Population density and diversity: an update to Schelling’s model

Alexander Tsiatas
University of California, San Diego

Abstract
Motivated by recent increases in population density and changes in urban development patterns, I give an update to Schelling’s landmark spatial segregation model. Inspired by the rapid construction of high-rise condominiums and apartments, I show a correlation between population density and racial diversity in San Diego neighborhoods. Using this as a basis, I give a modified high-density version of Schelling’s model which yields heterogeneous neighborhoods under the reasonable assumption that each person wants at least a few of its neighbors to share its type.

1 Introduction
In the 1970’s, social scientists such as Thomas Schelling were interested in the changing demographics that occurred in American cities in the previous decades of the 20th century. For example, social scientists noticed that many urban neighborhoods were becoming more racially homogeneous; for example, formerly mixed-race neighborhoods developing very high white or African American majorities. The development of white suburbs, Hispanic barrios, and African American inner-city neighborhoods is just a selection of phenomena that emerged during that time period.

For one example of this demographic shift, examine the data in Fig. 1 from a study of the racial composition of Chicago in 1940 and 1960 [3]. Note that in general, neighborhoods to the south of downtown became predominantly African American, and many neighborhoods to the north of downtown became predominantly white. Over those 20 years, much of Chicago homogenized into neighborhoods where one race represented over 75% of the population.
Undoubtedly, there are myriad reasons for this homogenization of American urban neighborhoods, as the human, socioeconomic, and political factors are exceedingly complex. To look at demographics completely independent of all these factors would reveal only limited understanding of human behavior. However, Schelling surmised that perhaps at least part of the reason for this homogenization was inherent in nature. Schelling gave a mathematical model [1, 5, 6] that showed that even if people have only loose restrictions on where they choose to live, homogeneous neighborhoods will naturally develop, even when there is a more diverse arrangement that would indeed satisfy those restrictions.

Schelling’s model of neighborhood development relied on modeling the world as a 2-dimensional grid, with one entity (which could be a family, person, household, etc.) at each grid point. This fixes the population density of the world, and this may be an appropriate model for the early part of the twentieth century before high-rise condominium and apartment buildings were conceived. Perhaps it is now time to revisit Schelling’s model and take into account the increases in population density that have occurred since then. In this paper, I will do exactly that, using simulation and census data to suggest that perhaps there is a correlation between population and racial diversity.
2 Schelling’s model

Schelling sought out to model an American city and its demographics. In his model [1, 5, 6], there are two types of people: red and blue. This can be considered analogous to modeling two different races, classes, or any other demographic characteristic. The world is an infinite 2-dimensional grid, and there is at most one person at each grid point. Thus, each person has 8 neighbors: north, east, south, west, and four diagonals. We say that a person is satisfied if a certain number \( n \) of its neighbors share its type. We will vary \( n \) when simulating the model. Note that while this is certainly a strong assumption if \( n = 8 \), we will concern ourselves with much weaker notions of satisfiability: \( n = 3 \) and \( n = 4 \).

The world is initialized randomly: each grid point is either a home to a red person or a blue person, or it is left vacant. This is the configuration of the world at time \( t = 0 \), and it can contain both satisfied and unsatisfied people. Then, during each discrete time interval \( t = 1, 2, 3, \ldots \), one unsatisfied person is selected at random (if such a person exists), and is moved randomly to a location where it will be satisfied (if such a place exists). This process is iterated a desired number of steps, until convergence, until there are no more unsatisfied people, or there are no vacant positions where an unsatisfied person can be moved.

Much of the analysis of Schelling’s model has been empirical in nature. While the model is mathematically well-defined, not very much headway has been made in rigorous mathematical analysis of the model [1, 3, 8, 9]. While there has been some limited mathematical analysis, most of the work done with this model has been based on simulation. One such framework for simulation can be found in an applet by Sean Luke [2].

When simulating Schelling’s model, there are a few decisions to be made. First of all, Schelling’s model invoked an infinite graph; a simulation will only want to examine a finite grid. Thus, the simulations in [1] generated by [2] studied only a 150-by-150 grid. This is both large enough to capture temporal trends and small enough to be computationally feasible. Of course, in a finite grid, the nodes on the boundary do not have their full complement of neighbors. To get around this, [2] simply lets the grid wrap around: nodes on the left boundary now have neighbors on the right boundary. This makes the world a torus.

Another question to consider when conducting a simulation is the initial distribution of the number of red people, blue people, and vacant spots. Perhaps this decision could be made based on the actual percentage of available housing in a certain neighborhood. In any case, these simulations [1] ini-
tialize the grid with 10,000 red people, 10,000 blue people, and 2500 vacant spots.

