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.

Computational Challenges for Neutral Redistricting

Daryl DeFord

MIT/Tufts
Metric Geometry and Gerrymandering Group

SCADA 2018
Weizmann Institute of Science
December 17, 2018

Outline

1. Introduction
2. “Fairness” in Redistricting
3. Previous MCMC Approaches
4. Hardness Results
5. Tree Based Methods
6. Conclusion

Collaborators

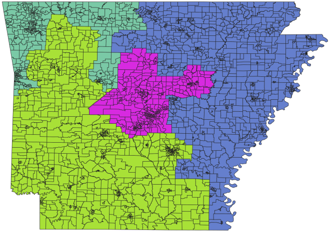
- Prof. Moon Duchin — **Tufts Math**
- Lorenzo Najt — **Wisconsin Math**
- Prof. Justin Solomon — **MIT CSAIL**

Collaborators

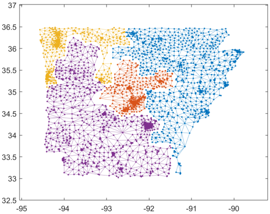
- Prof. Moon Duchin — **Tufts Math**
- Lorenzo Najt — **Wisconsin Math**
- Prof. Justin Solomon — **MIT CSAIL**

- Voting Rights Data Institute
  - 52 undergraduate and graduate students
  - 6–8 week summer program
  - mggg.org
  - github.com/{gerrymandr,mggg,mggg-states}

Political Partitioning

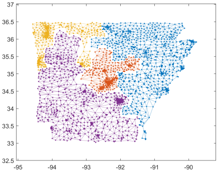


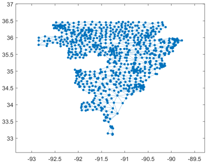
(a) Geography



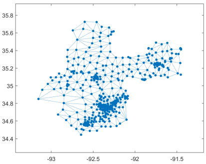
(b) Dual Graph

Arkansas Congressional Districts

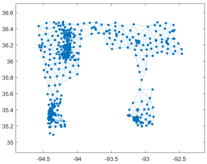




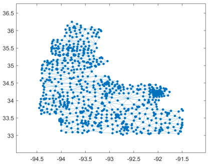
(b) District # 1



(c) District # 2



(d) District # 3



(e) District # 4

Districting Plans

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

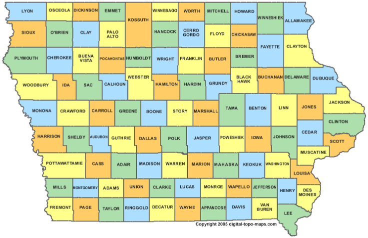
- Contiguity
- Population Balance
- Compactness
- Communities of Interest
- Municipal Boundaries
- Competitiveness/Responsiveness
- Incumbency Protection
- ...

Districting Plans

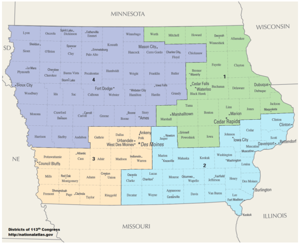
We want to partition a given geography (graph), at a given scale, into k pieces, satisfying some constraints:

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

Why Grids?



(a) Iowa Counties



(b) Congressional Districts

Figure: Still quintillions of possibilities...

Mathematical Formulation

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

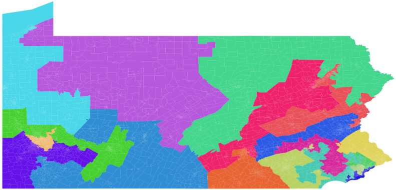
- A k-**partition** $P = \{V_1, V_2, \ldots, V_k\}$ of G is a collection of disjoint subsets $V_i \subseteq V$ whose union is V.
- A partition P is **connected** if the subgraph induced by V_i is connected for all i.
- The **cut edges** of P are the edges (u, w) for which $u \in V_i$, $w \in V_j$, and $i \neq j$
- 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

Ugly Shapes

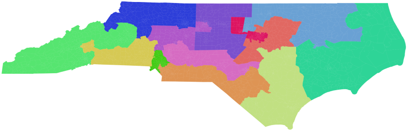
(a) NC12 #1

(b) NC12 #2

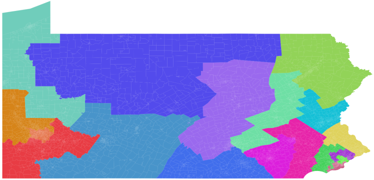
(c) NC12 #12



Partisan Imbalance



(e) NC16

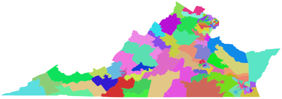


(f) PA TS-Proposed

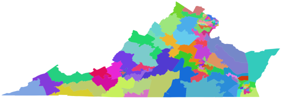
Partisan Fairness

- 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
- NJ
  - Controversial constitutional amendment
  - Competitiveness defined in terms of historical statewide averaging

