Èdouard Lucas:

The theory of recurrent sequences is an inexhaustible mine which contains all the properties of numbers; by calculating the successive terms of such sequences, decomposing them into their prime factors and seeking out by experimentation the laws of appearance and reproduction of the prime numbers, one can advance in a systematic manner the study of the properties of numbers and their application to all branches of mathematics.

Preview

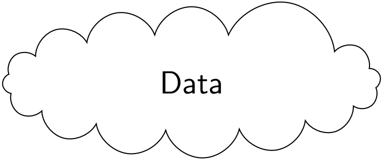
Total Variation Isoperimetric Profiles and Political Redistricting

Daryl DeFord

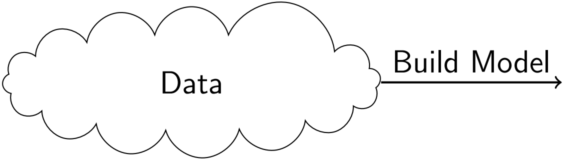
MIT – CSAIL
Geometric Data Processing Group

Math Department Colloquium
Dartmouth College
April 25, 2019

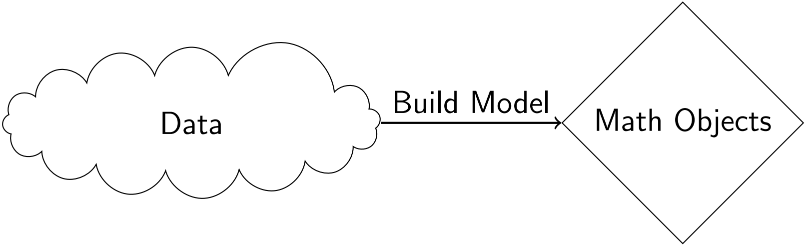
Philosophy



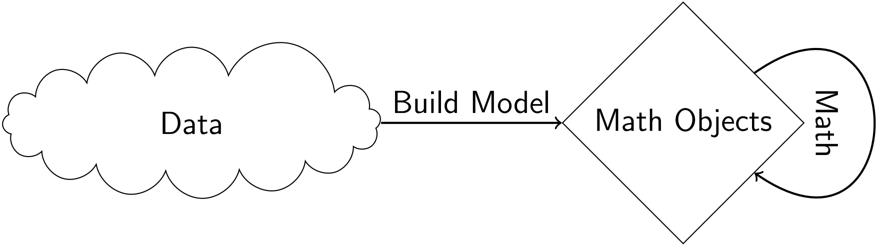
Philosophy



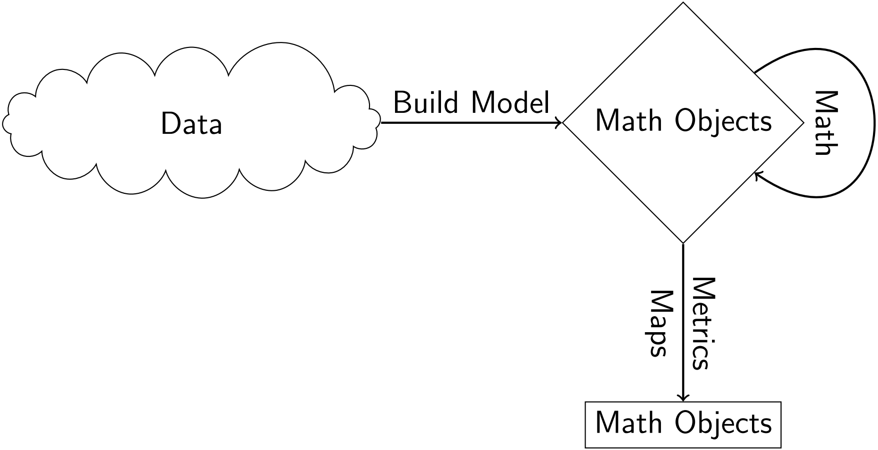
Philosophy



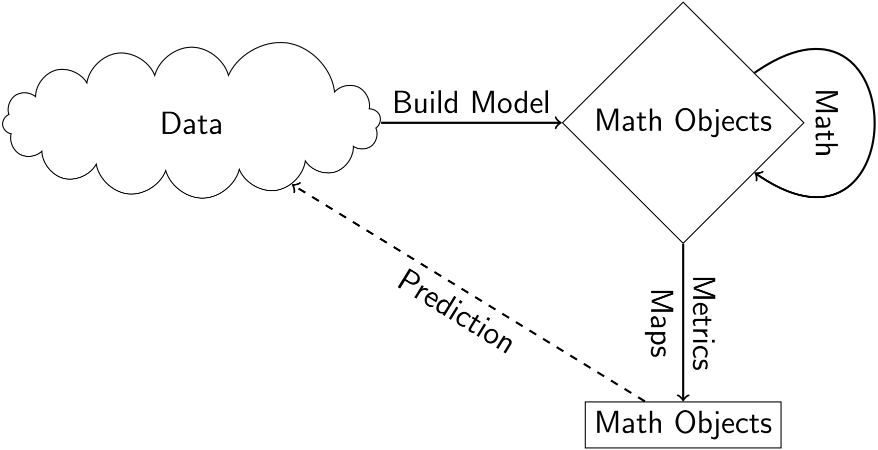
Philosophy



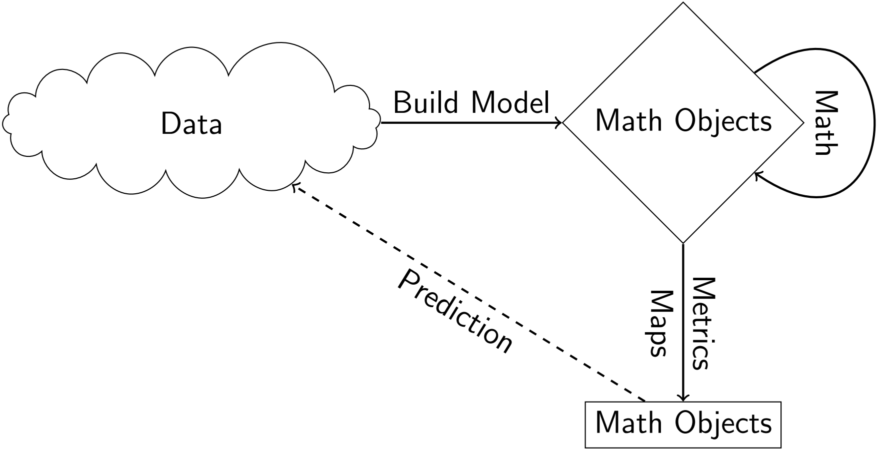
Philosophy


Data
Build Model
Math Objects
Math
Maps
Metrics
Math Objects

Philosophy


Data
Build Model
Math Objects
Math
Maps
Metrics
Math Objects
Prediction

Philosophy



people.csail.mit.edu/ddeford/Networks_Not_Graphs.pdf

Outline

Collaborators

- Prof. Moon Duchin — **Tufts Math**
- Prof. Justin Solomon — **MIT CSAIL**
- Hugo Lavenant — **Universitè Paris–Sud Math**
- Zachary Schutzman — **UPenn CIS**

Collaborators

- Prof. Moon Duchin — **Tufts Math**
- Prof. Justin Solomon — **MIT CSAIL**
- Hugo Lavenant — **Universitè Paris–Sud Math**
- Zachary Schutzman — **UPenn CIS**

- Voting Rights Data Institute
  - 52 undergraduate and graduate students
  - 8 week summer program
  - `mggg.org`
  - github.com/{gerrymandr,mggg,mggg-states}
  - `districtr.org`

MORAL #1:

MORAL #1:

Computational Redistricting is NOT a solved problem!

