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

Hardness results for sampling connected graph partitions with applications to redistricting

Daryl DeFord

MIT – CSAIL
Geometric Data Processing Group

ACM Seminar
Dartmouth Math Department
April 23, 2019

Outline

Collaborators

- Prof. Justin Solomon — **MIT CSAIL**
- Lorenzo Najt — **Wisconsin Math**
- Prof. Moon Duchin — **Tufts Math**

Collaborators

- Prof. Justin Solomon **MIT CSAIL**
- Lorenzo Najt **Wisconsin Math**
- Prof. Moon Duchin **Tufts Math**

- Voting Rights Data Institute
  - 52 undergraduate and graduate students
  - 6–8 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:

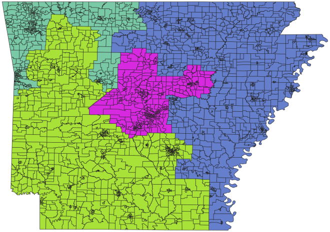
Computational Redistricting is NOT a solved problem!

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

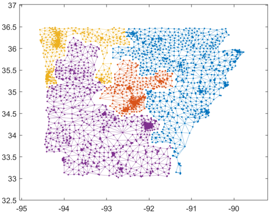
Research Projects:
tinyurl.com/gerryprojects

Research Papers:
mggg.org/work

Political Partitioning

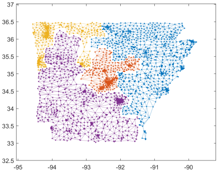


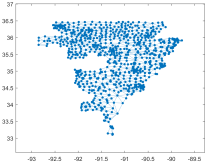
(a) Geography



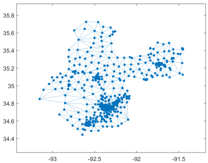
(b) Dual Graph

Arkansas Congressional Districts

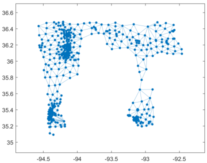




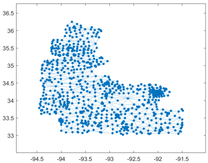
(b) District # 1



(c) District # 2



(d) District # 3



(e) District # 4

Permissible Districting Plans

We want to partition a given geography (graph), 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 partition a given geography (graph), at a given scale, into k pieces, satisfying some constraints:

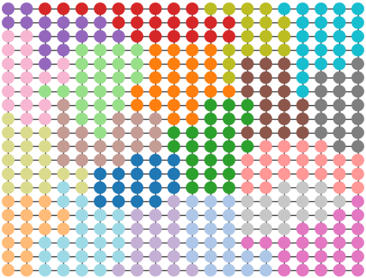
- **Contiguity**
- **Population Balance**
- Compactness
- Communities of Interest
- Municipal Boundaries
- Competitiveness/Responsiveness
- Incumbency Protection
- ...

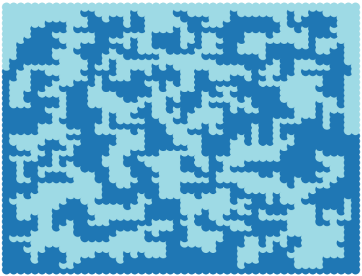
Mathematical Formulation

Given a (connected, planar) graph $G = (V, E)$:

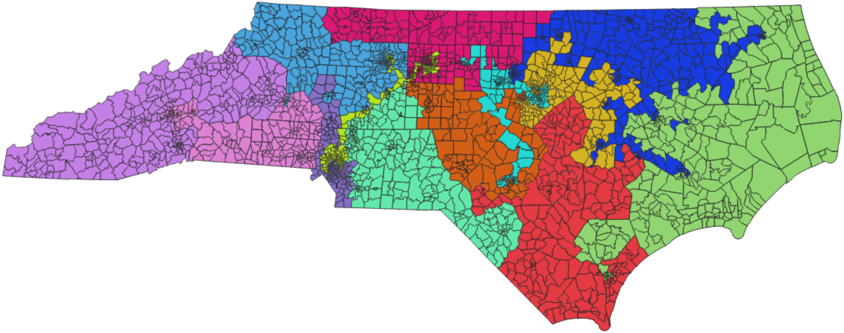
- A k-**partition** $P = \{V_1, V_2, \ldots, V_k\}$ of G is a collection of disjoint subsets $V_i \subseteq V$ whose union is V.
- A partition P is **connected** if the subgraph induced by V_i is connected for all i.
- The **cut edges** of P are the edges (u, w) for which $u \in V_i$, $w \in V_j$, and $i \neq j$
- A partition P is ε-**balanced** if $\mu(1 - \varepsilon) \leq |V_i| \leq \mu(1 + \varepsilon)$ for all i where μ is the mean of the $|V_i|$'s
- An **equi–partition** is a 0-balanced partition

(Discrete) Total Perimeter





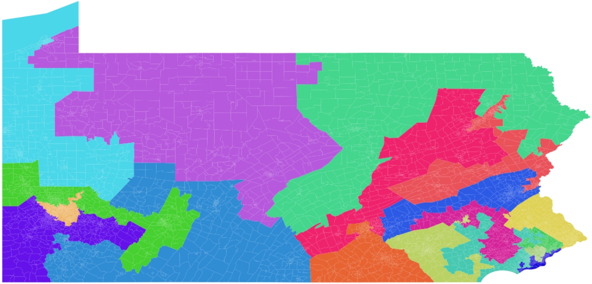
Ugly Shapes



Ugly Shapes

(a) NC12 #1 (b) NC12 #2 (c) NC12 #12

Ugly Shapes



Measurement Problems

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

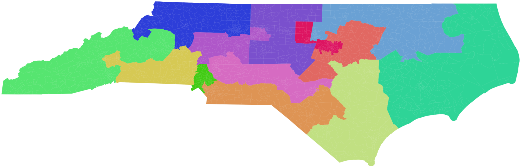
There is no local homeomorphism from the globe to the plane that preserves your favorite compactness measure.

