Using Algorithms to Detect Gerrymandering and Improve Legislative Redistricting

Kosuke Imai

Harvard University

Department of Political Science, Ohio State University

February 8, 2024

Joint work with Christopher Kenny, Cory McCartan, Tyler Simko, Shiro Kuriwaki, George Garcia III, Kevin Wang, and Melissa Wu
Motivation

- Today’s world for quantitative social science:
  1. increasing availability of granular data
  2. rapid methodological advancement
- Social scientists can and should solve problems of the real world!
- Redistricting as a major policy decision
- How can we use data and algorithms to evaluate redistricting plans?
  - traditional methods: comparison across states and time periods
  - confounded by state-specific political geography and rules
- Use of simulation algorithms
  1. obtain a representative sample of redistricting plans under constraints
  2. compare the enacted plan with this baseline distribution
- Technological solution to detecting gerrymandering
- Tool for analyzing redistricting
ALgorithm-Assisted Redistricting Methodology (ALARM)

What we do:
1. develop efficient and flexible simulation algorithms
2. build open-source software packages for the entire workflow
3. evaluate redistricting plans in the United States and elsewhere

Goal: empower researchers, policy makers, data journalists, and citizen data scientists with powerful tools
Redistricting Basics

- Classic gerrymandering strategies: packing and cracking

![Even distribution](image1)  
**Even distribution**  
2 red, 2 blue

![Packing](image2)  
Packing  
1 red, 3 blue

![Cracking](image3)  
Cracking  
3 red, 1 blue

- What has changed:
  - availability of granular data
  - mapping software (e.g., Maptitude, Dave’s Redistricting app)

- US Congressional redistricting
  - racial gerrymandering: *Allen v. Milligan*
  - partisan gerrymandering: *Rucho v. Common Cause*
Why Use Simulation Algorithm for Redistricting Evaluation?

- Traditional redistricting evaluation
  1. compute various fairness metrics
  2. compare them across states and over time
- Confounded by differences in political geography and redistricting rules

- Simulation-based redistricting evaluation
  1. generate many alternative plans under a set of redistricting criteria
  2. compare them with a proposed plan to evaluate its properties

- Benefits of simulation approach
  1. can control for state-specific political geography and redistricting rules
  2. transparency and ability to isolate a relevant factor
  3. mathematical properties as representative sample of alternative plans
Sequential Monte Carlo (SMC) Algorithm (McCartan and Imai, 2023)

- Start with a blank state **in parallel**, use the spanning tree approach to sample a district at a time, **resample with weights** at each step

- Advantage: unlike MCMC, sampled plans are nearly independent
- Limitation: hard to incorporate plan-wide or region-specific constraints
The SMC Algorithm

Splitting off a district using a spanning tree

1. random generation of spanning trees (Wilson’s algorithm)
2. computing the number of spanning trees

Target distribution:

$$\pi(\xi) \propto \tau(\xi)^\rho \exp(-J(\xi)) \times 1_{\xi \text{ connected}} \times 1_{\text{contiguity}} \times 1_{\text{equal population}}$$

for a given plan $\xi$ where

$$\tau(\xi)^\rho = \left[ \prod_{i=1}^{n} \tau(G_i(\xi)) \right]^\rho \approx C_1 \exp(-C_2 \rho \ \text{rem}(\xi))$$

fraction of edges removed
Reducing the Number of County Splits

1. Identify county borders
2. Draw a spanning tree in each of \( K \) counties
3. Create a quotient multigraph
4. Choose \( K - 1 \) edges to connect \( K \) spanning trees
SMC: 1,000 sampled plans of 11 districts on 2,465 units
`adapt_k_thresh`=0.985 • `seq_alpha`=0.5
`est_label_mult`=1 • `pop_temper`=0.01

Plan diversity 80% range: 0.82 to 0.98

R-hat values for summary statistics:

<table>
<thead>
<tr>
<th></th>
<th>pop_overlap</th>
<th>comp</th>
<th>dem</th>
<th>e_dem</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.0234</td>
<td>1.0112</td>
<td>1.0053</td>
<td>1.0042</td>
</tr>
</tbody>
</table>

Sampling diagnostics for SMC run 1 of 4 (250 samples)