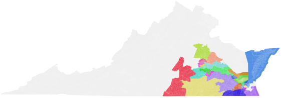
Virginia House of Delegates (https://mggg.org/VA-report.pdf)



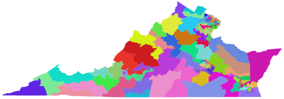
(a) Enacted



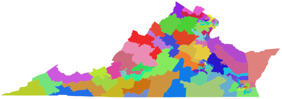
(b) Dem



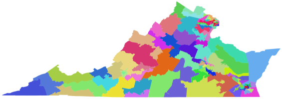
(c) Princeton



(d) GOP1

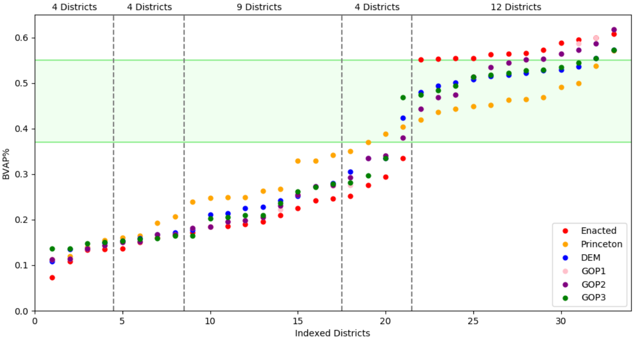


(e) GOP2

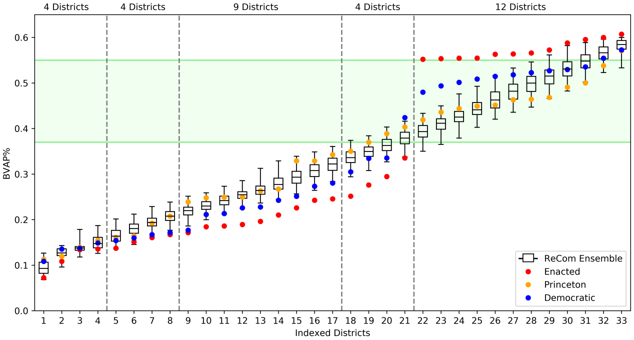


(f) GOP3

Racial Gerrymandering in Virginia



Racial Gerrymandering in Virginia



MORAL:

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

Outlier 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 do **outlier analysis** by comparing 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.

AR Outlier Example

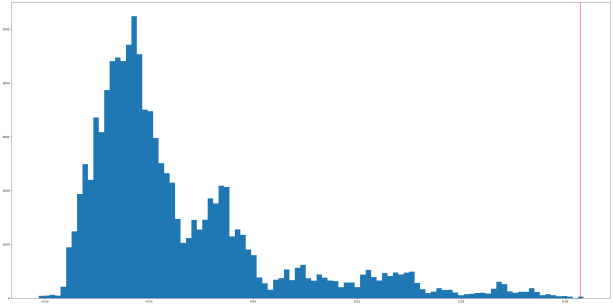
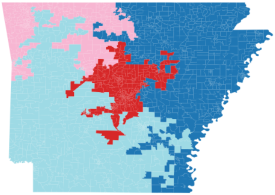
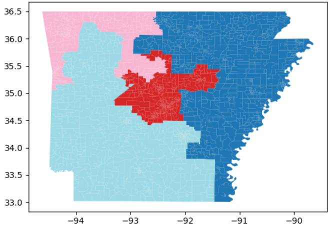


Figure: Mean–Median score using senate 2016 election data on 1,000,000 plans.

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

MCMC on partitions

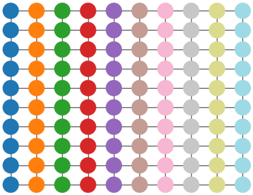
1. Set constraints to define the state space
2. Start with an initial plan
3. **Propose a modification**
4. Verify that the modification satisfies the constraints
5. Accept using MH criterion
6. Repeat

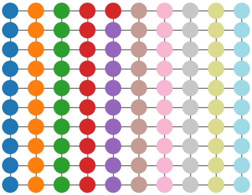
Why?