Problem (Barnes and Solomon 2018)

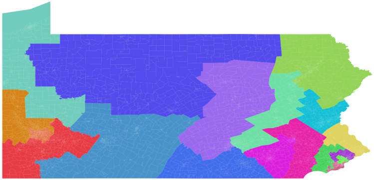
Compactness scores can be distorted by:

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

Partisan Imbalance



(a) NC16



(b) PA TS-Proposed

Partisanship Measures

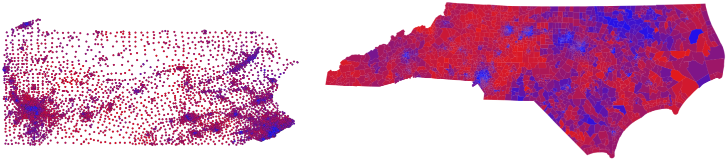


Figure: 2016 Presidential election votes by precinct in PA and NC.

Partisan Fairness

- MA
  - Duchin et al. (2018) Locating the representational baseline: Republicans in Massachusetts arXiv:1810.09051
  - Not all partisan outcomes are possible, given discretization
- MD
  - Two recent preprints claiming not gerrymandered
  - Court ruled one district unconstitutional
- NC/PA/WI
  - Heavy court involvement
  - Wide variance in partisan metrics

Partisan (a)Symmetry

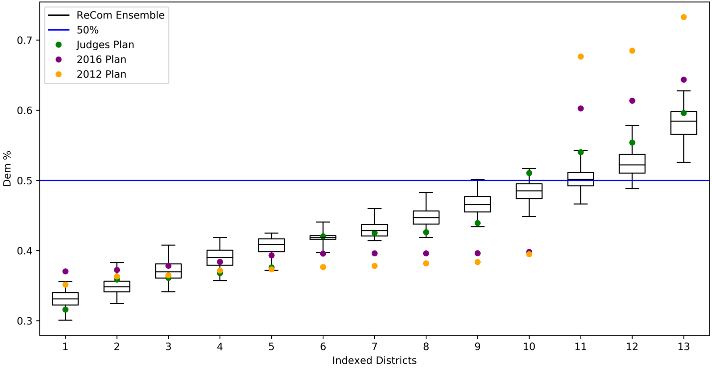
MORAL:

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

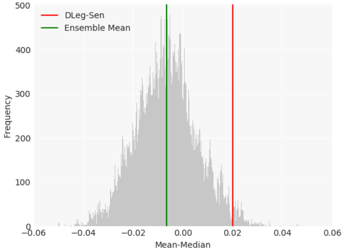
Ensemble Analysis

- The wide variety in rules applied to districting problems (even in the same state) means that any single measure of gerrymandering will be insufficient/exploitable
- Instead we want to compare to large ensembles of other feasible plans.
- This allows us to understand the impacts of the underlying political and demographic geography on a wide collection of metrics.

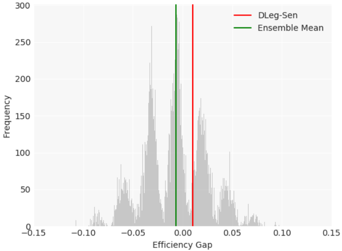
Outlier Example: NC



Baseline Example: VA

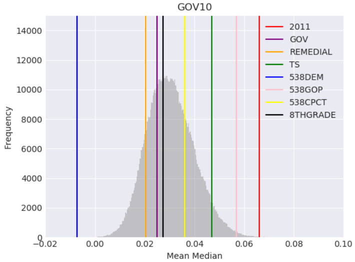


(a) Mean–Median

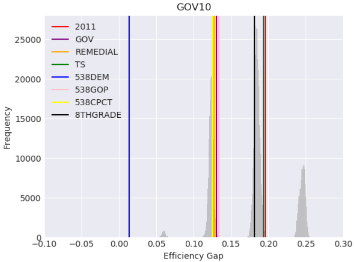


(b) Efficiency Gap

Baseline Example: PA

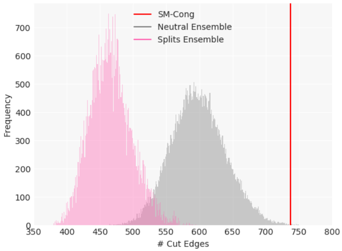


(a) Mean–Median

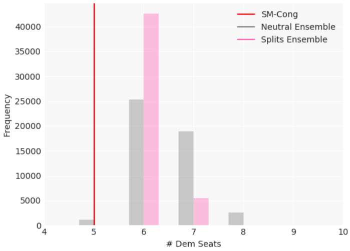


(b) Efficiency Gap

Reform Example: VA

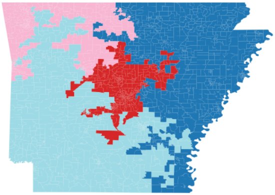


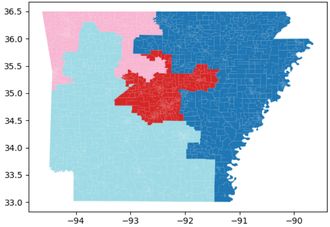
(a) Compactness



(b) Dem Seats

Which ensembles?





