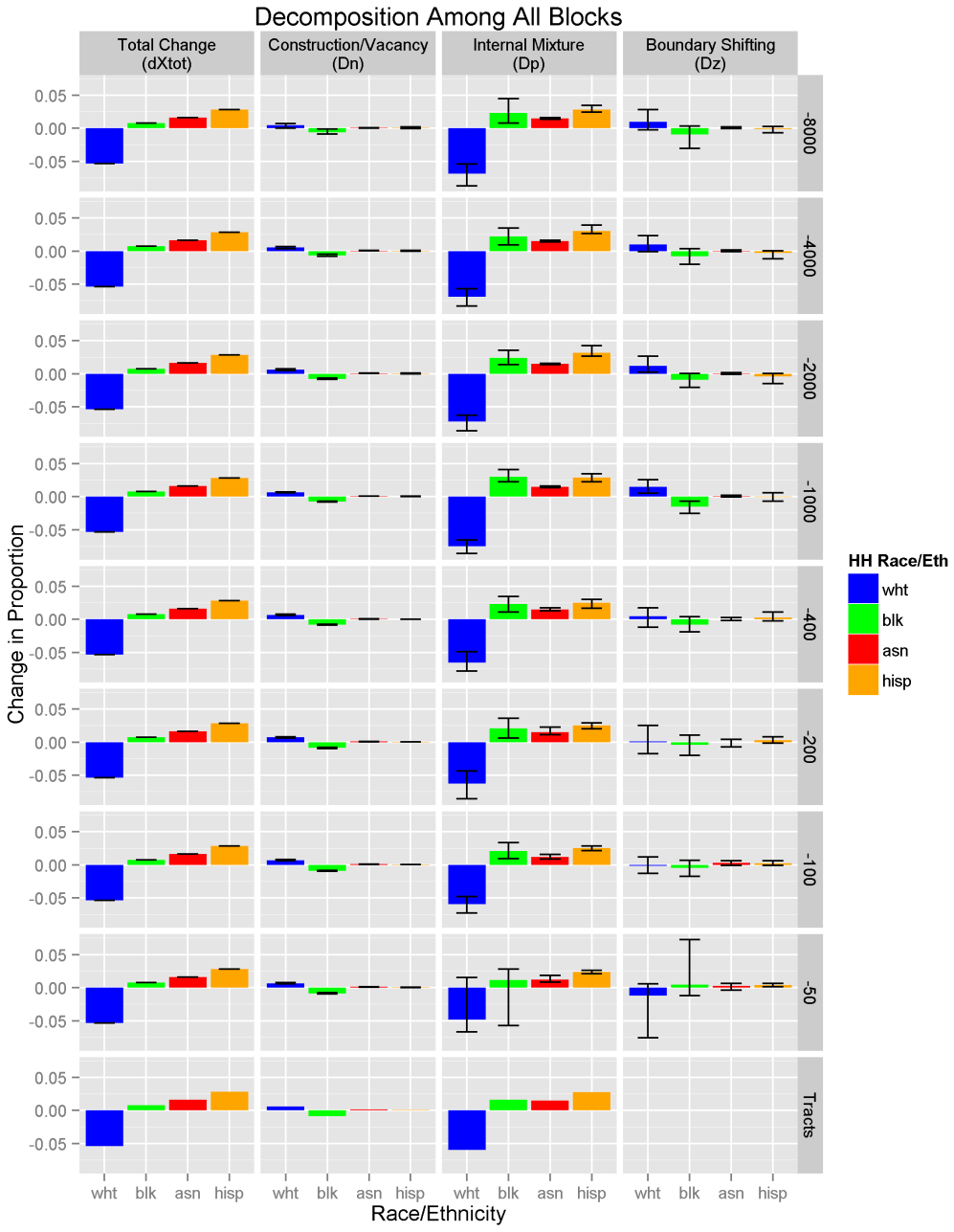

Figure A5: The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results for each α and tracts. Columns are components of the decomposition, and error bars are 2.5-97.5 credible intervals. $D^{(n)}$ represents the change due to within-block household numbers, $D^{(z)}$ due to cluster changes, and $D^{(p)}$ due to within-cluster level changes. Tracts fix $D^{(z)}$ at zero. N = 18,872 blocks.