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

Single Edge Flip Proposals

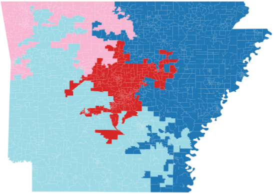
1. Uniformly choose a cut edge
2. Change one of the incident node assignments to the other

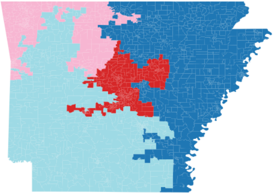




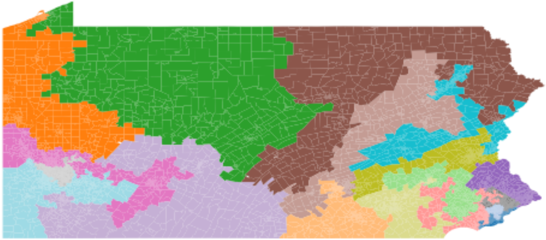
- Mattingly et al. (2017, 2018) Court cases in NC and WI.
- Pegden et al. Assessing significance in a Markov chain without mixing, PNAS, (2017). Court case in PA.

Single Edge Ensembles

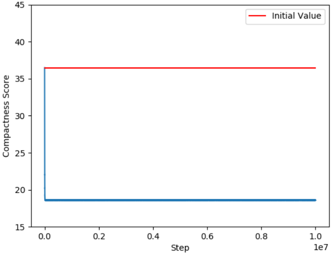




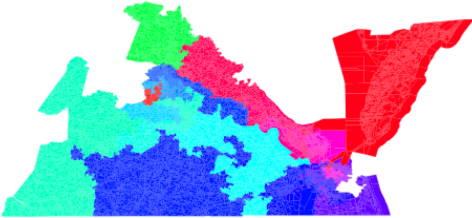
PA Single Edge Flip



Boundary Flip Distribution

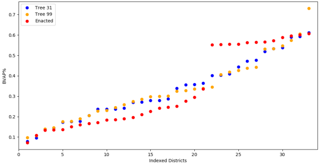


(a) Compactness

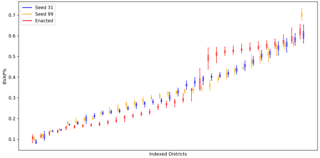


(b) 11996 cut edges

Boundary Flip Mixing

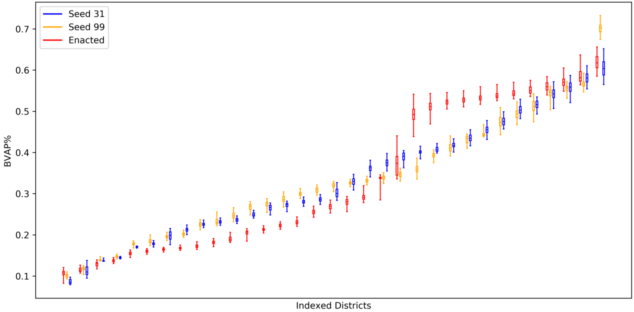


(a) Initial



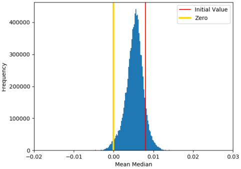
(b) 10,000,000 Flip Steps

Booundary Flip Mixing

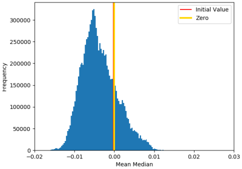


(b) 10,000,000 Flip Steps

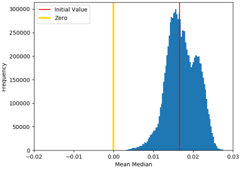
Boundary Flip Mean–Median



(a) Flip Seed31



(b) Flip Seed99



(c) Flip Enacted

Slowly Mixing Graph Families

Theorem

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:

$$H(Metagraph(G^{(d)}) = O(2^{-d})$$

Slow Mixing Example

Slow Mixing Example

Uniform Sampling of Contiguous Partitions

Theorem

Suppose that D is the class of connected planar graphs. If there is a polynomial time algorithm to sample uniformly from any of:

- *the connected 2-partitions of graphs in D,*
- *the connected, 0-balanced 2-partitions of graphs in D,*
- *or the connected, 0-balanced k-partitions of graphs in D.*

then $RP = NP$.

