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

Markov Chain Sampling for Connected Graph Partitions

Daryl DeFord

MIT – CSAIL
Geometric Data Processing Group

JMM 2020
Denver, CO
January 17, 2020

Collaborators

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

- *Complexity and Geometry of Sampling Connected Graph Partitions* (with L. Najt and J. Solomon), arXiv: 1908.08881.
- *ReCombination: A family of Markov chains for redistricting* (with M. Duchin and J. Solomon), arXiv:1911.05725
- Redistricting Reform in Virginia: Districting Criteria in Context (with M. Duchin), Virginia Policy Review, 12(2), 120-146, (2019).

MORAL:

MORAL:

Computational Redistricting is NOT a solved problem!

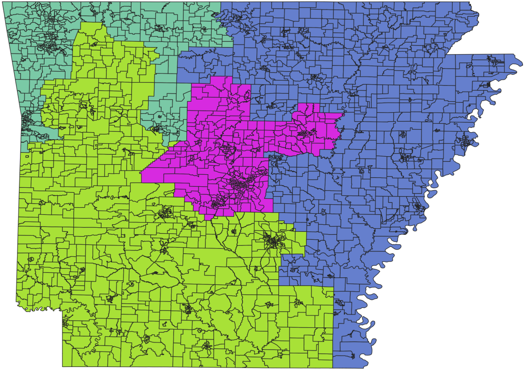
☹

MORAL:

Computational Redistricting is NOT a solved problem!

☺

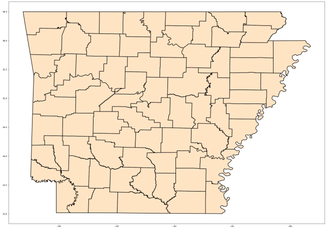
What is a district?

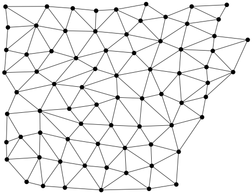


Permissible Districting Plans

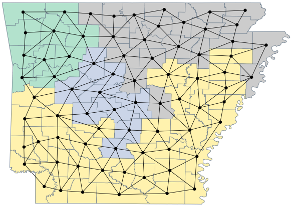
- Contiguity
- Population Balance
- Compactness
- Communities of Interest
- Municipal Boundaries
- Competitiveness/Responsiveness/Symmetry
- Incumbency Protection
- ...

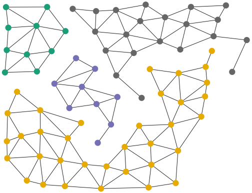
Discrete Partitioning





Discrete Partitioning



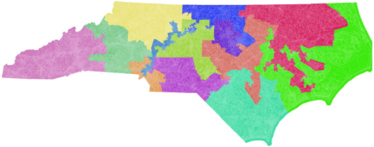


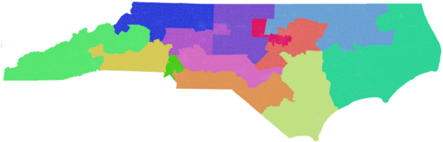
Mathematical Formulation

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. The full set of k–partitions of G will be denoted $\mathcal{P}_k(G)$.
- A partition P is **connected** if the subgraph induced by V_i is connected for all i.
- 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
- The (context dependent) collection of constraints will be denoted with a function $\mathcal{C}_\theta : \mathcal{P}_k(G) \mapsto \{\texttt{True}, \texttt{False}\}$. The set of permissible partitions will be $\mathcal{C}_\theta(G)$.

What is Gerrymandering?





Abstracted Problem Instances

Problem

Given a fixed G *and metric of interest* $f : \mathcal{P}(G) \mapsto \mathbb{R}^n$.

1. *Given a partition* P, *is it a statistical outlier*[a] *with respect to* f?
2. *Given* $\mathcal{C}_\theta$ *and* $\mathcal{C}_{\theta'}$ *how do the distributions* $f(\mathcal{C}_\theta(G))$ *and* $f(\mathcal{C}_{\theta'}(G))$ *compare?*

[a] gerrymander

Abstracted Problem Instances

Problem

Given a fixed G *and metric of interest* $f : \mathcal{P}(G) \mapsto \mathbb{R}^n$.