MORAL #2:

MORAL #2:

Computational Redistricting is NOT a solved problem!

MORAL #2:

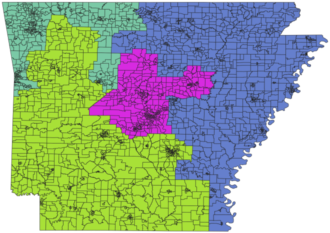
Computational Redistricting is NOT a solved problem!

More Background:
people.csail.mit.edu/ddeford/CAPR

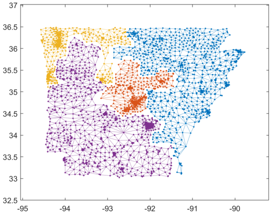
Research Projects:
tinyurl.com/gerryprojects

Research Papers:
mggg.org/work

Political Partitioning

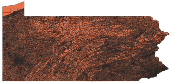


(a) Geography

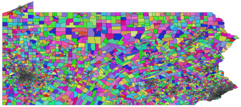


(b) Dual Graph

Geographic Units



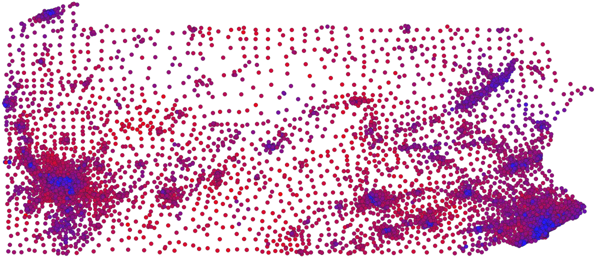
(a) Census Blocks



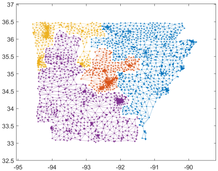
(b) Precincts

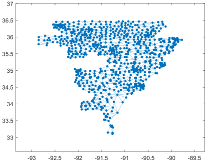
(c) Counties

Tabular Data

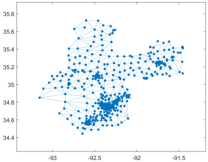


Arkansas Congressional Districts

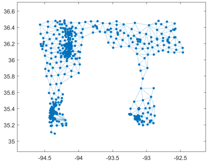




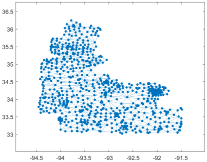
(b) District # 1



(c) District # 2



(d) District # 3



(e) District # 4

Permissible Districting Plans

We want to study partitions of a given geography, at a given scale, into k pieces, satisfying some constraints:

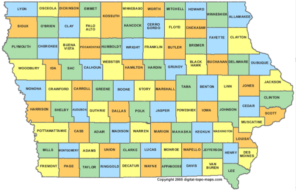
- Contiguity
- Population Balance
- Compactness
- Communities of Interest
- Municipal Boundaries
- Competitiveness/Responsiveness
- Incumbency Protection
- ...

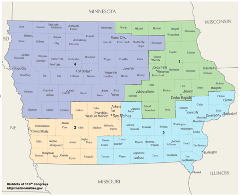
Permissible Districting Plans

We want to study partitions of a given geography, at a given scale, into k pieces, satisfying some constraints:

- Contiguity
- Population Balance
- **Compactness**
- Communities of Interest
- Municipal Boundaries
- Competitiveness/Responsiveness
- Incumbency Protection
- ...

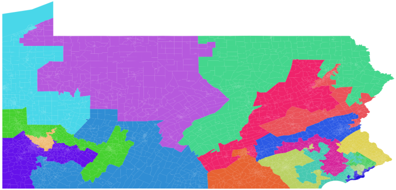
Example: Iowa

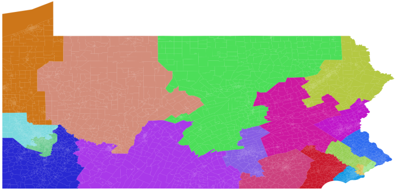




- 4 Congressional Districts, 100 House Districts, 50 Senate Districts
- House districts nest into Senate districts
- Congressional districts made out of counties
- Independent committee with legislative approval
- No partisan data allowed

Example: Pennsylvania



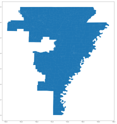


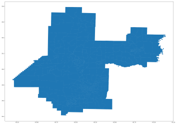
- 18 Congressional Districts, 203 House Districts, 50 Senate Districts
- Zero–balanced population
- Legislature draws congressional districts committee draws legislative districts
- Partisan behavior allowed

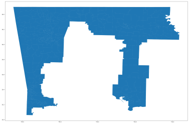
MORAL:

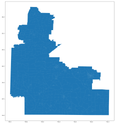
Computational Redistricting is **NOT** a solved problem!

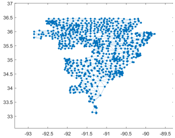
Single Valued Compactness Measures

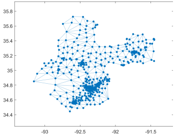


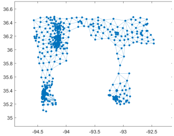


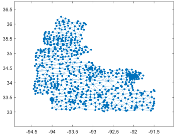




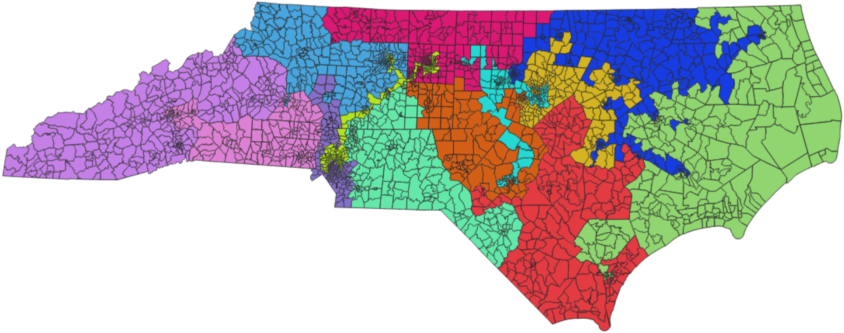








Ugly Shapes



Multiple Types of “Badness”


(a) NC12 #2


(b) NC12 #9


(c) NC12 #12

Measurement Problems

Theorem (Bar-Natan, Najt, and Schutzman 2019)

There is no local homeomorphism from a sphere to the plane that preserves your favorite compactness measure.