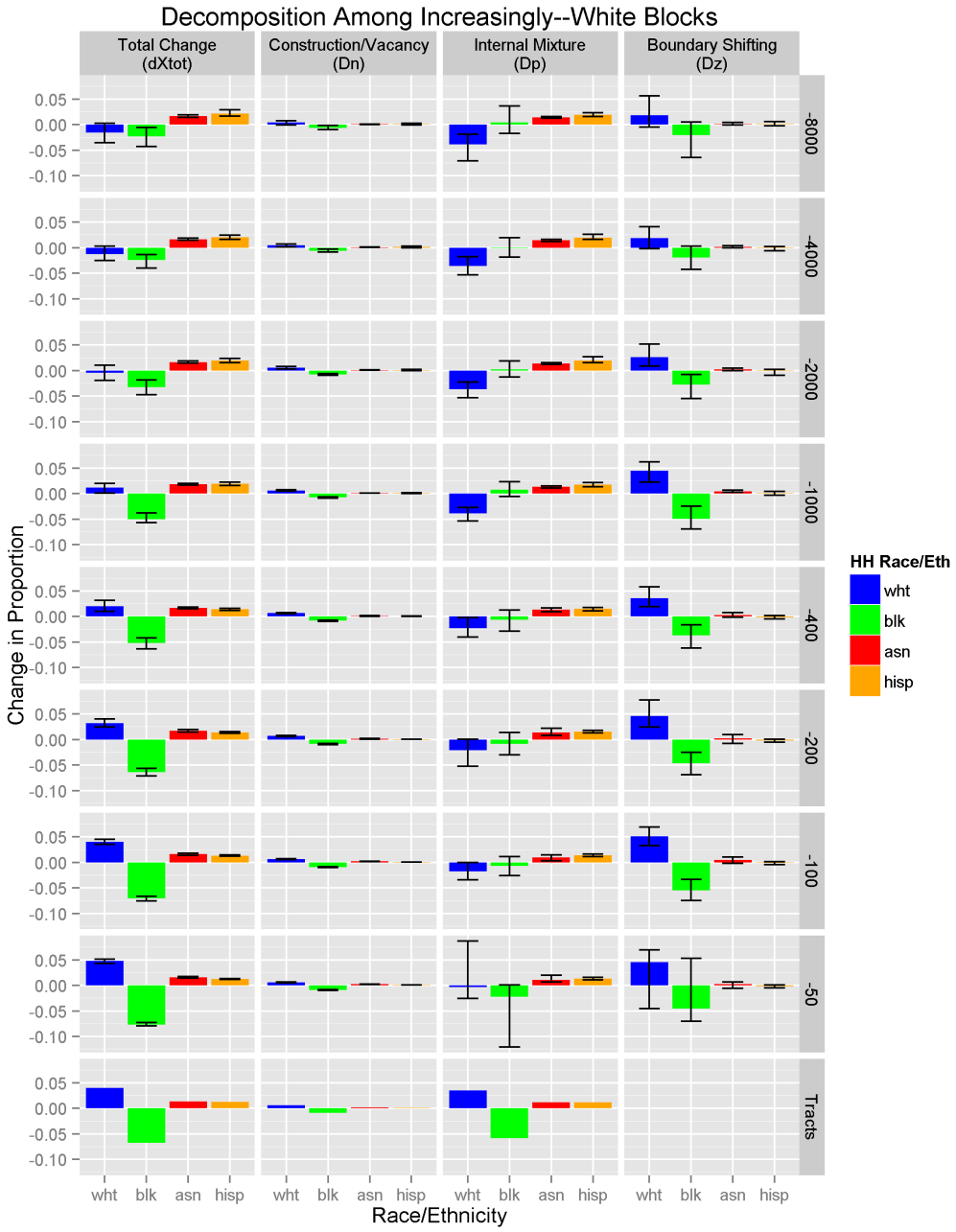

Figure A6: The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results for each α and tracts, limited to only blocks that experienced an increase in the proportion White. Columns are components of the decomposition, and error bars are 2.5-97.5 credible intervals. Tracts fix $D^{(z)}$ at zero. N = 3,829 blocks.