Ensembles in Practice

- The appeal of an ensemble method is that you get to control the input data very carefully
- However, just because a particular type of data was not considered doesn't mean that the outcome is necessarily "fair"
- There are lots of "random" methods for constructing districting plans
- Most don't offer any control over the distribution that you are drawing from

MCMC on partitions

1. Set constraints to define the state space
2. Start with an initial plan
3. Propose a modification
4. Verify that the modification satisfies the constraints
5. Accept using MH criterion
6. Repeat

MCMC on partitions

1. Set constraints to define the state space
2. Start with an initial plan
3. Propose a modification
4. Verify that the modification satisfies the constraints
5. Accept using MH criterion
6. Repeat

Why?

MCMC on partitions

1. Set constraints to define the state space
2. Start with an initial plan
3. Propose a modification
4. Verify that the modification satisfies the constraints
5. Accept using MH criterion
6. Repeat

Why?

- Control over sampling distribution and input data
- Possibility of local sampling
- Ergodic Theorem

MC**MC** on partitions

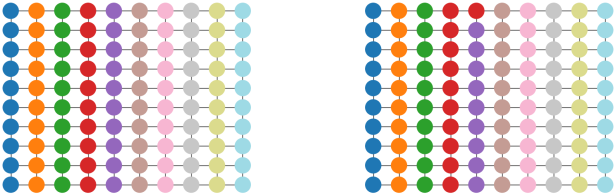
1. Set constraints to define the state space
2. Start with an initial plan
3. **Propose a modification**
4. Verify that the modification satisfies the constraints
5. Accept using MH criterion
6. Repeat

Why?

- Control over sampling distribution and input data
- Possibility of local sampling
- Ergodic Theorem

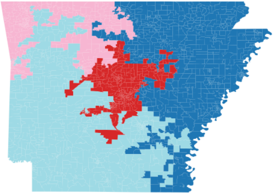
Single Edge Flip Proposals

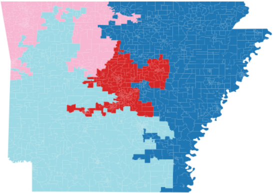
1. Uniformly choose a cut edge
2. Change one of the incident node assignments to the other



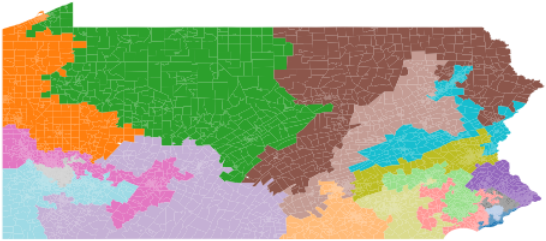
- Mattingly et al. (2017, 2018) Court cases in NC and WI.
- Pegden et al. Assessing significance in a Markov chain without mixing, PNAS, (2017). Court case in PA.

Single Edge Ensembles





PA Single Edge Flip



Slowly Mixing Graph Families

Theorem (D., Najt, and Solomon 2019)

Let G be any connected graph. Then let $G^{(d)}$ be the graph obtained by replacing each edge by a doubled d-star. Then the flip walk on partitions of family of graphs $G^{(d)}_{d\geq 1}$ is slowly mixing, in the sense the Cheeger constant is decaying exponentially fast. More specifically:

$$H(Partition\ Graph(G^{(d)}) = O(2^{-d})$$

Slowly Mixing Graph Families

Theorem (D., Najt, and Solomon 2019)

Let G be any connected graph. Then let $G^{(d)}$ be the graph obtained by replacing each edge by a doubled d-star. Then the flip walk on partitions of family of graphs $G^{(d)}_{d \geq 1}$ is slowly mixing, in the sense the Cheeger constant is decaying exponentially fast. More specifically:

$$H(Partition\ Graph(G^{(d)}) = O(2^{-d})$$

Remark

There are many similar constructions that give rise to equivalent mixing results.

Slow Mixing Example

Slow Mixing Example

Uniform Sampling of Contiguous Partitions

Theorem (D., Najt, and Solomon 2019)

Suppose that $\mathscr{C}$ *is the class of connected planar graphs and* $k \geq 2$*. If there is a polynomial time algorithm to sample uniformly from:*

- *the connected k-partitions of graphs in* $\mathscr{C}$*,*
- *or the connected, 0-balanced k-partitions of graphs in* $\mathscr{C}$*.*

then $RP = NP$*.*

Uniform Sampling of Contiguous Partitions

Theorem (D., Najt, and Solomon 2019)

Suppose that $\mathscr{C}$ is the class of connected planar graphs and $k \geq 2$. If there is a polynomial time algorithm to sample uniformly from:

- *the connected k-partitions of graphs in $\mathscr{C}$,*
- *or the connected, 0-balanced k-partitions of graphs in $\mathscr{C}$.*

then $RP = NP$.

