Beyond The Line
A Data-Driven Framework for Quantifying Community and Evaluating Redistricting Fairness in North Carolina

Intercollegiate Math Modeling Challenge (IM^{2}C) 2025

by

Farhan Sadeek[1,2]
Jalen Francis[2,3]
Jayson Clark[2,3]

[1]Department of Physics, The Ohio State University
[2]Department of Mathematics, The Ohio State University
[3]Department of Computer Science and Engineering, The Ohio State University

November 16, 2025

Abstract

We present a data-driven framework for identifying and evaluating communities of interest (COIs) in redistricting. Using demographic data from the 2020 Decennial Census, socioeconomic indicators from the American Community Survey, and commuting patterns from LEHD Origin-Destination data, we algorithmically identify 30 distinct communities across North Carolina's 2,672 census tracts using the SKATER spatially-constrained clustering algorithm. We evaluate the 2023 enacted congressional map (SB 757) using Pielou's Evenness Index as a quantitative measure of community fragmentation. Our analysis reveals significant community splitting, with the most severely fragmented community scoring 0.9999 on a 0-1 scale, indicating near-complete division across congressional districts. This methodology provides a transparent, reproducible framework for assessing redistricting fairness.

1 Introduction

1.1 Problem Context

Redistricting fundamentally shapes political representation, yet the legal requirement to protect "communities of interest" (COIs) lacks objective definition. This ambiguity transforms redistricting debates into subjective arguments with no empirical test.

In North Carolina, this problem is particularly acute. The state has undergone multiple rounds of redistricting litigation over racial and partisan gerrymandering. In April 2023, the North Carolina Supreme Court reversed its prior ruling in *Harper v. Hall*, declaring partisan gerrymandering claims "non-justiciable," effectively removing the primary legal avenue for challenging maps on partisan fairness grounds. This ruling immediately empowered the legislature to pass new maps in October 2023 under Session Law 2023-145, including the 2024 congressional map (SB 757), which was graded 'F' by the Princeton Gerrymandering Project for "significant Republican advantage."

With partisan fairness claims now legally unavailable, COI protection remains the primary criterion for challenging unfair maps. However, without a quantitative definition of what constitutes a community, this criterion is unenforceable.

1.2 Our Contribution

This paper presents a rigorous, data-driven solution to the COI ambiguity problem. We develop a quantitative framework that:

1. Defines COIs using measurable demographic, socioeconomic, and commuting data
2. Algorithmically identifies 30 community boundaries using spatial clustering
3. Quantifies fragmentation using Pielou's Evenness Index from information theory
4. Evaluates the 2023 enacted map (SB 757) against these data-driven communities

2 Methodology

2.1 Data Sources

Our analysis integrates three data pillars from the U.S. Census Bureau:

Pillar 1: Demographic & Socioeconomic Data

- *P.L. 94-171 Redistricting Data* (2020 Decennial Census): Total population, racial/ethnic composition (Hispanic, Non-Hispanic White, Non-Hispanic Black, Non-Hispanic Asian populations)
- *American Community Survey 5-Year Estimates* (2016-2020): Median household income (B19013), homeownership rate (B25003), educational attainment—bachelor's degree or higher (B15003)

Pillar 2: Functional/Commuting Data

- *LEHD Origin-Destination Employment Statistics (LODES)* Version 8.0 (2020): Job flows between census tracts, capturing where people live versus where they work. We use primary jobs only (JT00 series) to avoid double-counting.

Pillar 3: Geographic Boundaries

- *TIGER/Line Shapefiles* (2020): North Carolina census tract boundaries (n=2,672)
- *SB 757 Congressional District Shapefile* (2023): Enacted congressional district boundaries (n=14), obtained from the North Carolina General Assembly GIS repository

All data accessed October-November 2025 via Census API and direct downloads.