Uniform Sampling of Contiguous Partitions

Theorem

Suppose that D is the class of connected planar graphs. If there is a polynomial time algorithm to sample uniformly from any of:

- *the connected 2-partitions of graphs in D,*
- *the connected, 0-balanced 2-partitions of graphs in D,*
- *or the connected, 0-balanced k-partitions of graphs in D.*

then $RP = NP$.

Conjecture

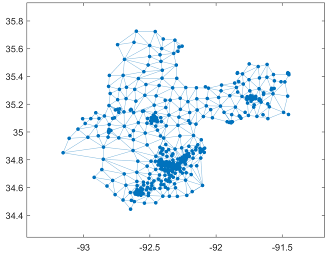
The same holds for uniform sampling of connected k-partitions.

Weighted Graphs

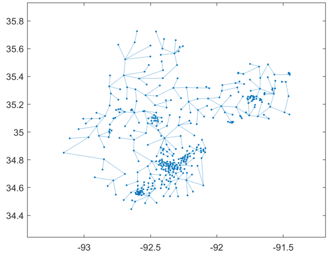
The case for node-weighted graphs also easily leads to hard problems. Our graphs come with weight functions $W : V \to \mathbb{R}$, and a partition is ε-balanced if $\mu_W(1-\varepsilon) \leq W(V_i) \leq \mu_W(1+\varepsilon)$ where $W(V_i) = \sum_{u \in V_i} W(u)$ and μ_W is the mean of the $W(V_i)$.

