Using Voter Precinct Data to Create a Partisan Gerrymandering Detection Metric

Sasha Hydrie, Jason Wang, Patrick Gallagher

March 27, 2021

Overview

1. Introduction to Gerrymandering
2. Existing Metrics and their Shortcomings
3. Proposed Metrics: The Eggshell and the Egg Yolk

Overview

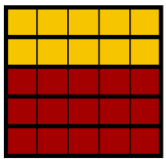
1. Introduction to Gerrymandering
2. Existing Metrics and their Shortcomings
3. Proposed Metrics: The Eggshell and the Egg Yolk

Definition

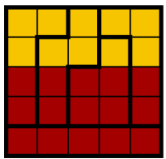
Definition

Gerrymandering is defined as the manipulation of election district borders to favor a particular group.

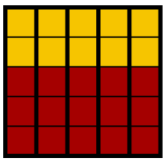
- The party in power is often the one drawing the districts, keeping themselves in power.



Maroon Wins, 3 - 2



Gold wins, 3 - 2

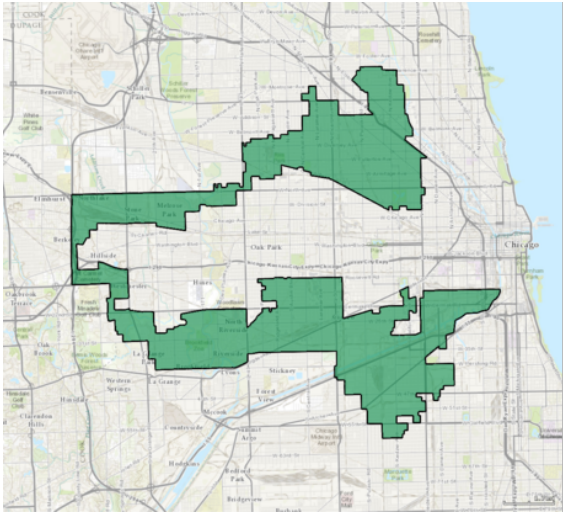


Maroon Wins, 5 - 0

Quintessential Examples, Historical

Figure: District designed by Elbridge Gerry, likened to a salamander. (Wikimedia Commons)

Quintessential Examples, Contemporary



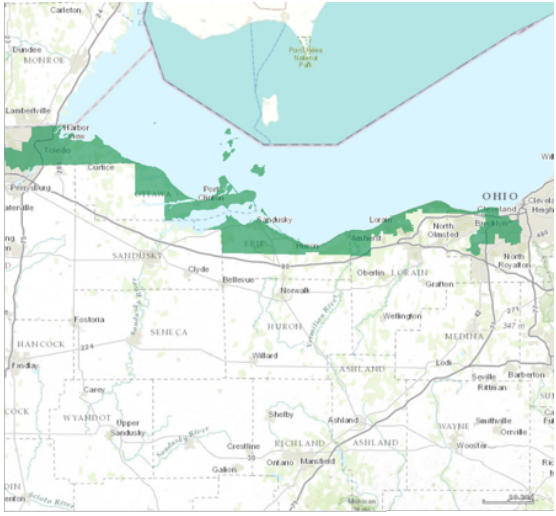


Figure: Current day Illinois's 4th (left) and Ohio's 9th (right) congressional districts. (National Atlas)

Packing

- By packing, the opposing party has many votes "wasted" on a few districts, making the other districts easier to win.

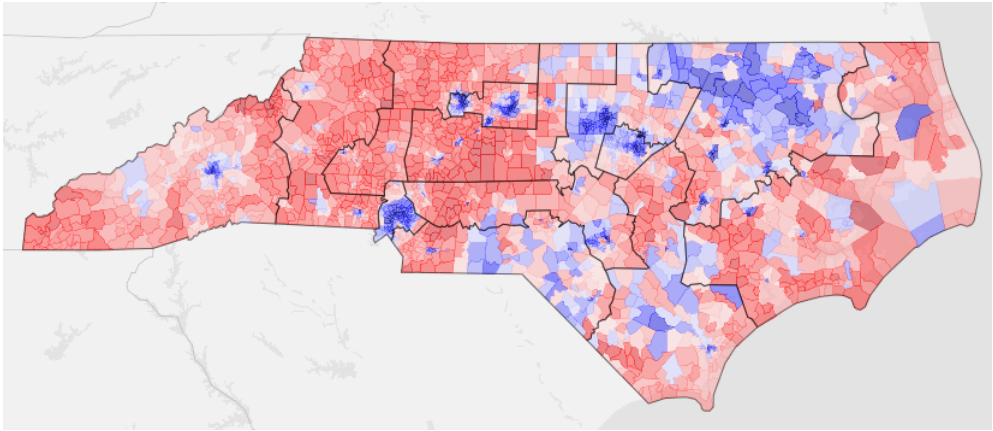


Figure: North Carolina's 2020 districts, greatly improved but still packed.