1. *Given a partition* P*, is it a statistical outlier*[a] *with respect to* f*?*
2. *Given* $\mathcal{C}_\theta$ *and* $\mathcal{C}_{\theta'}$ *how do the distributions* $f\left(\mathcal{C}_\theta(G)\right)$ *and* $f\left(\mathcal{C}_{\theta'}(G)\right)$ *compare?*

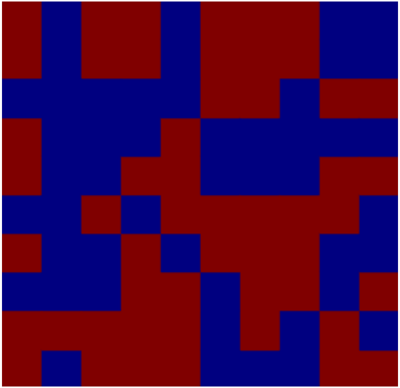
[a] gerrymander

Solution?

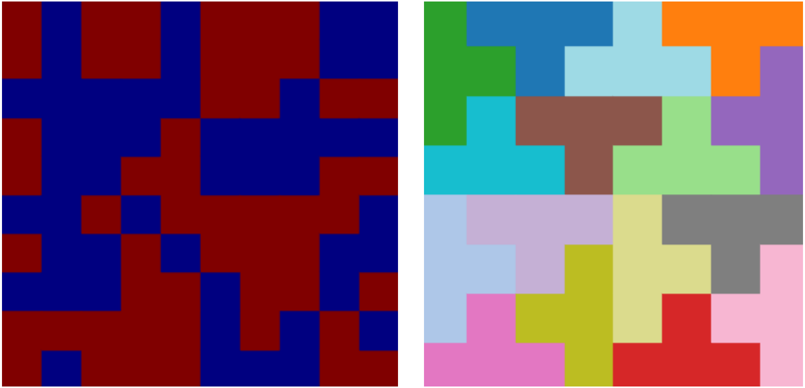
Draw (many) samples from $\mathcal{C}_\theta(G)$*!*

Other Partition Sampling Frameworks

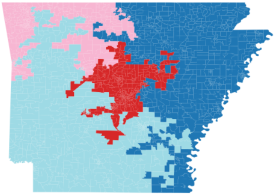
Other Partition Sampling Frameworks

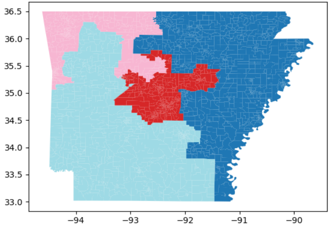


Other Partition Sampling Frameworks

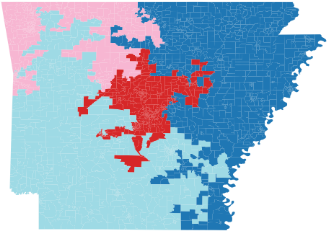


Which ensembles?





Single Node Flip Ensembles



Slowly Mixing Graph Families

Theorem (Najt, D., 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 (Najt, D., 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:

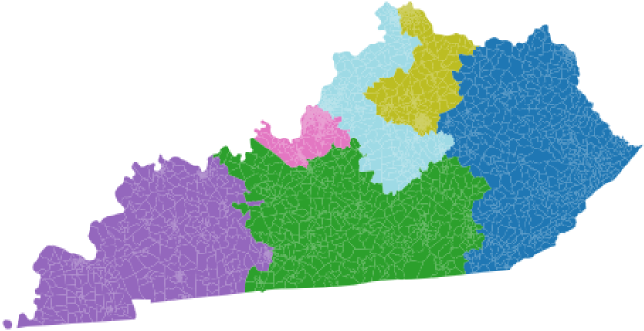
$$H(Partition\ Graph(G^{(d)}) = O(2^{-d})$$

Remark

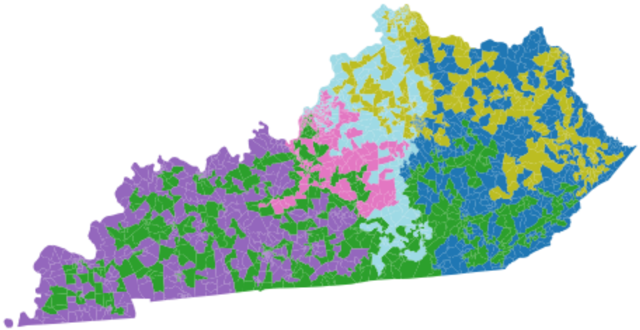
There are many similar constructions that give rise to equivalent mixing results.

