2023

Measuring Racial Segregation in Los Angeles County using Random Walks

Zarina Kismet Dhillon
Claremont McKenna College

Claremont McKenna College

Measuring Racial Segregation in Los Angeles County using Random Walks

submitted to
Professor Sarah Cannon

by
Zarina Dhillon

for
Senior Thesis
Spring 2023
24 April, 2023

Acknowledgements

I am forever grateful for the support system that uplifted me throughout my undergraduate adventure: To the 50+ brilliant professors who patiently guided me through the unfamiliar and daunting academic world, extending my gratitude especially to those who encouraged me to persist despite my unsteady background. To the even more patient tutors (both formal and informal) who took the time out of their days to explain a difficult problem to me. To faculty that handed me tissues, and staff that did mountains of unseen labor. To those in the world of dance who supported my transition and applauded my courage. To my absolutely brilliant research friends, whose dedication to abolishing oppressive academic structures inspires me every single day. To dear friends who forced me to stop studying to adventure together, to look up at the stars with them, and who insisted that ice cream breaks help productivity. Your unwavering support gets me out of bed every day. To the one who wrote me enough love letters to fill every textbook I've read, sent me care packages every finals week, and embraced me tight enough to transcend the distance. To my dear family, who poked (harmless!) fun at my love of math but hid notes in my calculator because they knew I'd find them there. For creating flowcharts for every life-changing decision, and never failing to give me the honest feedback I needed. For cutting up fruit for me while I studied, and covering me in blankets when I fell asleep doing so.

To myself for never giving up, and daring to ask “what if?”

Contents

List of Figures

Abstract

As of now there is no universal quantitative measure used to evaluate racial segregation in different regions. This paper begins by providing a history of segregation, with an emphasis on the impact of redlining in the early 20th century. We move to its effect on the current population distribution in Los Angeles, California, and then provide an overview of the mathematical concepts that have been used in previous measurements of segregation. We then introduce a method that we believe encompasses the most representative aspects of preceding work, proposed by Sousa and Nicosia in their work on quantifying ethnic segregation in cities using random walks: Given a graph with racial population information at each node, this method conducts a random walk and outputs the number of steps it takes to reach all racial groups in the system. In analyzing the amount of time the walk takes, we are able to form conclusions about the levels of segregation present in the region. We argue that this measure is more effective than those used previously because of its high explanatory power and preservation of graph structure. We applied this method to LA County with data collected from the U.S. 2020 Census, and found that there is a significant difference between the average step length of LA County vs. the average step length of an unbiased null model. The impacts of this study imply that LA County is racially segregated in comparison to a given neutral null model.

1 Introduction

With over ten million citizens, Los Angeles county is the largest county by population in the United States of America [4], but with its large community comes major instances of racial inequality. For example, Black citizens represent only nine percent of the general population in Los Angeles County, yet comprise 40 percent of the population experiencing homelessness [3]. This outcome, along with countless others that affect those in high-risk communities, can be traced back to the redlining practices appointed in the early establishment of the county – though they were implemented nearly a century ago their effects persist and harm racial minorities to this day, effectively contributing to instances of racial segregation.

While it is recognized that there exists racial segregation in regions across the U.S., as of now there is no universal agreement across researchers on any single quantitative segregation measure to implement uniformly. A method proposed by Sousa and Nicosia in their work on quantifying ethnic segregation in cities using random walks [17] provides the measure used in the proceeding pages: Given a graph with racial population proportions at each node, this method conducts a random walk (with equal probability) across nodes until all racial groups have been encountered and outputs the number of steps that this takes. This allows us to analyze the temporal behavior of different regions, where a large number of steps indicates more racial segregation and fewer steps indicate more racial integration.

The work done here claims that using a random walk algorithm provides a much more representative basis to analyze segregation than some of the mathematical measures that have been explored previously: This method effectively allows the researcher to compare segregation of different systems on equal grounds, while preserving the respective structures of each region. Noting that comparison based analysis works best with the implementation of a neutral model, we created a comprehensive null model built by taking the average across various sub-null models. In applying this method to Los Angeles County we discovered that the average number of steps for LA County versus that of our proposed null model differs by over a factor of two. Additionally, the average time (in number of steps) difference between LA County and a primary null model is 4.2 times longer than the average time difference between sub-null models. The impacts of this study imply that LA County is significantly racially segregated in comparison to an unbiased model.

We will first address the severely charged racial history of the country and its impact on the development of Los Angles County. An evaluation of previous measures of segregation will be discussed, followed by a full justification of the reasoning behind the method that was chosen. We then provide a detailed explanation behind the measure and our adaptations, following with an analysis of our results. We conclude with a discussion on limitations, further study, and a brief commentary on the implications behind this work.

2 Background

2.1 A (Brief) History of Segregation in the United States

In order to comprehensively understand how racial segregation has been developed and maintained in specific regions, we will first explore the background and causes more broadly throughout the United States. This will allow us to recognize historical patterns of discrimination, form connections between the instances we see today, and develop a broader context about the impacts these have on citizens. It is acknowledged that racial segregation can occur for many different reasons, often unintentional (a desire to form community and therefore residing close to family/people with similar racial backgrounds, for example). However, discrimination in racial segregation involves the use of intentional, systemic practices to enforce the separation of groups, and traditionally affects those who belong in a racial minority. The patterns of today's residential segregation closely follow the discriminatory practices developed in response to early racist ideals against marginalized groups [16].