<table>
<thead>
<tr>
<th></th>
<th>Eff. samples (%)</th>
<th>Acc. rate</th>
<th>Log wgt. sd</th>
<th>Max. unique</th>
<th>Est. k</th>
</tr>
</thead>
<tbody>
<tr>
<td>Split 1</td>
<td>242 (97.0%)</td>
<td>20.6%</td>
<td>0.36</td>
<td>245 (98%)</td>
<td>10</td>
</tr>
<tr>
<td>Split 2</td>
<td>240 (95.8%)</td>
<td>31.2%</td>
<td>0.43</td>
<td>193 (77%)</td>
<td>6</td>
</tr>
<tr>
<td>Split 3</td>
<td>233 (93.4%)</td>
<td>21.8%</td>
<td>0.49</td>
<td>199 (80%)</td>
<td>8</td>
</tr>
<tr>
<td>Split 4</td>
<td>231 (92.3%)</td>
<td>29.9%</td>
<td>0.56</td>
<td>196 (78%)</td>
<td>5</td>
</tr>
<tr>
<td>Split 5</td>
<td>219 (87.6%)</td>
<td>36.1%</td>
<td>0.62</td>
<td>195 (78%)</td>
<td>3</td>
</tr>
<tr>
<td>Split 6</td>
<td>213 (85.0%)</td>
<td>44.9%</td>
<td>0.67</td>
<td>191 (76%)</td>
<td>2</td>
</tr>
<tr>
<td>Split 7</td>
<td>224 (89.7%)</td>
<td>15.9%</td>
<td>0.59</td>
<td>189 (76%)</td>
<td>7</td>
</tr>
<tr>
<td>Split 8</td>
<td>227 (90.8%)</td>
<td>24.2%</td>
<td>0.59</td>
<td>192 (77%)</td>
<td>4</td>
</tr>
<tr>
<td>Split 9</td>
<td>227 (90.9%)</td>
<td>16.9%</td>
<td>0.60</td>
<td>181 (72%)</td>
<td>3</td>
</tr>
<tr>
<td>Split 10</td>
<td>228 (91.3%)</td>
<td>3.8%</td>
<td>0.58</td>
<td>174 (70%)</td>
<td>2</td>
</tr>
<tr>
<td>Resample</td>
<td>166 (66.4%)</td>
<td>NA%</td>
<td>0.59</td>
<td>183 (73%)</td>
<td>NA</td>
</tr>
</tbody>
</table>
Validation

- Divide a $6 \times 6$ grid into 6 equal-sized districts
- Enumerate 451,206 plans (out of 356 billion)
- Number of edge removed as a target statistic

100 samples per run

1,000 samples per run

10,000 samples per run
50 State Redistricting Simulations Project

Comprehensive project to simulate alternative congressional redistricting plans for all fifty states.

- tidied 2020 Census plus statewide election data from the VEST
- collect state-specific redistricting requirements
- construct algorithmic constraints based on these and traditional redistricting criteria
- 5,000 simulation plans based on SMC
- code and data are available at the Harvard Dataverse
Georgia Example

- 14 Congressional districts
- According to Georgia’s House Legislative and Congressional Reapportionment Committee, districts must:
  1. be contiguous
  2. have equal populations
  3. be geographically compact
  4. preserve county and municipality boundaries as much as possible
  5. avoid the unnecessary pairing of incumbents
- We attempted to account for everything except incumbency constraint
- Voting rights act (VRA) compliance is tricky
Widespread Partisan Gerrymandering Cancels Nationally

Control of redistricting

- Democrats
- Mixed
- Republicans
- Commission
- Court

Difference of seats won under enacted plan compared to non-partisan baseline

- D+2: No difference
- R+2
- R+4

States, ordered by seat difference:

- Texas (38)
- Fla. (28)
- Tenn. (9)
- Utah (4)
- Ohio (15)
- S.C. (7)
- Va. (11)
- Ga. (14)
- Iowa (4)
- Okla. (5)
- Kan. (4)
- Ariz. (9)
- Kan. (4)
- Okla. (5)
- Iowa (4)
- Ga. (14)
- Va. (11)
- S.C. (7)
- Ohio (15)
- Utah (4)
- Tenn. (9)
- Fla. (28)
- Texas (38)
Map of Partisan Gerrymandering
Partisan Gerrymandering Reduces Competitiveness

