FAWWAZ UDDIN MOHAMMED

A20560217

OSNA HW 4

Dunning Community Area Census Data Analysis Final Report

This report is the detailed analysis of the Dunning Community Area of Chicago that incorporates decennial census comparisons (2010-2020) with the American Community Survey (ACS) perspectives along with network-based community detection. The results indicate that the neighborhood is undergoing a significant demographic diversification process yet there are no significant increases in population and distinct external borders. However, under this consistency, there is a sense of disunity internally. A statistical network analysis acknowledges three sub-communities, which are more in tune with realistic demographic and socioeconomic conditions than the single official boundary, that solicits a finer perspective of how resources and outreach can be structured.

1. Introduction & Methodology

The paper incorporates several reliable sources of information. The long horizon comparisons of population and demographic composition are based on decennial census block group data of 2010 and 2020. The 2018-2022 five-year estimates based on ACS provide more socioeconomic texture, especially income, household size, housing tenure, at the block group level. Spatial delineation occurs according to official Chicago Community Area boundaries. Every spatial layer received a projection into a common coordinate reference system (WGS 84) and became harmonized to work on it. The pipeline used in the workflow is as follows: the Census API is used to obtain data, the task is converted to a spatial format and overlayed, feature engineering and standardization is performed, and lastly, a graph is built and the community identified.

1.1 Study Area Composition

The hub of the analysis is Dunning (Community Area 17). It comprises eleven census tracts and thirty-five census block groups that are employed throughout the decennial comparison. The community borders off Portage Park, Montclare, Belmont Cragin, and O'Hare. In case of the network component, the study extends the scope of Dunning to a wider scope of 242 block groups to have adequate scale in similarity modeling and community detection.

1.2 Analytical Framework

There are descriptive statistics, spatial visualization, and network science that are woven together into the analytical strategy. Descriptive summaries characterize the demographic and socioeconomic status and their time geography. Spatial differences and changes are revealed using choropleth maps and boundary comparisons. Similarity structure between block groups is characterized by a k-nearest neighbors graph (k=5) constructed on standardized demographic characteristics. The Louvain algorithm subsequently splits the graph into communities that give a maximum modularity, as an empirical perspective of sub-areas that operate as coherent neighborhoods.

2. Comparison of Decennial Census (2010-2020)

2.1 Population Dynamics

Dunning population levels had not fluctuated significantly during the 10-year period but increased between 41,932 in 2010 and 43,147 in 2020, the net change in population being 1,215 people, or 2.9. The comparison is made by thirty-five block groups, which offer constant geographic units, which favors a like-for-like

analysis. The continuity in spatial distribution of population also occurs: the gradient of population density that can be seen in 2010 is continued in 2020, and there are no significant changes in the settlement patterns throughout the community.

2.2 Demographic Transformation

Dunning underwent significant compositional change, below the stable totals. The proportion of White residents dropped to 62.1, which is a 6.1-point difference. The number of Black population grew to 6.8 and the number of Hispanic population grew to 27.9. These changes indicate a highly evident diversification trend: the neighborhood has become more heterogeneous, and the growth of the population of Hispanic representatives in particular and the relative increase in the Black population, based on a smaller base, are of particular significance. Dynamic rebalancing is therefore concealed within the overall stability of the population.

2.3 Patterns of Spatial Distribution

The patterns of the demographic variables indicate an overall logical spatial patterns. Different urban populations are also concentrated in the southeast whereas the northwest has a more concentration of White residents. The population density in Central Dunning remains to be higher as compared to the periphery. The movement across the border is more of a gradual change rather than a sharp reinvention and implying diffusion areas instead of strict borders.

3. Variable Selection & Comprehensive Analysis of ACS.

To produce a more multidimensional picture, the ACS analysis considers seven block group level variables to include total population, Whites population, Black population, Hispanic population, median household income, average household size, and owner-occupied housing units. Collectively these indicators shed light on the scale and composition of the demographics, economic stratification, family formation, and residential stability.

3.1 Data Quality Assessment

The completeness of the ACS-inputs is high, and the variables do not miss more than two percent. Distributions indicate that the average household per block group is healthy and can be modeled: it is approximately 1,200 with a coefficient of variation of approximately 45 percent; household income is distributed with significant dispersion that would indicate economic diversity of importance. The homeownership is wide-ranging; approximately between the upper-twenties and mid-nineties percent, which shows that the change in tenure is quite significant. The racial and ethnic shares differ considerably among block groups, which offers the contrast to the similarity modeling.

3.2 Spatial Variation Patterns

Internal diversity is emphasized through spatial dispersion. Block group populations fall between two hundred and more than three thousand inhabitants each. The range of median household income is over about 32,000 to 145,000 that indicates strong economic gradients. The racial and ethnic composition is highly bi-modal as the percentages of the Whites and Hispanics show large ranges, and the percentage of the Blacks grows significantly in several clusters. Homeownership, in turn, is characterized by high variability, monitoring patterns of development of the past and local housing stocks.

4. Network Analysis Community structure.

4.1 Methodology of Network Construction.

The network is used to gauge all block groups by exploring their proximity in the form of Euclidean distance on the seven standardized variables, and so each block group is connected to its five nearest neighbors. Thirty-two block groups that were created through quality checks and standardization were used to create the final graph on which community detection occurred. Louvain algorithm revealed three communities and the modularity score that was achieved was 0.336 which shows that there exists a moderate level of cluster structure.