It would be vastly inaccurate to disregard the history of slavery in the United States when discussing current practices of segregation: Though present informally for most of the U.S.'s history, segregation became systematic primarily following the enactment of the 13th Amendment [1] in order for those in positions of power to satisfy their desires to offset the recent emancipation of slavery: Formerly enslaved people were denied the full rights and privileges of an average citizen so they would not constitute a threat to the previously upheld slave regime. This was not limited to law; these trends grew throughout housing, education, public accommodations, communities, businesses, and more, down to the subtle yet impactful details of separate doors, elevators, and drinking fountains [12]. Despite slavery being legally outlawed, several state segregation statutes came into being in order to uphold this notion of racial hierarchy. This practice was born from the idea (which has never been empirically supported) that one race is somehow superior to others. In analysis of early documentation during the period of legal enslavement, most laws contained explicit racial language, such as directly equating "slave" with African Americans, and using them interchangeably in written accounts [11]. The normalization of discriminatory language and practices ultimately contributed to the racial profiling that still exists to this day.

One of the most glaring representations of systemic segregation in minority communities occurred in the mid 1930s, involving the implementation of redlining. During this period the Home Owners' Loan Corporation (HOLC) collected a series of data that represented an area's likelihood of safe vs. risky mortgage security. Among the data was the neighborhood's quality of housing, the recent history of sale and rent values, and the racial and ethnic identity and class of residents [6]. The HOLC followed the guidelines of the federal housing administration's underwriting manual, which stated that "the infiltration of inharmonious racial groups will produce the same effects as those which follow the introduction of nonconforming land uses which tend to lower the levels of land values and to lessen the desirability of residential areas" [7]. To imply in a government document that "inharmonious" racial groups have a negative effect on housing values was detrimental to the establishment of integrated residency patterns. As time went on, this implementation created a self-fulfilling prophecy, where more immigrants/residents of color resulted in a lower neighborhood grade. These lower grades were the backbone of redlining; without access to mortgages these communities were unable to own property and build capital. These discriminatory ideas and practices ultimately led to the racial segregation that is still present today, and form the basis of the work conducted in this study.

2.2 Segregation Practices in Los Angeles County

Now that we have a stronger foundation on which to analyze racial segregation, we can apply this to particular regions. Specifically, we will examine these effects on Los Angeles (LA), CA. A bustling city with boundless opportunities, LA has historically drawn citizens with diverse interests and backgrounds. As of July 2021, the total population of LA County reached 9,829,544 people: 49.1% Hispanic or Latino, 25.3% White alone (not Hispanic or Latino), 15.6% Asian, 9.0% Black or African American, 1.5% American Indian and Alaskan Native, and 0.4% Native Hawaiian and other Pacific Islander (with note that 3.3% of those surveyed identify as two or more races) [5]. However, diversity does not equate to integration; like most other major U.S. cities, there exists a high degree of residential racial segregation [18] due to the practices mentioned in the previous section. As the HLOC crafted the security map of Los Angeles in 1939, both class hierarchies and racial segregation worked together to design this region of Southern California. We can see the original redlining patterns created by the HLOC in Figure 1.

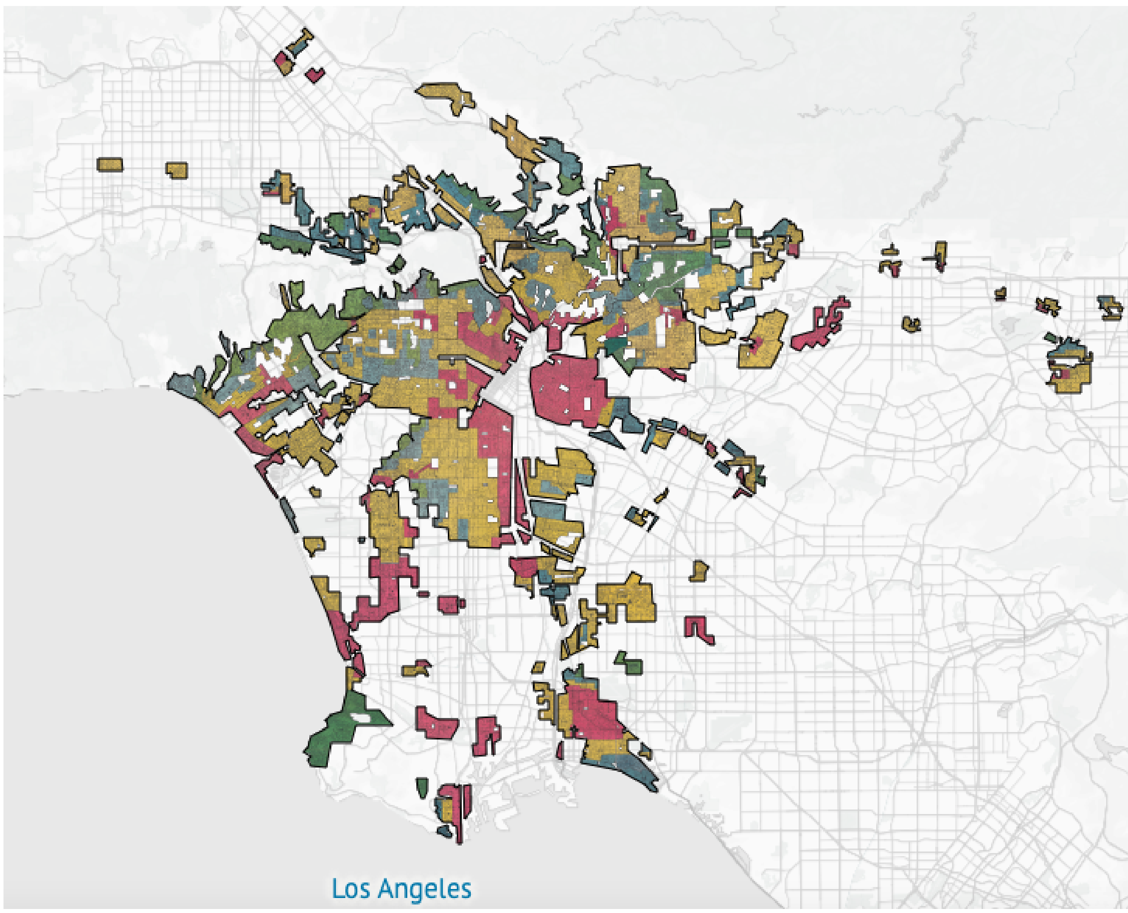

Figure 1: Original redlining map of Los Angeles, CA, created by the HLOC [6]. Areas colored red were graded "hazardous," yellow "definitely declining," blue "still desirable," and green "best."