Remark

This theorem has various interesting extensions, including:

- *Connectivity constraints on $\mathscr{C}$*
- *Degree bounds*
- *Distributions proportional to cut length*
- *TV distribution approximation*

Stronger Version Example

Theorem (D., Najt, and Solomon 2019)

Let $\mathscr{C}$ be the class of cubic, planar 3-connected graphs, with face degree bounded by $C = 60$. Let $\mu_x(G)$ be the probability measure on $P_k(G)$ such that a partition P is drawn with probability proportional to $x^{cut(P)}$. Fix some $x > 1/\sqrt{2}$, $\epsilon > 0$ and $\alpha < 1$. Suppose that there was an algorithm to sample from $P_2^\epsilon(G)$ according to a distribution $\nu(G)$, such that $||\nu_G - \mu_x(G)||_{TV} < \alpha$, which runs polynomial time on all $G \in \mathscr{C}$. Then $RP = NP$.

Proof Outline Sketch

Following technique of Jerrum, Valiant, and Vazirani[1].

1. Show that uniformly sampling simple cycles is hard on some class $\mathscr{C}$
   1. Choose a gadget that respects $\mathscr{C}$ and allows us to concentrate probability on long cycles
   2. Count the proportion of cycles as a function of length
   3. Reduce to Hamiltonian path on the graph class
2. Show closure of class under planar dual
3. Identify partitions with cut edges $\mapsto$ simple cycles (via planar duality)
4. Conclude that sampling partitions would allow you to sample from cycles which would allow you to find Hamiltonian cycles

□

[1] M. Jerrum, L. Valiant, and V. Vazirani, Random generation of combinatorial structures from a uniform distribution, Theoretical Computer Science, 43 (1986), 169–188.

Proof Sketch – Planar 2–Partitions

Still following technique of Jerrum, Valiant, and Vazirani.

1. Let $\mathscr{C}$ be the planar connected graphs
    1. Replace the edges with chains of dipoles
    2. Hamiltonian hardness for $\mathscr{C}$ given by [1]
2. $\mathscr{C}$ closed under planar duals
3. Identify partitions with cut edges (via planar duality)

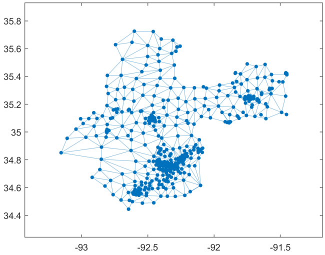
□

[1] M. Garey, D. Johnson, and R. Tarjan, The Planar Hamiltonian Circuit Problem is NP-Complete, SIAM Journal on Computing, 5, (1976), 704–714.

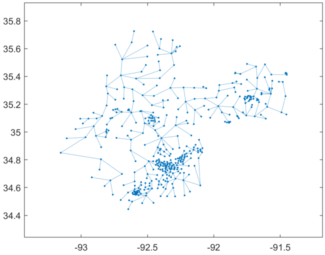
Weighted Graphs

The case for node-weighted graphs also easily leads to hard problems. Our graphs come with weight functions $W : V \to \mathbb{R}$, and a partition is ε-balanced if $\mu_W(1-\varepsilon) \leq W(V_i) \leq \mu_W(1+\varepsilon)$ where $W(V_i) = \sum_{u \in V_i} W(u)$ and μ_W is the mean of the $W(V_i)$. By considering the complete graph with proscribed integer node weights, finding a connected component of a given size is the `SUBSET-SUM` problem and finding a k–equi partition is `BIN PACKING`.

Tree based methods

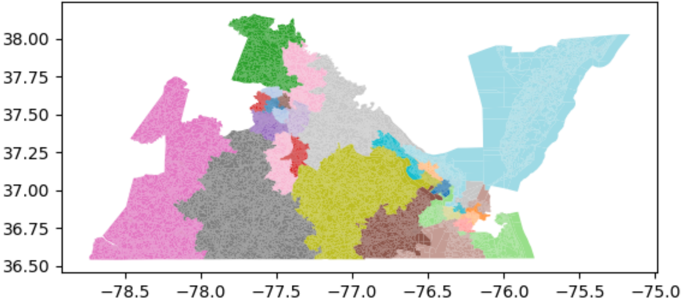


(a) District



(b) Spanning Tree

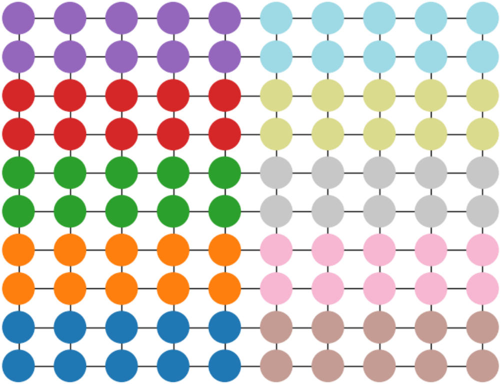
Tree Seeds Ensemble



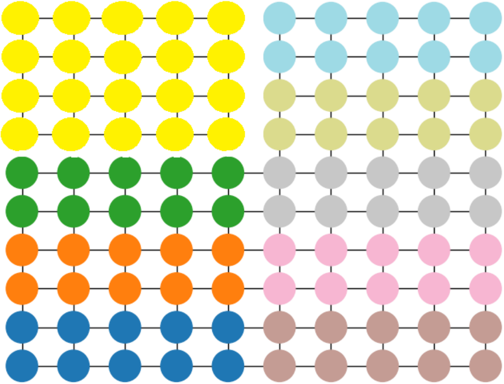
Recombination Steps

1. At each step, select two adjacent **districts**
2. Merge the subunits of those two districts
3. Draw a spanning tree for the new super–district
4. Delete an edge leaving two population balanced districts
5. Repeat
6. (Optional) Mix with single edge flips