Finally, to simulate Schelling’s model, we also need to select an appropriate value for $n$, the number of same-type neighbors that a person needs to have in order to be satisfied. As discussed earlier, setting $n$ to be too high is unrealistic, but simulations will show interesting behavior even for smaller $n$. First, we will see what happens when $n = 3$. This is not a very strong restriction; each node simply wants a few of its own kind as neighbors to be satisfied. No node is demanding to be part of the majority in an immediate neighborhood. I will note that there is an arrangement of blue and red people that satisfies this condition and still maintains heterogeneity on the entire grid: simply alternating red and blue people. But we will see that this does not arise naturally starting from a random starting configuration.

The results in Fig. 2 show that when $n = 3$, simulating Schelling’s model results in a configuration where all agents are satisfied, and they form relatively large homogeneous areas. These all-blue and all-red areas are interlocking with many tendrils, but they are still large homogeneous areas.

![Figure 2: Two simulations of the Schelling model with $n = 3$.](image)

When $n = 4$, the results shift to even more drastically homogeneous areas. In Fig. 3, it is apparent that after 800 time steps, there are two large, completely separate, and disjoint homogeneous areas, even though the starting configuration was entirely random. So it is apparent that under this model of human behavior, racial segregation and a lack of diversity can
occur independent of outside influences. Surely there are many factors that led to the homogenization of urban neighborhoods in the twentieth century, but Schelling gave some insight into the notion that perhaps this will occur naturally.

Figure 3: A simulation of the Schelling model with $n = 4$.

I will note that these simulations generated by [2] use a modified version of Schelling’s model. In these simulations, when an agent is unsatisfied and selected to move, it will select an available position at random, without first checking to see if that position will satisfy it. Later on, I perform my own simulations without this modification.

3 Effects of density on diversity

One aspect of Schelling’s model that I will explore deeper is the population density of the artificial world. The model placed at most one agent at each grid point, effectively limiting the population density. But many American urban neighborhoods are increasing in density, especially with the introduction of high-rise condominium and apartment development. Nowadays, it is likely that an urban resident has several hundred immediate neighbors, including many in the same physical building. One may ask the question: what demographic trends arise when density is increased?

On the surface, it appears as if some neighborhoods are actually becoming more racially diverse. For example, Oakland, CA and Newark, NJ are often thought of as African American cities, but in reality their populations
have become much more diverse in recent years. Part of this may be attributable to factors such as gentrification or demand for new land close to the commercial centers of New York and San Francisco, but perhaps there is some connection between increased population density and increased diversity, independent of outside factors.

To see if this idea has any merit, I investigated some data from the 2000 U.S. Census, as prepared by the San Diego Association of Governments [4]. I analyzed the racial demographics and population density of 30 San Diego ZIP codes as a small dataset to use for these investigations. The ZIP codes had populations ranging from 3828 to 74,388 (with an average of around 36,000), and areas of 3 to 22.1 square miles (with an average of 8.58 square miles). These areas represent all neighborhoods of the city of San Diego, and they span quite distinctive areas: wealthy quasi-rural neighborhoods in Carmel Valley and Rancho Peñasquitos, urban core neighborhoods of North Park and Hillcrest, suburban neighborhoods such as Clairemont and Mira Mesa, and beachfront communities such as Coronado and Ocean Beach.

The census data counts the number of people in each of six racial categories: white; black or African American; American Indian; Asian; Native Hawaiian and other Pacific Islander; some other race; and two or more races. In order to compare data from various ZIP codes quantitatively, we consider the percentages of each category; denote them \( p_i \) for \( i = 1, \ldots, 6 \). Then, in order to reduce the proportions down to one measure of diversity, I use Shannon’s diversity index [7]:

\[
\text{diversity} = -\sum_{i=1}^{6} p_i \ln p_i.
\]

This resulted in diversity scores ranging from 0.4 to 1.6. I plotted these diversity scores against population density (measured simply as the ratio of ZIP code population to ZIP code area), resulting in Fig. 4. It is apparent that despite looking at this data wholly removed from outside socioeconomic or political factors, there appears to be at least some correlation between population density and racial diversity, with a correlation coefficient of \( R^2 \approx 0.3 \).

It is important to note that this is a small dataset, and the measurements are fairly imprecise. The first issue is with the small selection of ZIP codes, limited only to the city of San Diego. One must ask if San Diego is representative of the U.S. city as a whole; it is possible that this correlation is limited only to such a small area. Indeed, San Diego is still a fast-growing young city, and it would be interesting to see if these trends
Figure 4: A relationship between San Diego ZIP code population density and racial diversity.
also occur in older cities such as Boston, contracting cities such as Detroit, or cities of explosive growth such as Phoenix or Las Vegas. Furthermore, the granularity of the census data used is rather coarse. The ZIP code areas are all several square miles (and in some cases more than twenty), and to really examine the demographics of neighborhoods, it would be beneficial to make comparisons using smaller areas such as census tracts, block groups, or blocks.