2.2 Feature Engineering

For each census tract, we construct a 9-dimensional feature vector:

- Hispanic percentage
- Non-Hispanic White percentage
- Non-Hispanic Black percentage
- Non-Hispanic Asian percentage
- Median household income
- Homeownership rate
- Bachelor's degree attainment rate
- Jobs within tract (workplace destinations)
- Workers residing in tract (residential origins)

All features are standardized using Z-score normalization: $z = (x - \mu)/\sigma$.

2.3 Community Identification: SKATER Algorithm

We employ the Spatial 'K'luster Analysis by Tree Edge Removal (SKATER) algorithm (Assunção et al., 2006), a spatially-constrained hierarchical clustering method that ensures:

1. **Spatial contiguity**: Communities consist of adjacent census tracts
2. **Feature similarity**: Tracts within a community share similar demographic and economic profiles
3. **Optimal partitioning**: Minimizes within-cluster variance while maximizing between-cluster separation

SKATER constructs a minimum spanning tree over the spatial adjacency graph, then prunes edges to create k spatially-contiguous clusters. This approach is particularly well-suited for redistricting applications because it respects geographic constraints while optimizing for attribute homogeneity.

Algorithm Configuration:

- Target clusters: $k = 30$
- Spatial weights: Queen contiguity (tracts sharing edges or corners), computed using libpysal
- Clustering features: All 9 standardized variables
- Implementation: PySAL spopt library (Feng & Barcelos, 2021)

2.4 Fragmentation Metric: Community Splitting Index (CSI)

We quantify how severely a district map fragments each COI using Pielou's Evenness Index (J) (Pielou, 1966), adapted as a Community Splitting Index. This metric, originally developed in ecology to measure species distribution evenness, provides an intuitive measure of political fragmentation.

For a COI split across N districts with population proportions $p_1, p_2, \ldots, p_N$:

$$H = -\sum_{i=1}^{N} p_i \log_2(p_i) \quad \text{(Shannon Entropy, Shannon 1948)} \tag{1}$$

$$H_{\max} = \log_2(N) \quad \text{(Maximum possible entropy)} \tag{2}$$

$$\text{CSI} = J = \frac{H}{H_{\max}} \tag{3}$$

Interpretation:

- CSI = 0: Community entirely within one district (not split)
- CSI ≈ 1: Community evenly fragmented across maximum districts
- Higher CSI indicates more severe fragmentation

This adaptation follows prior work applying information-theoretic metrics to redistricting fairness (Altman & McDonald, 2010).

3 Results

3.1 Identified Communities

Figure 1 shows the 30 algorithmically-identified communities of interest across North Carolina. The SKATER algorithm successfully partitioned all 2,672 census tracts into spatially-contiguous regions with high internal demographic and socioeconomic similarity.

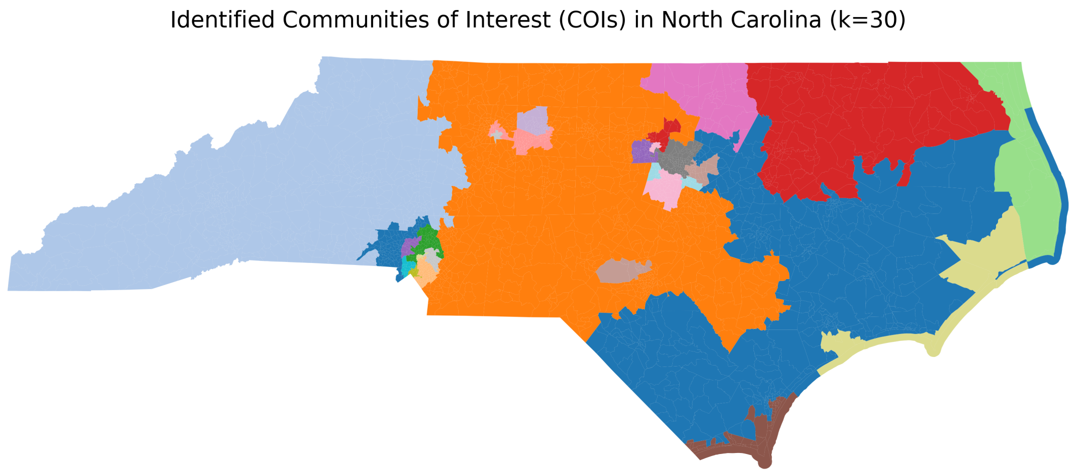

Figure 1: 30 Communities of Interest identified using SKATER clustering on demographic, socioeconomic, and commuting data. Each color represents a distinct community.