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

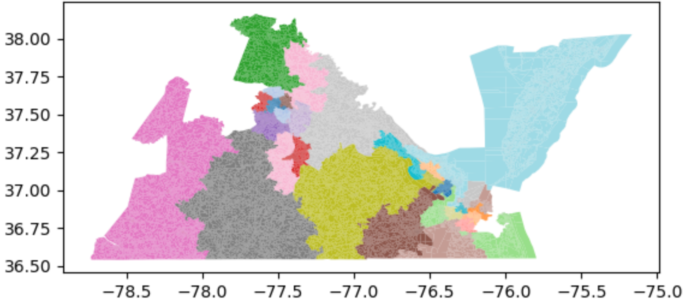


(a) District



(b) Spanning Tree

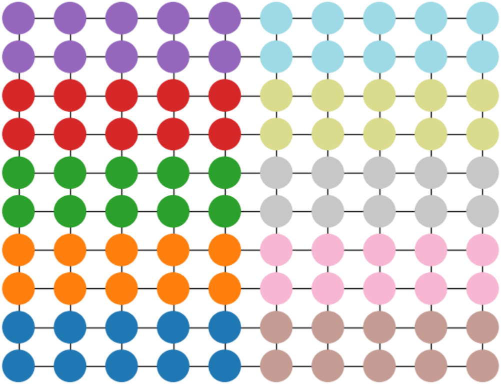
Tree Seeds Ensemble



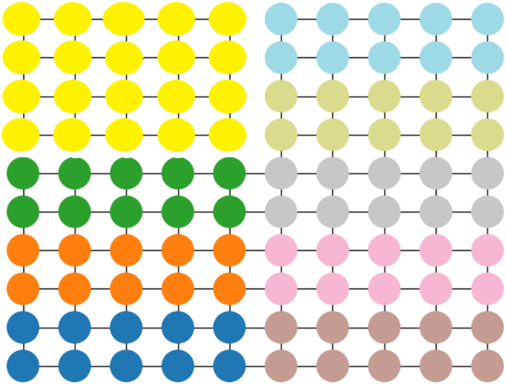
Recombination Steps

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

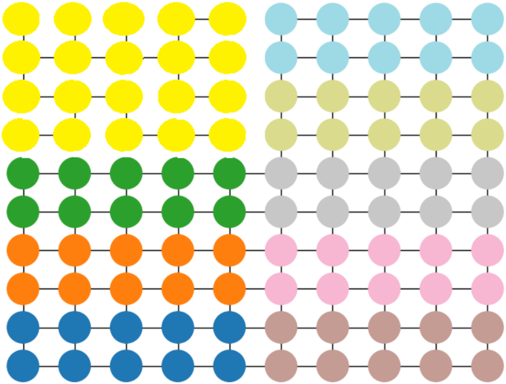
Recombination Step Example



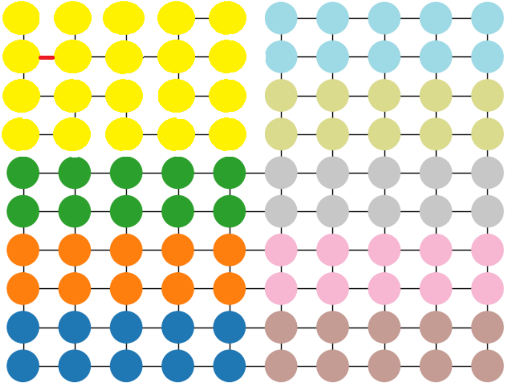
Recombination Step Example



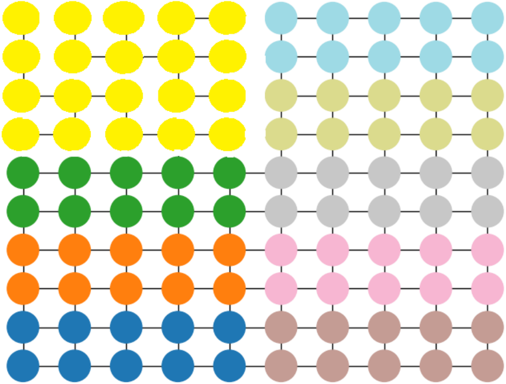
Recombination Step Example



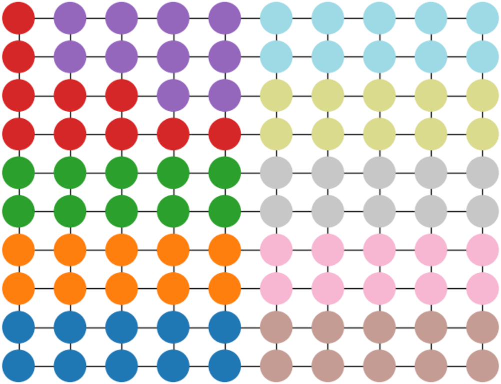
Recombination Step Example



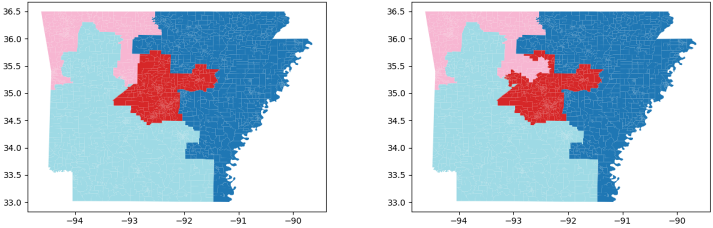
Recombination Step Example



Recombination Step Example

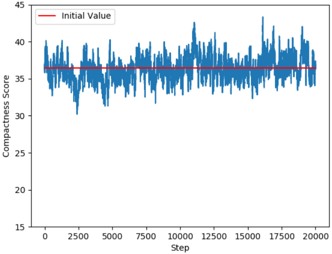


AR Ensembles

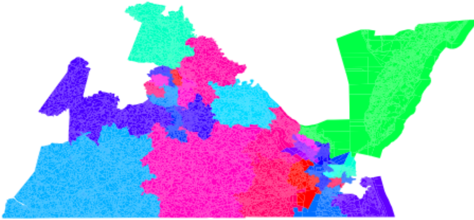


PA Recombination Steps

Recombination Distribution

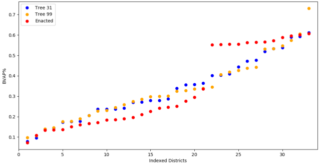


(a) Compactness

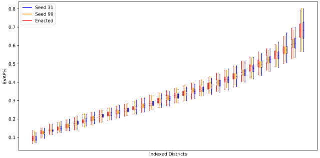


(b) 5702 cut edges

Recombination Mixing

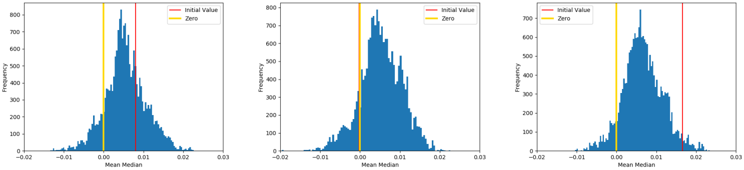


(a) Initial



(b) 20,000 Recombination Steps

Recombination Mean–Median



(a) ReCom Seed31 (b) ReCom Seed99 (c) ReCom Enacted

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?

MORAL:

Computational Redistricting is NOT a solved problem!

The End

Thanks!