This was perpetuated in early years primarily through the implementation of zoning and restrictive covenants (including racially restrictive covenants), and Contracts, Conveyances, and Restrictions (CCRs). The 1924 Code of Ethics Article 34 defined the ideal neighborhood as segregated by race and class: "A realtor should never be instrumental in introducing into a neighborhood a character of property or occupancy, members of any race or nationality, or any individuals whose presence will clearly be detrimental to property values in the neighborhood" [2]. Ultimately, these practices promoted the prohibition of certain racial or ethnic groups from either owning or occupying a property, until the Supreme Court later deemed it unlawful [13]. This impacted integrated development throughout the years, as the severe aforementioned restrictions on mortgages and loans took their toll on the racial diversity/homogeneity of the city. The result was minority populations densely concentrated in neighborhoods with histories of extreme disadvantage [9]. The effects of redlining are still evident in current residential trends: consider the racial distribution of LA County (as of 2020) seen in Figure 2, where the clustering patterns of racial minority groups closely follow that of the original HLOC guidelines. One example (shown in Figure 3) is that the most prominent "hazardous" region, the downtown Los Angeles business district, was located just east of present day Inglewood, a majority Black/African American region.

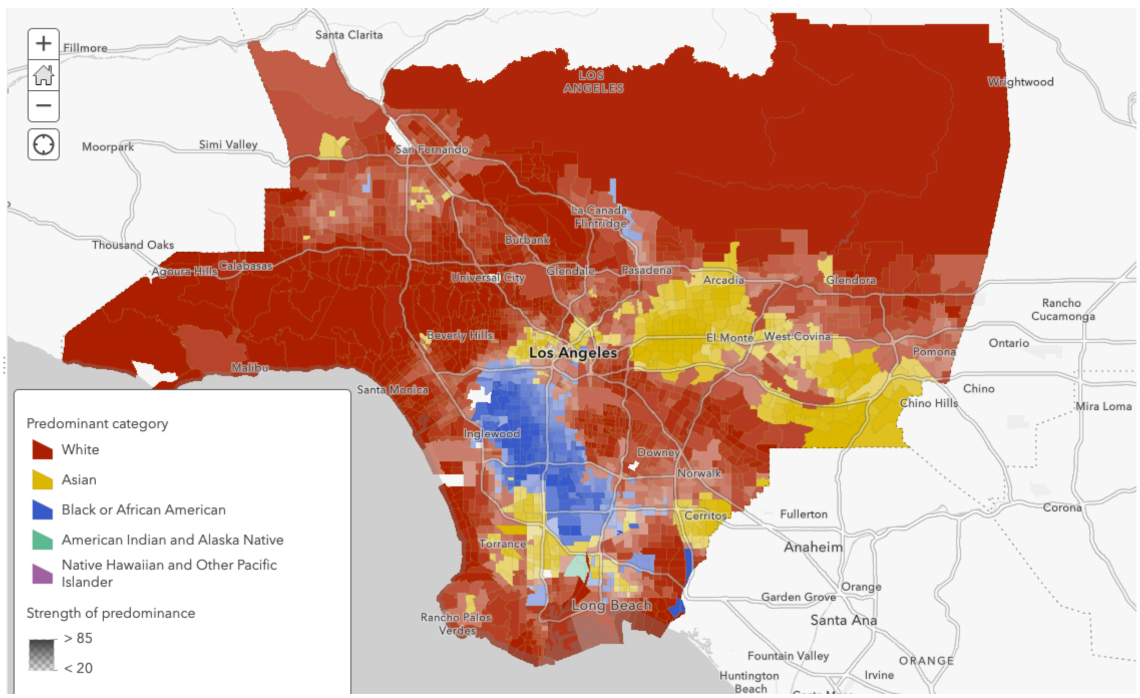

Figure 2: Census tracts within Los Angeles County (decennial redistricting data, 2020 (PL 94-171)) by race, created with ArcGIS. Note that the Hispanic/Latino category is considered by the U.S. Census to be an ethnicity rather than a race.

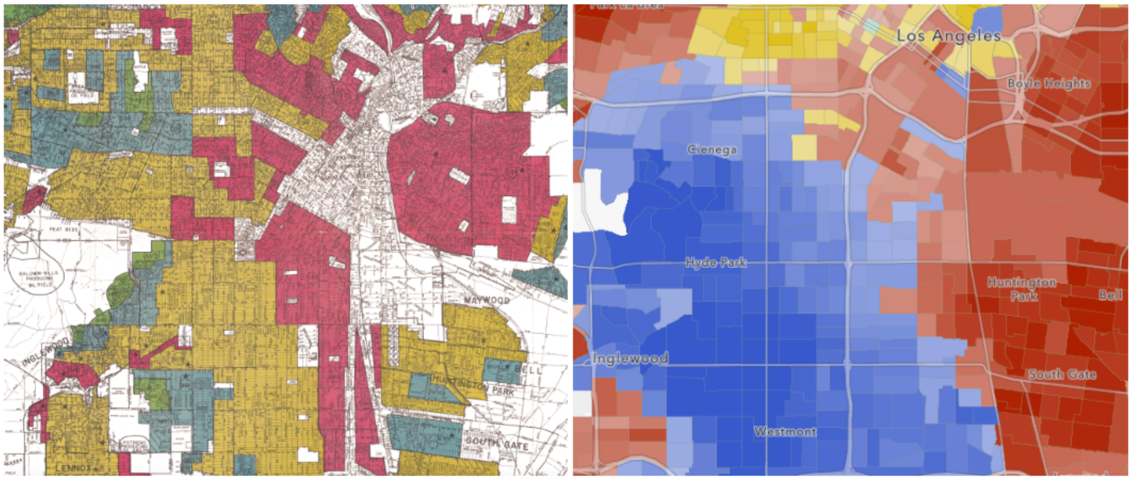

Figure 3: On the left: The Los Angeles business district, a subset of LA County deemed by the HLOC to be dangerously hazardous. On the right: Approximately the present-day configuration of the Los Angeles business district, a majority Black/African American region. Coloring in the left and right panels aligns with those in Figure 1 and Figure 2, respectively.