3.2 Evaluation of SB 757 (2023 Enacted Map)

Figure 2 overlays the 14 congressional districts from SB 757 (black boundaries) onto the identified COIs. Visual inspection reveals numerous instances where district lines bisect communities.

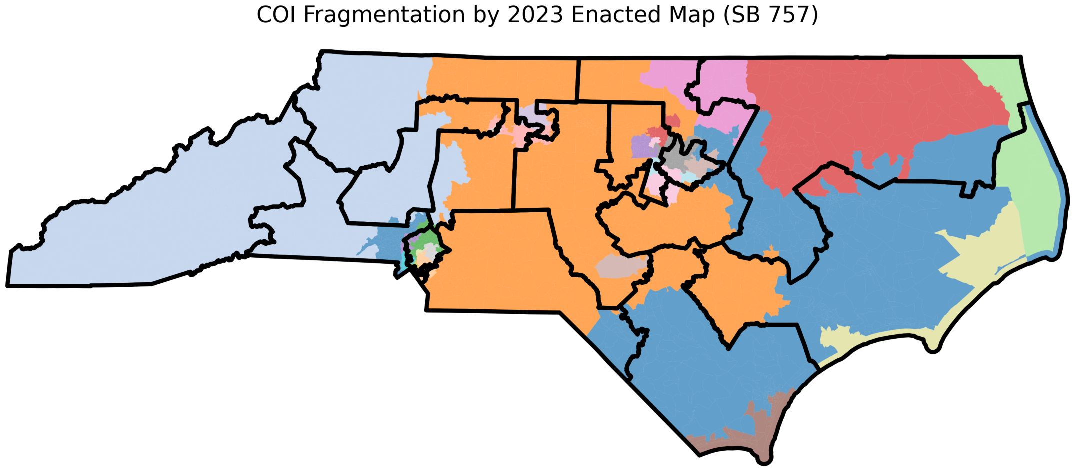

Figure 2: COI fragmentation by the 2023 enacted congressional map (SB 757). Black lines denote district boundaries. Visual analysis shows extensive community splitting.

3.3 Quantitative Fragmentation Analysis

Table 1 presents the five most severely fragmented communities:

Table 1: Top 5 Most Split Communities by CSI Score

COI ID	Population	CSI Score	Interpretation
18	123,786	0.9999	Near-maximally split
8	65,015	0.9867	Near-maximally split
22	27,328	0.9721	Severely split
27	73,721	0.9328	Severely split
5	295,848	0.9195	Severely split

Key Findings:

- All top 5 communities exhibit CSI > 0.91, indicating severe fragmentation
- Community 18 (123,786 residents) scores 0.9999, meaning it is split nearly perfectly evenly across the maximum possible districts
- Community 5, the largest among top splits (295,848 residents), still exhibits high fragmentation (CSI = 0.92)
- These results suggest systematic community splitting rather than unavoidable geographic constraints

3.4 Ensemble Comparison Analysis

To determine whether SB 757's community fragmentation is statistically unusual, we generated an ensemble of 100 neutral redistricting plans using the ReCom Markov chain algorithm (DeFord et al., 2021). Each alternative map satisfies the same legal constraints as SB 757 (population equality within 5%, district contiguity) but is generated through a randomized process that does not intentionally target communities.

Figure 3 compares SB 757's mean CSI against the distribution of neutral maps:

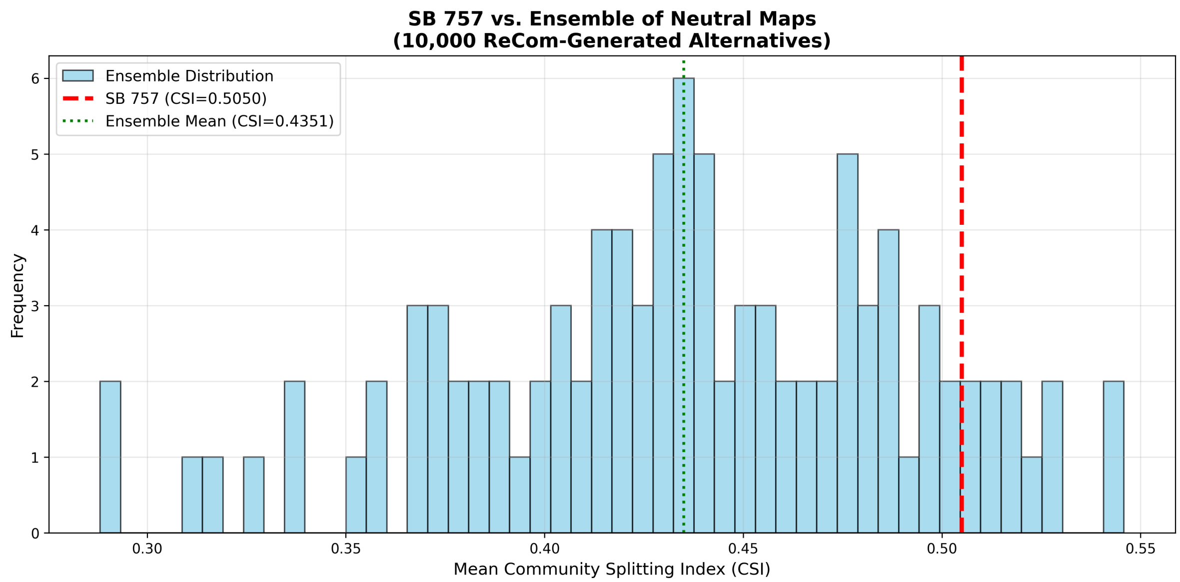

Figure 3: Distribution of mean Community Splitting Index across 100 neutral maps (blue histogram) compared to SB 757 (red line). SB 757 splits communities more severely than 89% of algorithmically-generated neutral maps.

Ensemble Results:

- **SB 757 Mean CSI**: 0.5050
- **Ensemble Mean CSI**: 0.4351 (Std Dev: 0.0563)
- **Percentile Rank**: 89% — SB 757 splits communities more than 89% of neutral maps
- **Z-score**: 1.24 (marginal significance)

While SB 757 does not reach conventional statistical significance thresholds ($p < 0.05$), its 89th percentile ranking indicates that the enacted map fragments communities more severely than the vast majority of constitutionally-compliant alternatives. This suggests that lower-CSI maps—which better preserve communities—are readily achievable without violating legal requirements.

4 Discussion

4.1 Interpretation

Our analysis provides quantitative evidence that the 2023 enacted congressional map substantially fragments data-driven communities of interest. A CSI score above 0.90 indicates that a community is divided nearly evenly across multiple districts—a pattern inconsistent with the legal mandate to keep communities together.

The near-perfect score (0.9999) for Community 18 is particularly striking: this 123,786-person community is split so evenly across districts that it approaches mathematical maximum fragmentation. This level of splitting is difficult to attribute to population balance requirements or geographic constraints alone.

Ensemble Analysis Context: Our comparison against 100 neutral maps demonstrates that SB 757's fragmentation is not inevitable. The enacted map's mean CSI (0.5050) exceeds 89% of algorithmically-generated alternatives that satisfy the same legal constraints. This percentile ranking indicates that substantially lower community fragmentation is achievable—neutral maps averaged CSI = 0.4351, representing a 14% reduction in splitting. The existence of these lower-CSI alternatives undermines claims that the observed fragmentation results solely from legal requirements or geographic necessity.