A further complication comes from the measure of diversity within each ZIP code area. The main concern here is the inclusion of people of Hispanic origin. “Hispanic” is not a category on its own; the census leaves it up to the population to make a racial decision. This results in the Hispanic population being spread primarily among “white” and “some other race”, with others responding as “black or African American” or “two or more races”. This may be a controversial subject in reality, so ignored it completely by sticking with the strict census categories, perhaps losing information about Hispanic populations. It is also unclear that Shannon’s diversity index is appropriate for this purpose, but it is easy to compute and seemed logical to use for these investigations.

It is also important to note that due to the high granularity of the data, the population density figures are sometimes skewed. Population density should be an indication of how closely spaced residents are, but the land area of several ZIP codes include sizable areas that are closed to residential use. For example, ZIP code 92107 includes U.S. Navy land and a national monument on the Point Loma peninsula, 92101 includes the San Diego airport, 92118 includes a large naval base on Coronado Island, 92109 includes the large Mission Bay recreational area, and nearly every ZIP code contains a fair amount of parkland and open space, some more than others. With a more accurate measurement of the actual livable area, perhaps there would be even more correlation between density and population diversity.

4 An updated model

To investigate this correlation between population density and diversity, I analyze a variant of Schelling’s model that accounts for higher population density. In this high-density Schelling model, the world is still a 2-dimensional grid, but there are now 10 agents at each node. There are still two types of people (red and blue), and each grid point can have any mix of red and blue people, as well as vacant slots. Thus, each agent now has up to 89 neighbors, and the population density has been increased by a factor
of 10.

To keep my study similar to the simulations in [1], I let the initial configuration have 100,000 red people, 100,000 blue people, and 25,000 empty spaces. This was achieved simply by increasing all categories by a factor of 10. Thus, the proportions are the same as in the previously-discussed simulations. What remains is choosing the threshold $n$ for determining whether or not an agent is satisfiable. Choosing a good $n$ depends on which of the following statements is more accurate about human behavior:

- Every person wants at least a few of its neighbors to be of the same type.
- Every person wants at least a certain percentage of its neighbors to be of the same type.

The simulations with one agent at each grid point and $n = 3$ or 4 do not really differentiate between the two cases, because 3 and 4 both qualify as “a few people” and ”a certain percentage” of neighbors (37.5% and 50%). However, I can simulate both cases with different values of $n$.

First, I simulate this modified high-density model with $n = 45$, representing the case where each person wants at least half of its neighbors to be of the same type. The results here (Fig. 5) show that small neighborhoods of varying degrees of homogeneity start to develop, but unfortunately every simulation ran into the case where an unsatisfied agent had nowhere to move to become satisfied. It is possible that using XXX’s variation (where that agent would just move to a random vacant space) would yield results more similar to the traditional Schelling model. However, the results for $n = 30$ (Fig. 6) are more conclusive. This still represents a sizable percentage of the neighbors (about one-third). In this case, we get results resembling the traditional Schelling model with $n = 3$: there are interlocking large homogeneous swathes of red and blue.

When $n$ is lowered all the way to 10, the results are dramatically different from Schelling’s original model. This represents the case where every person wants at least a few of its neighbors to be of the same type, but not necessarily a large percentage. In this model, the large number of neighbors makes it much easier to satisfy the agents, and the simulation terminates with all agents satisfied in a very heterogeneous configuration (Fig. 7).

This indicates that if human behavior is such that people are content to be living in a location where at least a few of their neighbors share their characteristics (and not necessarily a large percentage), heterogeneous neighborhoods are much more likely to arise naturally in densely populated
Figure 5: A simulation of the high-density Schelling model with $n = 45$ (about $1/2$ of the neighbors).

Figure 6: A simulation of the high-density Schelling model with $n = 30$ (about $1/3$ of the neighbors).
regions. This could at least be a partial explanation for a possible correlation between population density and racial diversity in San Diego if not elsewhere in American cities.

5 Conclusions

I am hesitant to draw any wide-reaching conclusions about population density and demographics based on this investigation, but the possibility is there for a positive correlation between density and diversity. There are just so many factors that contribute to demographics that the dynamics of human behavior may be too complex for a simple model to be relevant. Furthermore, it would be valuable to see if these trends found among ZIP codes in San Diego carry over when populations are measured over a finer granularity and from different cities. It would also be interesting to compare population diversity and density data from different time periods; perhaps the upcoming 2010 census will lend itself to such comparisons.

Still, this investigation seems to show that there could be a correlation here, and my personal experiences living in various parts of the country do fall in line with these experiments. Perhaps eventually it will become clear that higher-density development is a way to encourage more heterogeneous urban neighborhoods to develop in the future.
References