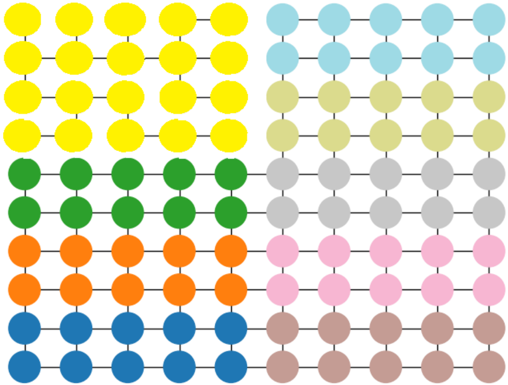
Recombination Step Example



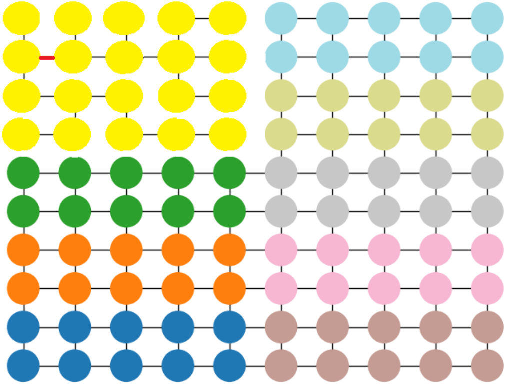
Recombination Step Example



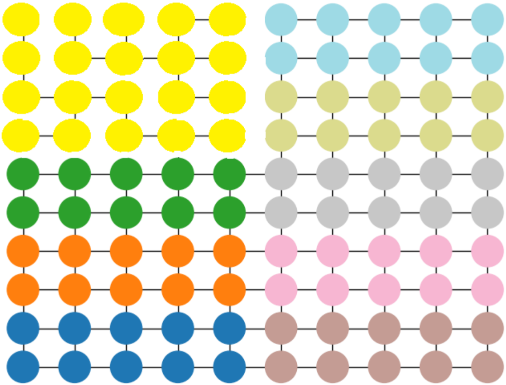
Recombination Step Example



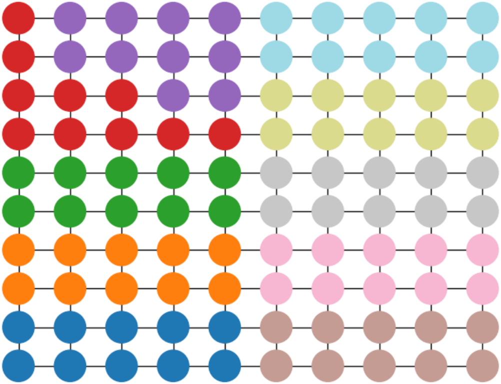
Recombination Step Example



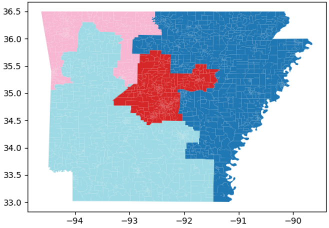
Recombination Step Example

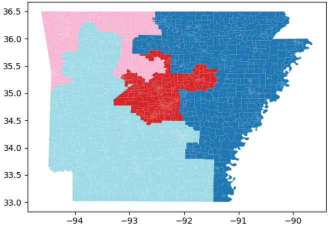


Recombination Step Example



AR Ensembles





PA Recombination Steps

General Tree Proposals

1. Form the induced subgraph on the complement of the cut edges
2. Add some subset of the cut edges
3. Uniformly select a maximal spanning forest
4. Apply a Markov chain on trees
5. Partition the spanning forest into k population balanced pieces

Special Cases

- Uniform Trees: Add all cut edges
- k–edges: Uniformly add k cut edges
- Recombination: Add all cut edges between one pair of districts.
- Super-Recombination: Take a maximal matching on the dual graph to the districts and add all cut edges between matched districts.
- Bounce Walk: Add a single cut edge between enough pairs of districts to make a tree in the dual graph of districts.

Special Cases

- Uniform Trees: Add all cut edges
- k–edges: Uniformly add k cut edges
- Recombination: Add all cut edges between one pair of districts.
- Super-Recombination: Take a maximal matching on the dual graph to the districts and add all cut edges between matched districts.
- Bounce Walk: Add a single cut edge between enough pairs of districts to make a tree in the dual graph of districts.

Question

What are the steady state distributions (and mixing times) of these walks?

Tree Partitioning Questions

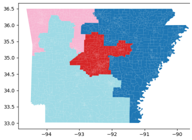
- Characterizing the distribution on partitions defined by cutting trees!
- How bad is the best cut?
- Criteria for determining when a tree is ε cuttable?
- Criteria for determining when all spanning trees of a graph are ε cuttable?
- How hard is it to find the mininum ε for which a cut exists?
- As a function of ε what proportion of spanning trees are cuttable?
- As a function of ε what proportion of edges in a given tree are cuttable?
- What is the fastest way to sample uniformly from $k-1$ balanced cut edges?