4.2 Evidence of Dunning being a Cohesive Community.

Although the internal variation was present, the analysis does reveal that Dunning has very distinct external boundaries. Internal similarity averages 0.314, and the similarity to the neighboring areas is practically zero in the modeled feature space. This produces an infinite cohesion ratio which means that Dunning has internal ties that are meaningfully stronger than the external ties. Comparisons of variances also indicate that variability inside Dunning is, relative, less than the set of neighboring areas that are used as a reference.

4.3 Evidence to the contrary of the Single-Community Concept.

Meanwhile, the network does not represent Dunning as one, solid cluster. The block groups by Dunning are spread among all the three identified communities and not one network community covers the whole area. There are block groups that are more similar to places beyond their official territory than they are to other regions of Dunning. These internal fault lines are consistent with income disparities, racial frame work and tenure and they generate sub-areas that act as separate neighborhoods within the broader society.

4.4 Data-Based Alternative Community Proposal.

The identified structure contributes to the three-part organization. There is one sub-community with twelve block groups, moderate incomes and mixed composition; and there is a second group with less than twelve block groups, higher income and homeownership and a third group with less than nine block groups, more diverse and variable in income and tenure. Both of the suggested sub-communities are more internally similar than the existing single area definition which implies a more statistically valid definition and more nearly resembles patterns of neighborhoods actually lived in.

5. Spatial Patterns & Visual Analytics.

Findings in the network are reinforced by maps. The southeast has more mixed and poorer populations whereas the northwest is more owner-occupied and is higher-income. An evident northwest to southeast income gradient is in line with demographic composition variance. The visualization of the network shows that there are three different clusters with limited number of cross-links and the spatial arrangement of the cluster membership follows the demographic gradients of the maps. Boundary strength is particularly apparent in that cluster membership changes abruptly at the Dunning edge that proves the external boundaries of the community are out of place.

6. AI Transcript

Q: "I'm trying to load the community area boundaries but keep getting CRS warnings. The shapefile loads but when I try to plot with census data, nothing aligns properly."

AI: "The issue is coordinate reference system mismatch. Your community areas are in WGS 84 (EPSG:4326) while the census data might be in a different projection. Use st_transform() to ensure both datasets use the same CRS:"

```
dunning_sf <- st_transform(dunning_sf, st_crs(cook_bg_2020))
cook_bg_2020 <- st_transform(cook_bg_2020, st_crs(dunning_sf))
```

Q: My network analysis is failing because some block groups have missing data. The similarity matrix has NA values that break the distance calculation."

AI: "You need to handle missing values before the network construction. Either impute or remove incomplete cases:"

```
analysis_data_clean <- analysis_data %>%
  select(GEOID, total_popE, pct_white, pct_black, pct_hispanic,
         income_scaled, hh_sizeE, pct_owner, area_type) %>%
  drop_na()

analysis_data_imputed <- analysis_data %>%
  mutate(across(where(is.numeric), ~ifelse(is.na(.), median(., na.rm = TRUE), .)))
```

Q: When I combine Dunning with neighbors, I'm getting duplicate block groups along the boundaries."

AI: "This is a common issue with boundary overlaps. Use st_intersection() carefully and remove duplicates:"

```
# Clean approach for combining areas
study_bg_2022 <- neighbor_bg_2022 %>%
  st_intersection(study_area) %>%
  distinct(GEOID, .keep_all = TRUE)  # Remove duplicate
```

Q: I'm getting inconsistent results because some block groups have zero population but still have demographic counts. How should I handle these cases?"

AI: "You need to implement robust data cleaning for demographic percentages. Here's a safe approach:"

```
dunning_bg_acs_2022 <- dunning_bg_acs_2022 %>%
  mutate(
    # Check for valid population before calculations
```

```
  valid_population = total_popE > 0,
  pct_white = ifelse(valid_population, (race_whiteE / total_popE) * 100, NA),
  pct_black = ifelse(valid_population, (race_blackE / total_popE) * 100, NA),
  pct_hispanic = ifelse(valid_population, (hispanicE / total_popE) * 100, NA),
  # Handle edge cases with small populations
  pct_white = ifelse(pct_white > 100, 100, pct_white),
  pct_black = ifelse(pct_black > 100, 100, pct_black),
  pct_hispanic = ifelse(pct_hispanic > 100, 100, pct_hispanic)
 ) %>%
 # Remove block groups with no population or invalid data
 filter(valid_population & !is.na(med_incomeE) & med_incomeE > 0)

cat("Removed", nrow(raw_data) - nrow(dunning_bg_acs_2022), "invalid block groups\n")
```

7. References

[1] U.S. Census Bureau. (2020). Decennial Census 2010 and 2020 [Data set].

[2] U.S. Census Bureau. (2022). American Community Survey 2018-2022 5-year estimates [Data set].

[3] City of Chicago. (2023). Community Area Boundaries [GIS data].

[4] Csardi, Gabor & Nepusz, Tamas. (2005). The Igraph Software Package for Complex Network Research. InterJournal. Complex Systems. 1695.

[5] Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis.