Community Preservation in Iowa: Using Clustering to Identify Communities of Interest

Jayson Kunkel, Dr. Eric Autry

Grinnell College

Introduction

In this work, we consider Communities of Interest (COIs) in Iowa at the census tract level. We define COIs to be regions with similar demographic, social, or economic characteristics, and those with shared services, resources, or amenities, which should be preserved during electoral redistricting. We initially aim to represent COIs by collecting census data and clustering census tracts with similar statistics. We then generate heat maps of census tracts that appear in multiple clusters, like those created by Chambers et al., to identify the most similar regions.

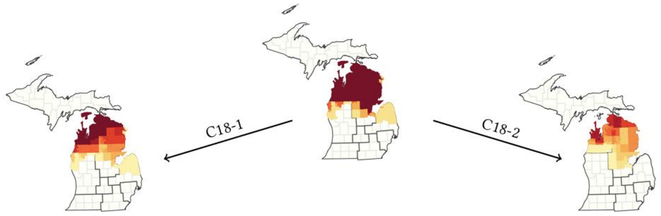

Figure 1: A heat map of COIs in Michigan split into subclusters. (Chambers et al., 2022)

Summary Statistics

Referencing the Delaware Valley Regional Planning Commission (DVRPC) paper, we selected statistics from the American Community Survey (ACS) which reasonably describe a community, including percentages for race and ethnicities, age and sex cohorts, educational attainment, families below the poverty level, and industry; median estimates for household income and benefits, monthly rent, and home value; and mean commute time in minutes. Due to time constraints, we did not analyze every available statistic.

Identifying Regional Structures

For the initial clustering, we performed the following steps:

1. Create a dual graph of the census tracts, and for each tract, generate a list of its neighbors. Only neighboring tracts will be clustered together.

2. Following the DVRPC paper, gather summary statistics from the most recent ACS 5-Year data at the census tract level.

3. For statistics with a roughly Gaussian distribution, calculate the z-score for each observation. For a rough power law distributions, find p, the 85th percentile of the statistic, and label observations greater than p as 1, and lower than p as 0. In both cases, perform single linkage flat clustering on these observations with a threshold of 0.5.

Our initial clustering produced maps of similar regions based on the given statistic, where each color is a distinct cluster. We removed large statewide clusters as these do not accurately represent communities of interest.

Clustering Results for Regional Structures

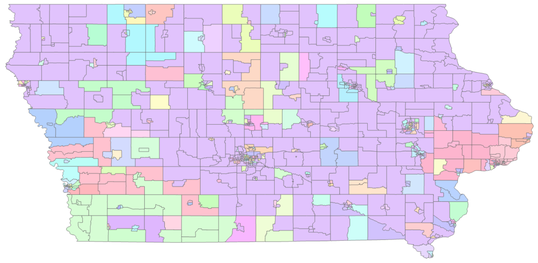

Figure 2: Iowa flat clusters for industry: Transportation, Warehousing, and Utilities.

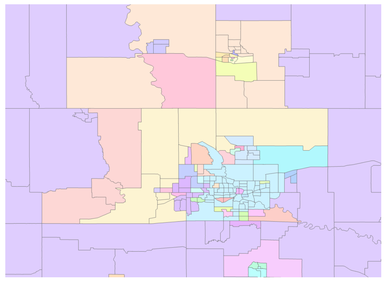

Figure 3: Des Moines area flat clusters for median household income.

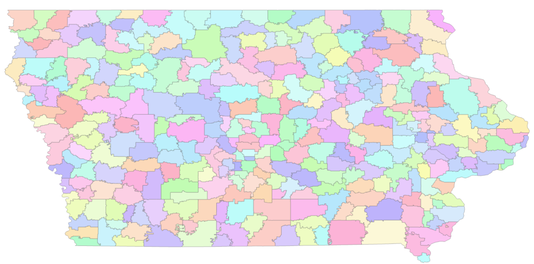

Figure 4: A map of Iowa school districts.

Clustering Regions into Heat Maps

To generate a heat map, we performed the following steps:

1. Obtain geographical data such as school districts, electrical service boundaries, and municipalities that represent communities.

2. Record the census tracts which overlap with polygons representing each initial statistic cluster and spatially-defined region.

3. Perform complete-linkage flat clustering using a Modified Hausdorff Distance, the average of the pairwise minimum path distances in the dual graph between the tracts contained within each region.

4. Create a map of an individual cluster. Add a graduated symbology where each tract is colored according to the number of times it appears in the cluster. This creates a heat map where more popular tracts are darker.

Results of Heat Map Clustering

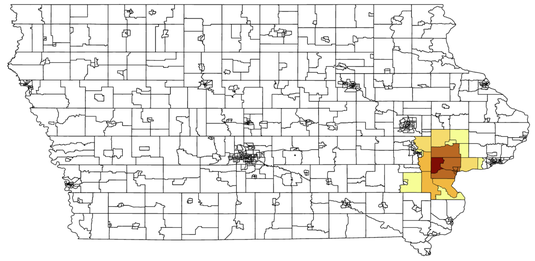

Figure 5: The Muscatine cluster.

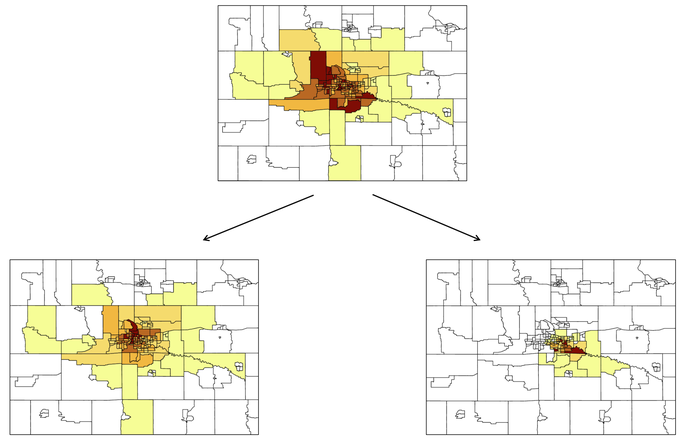

Figure 6: A heat map of Des Moines split into subclusters.

Future Work

We aim to include more summary statistics and geographically defined regions for a more comprehensive representation of COIs. We will consider the current statistics which do not roughly follow the Gaussian or power law distribution. We also plan to refine our process to work with smaller geographies like census blocks or block groups to better identify and represent COIs. Finally, we will experiment with different clustering thresholds, allowing us to explore subcommunities and better define COIs.

Acknowledgments

We thank Dr. Gregory Herschlag for his contributions to our research, and to Grinnell College for providing computing resources.

References

Erin Chambers, Moon Duchin, Ranthony A.C. Edmonds, Parker Edwards, JN Matthews, Anthony E. Pizzimenti, Chanel Richardson, Parker Rule, and Ari Stern (2022). “Aggregating Community Maps.” doi:10.1145/3557915.3560961

Delaware Valley Regional Planning Commission (2020). “Demographic Multipliers and Summary Statistics for Residents of Multifamily Housing (2016).” https://catalog.dvrpc.org/dataset/demographic-multipliers-and-summary-statistics-for-residents-of-multifamily-housing-2016