2.3 Previous Mathematical Measures of Segregation

Although several measures to evaluate segregation have been explored, there is no consensus among experts on a single method to implement uniformly. The measures we explore in this section are just a subset of those that have been studied throughout various fields of mathematics, but serve as a basis for the reasoning behind the eventual method that was chosen.

A common theme among this topic is the formulation of segregation scores, an equation to input regional data and output a scale of how segregated the area of interest is. One such score is a clustering propensity score, designed to measure the clustering level of one or more subgroups within a population. Given a partition into geographic units, a dual graph can be constructed with information at each vertex v of a given region, as done by Alvarez, Duchin, Meike, and Mueller [8]. The dual graphs are then used to measure the extent to which people of one demographic tend to live next to each other (rather than next to those of a different demographic). This is completed by expanding each node into a complete graph (the authors call this an *exploded graph* as it disaggregates nodes by replacing each one with a subgraph) to more sufficiently represent the edges between members of respective clusters. The clustering propensity scores are calculated as a result of these exploded graphs, as they have a clear probabilistic interpretation. This approach allows for the analysis of both within-unit adjacencies and neighboring-unit adjacencies using a graph theoretic framework. However, a severe fallback is that it ignores local adjacencies – By replacing each node with a complete graph, it effectively assigns the same neighborhood weights to each new node, meaning that areas that are disconnected in reality are now all related equally in the graph.

Other segregation measures were explored through a comprehensive overview conducted by Rodriguez and Vorsatz [14]. They note that the main mathematical devices developed to measure segregation are segregation indices, formally defined as some function $S : N \to [0, 1]$ that maps a distribution N into the unit interval (where the maximum value of segregation occurs when each unit contains individuals from a single group). Put more simply, standard indices output some value between 0 (minimal segregation) and 1 (maximum segregation) depending on how segregated a given region is. They also note that the most widely used segregation measure is the dissimilarity index, which measures how closely subarea demographic proportions match the demographic proportions of the larger area. A problem with this method, as noted by Alvarez et.al [8], is that the dissimilarity score is given by summing over the nodes without reference to adjacency, so it does not take into account the spatial relationship between units. Thus this score equates neighboring units to those on opposite sides of the region of interest, potentially omitting crucial spatial information.

Researchers in the field of geography aim to adapt these scores in order to account for the potentially crucial spatial relationships described above. The final measure we will discuss is the standard spatial statistic used in geography literature, Moran's I. Given numerical values (such as population) associated with the nodes of a dual graph, Moran's I returns a real number between -1 and 1 [10]. The interpretation is that values near 1 indicate extreme segregation, values near zero indicate no pattern, and negative values flag "anti-segregation," where people are more likely to live next to those of a different race. Moran's I essentially examines where values are positive or negative and analyzes the patterns between such areas. Given a deviation vector $\bar{x}$ (where values $[x_1,\ x_2,\ .\ .\ .\ ,\ x_n]$ describe how much each unit i's population proportion deviates from the overall average population proportion in the region) and a graph $G(n, m)$, the equation for Moran's I can be calculated as follows:

$$I(\bar{x}) = \frac{n}{m} \cdot \frac{\sum_{i \sim j} x_i x_j}{\sum_i x_i^2},$$

where $i \sim j$ if the two nodes are adjacent. However, a major concern with using Moran's I is that changing the aggregation level has a drastic impact on the output (frequently referred to as the Modifiable Areal Unit Problem) – the result depends heavily on the choice of geographical units, which is a concern when attempting to relate scores of different regions that may be separated into varying units.

Across literature many researchers have concluded that in analyzing different measures of segregation, the most effective way to maintain the structure of a region is to approach the problem from a graph theoretical standpoint rather than purely statistical one. However, as we have seen, despite the benefits there are still valid weaknesses across the methods discussed here.

3 Methods

3.1 Random Walk Method: Sousa and Nicosia (2022)

The measure used in this analysis closely follows that of Sousa and Nicosia in their work on quantifying ethnic segregation in cities using random walks [17]. The method essentially conducts a random walk on the region and evaluates the number of steps needed to encounter all racial classes present in the system. This approach aids our analysis of segregation by providing a quantitative measure from which to compare different regions, allowing us to evaluate how separated different racial classes are from each other. We are thus able to quantify the heterogeneity of the distribution of ethnicities across an area. Analyzing the resulting levels of disparity from an objective, numerical point of view can also reveal characteristics about the overall structure of the region.

The method itself involves consideration of a graph $G(V, E)$ in place of the region of interest. By placing each vertex (or node) $v \in V$ at a specific partition (in our case, Census tracts) of the region, an edge $e \in E$ between nodes represents a common border. Each node contains information about some variable of interest – in the case of this problem, $\bar{x}_i$ is a vector containing the racial class population at each node i. Here we are interested in the spatial distribution of $\bar{x}_i$, which will reveal to what extent nodes being spatially close in the graph also have similar values encoded in their vectors, thus creating clusters with similar racial groups. By applying a random walk from any v_i (where each "step" goes to a neighboring node of v_i with equal probability), information is gathered as the walk goes on. This specific method measures the length of time t (in steps of the walk) it takes to reach a certain fraction c of total classes in the system. The idea is that greater values of t represent more segregated areas, and smaller values of t indicate more racial integration. The principal proposal of their paper is that the level of segregation of an area can be represented by Class Coverage Time (CCT) of a random walk on a corresponding dual graph G. CCT is defined as the expected number of steps needed by the walker, starting at a generic node, to visit some fraction c of total classes in the system. If a system is well dispersed, the starting node is irrelevant, and vice versa.

In relation to previous mathematical measures of segregation, the explanatory power of this method is much higher than other indices, and was chosen for several reasons: The measure only depends on the structural characteristic of the graph in question and the distribution of node properties, which therefore preserves the overall structure of the dual graph. This effectively allows researchers to compare the segregation of different systems on equal grounds, as the focus on preserving structure acts as a normalization aspect to relate contrasting levels of disparity across regions to one another. These aspects combine to let us analyze patterns in the structures of different regions, and objectively compare the overall levels of segregation across systems. Additionally, outputting specific lengths of time (rather than a single statistic) allows us to easily determine any glaring outliers in the system. Ultimately, the concept of using random walks to measure segregation was chosen because it allows us to compare different systems in equal ways, as the approach is not parameter-based. Random walks also provide a sufficiently arbitrary basis from which to evaluate these different regions – in using this method Sousa and Nicosia provide a consistent estimation regardless of any outside qualities of the system.

3.2 Data and Code

All data used in this analysis was collected from the U.S. Census database. Clean versions of the county-level data for the random walk analysis were provided by authors Sousa and Nicosia in their GitHub repository [15]. This data consisted of an edge-list of LA County along with a list of class population information for each Census tract [15]; Our results are built from a total of 3,923 Census tracts with 64 classes each. Maps were created using the geographic information system visualization tool ArcGIS using decennial data from 2020 at the Census tract level.

The code used to analyze this problem was adapted from Sousa and Nicosia's approach in their work "Quantifying ethnic segregation in cities through random walks," accessed from their GitHub repository [15]. Upon input (an edge-list of LA County Census tracts, along with a list of class data at each node) the code outputs the time that it takes to reach 100% of the total classes in the region. Specification of the number of iterations is required, and in these results we conducted 10 iterations each starting from a randomly selected node. The range of total nodes in the walk can be modified (i.e., 1-3 would consist of a walk between only three nodes in the graph), however in this analysis we ranged across all 3,923 nodes. The resulting output is a vector of length 100: the first number is the time until the walker has seen 1% of the total classes in the system, the second number the time until the walker has seen 2% of the total classes in the system, ... , the nth number is the time until the walker has seen n% of the total classes in the system, until the final number outputs the total time needed to see every class in the system. An additional feature to aid the analysis of averages was implemented, where after the trajectories are calculated the function outputs the average across the range of nodes. A complete documentation of the code and data used in this specific paper is available on GitHub[19].

3.3 Adapted Measure: Null Model

In order to have a basis from which to analyze the resulting output of Los Angeles County data, it was deemed necessary to implement a null model. Due to complications in recreation of Sousa and Nicosia's null model, we ultimately created our own design that served the purposes needed in this analysis. We crafted a comprehensive null model as follows: original class-population data from each LA County Census tract was randomly permuted (across all 3,923 tracts) ten separate times. Each permutation was input into the above code to determine the average time it takes to reach 100% of the racial groups. This average time was then averaged across the ten outputs to give us the single, primary null model used in the analysis. By swapping the racial population proportions randomly between nodes, we believe that this effectively takes any existing biases in the structure of LA County and eradicates them. Therefore this method provides us with a sufficiently randomized, arbitrary structure from which to compare LA County to.

4 Results

The output of LA County Census data along with the constructed null model show that it takes LA County a consistently higher number of steps to reach the same percent of classes as the null model, as shown in Figure 4. Here we see that the time (in number of steps) is somewhat consistent until roughly 90%, where their difference increases significantly. Note that in particular it takes on average more than twice as long to reach 100% of the total classes in LA County than in the null model (an average of 28,068 steps vs. an average of 12,803 steps, respectively). Throughout the measurement, LA County is consistently above the null model for the majority of steps in the walk. Because of the large range of values, and to assist in seeing the steepness of this curve, the natural log of the CCT is shown in Figure 5, where we can see that the walk progresses to trends exhibited in exponential growth. From these plots we observe that the difference between the two models is directly related to the percentage of classes seen (a higher percentage corresponding to a bigger difference, and vice versa): We note that this is likely an effect of having the total fraction of classes being equal to 1, as the overall coverage time will increase when the system contains rare classes – as is the case in LA County, where multiple racial groups contain only a few citizens each. To clearly see their differences at earlier times, more focused time boundaries of this output at various class percentages can be seen in Figures 6, 7, and 8.