Measurement Problems

Theorem (Bar-Natan, Najt, and Schutzman 2019)

There is no local homeomorphism from a sphere to the plane that preserves your favorite compactness measure.

Problem (1, 2)

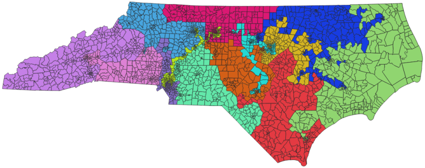
Compactness scores can be distorted by:

- *Data resolution*
- *Map projection*
- *State borders and coastline*
- *Topography*
- *...*

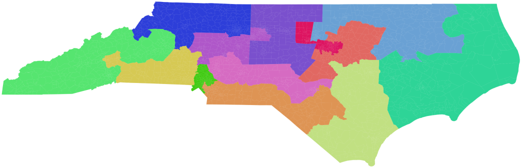
[1] M. Duchin and B. Tenner: Discrete geometry for electoral geography, https://arxiv.org/abs/1808.05860, (2018).

[2] R. Barnes and J. Solomon: Gerrymandering and Compactness: Implementation Flexibility and Abuse, https://arxiv.org/abs/1803.02857, (2018).

Bigger Problem



(a) NC12



(b) NC16

Polsby–Popper

Theorem (Isoperimetry)

Let Ω be a bounded open subset of $\mathbb{R}^2$ with finite perimeter. Then:

$$4\pi A \leq P^2.$$

Definition (Polsby–Popper)

The Polsby–Popper score of a district is:

$$PP(\Omega) = \frac{4\pi A}{P^2}$$

Polsby–Popper

Theorem (Isoperimetry)

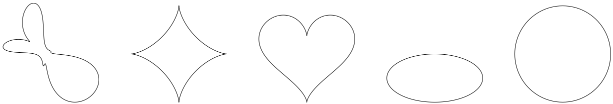
Let Ω be a bounded open subset of $\mathbb{R}^2$ with finite perimeter. Then:

$$4\pi A \leq P^2.$$

Definition (Polsby–Popper)

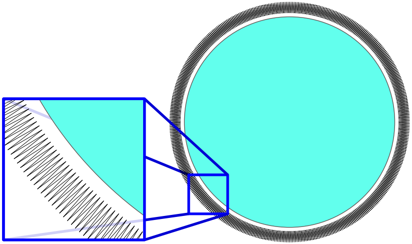
The Polsby–Popper score of a district is:

$$PP(\Omega) = \frac{4\pi A}{P^2}$$

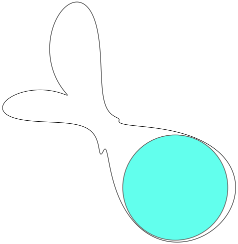


$$\frac{4\pi A}{L^2} = \quad 0.359 \quad 0.411 \quad 0.680 \quad 0.841 \quad 1.000$$

Boundary Perturbation



$$\frac{4\pi A}{L^2} = 0.004$$



$$\frac{4\pi A}{L^2} = 0.359$$

Multiscale Desiderata

- Disambiguate different types of “badness”
- Stability under practical constraints
- Interpolate well–studied single measures
- Continuous and discrete versions
- Internal vs. external

Isoperimetric Profile

Definition (Isoperimetric Inequality)

Let $\Omega \subseteq \mathbb{R}^n$ to be a compact region whose boundary $\partial\Omega \subseteq \Omega$ is an $(n-1)$-dimensional hypersurface in $\mathbb{R}^n$

$$n \cdot \mathrm{vol}(\Omega)^{\frac{(n-1)}{n}} \cdot \mathrm{vol}(B(1, \mathbf{0}))^{\frac{1}{n}} \leq \mathrm{area}(\partial\Omega).$$

Isoperimetric Profile

Definition (Isoperimetric Inequality)

Let $\Omega \subseteq \mathbb{R}^n$ to be a compact region whose boundary $\partial\Omega \subseteq \Omega$ is an $(n-1)$-dimensional hypersurface in $\mathbb{R}^n$

$$n \cdot \mathrm{vol}(\Omega)^{\frac{(n-1)}{n}} \cdot \mathrm{vol}(B(1, \mathbf{0}))^{\frac{1}{n}} \leq \mathrm{area}(\partial\Omega).$$

Definition (Isoperimetric Profile)

With Ω as above and $t \in [0, \mathrm{vol}(\Omega)]$ we ask for the smallest surface area needed to enclose volume t completely within Ω:

$$I_\Omega(t) := \min\{\mathrm{area}(\partial\Sigma) : \Sigma \subseteq \Omega \text{ and } \mathrm{vol}(\Sigma) = t\}.$$

Geometric Properties

Theorem (Flores and Nardulli (2016) [1])

] Let M^n be a complete smooth Riemannian manifold with $\mathrm{Ric}_M \geq (n-1)k$, with $k \in \mathbb{R}$ and $V(B(p,1)) \geq v_0 > 0$. Then the isoperimetric profile is continuous on $[0, V(M)[$

[1] A. Flores and S. Nardulli: Continuity and differentiability properties of the isoperimetric profile in complete noncompact Riemannian manifolds with bounded geometry, https://arxiv.org/abs/1404.3245.

Geometric Properties

Theorem (Flores and Nardulli (2016) [1])

] Let M^n be a complete smooth Riemannian manifold with $\mathrm{Ric}_M \geq (n-1)k$, with $k \in \mathbb{R}$ and $V(B(p,1)) \geq v_0 > 0$. Then the isoperimetric profile is continuous on $[0, V(M)[$

Question

Identify a polynomial-time algorithm or NP-hardness result for computing isoperimetric profiles. The simplest open problem is computing the isoperimetric profile of a polygon in the plane $\mathbb{R}^2$.

[1] A. Flores and S. Nardulli: Continuity and differentiability properties of the isoperimetric profile in complete noncompact Riemannian manifolds with bounded geometry, https://arxiv.org/abs/1404.3245.

Total Variation

Definition (Three formulations of TV)

$$\mathrm{TV}[f] =$$

1. $$\sup\left\{\int_{\mathbb{R}^n} [f(x)\nabla\cdot\phi(x)]\,dx : \phi \in C^1_c(\mathbb{R}^n \to \mathbb{R}^n) \text{ and } \|\phi\|_\infty \leq 1\right\}$$
2. $$\int_{\mathbb{R}^n} \|\nabla f\|_2\,dx$$
3. $$\int_0^{+\infty} \mathrm{area}(\partial\{f \geq s\})ds$$

Perimeter as Total Variation

Definition

For a region $\Sigma \subseteq \mathbb{R}^n$, denote its *indicator function* $\mathbb{1}_\Sigma$ via

$$\mathbb{1}_\Sigma(x) := \begin{cases} 1 & \text{if } x \in \Sigma \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Then, a consequence of the co-area formula is that

$$\text{area}(\partial\Sigma) = \text{TV}[\mathbb{1}_\Sigma]. \tag{2}$$

TV Relaxation

Definition (Isoperimetric Profile)

$$I_\Omega(t) = \begin{cases} \inf_{f \in L^1(\mathbb{R}^n)} & \mathrm{TV}[f] \\ \text{subject to} & \int_{\mathbb{R}^n} f(x)\,dx = t \\ & 0 \le f \le \mathbb{1}_\Omega \\ & f(x) \in \{0,1\} \ \forall x \in \mathbb{R}^n. \end{cases}$$