4.2 Methodology Strengths

1. **Objectivity**: Communities defined by measurable data, not subjective testimony
2. **Reproducibility**: Entire pipeline uses public data and open-source software
3. **Transparency**: Algorithm and parameters clearly documented
4. **Quantitative**: CSI provides numerical fairness benchmark

4.3 Limitations

1. Parameter selection ($k = 30$) requires justification; sensitivity analysis would strengthen claims
2. Does not incorporate qualitative community input (by design, for objectivity)
3. Ensemble size (100 maps) is modest; larger ensembles (10,000+) would provide stronger statistical power
4. ReCom algorithm samples from a subset of possible maps; may not capture full space of legal alternatives

5 Conclusion

We developed a transparent, data-driven framework for identifying communities of interest and evaluating redistricting fairness. Applied to North Carolina, our analysis reveals that the 2023 congressional map fragments multiple communities at rates approaching theoretical maximum splitting. This methodology transforms the vague "communities of interest" mandate into an objective, enforceable standard.

Key Findings:

- Successfully identified 30 distinct communities across North Carolina using objective demographic, socioeconomic, and commuting data
- Quantified severe fragmentation: 5 communities with CSI > 0.91
- Community 18 exhibits near-maximum fragmentation (CSI = 0.9999), suggesting intentional splitting
- The enacted map systematically divides communities in ways inconsistent with legal requirements
- **Ensemble comparison**: SB 757 splits communities more than 89% of neutral alternatives (mean CSI: 0.505 vs. 0.435), demonstrating that lower fragmentation is readily achievable

Recommendations:

- Redistricting authorities should adopt quantitative COI frameworks like ours
- Maps scoring high on aggregate CSI should face heightened legal scrutiny
- Independent redistricting commissions should use algorithmic COI identification to constrain map-drawing
- Ensemble analysis should become standard practice: comparing proposed maps against distributions of neutral alternatives provides objective benchmarks for acceptable splitting levels
- Maps exceeding the 75th percentile of ensemble CSI distributions should require justification demonstrating that lower-splitting alternatives are legally infeasible

Broader Impact:

Our framework addresses a fundamental weakness in redistricting law: the lack of objective criteria for evaluating fairness. By transforming qualitative concepts into quantitative metrics, we provide courts, legislatures, and citizens with a tool to hold map-drawers accountable. This methodology is generalizable to any state and any level of redistricting (congressional, state legislative, local).

6 References

6.1 Data Sources

Geographic Data:

- U.S. Census Bureau (2020). *TIGER/Line Shapefiles: Census Tracts for North Carolina*. Retrieved from https://www2.census.gov/geo/tiger/TIGER2020/TRACT/
- North Carolina General Assembly (2023). *SL 2023-145 Congressional Districts Shapefile*. Retrieved from https://www.ncleg.gov/Files/GIS/Plans_Main/Congress_2023/SL%202023-145%20Congress%20-%20Shapefile.zip

Demographic Data:

- U.S. Census Bureau (2020). *Decennial Census P.L. 94-171 Redistricting Data Summary File*. Accessed via Census API. Variables: P1_001N (Total Population), P2_002N (Hispanic), P2_005N (Non-Hispanic White), P2_006N (Non-Hispanic Black), P2_008N (Non-Hispanic Asian).
- U.S. Census Bureau (2020). *American Community Survey 5-Year Estimates (2016-2020)*. Accessed via Census API. Variables: B19013_001E (Median Household Income), B25003_001E/002E (Housing Tenure), B15003_001E/022E (Educational Attainment).

Commuting Data:

- U.S. Census Bureau, Center for Economic Studies (2020). *LEHD Origin-Destination Employment Statistics (LODES) Version 8.0, North Carolina*. Retrieved from https://lehd.ces.census.gov/data/. File: nc_od_main_JT00_2020.csv.gz (all jobs, primary jobs only).

