A Comparison of Precinct and District Voting Data Using Persistent Homology to Quantify Gerrymandering in Texas & California

Shah, Ananya (School: Edgemont High School)

Gerrymandering poses a significant threat to American democracy. Current identification methods, e.g. Polsby-Popper and Reock scores are highly sensitive to physical geography and unreliable as indicators of gerrymandering. I introduce a topology-based metric that detects gerrymandering through voting data rather than district geometry. Using publicly available precinct-level voting data and open-source computational tools, I extend Feng & Porter's 2019 topological framework, which applies the levelset method to transform geographic shapefiles into filtered simplicial complexes. I compare the persistent features of Democratic voting regions at the precinct level against those same votes aggregated to proposed district boundaries, identifying when areas have been 'cracked' (divided across multiple districts) or 'packed' (concentrated into a single district) for partisan gain. Because precincts are too small to be gerrymandered, discrepancies between precinct and district-level topological features serve as an indicator of manipulation. This framework was applied to the Texas and California mid-cycle district maps. The Texas plan was ruled a racial gerrymander but was upheld by the Supreme Court. The California map was passed by voter referendum (Proposition 50). Comparing persistence barcodes via Bottleneck and Wasserstein distance (standard metrics for comparing topological features) reveals that in both states, the proposed district boundaries produce significantly different topological features than the underlying precinct data, demonstrating that discrepancies result from manipulation rather than changes in voter behavior. This research provides courts, commissions, & citizens with a tool for evaluating the fairness of their legislative districts.

Awards Won:
Fourth Award of $600