TV Relaxation

Definition (Isoperimetric Profile)

$$I_\Omega(t) = \begin{cases} \inf_{f \in L^1(\mathbb{R}^n)} & \mathrm{TV}[f] \\ \text{subject to} & \int_{\mathbb{R}^n} f(x)\,dx = t \\ & 0 \le f \le \mathbb{1}_\Omega \\ & f(x) \in \{0,1\}\ \forall x \in \mathbb{R}^n. \end{cases}$$

Definition (TV Profile)

$$I_\Omega^{\mathrm{TV}}(t) := \begin{cases} \min_{f \in L^1(\mathbb{R}^n)} & \mathrm{TV}[f] \\ \text{subject to} & \int_{\mathbb{R}^n} f(x)\,dx = t \\ & 0 \le f \le \mathbb{1}_\Omega. \end{cases}$$

Example: Circle

Proposition

For all (Ω, t)*, we have* $I_\Omega^{\mathrm{TV}}(t) \leq I_\Omega(t)$*.*

Example (Circle)

Suppose $\Omega \subset \mathbb{R}^2$ *is a circle of radius* R*, and take* $t = \pi r^2$ *for* $r \in (0, R)$*. In this case, by the isoperimetric inequality we know* $I_\Omega(t) = 2\pi r$*. But suppose we take* $f(x) \equiv \dfrac{r^2}{R^2}$*. By the co-area formula*

$$I_\Omega^{\mathrm{TV}}(t) \leq \mathrm{TV}[f] = 2\pi R \cdot \frac{r^2}{R^2} = 2\pi r \cdot \frac{r}{R} < I_\Omega(t).$$

Hence, our relaxation is not tight.

Isoperimetry and Convexity

Proposition (Isoperimetry)

Suppose $B \subset \mathbb{R}^n$ is a ball whose volume matches $\mathrm{vol}(\Omega)$. Then, for all $t \in [0, \mathrm{vol}(\Omega)]$, we have $I_B^{\mathrm{TV}}(t) \leq I_\Omega^{\mathrm{TV}}(t)$, and if the equality holds for some $t > 0$ then Ω is a ball.

Proposition (Convexity)

$I_\Omega^{\mathrm{TV}}(t)$ is a convex function of t.

Proposition (Convex Envelope)

The function I_Ω^{TV} is the lower convex envelope of I_Ω.

Dual Optimization

Dual Formulation:

$$I_\Omega^{TV}(t) = \begin{cases} \sup_{\phi \in C_c^1(\mathbb{R}^n \to \mathbb{R}^n), \lambda \in \mathbb{R}} & \lambda t - \int_\Omega \max(\lambda - \nabla \cdot \phi(x), 0)\, dx \\ \text{subject to} & \|\phi\|_\infty \leq 1 \end{cases}$$

Proof.

With the dual in hand, the convexity results follow from constructing an auxilliary function:

$$h(\lambda) = \inf_{\|\phi\|_\infty \leq 1} \int_\Omega \max(\lambda - \nabla \cdot \phi(x), 0)\, dx.$$

and computing some Legendre transforms. □

Minimizer Structure

Proposition (Distinguished Solutions)

There exists a family $(f_t)_{t\in[0,1]}$ such that:

- *For any $t \in [0,1]$, the function $f_t \in L^1(\mathbb{R}^n)$ satisfies $0 \leq f_t \leq \mathbb{1}_\Omega$, $\int_{\mathbb{R}^n} f_t(x)\,dx = t$ and $\mathrm{TV}(f_t) = I_\Omega^{\mathrm{TV}}(t)$.*
- *For any $t \in [0,1]$, there exist $v_t \in (0,1)$ such that f_t takes its values in $\{0, v_t, 1\}$.*
- *For a.e. $x \in \Omega$, the function $t \to f_t(x)$ is increasing.*

NC 12 # 9

NC 12 # 2

NC 12 # 12

Cheeger Sets

Definition (Cheeger Constant)

The Cheeger constant of Ω, denoted by $h_1(\Omega)$, is defined as

$$h_1(\Omega) := \inf_{\tilde{\Sigma} \subseteq \Omega} \frac{\text{area}(\partial \tilde{\Sigma})}{\text{vol}(\tilde{\Sigma})},$$

and a subset $\Sigma \subseteq \Omega$ such that $h_1(\Omega) = \frac{\text{area}(\partial \Sigma)}{\text{vol}(\Sigma)}$ is known as a Cheeger set of Ω.

Proposition (Small t)

Let Ω be compact, let $h_1(\Omega)$ be the Cheeger constant of Ω, and let Σ be a Cheeger set of Ω. Then for any $t \leq \text{vol}(\Sigma)$, we have $I_\Omega^{\text{TV}}(t) = h_1(\Omega)t$, and a solution f is given by $f := \frac{t}{\text{vol}(\Sigma)} \cdot \mathbb{1}_\Sigma$.

Cheeger Proof

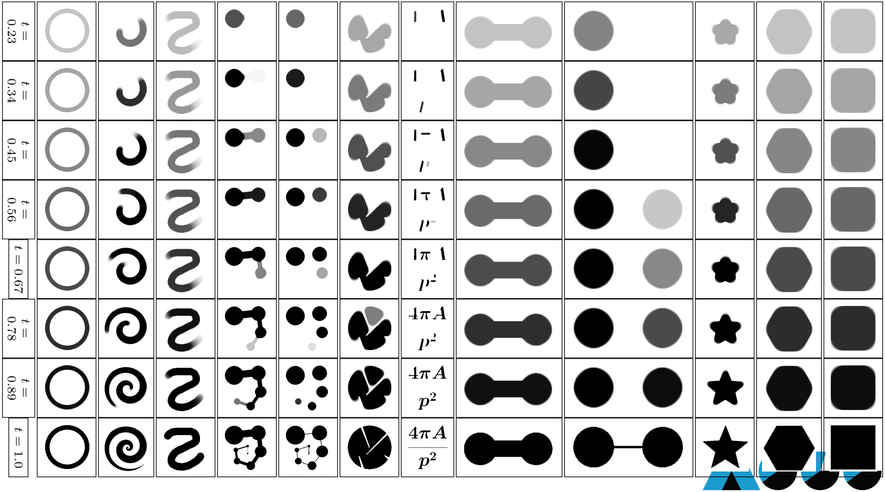
Proof.