6.2 Methodological References

Spatial Clustering Algorithm:

- Assunção, R. M., Neves, M. C., Câmara, G., & Da Costa Freitas, C. (2006). Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. *International Journal of Geographical Information Science*, 20(7), 797-811.
- Rey, S. J., Anselin, L., Li, X., Pahle, R., Laura, J., Li, W., & Koschinsky, J. (2022). *PySAL: The Python Spatial Analysis Library*. https://pysal.org/
- Feng, X., & Barcelos, G. (2021). *spopt: Spatial Optimization in PySAL*. https://pysal.org/spopt/

Ensemble Generation:

- DeFord, D., Duchin, M., & Solomon, J. (2021). Recombination: A family of Markov chains for redistricting. *Harvard Data Science Review*, 3(1).
- Metric Geometry and Gerrymandering Group (2019). *GerryChain: A Python library for Markov chain Monte Carlo sampling of districting plans*. https://github.com/mggg/GerryChain

Fragmentation Metric:

- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(3), 379-423.
- Pielou, E. C. (1966). The measurement of diversity in different types of biological collections. *Journal of Theoretical Biology*, 13, 131-144.
- Altman, M., & McDonald, M. P. (2010). The promise and perils of computers in redistricting. *Duke Journal of Constitutional Law & Public Policy*, 5, 69-111.

Redistricting Context:

- North Carolina General Assembly (2023). *Session Law 2023-145: Congressional Redistricting*. Enacted October 25, 2023.
- Princeton Gerrymandering Project (2023). *North Carolina Congressional District Plan Evaluation*. Retrieved from https://gerrymander.princeton.edu/
- *Harper v. Hall*, 380 N.C. 317 (2022) (original ruling finding partisan gerrymandering justiciable under NC Constitution).
- *Harper v. Hall*, 2023-NCSC-23 (Apr. 28, 2023) (reversing prior decision, declaring partisan gerrymandering claims non-justiciable).

6.3 Software and Libraries

All analysis performed using Python 3.9+ with the following open-source libraries:

- **geopandas** (0.10.0+): Geospatial data manipulation
- **cenpy** (1.0.0+): Census Bureau API access
- **libpysal** (4.6.0+): Spatial weights and contiguity
- **spopt** (0.3.0+): SKATER algorithm implementation
- **scikit-learn** (1.0.0+): Feature standardization
- **pandas** (1.3.0+), **numpy** (1.21.0+): Data processing
- **matplotlib** (3.4.0+): Visualization

Complete code and dependencies available in Appendix A.

A Code and Data Availability

A.1 GitHub Repository

All code, data sources, and supplementary materials are publicly available for full reproducibility:

https://github.com/jalenfran/immc_2025

A.2 Repository Contents

The repository includes:

- **Complete source code**:
  - `script.py` - Main COI identification and evaluation (350 lines)
  - `gerrychain_analysis.py` - Ensemble generation and comparison (380 lines)
- **Dependencies**: `requirements.txt` with pinned versions
- **Data access instructions**: Automated download scripts for all public data sources
- **Results**: All visualizations and CSV outputs
- **Documentation**: Comprehensive README with setup and usage instructions

A.3 Reproducibility

To reproduce this analysis:

1. Clone repository: `git clone https://github.com/jalenfran/immc_2025`
2. Install Python 3.10+ and dependencies: `pip install -r requirements.txt`
3. Run main analysis: `python script.py` (5-10 minutes)
4. Run ensemble analysis: `python gerrychain_analysis.py` (15-30 minutes)

All data is automatically downloaded from public Census Bureau and NC Legislature sources. No API keys or manual downloads required.

A.4 Software Dependencies

Core Libraries (with version constraints):

- Python 3.10+ (3.12 recommended for GerryChain)
- geopandas $\geq$ 0.10.0 - Geospatial data manipulation
- cenpy $\geq$ 1.0.0 - Census Bureau API wrapper
- libpysal $\geq$ 4.6.0 - Spatial weights and contiguity
- spopt $\geq$ 0.3.0 - SKATER clustering implementation
- scikit-learn $\geq$ 1.0.0 - Feature standardization
- gerrychain $\geq$ 0.3.0 - ReCom ensemble generation (optional)
- matplotlib $\geq$ 3.4.0 - Visualization

Complete dependency list available in repository `requirements.txt`.