General Merge Proposals

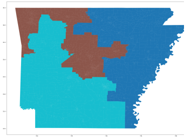
1. At each step, select two adjacent **districts**
2. Merge the subunits of those two districts
3. Bipartition the new super–district
4. Repeat
5. (Optional) Mix with single edge flips

General Merge Proposals

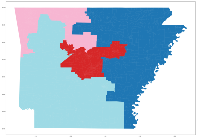
1. At each step, select two adjacent **districts**
2. Merge the subunits of those two districts
3. Bipartition the new super–district
4. Repeat
5. (Optional) Mix with single edge flips



(a) Before



(b) During



(c) After

Bipartitioning Methods

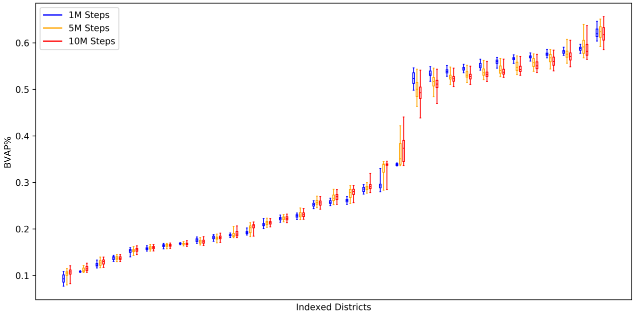
- Trees!
- Flood Fills
- Path Fills
- Agglomerative/Hierarchical
- Spectral
- Min Cut

More details (and colorful figures) at:
https://www.overleaf.com/read/zpmyzqmpvmnx

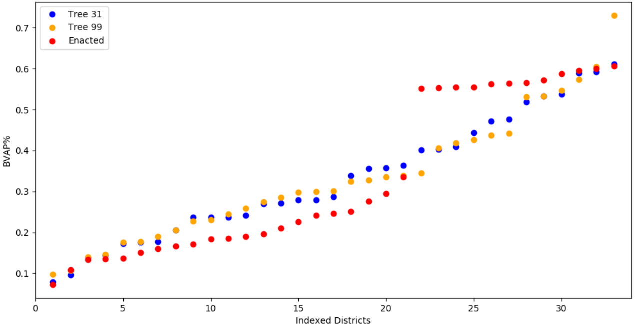
MORAL:

Computational Redistricting is NOT a solved problem!