We start with $\hat{f} = \frac{t}{\mathrm{vol}(\Sigma)} \cdot \mathbb{1}_\Sigma$ which satisfies the constraints of the problem defining $I_\Omega^{\mathrm{TV}}(t)$ as soon as $t \leq \mathrm{vol}(\Sigma)$, which ensures $0 \leq \hat{f} \leq \mathbb{1}_\Sigma \leq \mathbb{1}_\Omega$. Hence, $I_\Omega^{\mathrm{TV}}(t) \leq h_1(\Omega)t$. On the other hand, using the co-area formula, if f is any competitor then

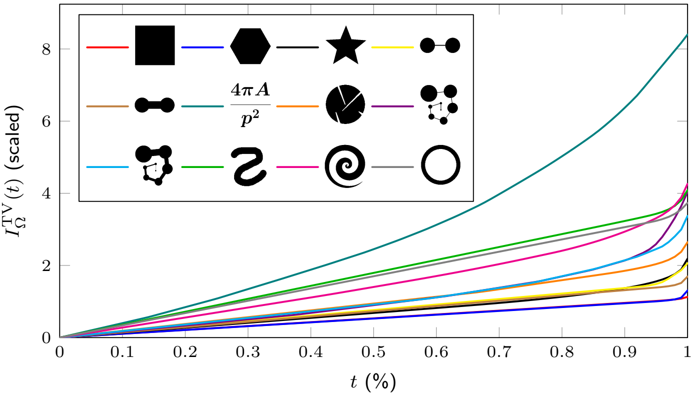
$$\begin{aligned}
\mathrm{TV}(f) &= \int_0^{+\infty} \mathrm{area}(\partial\{f \geq s\}) ds \\
&= \int_0^{+\infty} \mathrm{vol}(\{f \geq s\}) \cdot \underbrace{\frac{\mathrm{area}(\partial\{f \geq s\})}{\mathrm{vol}(\{f \geq s\})}}_{\geq h_1(\Omega) \text{ by definition}} ds \\
&\geq h_1(\Omega) \int_0^{+\infty} \mathrm{vol}(\{f \geq s\}) ds = h_1(\Omega) \int_{\mathbb{R}^d} f(x) dx = h_1(\Omega)t.
\end{aligned}$$

Hence, for $t \leq \mathrm{vol}(C)$, we have $I_\Omega^{\mathrm{TV}}(t) = h_1(\Omega)t$. □