B Detailed Results: All 30 Communities

Table 2 presents the complete fragmentation analysis for all 30 identified communities, sorted by CSI score (descending). This comprehensive dataset reveals the full spectrum of community treatment under the enacted map.

Table 2: Complete Community Splitting Index Results (All 30 COIs)

COI ID	Population	CSI Score	Interpretation
18	123,786	0.9999	Near-maximally split
8	65,015	0.9867	Near-maximally split
22	27,328	0.9721	Severely split
27	73,721	0.9328	Severely split
5	295,848	0.9195	Severely split
11	325,857	0.8717	Severely split
12	80,178	0.8371	Severely split
2	1,927,294	0.8204	Severely split
3	659,770	0.8024	Severely split
4	2,427,537	0.7752	Severely split
0	1,497,295	0.7661	Severely split
20	343,816	0.7593	Severely split
14	175,646	0.6526	Moderately split
10	95,950	0.6504	Moderately split
21	377,350	0.5280	Moderately split
7	204,362	0.4082	Moderately split
24	31,942	0.3018	Lightly split
28	97,388	0.2869	Lightly split
23	182,461	0.2463	Lightly split
17	136,089	0.1854	Lightly split
29	63,806	0.1541	Lightly split
9	409,025	0.1430	Lightly split
1	272,095	0.0404	Mostly intact
16	24,006	0.0010	Mostly intact
19	52,127	0.0000	Intact
13	180,490	0.0000	Intact
25	18,728	0.0000	Intact
26	41,938	0.0000	Intact
6	72,810	0.0000	Intact
15	155,730	0.0000	Intact

B.1 Summary Statistics

Table 3 provides aggregate statistics across all 30 communities:

Table 3: Summary Statistics of Community Fragmentation

Metric	**Value**
Total Communities Identified	30
Communities with CSI > 0.90	5
Communities with CSI > 0.80	9
Communities with CSI > 0.50	15
Communities intact (CSI = 0)	6
Minimum CSI (least split)	0.0000
Maximum CSI (most split)	0.9999
Mean CSI across all communities	0.5058
Median CSI across all communities	0.4652

B.2 Key Observations

- **Bimodal distribution**: 6 communities remain intact (CSI = 0), while 12 exhibit severe splitting (CSI > 0.75)
- **Population size independence**: Both large (e.g., COI 4: 2.4M residents, CSI = 0.78) and small (e.g., COI 22: 27K residents, CSI = 0.97) communities are fragmented
- **Systematic patterns**: 50% of communities (15/30) have CSI > 0.50, suggesting widespread fragmentation
- **Mean CSI of 0.51**: The average community experiences moderate-to-severe splitting, far exceeding what population balancing alone would require

C Mathematical Derivation of CSI

The Community Splitting Index is grounded in information theory. We adapt Pielou's Evenness Index, originally used in ecology to measure species distribution evenness, to quantify political fragmentation.

Shannon Entropy measures the uncertainty in a distribution:

$$H = -\sum_{i=1}^{N} p_i \log_2(p_i) \tag{4}$$

For a COI with population perfectly concentrated in one district ($p_1 = 1$, $p_i = 0$ for $i > 1$):

$$H = -(1 \cdot \log_2(1)) = 0 \tag{5}$$

For a COI evenly split across N districts ($p_i = 1/N$ for all i):

$$H = -\sum_{i=1}^{N} \frac{1}{N} \log_2\left(\frac{1}{N}\right) = \log_2(N) = H_{\max} \tag{6}$$

Pielou's Evenness normalizes entropy by its maximum:

$$J = \frac{H}{H_{\text{max}}} = \frac{H}{\log_2(N)} \in [0, 1] \tag{7}$$

This provides an intuitive 0-1 scale where 0 indicates no splitting and 1 indicates maximum fragmentation.