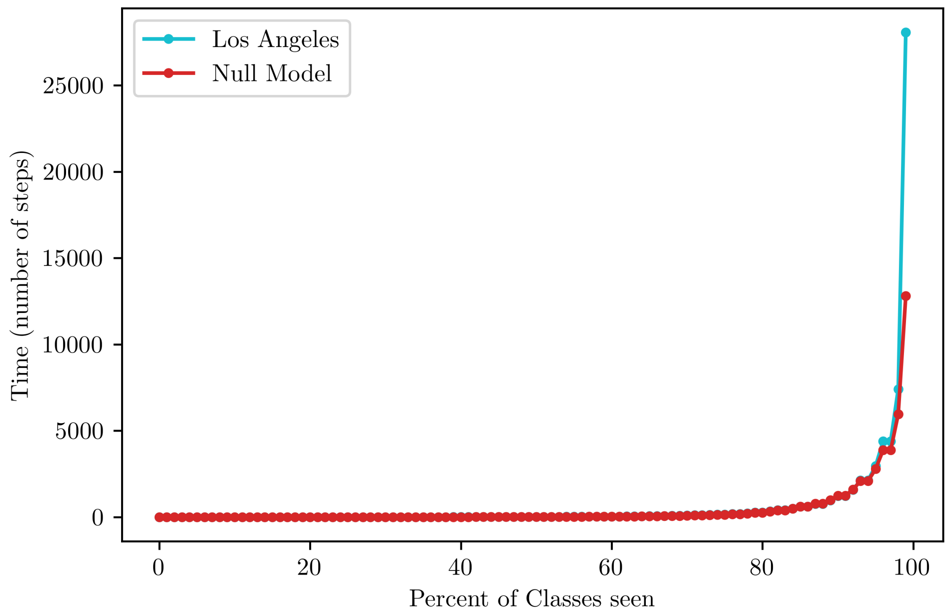

Figure 4: Complete output of CCT for LA County and the null model

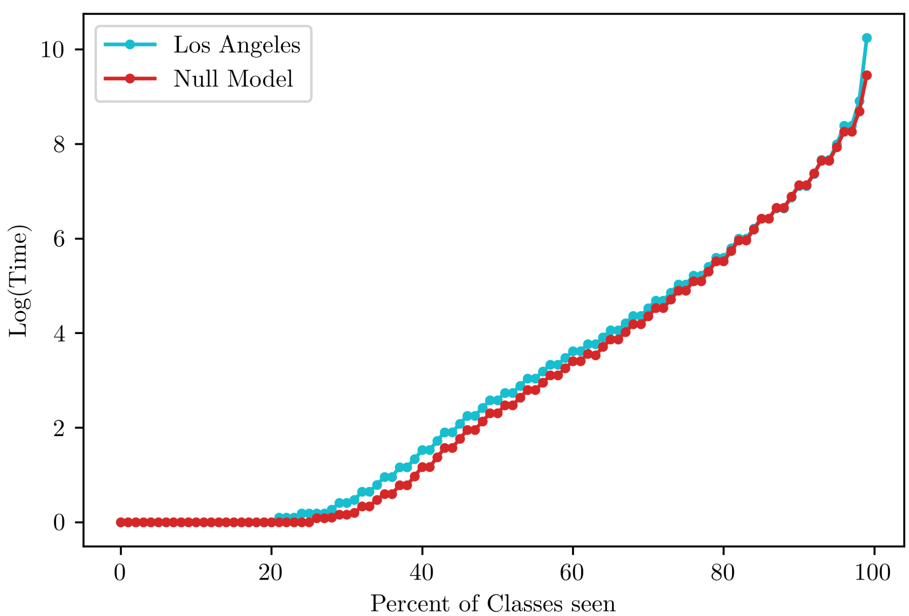

Figure 5: Natural log of the CCT for LA County and the null model

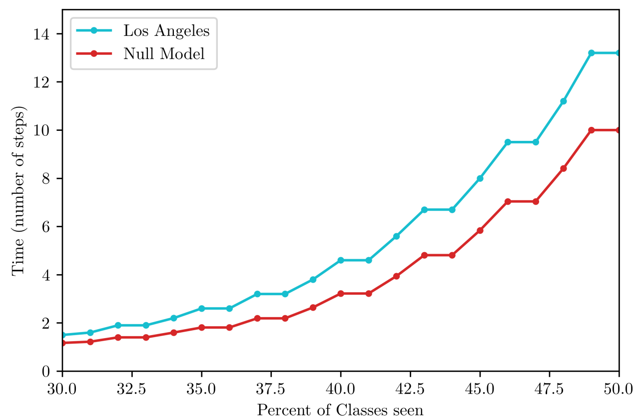

Figure 6: CCT for LA County and the null model from range 30% to 50% of total classes

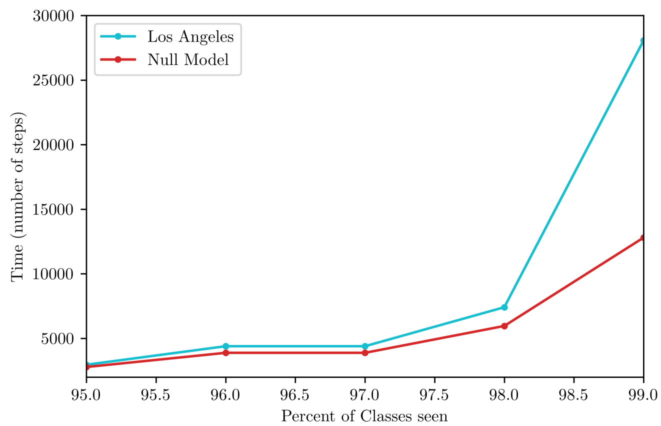

Figure 7: CCT for LA County and the null model from range 95% to 99% of total classes

Because the null model is comprised of the average of 10 sub-null models, Figure 9 allows us to properly visualize the spread across the range of trials. The minimum value is 9.8 steps for 50% and 11.7 steps for 51%, while the maximum is 10.1 steps for 50% and 12 steps for 51%. With respect to the averages, the primary null model is 10 steps at 50% and 11.86 steps at 51%, compared to LA County's 13.2 steps and 15.4 steps, respectively. Notice that the deviation between sub-null models and the primary null model here is at most $\pm 0.2 \implies \pm 2\%$ (compared to the 32% deviation between LA County and the null model). This indicates that there are minimal levels of variation across sub-null model runs, which also implies that LA is consistently at higher values than all the null models across the range of trials, and not solely the average.

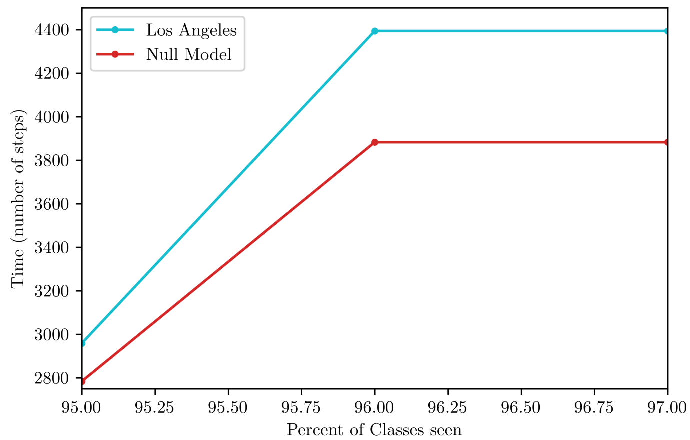

Figure 8: CCT for LA County and the null model from range 95% to 97% of total classes. Note that what appears to be only a slight difference in Figure 7 is nontrivial when examined at a more concentrated time boundary.

In their analysis, Sousa and Nicosia defined spatial heterogeneity $\Delta\mu$ as the average deviation between a given region ($\mu(t)$) and its corresponding null model ($\mu(t)^{null}$), as follows:

$$\Delta\mu = \int_0^1 dc|\mu(t) - \mu(t)^{null}|$$

where c is the prescribed fraction of total classes present in the system. Though we constructed a different null model than in their analysis, we applied the same approach to quantify the difference between our primary null model and LA County. We computed the average deviation as a discrete summation (as our walk contains distinct steps) for some integer $1 \leq k \leq 100$:

$$\Delta\mu = \sum_{k=1}^{100} c\left|\mu(t)\left(\frac{k}{100}\right) - \mu(t)^{null}\left(\frac{k}{100}\right)\right|$$

where $c = 0.01$.

For the deviation between LA County and its null model, we calculated a spatial heterogeneity score of $\Delta\mu = 185.0858$. However, because of variations between the respective creations of our null models, we are unable to compare our value to those in Sousa and Nicosia's analysis in any meaningful way. Instead, we hope to supply additional context by providing spatial heterogeneity scores between the different supplementary

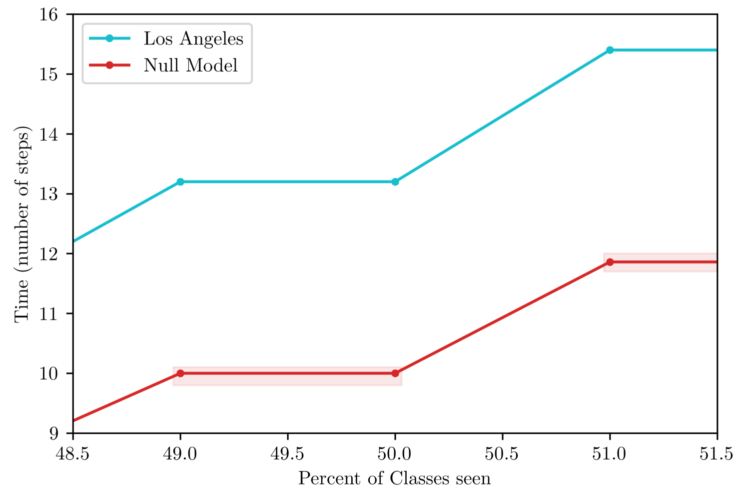

Figure 9: CCT for LA County and the null model, where the shaded bands indicate the min/max ranges across the sub-null model runs. Note that the range of sub-null models has only a slight deviation from the primary null model, which is their average.

null models (the 10 sub-null models that went into the eventual primary null model).

To analyze each difference pairwise, there were a total of $\binom{10}{2} = 45$ different scores. For sub-null models n_i, n_j, where $1 \leq i < j \leq 10$ and some integer $1 \leq k \leq 100$, spatial heterogeneity scores were calculated as above:

$$\Delta\mu_{nulls} = \sum_{k=1}^{100} c\left| n_i\left(\frac{k}{100}\right) - n_j\left(\frac{k}{100}\right)\right|.$$

The average across these scores was then calculated

$$\Delta\mu_{null} = \frac{1}{45}\sum_{\substack{i,j \\ i\neq j}}^{10} \Delta\mu_{nulls},$$

where i, j span all possible pairwise configurations of the 45 sub-null models. This resulted in a spatial heterogeneity value of $\Delta\mu_{null} = 43.428$, which is roughly 4.2 times less than LA County's spatial heterogeneity score, indicating lower levels of deviation between the sub-null models vs. that of LA County and the primary null model. This can be visualized in Figure 10 as the natural log of the difference values $(\mu(t) - \mu(t)^{null})$ along with the natural log of the average difference across all sub-null models. We acknowledge that the average difference between sub-null models is higher than the difference between LA County and the null model between 83% to 96% of total classes seen. We cannot conclude any specific reasoning for this behavior, however in the original CCT at this range LA County is still greater than or equal to the null model, which confirms the validity of our conclusions.

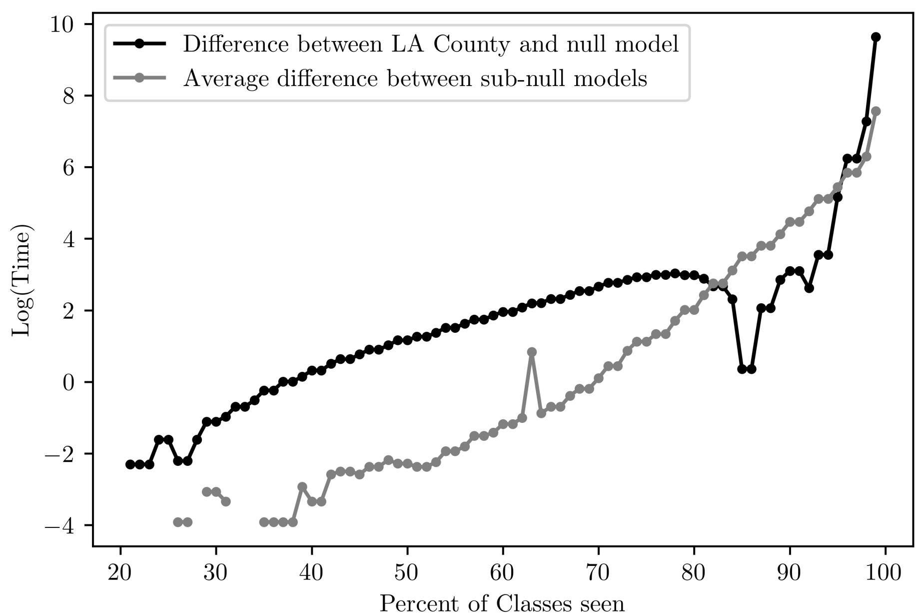

Figure 10: Natural log of the difference between LA County and the null model and the average difference between sub-null models (note that the difference at 0%-25%, 28%, and 32%-34% for the sub-null models is 0).

Our analysis shows that the difference between LA County and the null model is significantly larger than the average difference between each of the sub-null models. Recall that as greater values of t represent more segregated areas and smaller values of t represent more racial integration, the results of this section indicate a higher level of racial segregation in LA County than in the null model.

5 Conclusion

There is a level of consistency in our results, across all trials, that is nontrivial to the larger implications of this study – It can be concluded that the racial history contributing to the construction of LA County is still present in the population distribution centuries later. This was discovered by comparing LA County to a sufficiently unbiased county structure of the same racial makeup and analyzing their similarities and differences.

However, we acknowledge that there are some limitations to this study: in exploring this problem further, it would be beneficial to have a stronger metric on what makes a null model truly unbiased. One way to accomplish this would be by constructing more iterations of sub-null models and determining at what number of trials their average seems to converge. Another setback, as mentioned before, occurs when there is a large number of racial and ethnic classes present in the system, which effectively lengthens the overall CCT of the walk. This could be mitigated by performing the analysis across a smaller range of percent of classes encountered (i.e., eliminating when $c = 1$), or by potentially omitting especially rare classes from the analysis in order to get more representative results. One final limitation to note is the missing context needed to evaluate the spatial heterogeneity scores; without a meaningful interpretation of the values, it is difficult to fully understand the severity of the regional differences. A potential way to reduce this would be to compare the CCT of LA County across difference time periods, and assessing the relationship between number of steps and the social, political, and racial context of the time.

The work done here is just a fraction of the possibilities available towards creating a more just and equitable society, and readers are encouraged to continue their knowledge of intersectional issues spanning race, political structures, and quantitative analysis. Being able to objectively provide evidence that there is inequality present in a system in a crucial step towards eradicating it, and this cannot be done without acute levels of awareness across fields and a desire to enact change.

References

[1] 13th Amendment to the U.S. Constitution: Abolition of Slavery. https://www.archives.gov/milestone-documents/13th-amendment#:~:text=Section%201.,this%20article%20by%20appropriate%20legislation., 1865.

[2] National Association of Real Estate Boards, "Code of Ethics". https://www.nar.realtor/about-nar/history/1924-code-of-ethics, 1924.

[3] Healing LA Neighborhoods: A once-in-a-generation opportunity to create thriving and inclusive communities across Los Angeles. https://www.preventioninstitute.org/publications/healing-la-neighborhoods-once-generation-opportunity-create-thriving-and-inclusive, 2020.

[4] Large counties in the United States by population. https://ballotpedia.org/Large_counties_in_the_United_States_by_population, 2021.

[5] Los Angeles County, California - U.S. Census Bureau QuickFacts. https://www.census.gov/quickfacts/fact/table/losangelescountycalifornia/RHI125221#RHI125221, 2022.

[6] Mapping Inequality, Redlining in New Deal America. https://dsl.richmond.edu/panorama/redlining/#loc=10/34.005/-118.486&maps=0&city=los-angeles-ca, Accessed 2023.

[7] Federal Housing Administration. Underwriting Manual: Underwriting Analysis under Title II, Section 203 of the National Housing Act, 1936.

[8] Emilia Alvarez, M Duchin, E Meike, and M Mueller. Clustering propensity: A mathematical framework for measuring segregation. *preprint*, 2018.

[9] Camille Zubrinksy Charles. *Won't you be my neighbor: Race, class, and residence in Los Angeles.* Russell Sage Foundation, 2006.

[10] Moon Duchin and Olivia Walch. Political geometry, 2021.

[11] Olaudah Equiano. *The Interesting Narrative of the Life of Olaudah Equiano, or Gustavus Vassa, the African.* 1789.

[12] John Hope Franklin. History of Racial Segregation in the United States. *The ANNALS of the American Academy of Political and Social Science*, 304(1):1–9, Mar 1956.

[13] Laura Redford. The Intertwined History of Class and Race Segregation in Los Angeles. *Journal of Planning History*, 16(4):305–322, 2017.

[14] Antonio Rodriguez-Moral and Marc Vorsatz. An overview of the measurement of segregation: classical approaches and social network analysis. *Complex networks and dynamics: Social and economic interactions*, pages 93–119, 2016.

[15] @sandrofsousa. GitHub - segregation-rw/ethnic-segregation-rw: Software to compute segregation based on random walks as in the paper "Quantifying ethnic segregation in cities through random walks". `https://github.com/segregation-rw/ethnic-segregation-rw`, Jul 2022.

[16] Robin Smyton. How Racial Segregation and Policing Intersect in America. *Tufts Now*, 17, 2020.

[17] Sandro Sousa and Vincenzo Nicosia. Quantifying ethnic segregation in cities through random walks. *Nature Communications*, 13(1):5809, 2022.

[18] Taeuber, Karl E, Taeuber, Alma F. Negroes in Cities, 1965.

[19] @zarinad. GitHub - zarinad/Senior-Thesis: Code and data used in Zarina Dhillon's Claremont McKenna College Senior Thesis, Spring 2023 . `https://github.com/zarinad/Senior-Thesis`, Apr 2023.