Boundary Flip Mixing – Length

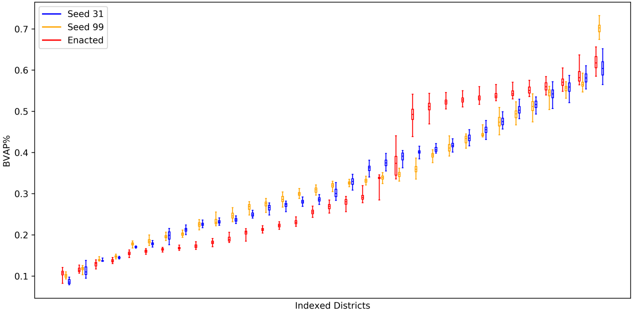


Initial Seeds



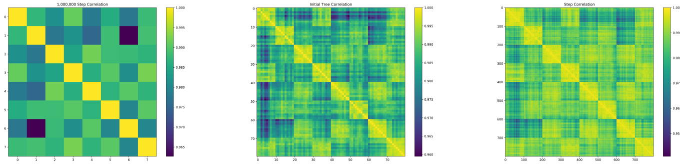
(a) Initial

Boundary Flip Mixing – Seeds



(a) 10,000,000 Flip Steps

Initial State Correlation

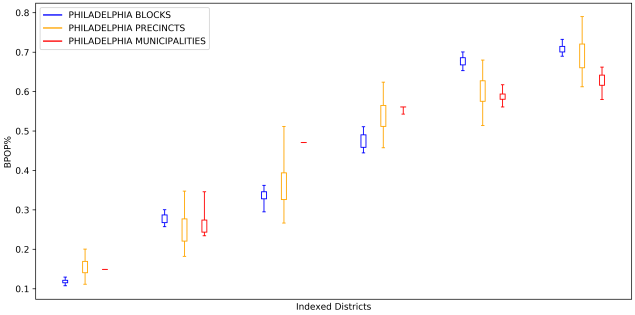


(a) Seeds (b) 100k Steps (c) 10K Steps

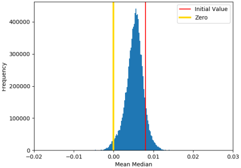
Philadelphia Units

(a) Blocks (b) Precincts (c) Wards

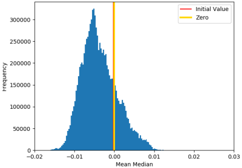
Boundary Flip Mixing – Resolution



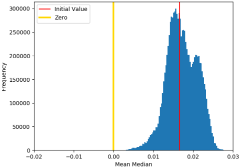
Boundary Flip Mean–Median



(a) Flip Seed31

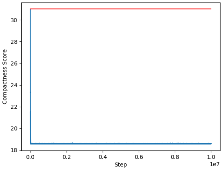


(b) Flip Seed99

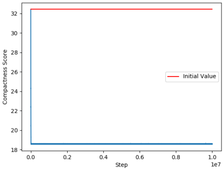


(c) Flip Enacted

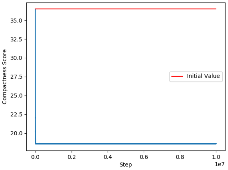
Compactness Comparison



(a) Tree 31

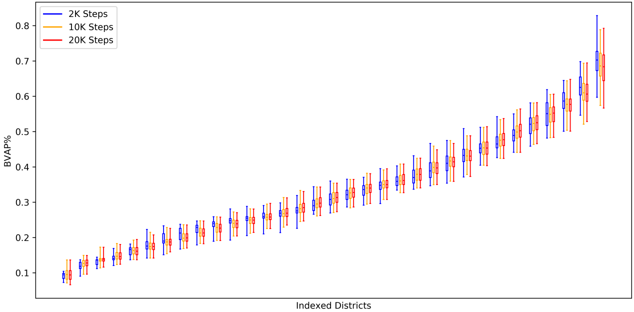


(b) Tree 99



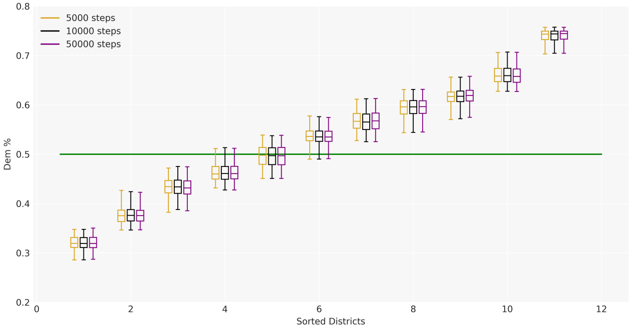
(c) Enacted

Recombination Mixing – Length



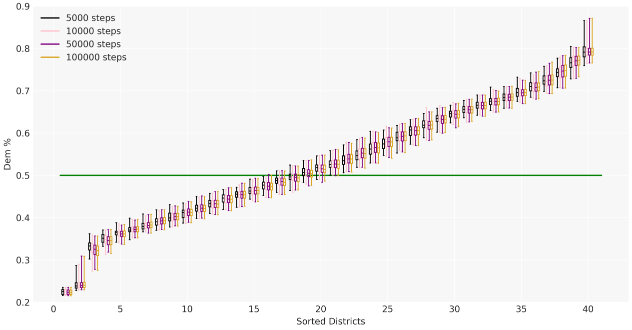
(a) 20,000 Recombination Steps

Recombination Mixing – Length



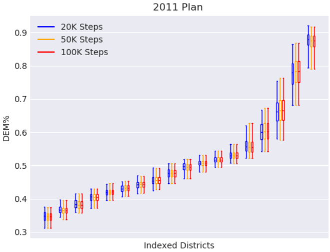
(a) 50,000 Recombination Steps

Recombination Mixing – Length

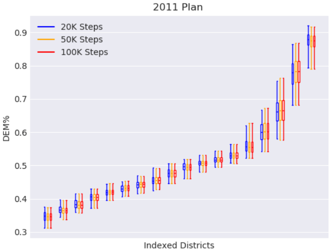


(a) 50,000 Recombination Steps

Recombination Mixing – Length

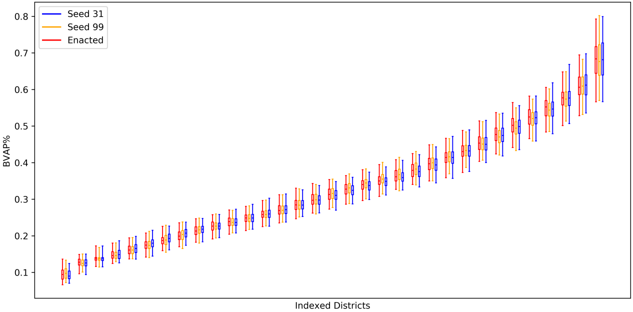


(a) 2011 Seed



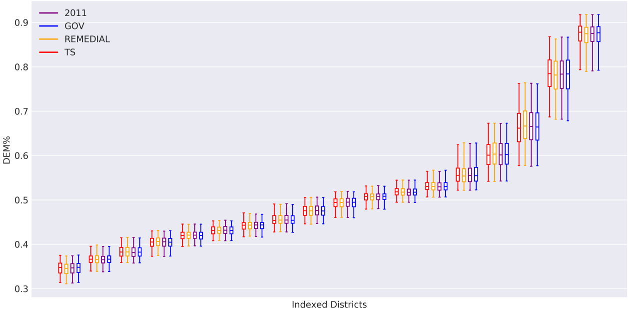
(b) GOV Seed

Recombination Mixing – Seeds



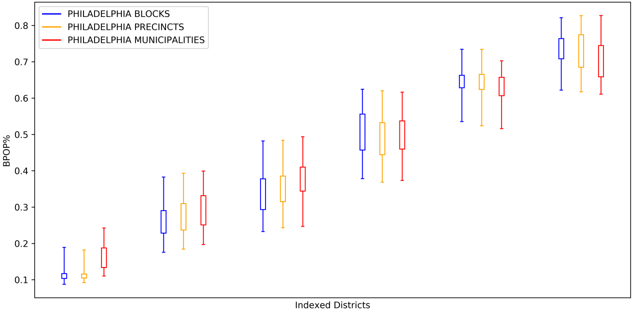
(a) 20,000 Recombination Steps