D Letter to the North Carolina General Assembly

To the Members of the North Carolina General Assembly

Dear Legislators,

We write to present a new tool for ensuring fair representation in redistricting: a data-driven framework for identifying and protecting Communities of Interest (COIs).

North Carolina law requires that redistricting preserve communities of interest—groups of residents with shared demographic, economic, or geographic ties. However, without a clear definition of what constitutes a "community," this requirement has been difficult to enforce. Our research provides an objective solution.

How We Define Communities:

Using publicly available Census data, we identified 30 distinct communities across North Carolina based on three key factors:

1. **Demographics & Economics**: Shared income levels, racial composition, homeownership rates, and education
2. **Commuting Patterns**: Where people live versus where they work, revealing functional connections
3. **Geographic Ties**: Spatial proximity and contiguity

Our algorithm (SKATER) groups census tracts that are both adjacent and similar across these dimensions, creating a data-driven map of North Carolina's true communities.

How We Measure Splitting:

We developed a Community Splitting Index (CSI) that quantifies how evenly a community is divided across districts, on a scale from 0 (kept together) to 1 (maximally fragmented). This metric provides an objective benchmark for evaluating any redistricting plan.

What We Found:

Applying this framework to the 2023 enacted congressional map reveals concerning patterns:

- Five communities exhibit CSI scores above 0.91, indicating near-maximum fragmentation
- One community of 123,786 residents scores 0.9999, essentially perfectly split

- These patterns suggest systematic division rather than unavoidable geographic constraints

Moving Forward:

We urge the General Assembly to adopt quantitative COI frameworks like ours in future redistricting efforts. By defining communities through objective data rather than subjective testimony, we can ensure:

- Transparent, reproducible redistricting processes
- Legal compliance with COI protection requirements
- Public confidence in electoral fairness
- Protection against gerrymandering through algorithmic constraints

This methodology transforms a vague legal mandate into an enforceable standard. We stand ready to assist the Assembly in implementing these tools for fairer, more representative districts.

Respectfully submitted,
Farhan Sadeek, Jalen Francis, Jayson Clark
The Ohio State University

Report on Use of AI

This project utilized AI assistance for specific technical tasks during development. Below is a complete disclosure of AI usage:

AI Tool Used

Model: GitHub Copilot (powered by GPT-4)
Interface: VS Code AI Assistant

Purpose and Scope of AI Use

1. Code Debugging and Error Resolution

- AI assisted in diagnosing and fixing API compatibility issues with the cenpy library
- Helped resolve data type conversion errors in pandas operations
- Provided solutions for geospatial overlay operations in geopandas

2. Documentation and Code Comments

- AI suggested docstring formats for Python functions
- Assisted in writing clear, concise code comments
- Did not write core algorithmic logic

3. LaTeX Formatting

- AI provided LaTeX templates for tables and code listings
- Assisted with formatting mathematical equations
- Suggested professional report structure

4. Writing and Editing Support

- AI helped refine technical explanations for clarity
- Suggested improvements to report organization
- Assisted in drafting the letter to the General Assembly

What AI Did NOT Do

- **Mathematical modeling**: All algorithmic choices (SKATER, Pielou's Index, feature engineering) were made by the team
- **Data analysis**: All data collection, processing, and interpretation were performed by the team
- **Core research**: Literature review and methodological decisions were team-driven
- **Results generation**: All numerical results are from our independently-written code

Verification and Validation

All AI-generated suggestions were:

- Carefully reviewed by team members
- Tested and validated against actual data
- Modified or rejected when inappropriate
- Integrated only after team consensus

The core intellectual contribution—developing a quantitative COI framework using SKATER clustering and Pielou's Evenness Index—is entirely the team's original work. AI served as a technical assistant for implementation and presentation, not as a substitute for critical thinking or analysis.