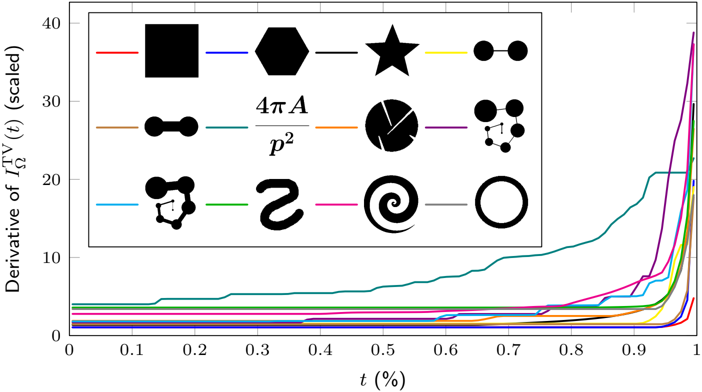
Synthetic Examples



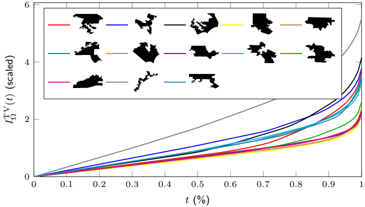
Synthetic Profiles



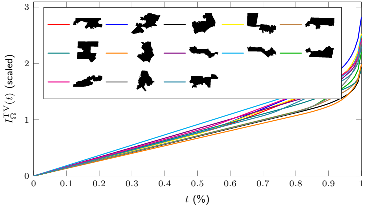
Synthetic Derivatives



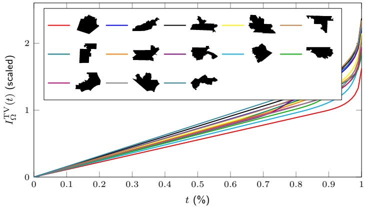
North Carolina



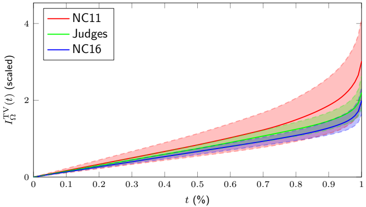
(a) 2011 districts



(b) 2016 districts

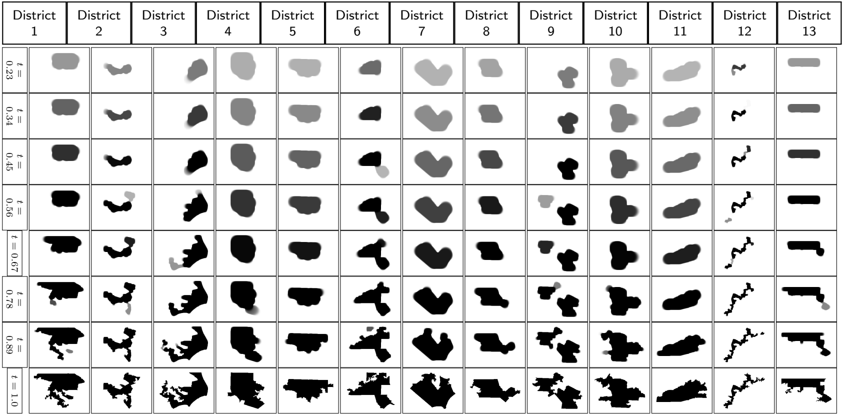


(c) Judges plan

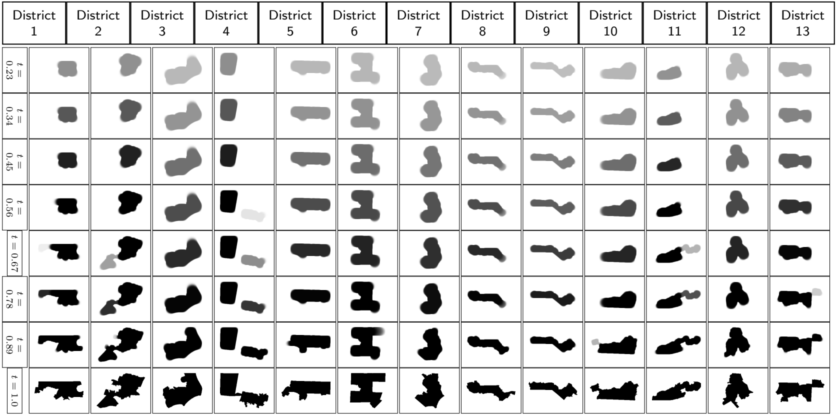


(d) Comparison

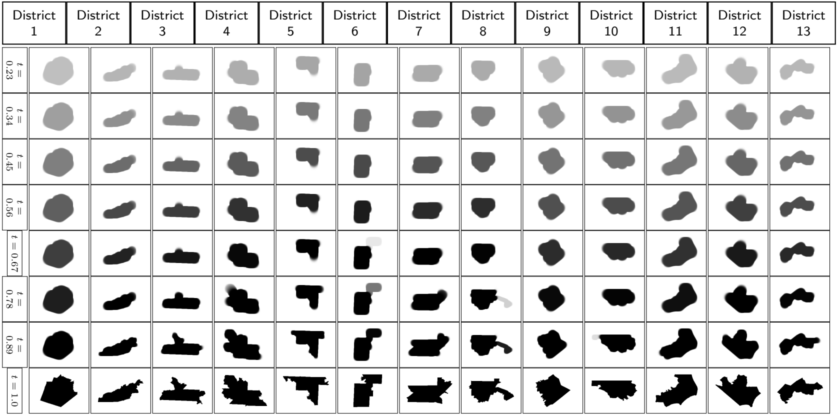
NC 2011 Districts



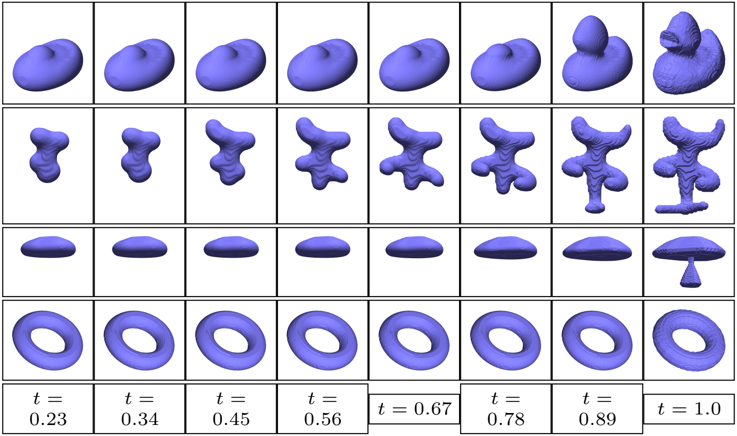
NC 2016 Districts



Judge's Plan


District 1
District 2
District 3
District 4
District 5
District 6
District 7
District 8
District 9
District 10
District 11
District 12
District 13
t = 0.23
t = 0.34
t = 0.45
t = 0.56
t = 0.67
t = 0.78
t = 0.89
t = 1.0

Higher Dimensions



Other Formulations

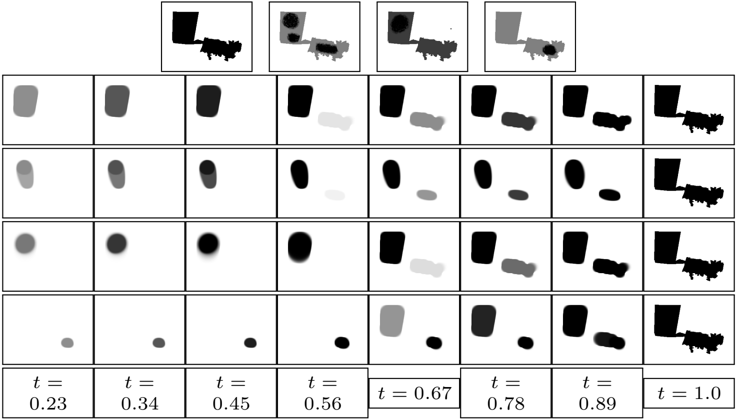
Definition (Population Measure)

$$I^{\mathrm{TV}}_{\Omega,\rho}(t) := \begin{cases} \min_{f \in L^1(\mathbb{R}^n)} & \mathrm{TV}[f] \\ \text{subject to} & \int_{\mathbb{R}^n} f(x)\, d\rho(x) = t \\ & 0 \le f \le \mathbb{1}_\Omega. \end{cases}$$

Definition (Discrete)

$$I^{\mathrm{TV}}_{V_0}(t) := \begin{cases} \min_{f \in \mathbb{R}^V} & \sum_{(v,w)\in E} |f(v) - f(w)| \\ \text{subject to} & \sum_{v \in V_0} f(v) = t|V_0| \\ & f(v) = 0 \ \forall v \notin V_0 \\ & f(v) \in [0,1] \ \forall v \in V. \end{cases}$$

Synthetic Cities



Discrete Animation

Mean Curvature Flow (VRDI)

Mean Curvature Flow (VRDI)

Multiscale Wrapup

Open questions:

- How much can we learn about the full profile from the relaxed version?
- Can the medial axis be computed from the TV-Profile?
- What is the right way to compare regions of the profiles?

- Spectral Versions (i.e. how to make the heat kernel useful)
- Random walk versions (absorbing boundary nodes)
- Distance based measures
- ...

Seats–Votes Curves

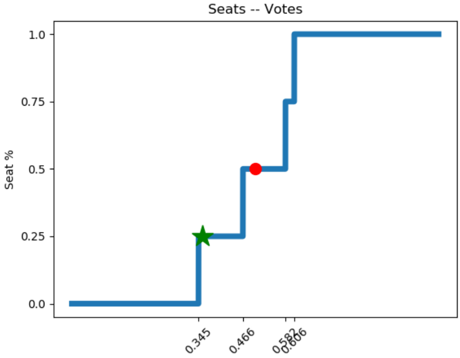


Figure: Dem %: [.249,.389,.273,.51]

Seats–Votes Curves

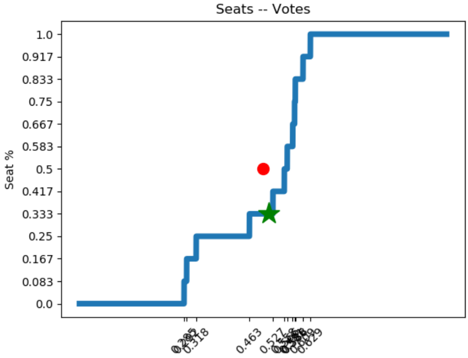


Figure: Dem %: [.698,.458,.724,.43,.435,.428,.553,.489,.407,.387,.731,.45]

Seats–Votes Curves

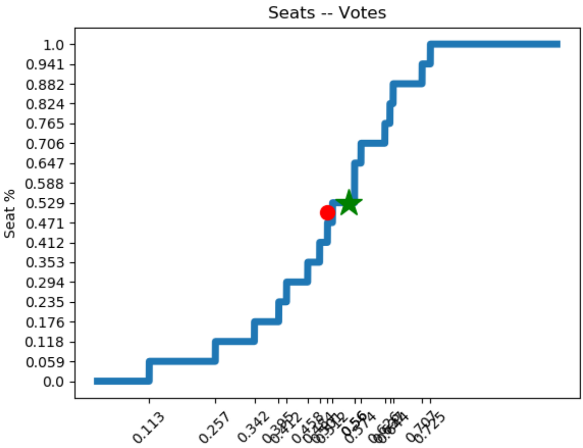


Figure: Dem %: [.487,.79,.934,.635,.652,.589,.535,.546,.403,.487,.41,.34,.705,.421,.322,.473,.563]

Partisan Metrics

Definition (Mean–Median)

Horizontal distance between $(.5, .5)$ and the seats votes curve).

Definition (Partisan Bias)

Vertical distance between $(.5, .5)$ and the seats votes curve

Definition (Partisan Asymmetry)

Integral of the difference between the seats votes curve and its reflection around $(.5, .5)$.