Slow Mixing Example

Starting Partition



Generic Partition



Uniform Sampling of Contiguous Partitions

Theorem (Najt, D., 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 (Najt, D., 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* $\mathscr{C}$*.*

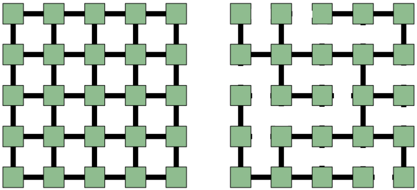
then $RP = NP$*.*

Remark

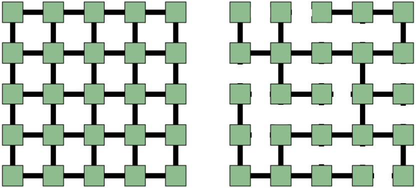
This theorem has various interesting extensions, including:

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

New Proposal: Spanning Trees



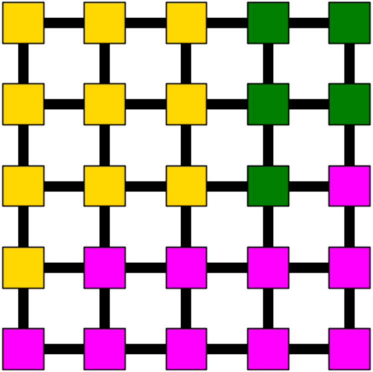
New Proposal: Spanning Trees



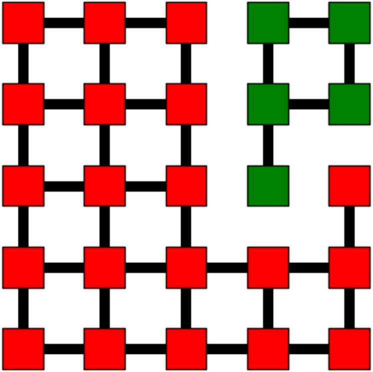
ReCombination

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

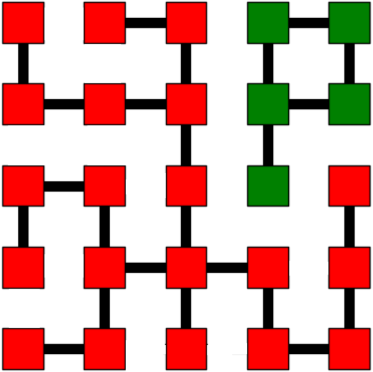
ReCombination Example



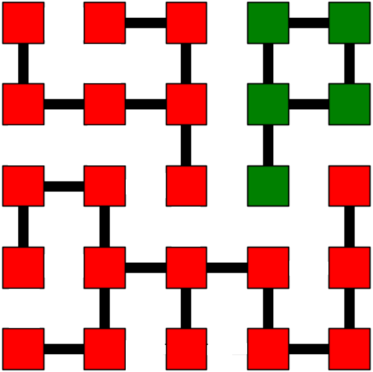
ReCombination Example



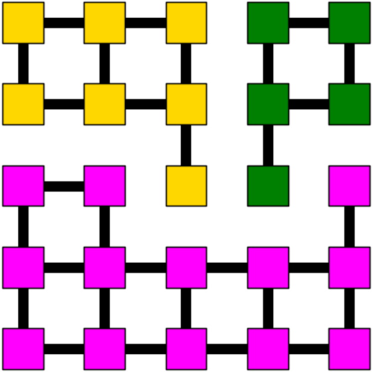
ReCombination Example



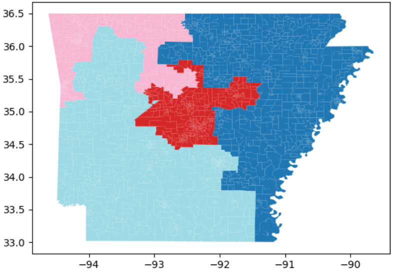
ReCombination Example



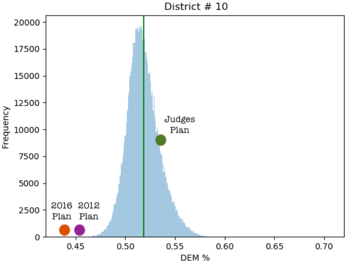
ReCombination Example

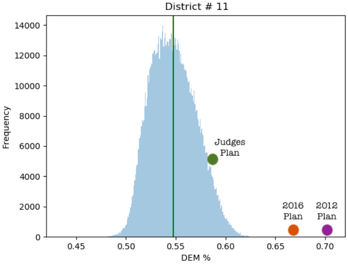


Tree Ensembles



Amicus Brief





The End

Thanks!

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**
  - https://github.com/mggg/gerrychain
  - Flexible, modular software for sampling graph partitions
  - Handles the geodata processing as well as the MCMC sampling
  - Templates to get started: https://github.com/drdeford/GerryChain-Templates
  - Detailed documentation: http://people.csail.mit.edu/ddeford/GerryChain_Guide.pdf
- Data is available for your favorite state!
  - Census dual graphs with demographic information:
  - https://people.csail.mit.edu/ddeford/dual_graphs
  - Precincts with electoral results
  - https://github.com/mggg-states