Cracking

- Cracking involves spreading out voters of a certain party across multiple districts such that they narrowly lose the vote in each district

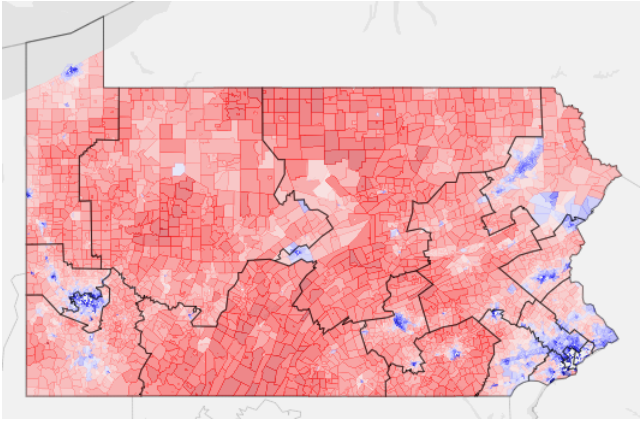


Figure: Pennsylvania's 2020 districts, with cracking in both the east and west.

Overview

1. Introduction to Gerrymandering
2. Existing Metrics and their Shortcomings
3. Proposed Metrics: The Eggshell and the Egg Yolk

Existing Metrics: Shape-Based

- **Polsby-Popper** ($PP(D)$): Ratio of area of district A_D to area of circle with equal perimeter P_D
- **Schwartzberg** ($S(D)$): Ratio of perimeter of district P_D to circumference of circle with equal area A_D.

$$PP(D) = \frac{1}{S(D)^2} = \frac{4\pi A_D}{P_D^2}$$

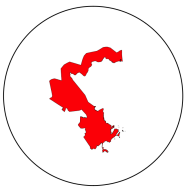


Figure: A visualization of the Polsby-Popper (left) and Schwartzberg (right) metrics. (Fisher)

Existing Metrics: Shape-Based

- **Reock:** Ratio of the area of the district to the area of the smallest circle that encloses it.
- **Convex Hull:** Ratio of the area of the district to the area of the convex hull of the district

Figure: A visualization of the Reock (left) and Schwartzberg (right) metrics. (Fisher)

Shape-Based Metric Shortcomings

- Resolution can affect perimeter-based geometry metrics due to coastline paradox
- Awkward population distributions and geographic boundaries can lead to gerrymandering being detected for completely reasonable districts
- Example: Alaska **at large**; Alaska scored terribly when ranked and compared with other congressional districts

- **430th/438** districts in Polsby-Popper
- **430th/438** districts in Schwartzberg
- **435th/438** districts in Convex Hull

Existing Metrics: Vote-Based

Efficiency Gap (Stephanopoulos & McGhee, 2014)

- Count votes for the losing party of each district as "wasted"
- Count *excess* votes for the winning party of each district as "wasted"
- Sum wasted votes by party and see if more votes are wasted for a certain party.

Existing Metrics: Vote-Based

Declination (Warrington, 2018)

- Measures the difference in margins of victory by party
- Motivation: in gerrymandered states, a disproportionate number of districts will have parties having approximately 50% of the vote.
  - Lots of victories with just over 50% of the vote implies cracking
  - Landslide victories (with much more than 50% of the vote) implies packing

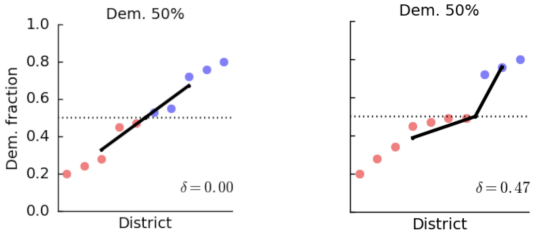


Figure: A visualization of Declination (Warrington)

Vote-Based Metric Shortcomings

- Efficiency gap does not respect proportionality, e.g. landslides.
- Declination is undefined when only one party wins all of the districts.
- Neither can determine *intent*; supposedly cracked districts may be limited by geography.

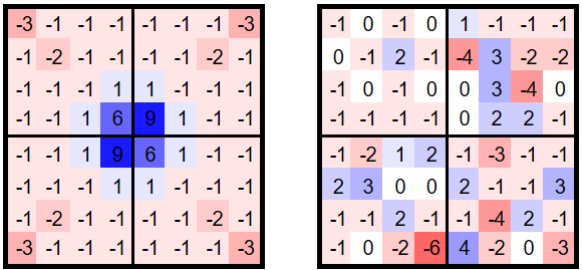


Figure: Two equivalent maps, according to vote-based metrics.

Overview

1. Introduction to Gerrymandering
2. Existing Metrics and their Shortcomings
3. Proposed Metrics: The Eggshell and the Egg Yolk

Motivation

Can we take both voting statistics and shape into account?

Motivation

Can we take both voting statistics and shape into account? Yes, through voter precinct data!