Definition (Efficiency Gap)

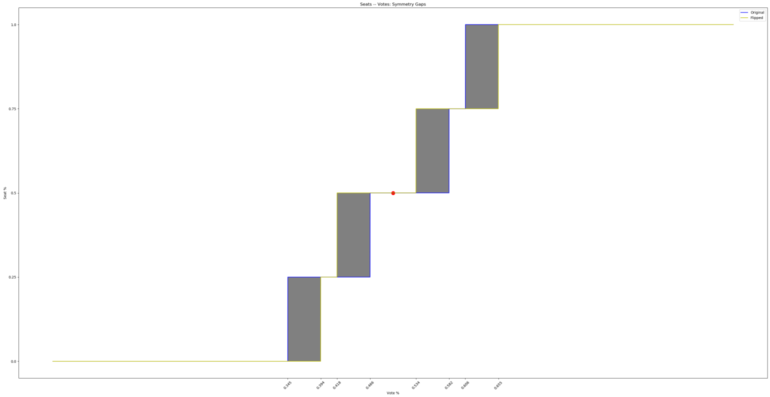
$$\frac{\text{Wasted R Votes} - \text{Wasted D Votes}}{\text{Total Votes}}$$

With equal turnout: twice the seat margin minus the vote margin.

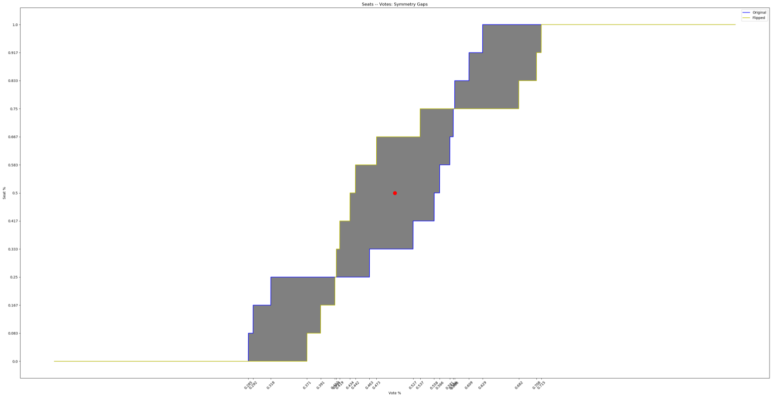
Partisan Examples

- Utah
  - Mean–Median: -.024
  - Efficiency Gap: -.039
  - Asymmetry: .048
- Pennsylvania
  - Mean–Median: .011
  - Efficiency Gap: .063
  - Asymmetry: .050
- North Carolina
  - Mean–Median: .062
  - Efficiency Gap: .198
  - Asymmetry: .093

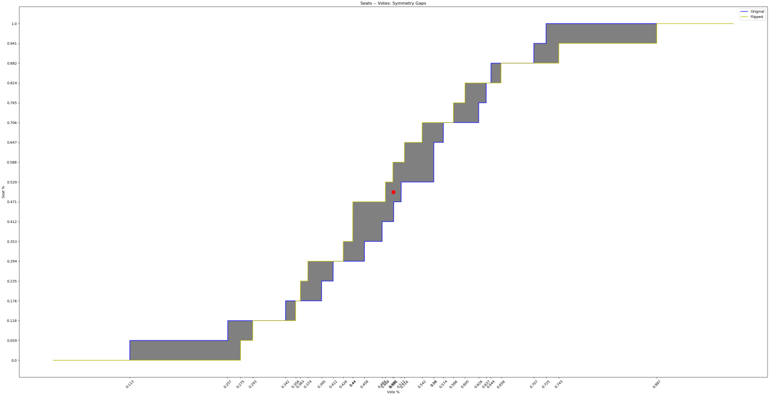
Seats–Votes Asymmetry



Seats–Votes Asymmetry



Seats–Votes Asymmetry



Alternative Formulations

Dem seat share	$S = \#\{i : D_i > R_i\}/k$
Efficiency gap[1]	$EG = S - 2V + \frac{1}{2} + \frac{S(1-S)(1-\rho)}{S(1-\rho)+\rho}$
Mean-median gap	$MM = \text{median}(\{V_i\}) - \text{mean}(\{V_i\})$

In the efficiency gap formula ρ is the turnout ratio: the average number of votes cast in Democratic-won districts divided by the average in Republican-won districts.

[1] E. Veomett, Efficiency Gap, Voter Turnout, and the Efficiency Principle, Election Law Journal, 17(4), 249–263, (2018).

Redistricting Analysis in Practice

The shape and partisan imbalance metrics that I discussed above

Redistricting Analysis in Practice

The shape and partisan imbalance metrics that I discussed above

- Outlier Analysis

Redistricting Analysis in Practice

The shape and partisan imbalance metrics that I discussed above

- Outlier Analysis
- Political Geography Baselines

Redistricting Analysis in Practice

The shape and partisan imbalance metrics that I discussed above

- Outlier Analysis
- Political Geography Baselines
- Reform Legislation Consequences

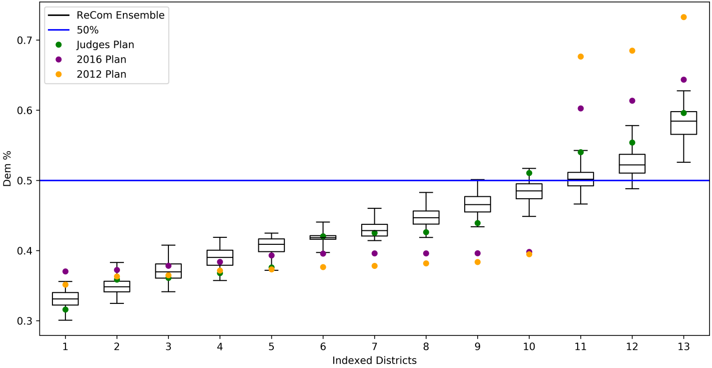
Redistricting Analysis in Practice

The shape and partisan imbalance metrics that I discussed above

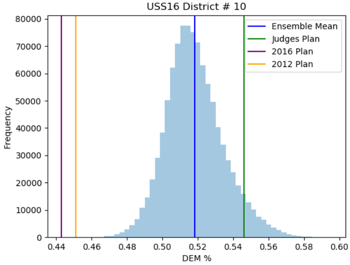
- Outlier Analysis
- Political Geography Baselines
- Reform Legislation Consequences
- More Case Studies:
  `http://people.csail.mit.edu/ddeford/IAP_2019_Part4.pdf`