Recombination Mixing – Seeds

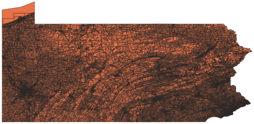


(a) 20,000 Recombination Steps

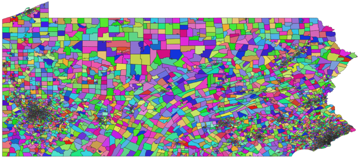
Recombination Mixing – Resolution



Pennsylvania Units

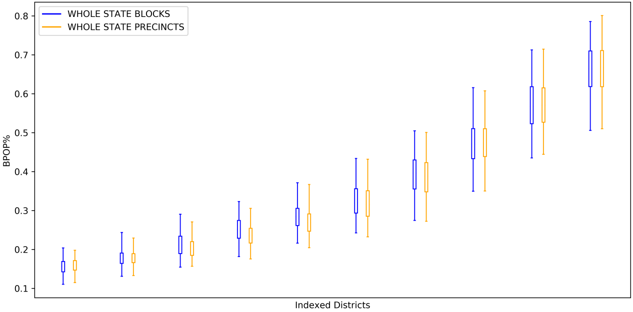


(a) Pennsylvania

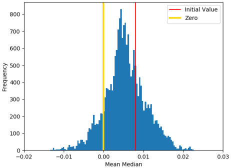


(b) Pennsylvania

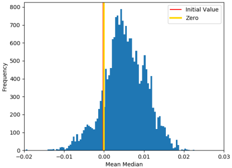
Recombination Mixing – Resolution



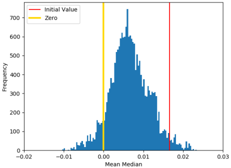
Recombination Mean–Median



(a) ReCom Seed31

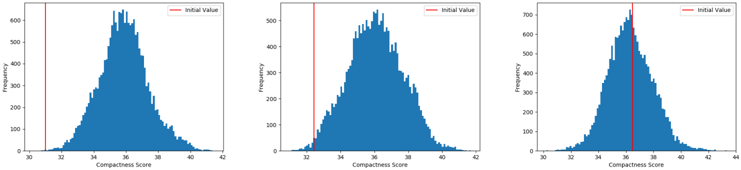


(b) ReCom Seed99



(c) ReCom Enacted

Compactness Comparison



(a) Tree 31 (b) Tree 99 (c) Enacted

Try it at home!

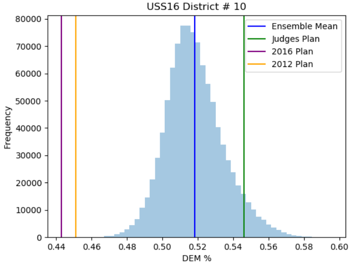
- Draw your own districts with **Districtr**
  - `https://districtr.org`
  - Easy to generate complete districting plans in browser or on a tablet
  - Measures district demographics and expected partisan performance
  - Identifies communities of interest
- Generate your own ensembles with **GerryChain**[1]
  - `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.

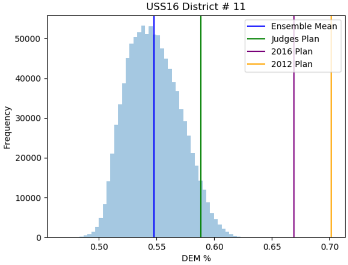
[1] Originally RunDMCMC

The End

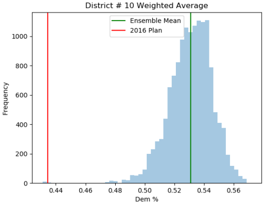
Thanks!

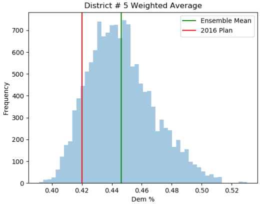
Ensemble Example: NC

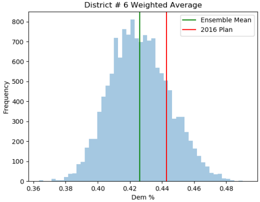


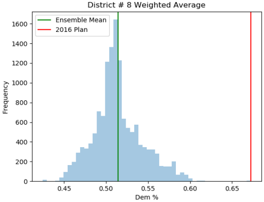


Ensemble Example: NC

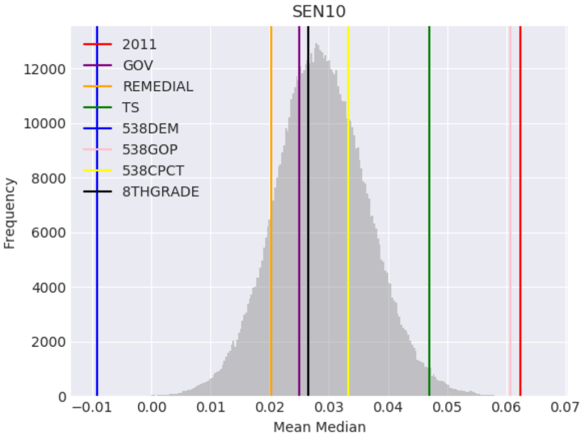








Ensemble Example: PA



Ensemble Example: PA