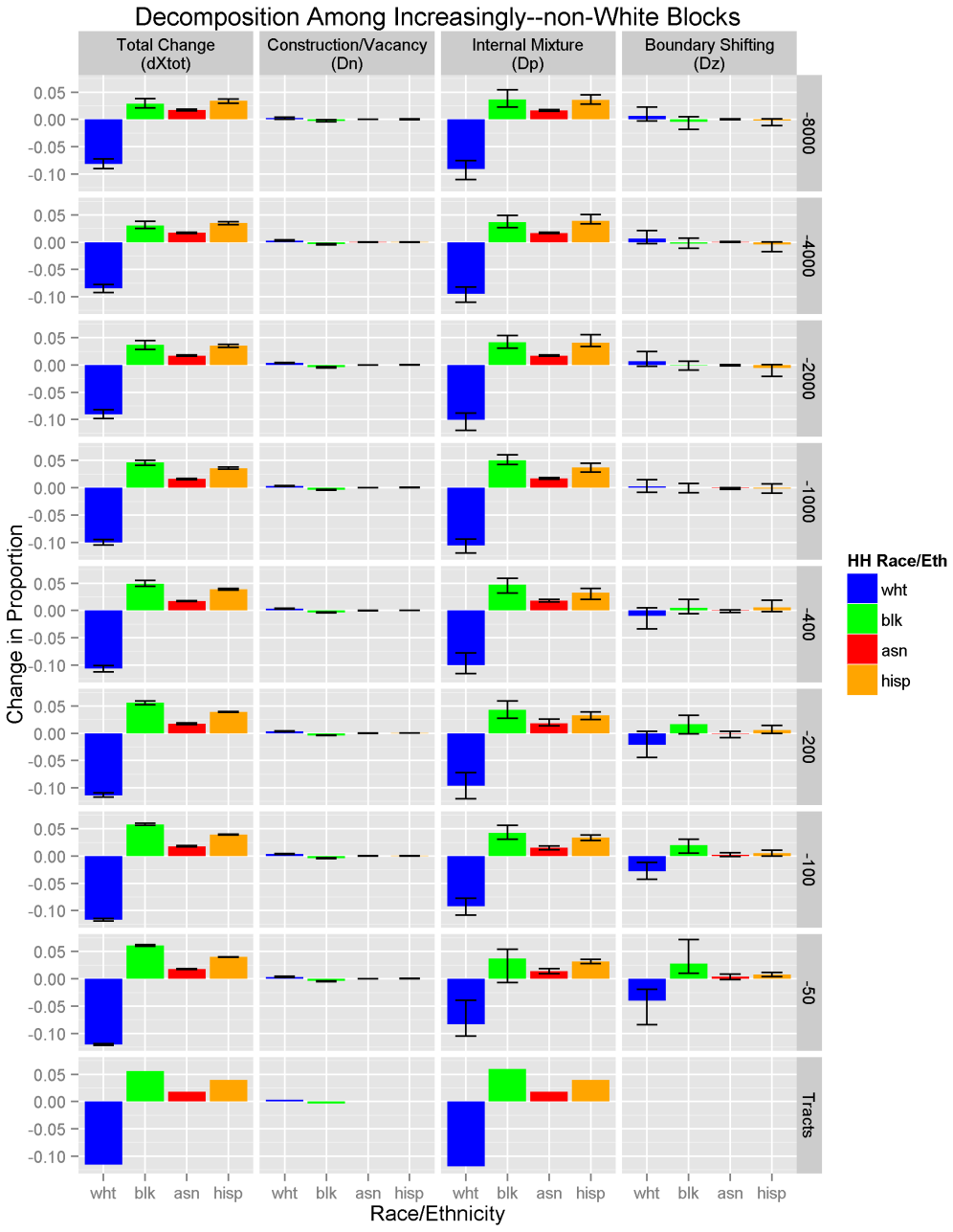

Figure A7: The cluster-based decomposition of Household ethnoracial change in Philadelphia from 2000-2010 for cluster results for each α and tracts, limited to only blocks that experience an increase in the proportion non-White. Columns are components of the decomposition. Columns are components of the decomposition, and error bars are 2.5-97.5 credible intervals. Tracts fix $D^{(z)}$ at zero. N = 7,413 blocks.