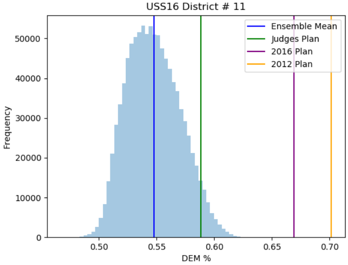
Ensemble Generation

Outlier Example: NC

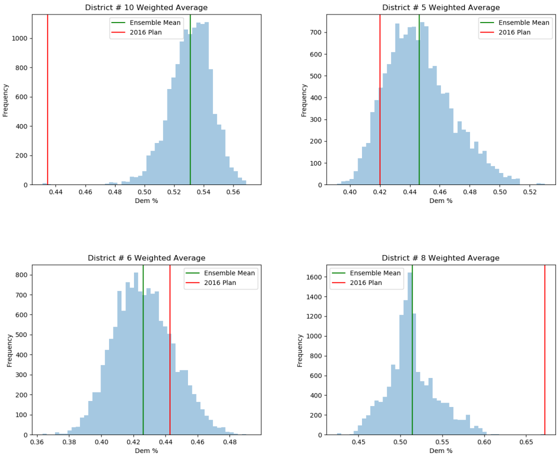


Outlier Example: NC

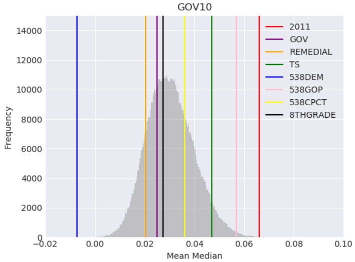




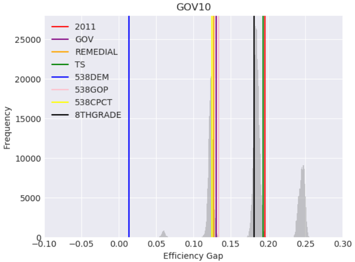
Outllier Example: NC



Outlier Example: PA

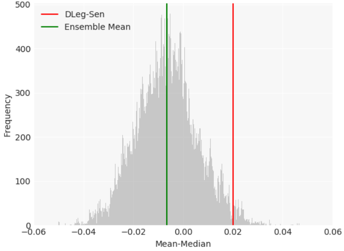


(a) Mean–Median

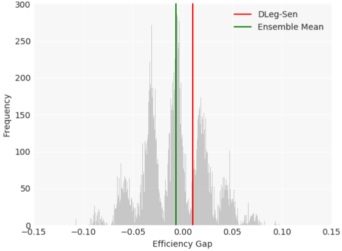


(b) Efficiency Gap

Baseline Example: VA

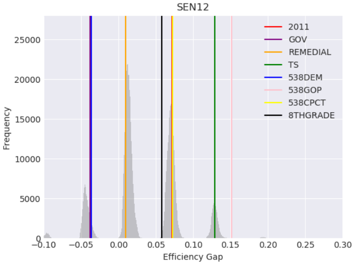


(a) Mean–Median

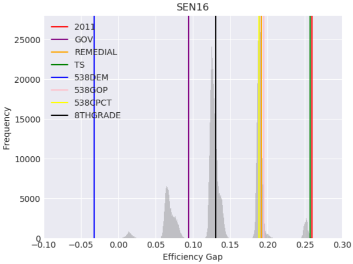


(b) Efficiency Gap

Baseline Example: PA

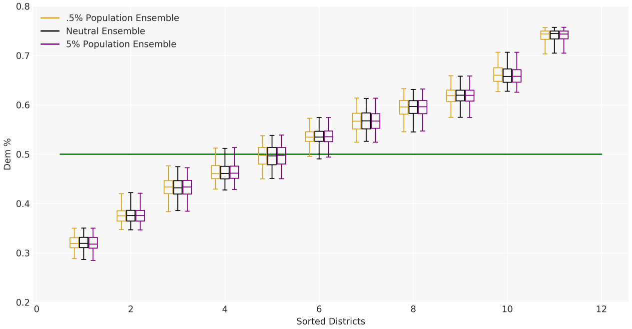


(a) 2012 Senate



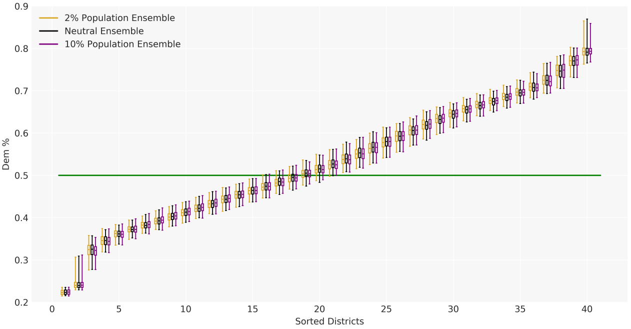
(b) 2016 Senate

Reform Example: VA



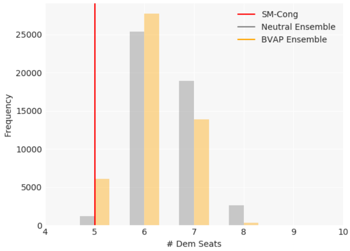
(a) Congress

Reform Example: VA

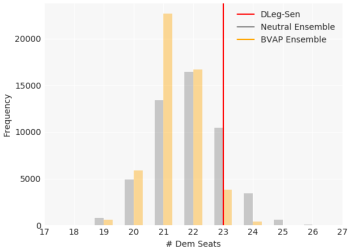


(b) Senate

Reform Example: VA

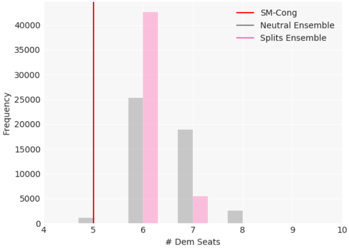


(a) Congress

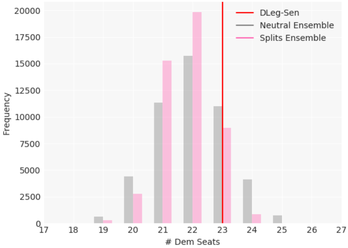


(b) Senate

Reform Example: VA

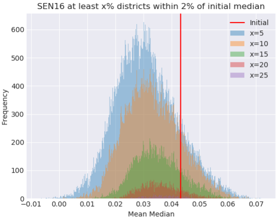


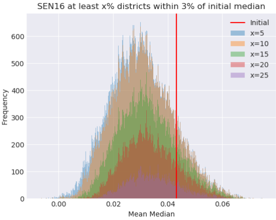
(a) Congress

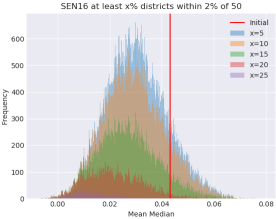


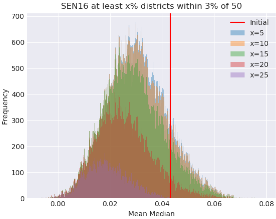
(b) Senate

Reform Example: PA









Try it at home!

- Draw your own districts with **Districtr**
  - `https://districtr.org`
  - Easy to generate complete districting plans in browser or on a tablet
  - Measures district demographics and expected partisan performance
  - Identifies communities of interest
- Generate your own ensembles with **GerryChain**[1]
  - `https://github.com/mggg/gerrychain`
  - Flexible, modular software for sampling graph partitions
  - Handles the geodata processing as well as the MCMC sampling
  - Current support for a
  - Successfully applied in VA, NC, PA, etc.

[1] Originally RunDMCMC

MORAL:

Computational Redistricting is **NOT** a solved problem!

The end!

Thanks!