Motivation

Can we take both voting statistics and shape into account? Yes, through voter precinct data!

- Breaks districts into components
- Higher-resolution information
- Allows for granular judgement of boundaries

Motivation

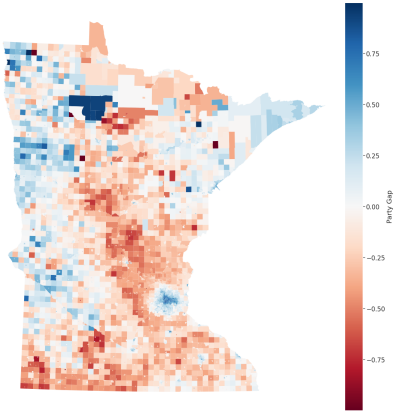
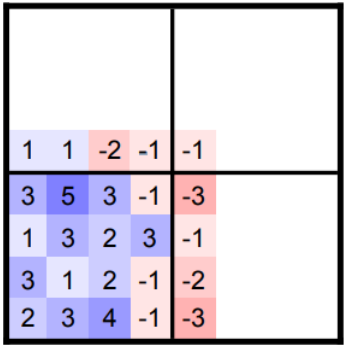
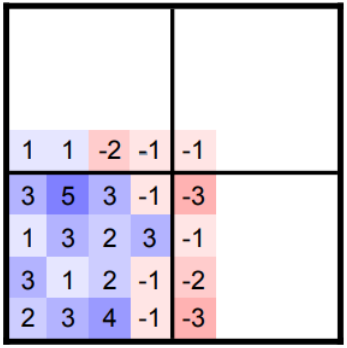


Figure: A map of Minnesota depicting the percentage of victory of each party in each precinct. Observe the precision of data, much more than the results from 8 districts can provide.

Eggshell Metric



Eggshell Metric



$$\epsilon_D = \sum_{p \in D} \sum_{q \in O_D} \frac{g(p) - g(q)}{|\overrightarrow{G_p G_q}|}$$

Claims

- Precinct-resolution detail reveals internal structure.
- Scaling by distance accounts for intent
- Using all precincts prevents "insulation."

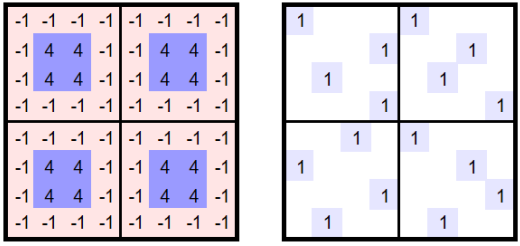


Figure: An example of two states that would appear identically at district-level resolution.

Motivation

How do we design a metric to detect cracking?

Motivation

How do we design a metric to detect cracking?

- Division of population centers
- Consider multiple districts simultaneously
- State-wide approach

Egg Yolk Algorithm

1. Find the n precincts with the greatest margins of victory
2. Iteratively generate clusters of said precincts such that each member of the cluster is within r meters of all other members, for parameter r
3. **Compare the number of districts the cluster is split into to the expected value**
4. Scale by cluster population
5. Sum each cluster's score to create a state-wide cracking score

Expectation Heuristic

The expected value is a key component of the Egg Yolk metric.

- Many factors can influence cluster division.
- Methods like Monte-Carlo Markov Chain generation to create random samples.
- Could also work with empirical examples, such as current districts. **May introduce bias!**
- Heuristic lets us tune cracking and packing.

Claims

- Clusters split across districts are evidence of cracking.
- Scaling by population properly weighs large clusters.
- Measures packing after heuristic tuning.
- Simple!

Findings, MN Redistrictings

Results from running Egg Yolk ($n = 200$ precincts, $r = 8000$ m) on various districtings of Minnesota.

- The 2016 districts: **2.9**
- The most compact districting: **3.8**
- MN heavily gerrymandered so that Republicans win: **4.8**

Most compact districting minimizes the average distance of the population from the geographic center of the district
Notably, most compact was more cracked than the 2016 districts by our metric!

Findings, Correlation

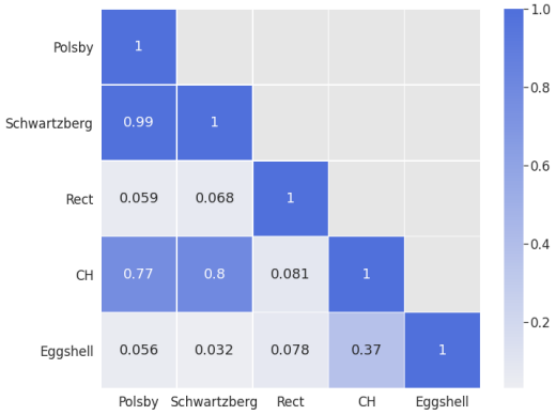


Figure: Comparison of Eggshell against geometry-based metrics.

Conclusion

- Voter precinct data is useful.
- More research needs to be done.
- More data needs to be made available.

The End