The graph shows the number of seats won by different average Democratic district vote shares under three scenarios:

1. **Non-partisan simulations**
2. **Enacted plan**
3. **Average Democratic district vote share**

The enacted plan shows a significant skew towards higher vote shares, indicating reduced competitiveness compared to the non-partisan simulations.
Application in the Court: Ohio Congressional Redistricting

- In 2018, Ohio voters passed the constitutional amendment
- In 2020, 16 districts: 4 Democrats and 12 Republicans
- After 2020 Census, the number of seats is reduced to 15 districts
- In 2022, 15 districts: 5 Democrats and 10 Republicans

Simulation analysis

- 5,000 alternative plans
- contiguous and compact districts
- compliant with the Voting Rights Act (Cleveland)
- several complicated splitting constraints
- Section 2(B)(5): out of Ohio’s 88 counties,
  - at least 65 counties should not be split
  - no more than 18 counties can be split no more than once
  - no more than 5 counties can be split no more than twice
The Enacted and Example Simulated Plans

Two-party share

30.0% 40.0% 50.0% 60.0%
Polsby-Popper: the ratio of the district area to the area of a circle with the same perimeter

Edge-removal
# Administrative Boundary Splits

<table>
<thead>
<tr>
<th>Plans of the Form</th>
<th>Counties split once</th>
<th>Counties split twice</th>
<th>Total counties split</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enacted</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Count</th>
<th>Percentage of Plans</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0%</td>
</tr>
<tr>
<td>1</td>
<td>20%</td>
</tr>
<tr>
<td>2</td>
<td>30%</td>
</tr>
<tr>
<td>3</td>
<td>40%</td>
</tr>
<tr>
<td>4</td>
<td>50%</td>
</tr>
<tr>
<td>5</td>
<td>60%</td>
</tr>
<tr>
<td>6</td>
<td>70%</td>
</tr>
<tr>
<td>7</td>
<td>80%</td>
</tr>
<tr>
<td>8</td>
<td>90%</td>
</tr>
<tr>
<td>9</td>
<td>100%</td>
</tr>
</tbody>
</table>

Plan 19 / 26
Expected Number of Republican Seats

The bar chart shows the expected number of Republican seats for various fractions of plans. The y-axis represents the fraction of plans, while the x-axis shows the number of expected Republican seats. The chart indicates that the majority of plans (80%) result in 8 Republican seats, with a smaller fraction resulting in 9 seats. There are no plans that result in 10 or 11 Republican seats.
Cracking: Hamilton County (Cincinnati Area)

Enacted plan

Average across simulated plans

Two-party vote share

30%

40%

50%

60%
Enacted plan

Average across simulated plans

Two-party vote share

30%
40%
50%
60%
Ohio Supreme Court Strikes Down the Enacted Map
Id. at Section 1(C)(3)(a). The above evidence, particularly Dr. Imai’s conclusion that the enacted plan will result in, on average, 2.8 more Republican seats than are warranted, shows that the General Assembly’s decision to shift what could have been—under a neutral application of Article XIX—Democratic-leaning areas into competitive districts, i.e., districts that give the Republican Party’s candidates a better chance of winning than they would otherwise have had in a more compactly drawn district, resulted in a plan that unduly favors the Republican Party and unduly disfavors the Democratic Party.
South Carolina racial gerrymandering case argued on Oct 11, 2023
Served as an expert witness for the plaintiffs
Used simulation to provide evidence that a disproportionately large number of Black voters are packed into District 6

Justice Alito: Did Dr. Imai run a simulation using the political data as well but then decide to shelve it when the results were not favorable to your client?

Ms. Aden: That is not what I believe the record reflects, Your Honor.

Justice Alito: It just never occurred to him that politics might have something to do with this?
Concluding Remarks

- Redistricting matters
  - fair representation and policy outcomes
  - competitiveness of districts and responsiveness
  - political polarization
  - state and local offices, education districts, non-US contexts

- How should we stop gerrymandering?
  - independent commission (e.g., Michigan)
  - use of algorithms to detect gerrymandering

- Role of experts
  - legislative process
  - court testimony
  - work with non-partisan groups

- Open problems
  - large-scale redistricting problems (e.g., state legislatures)
  - algorithm-generated redistricting plan proposals
  - communities of interest, impact of redistricting rules