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 Approaches for Political Redistricting

Daryl DeFord

CSAIL – GDP Group

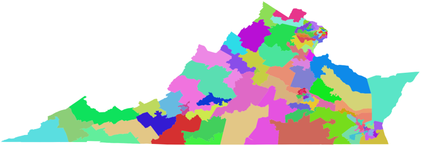
IAP 2019
Massachusetts Institute of Technology
January 8, 2019

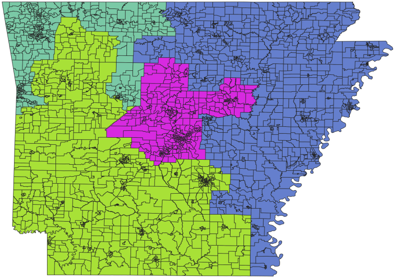
MORAL:

Computational Redistricting is NOT a solved problem!

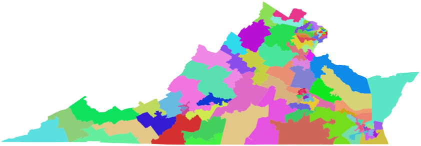
Outline

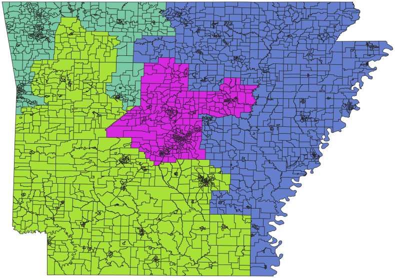
Electoral Districts





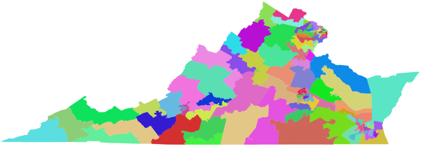
Electoral Districts

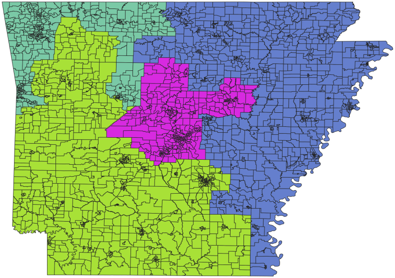




Example (Which states are these?)

Electoral Districts





Example (Which states are these?)

1. Virginia
2. Arkansas

Representative Democracy

Example (What types of districts?)

- Congressional Representatives

Representative Democracy

Example (What types of districts?)

- Congressional Representatives
- State House and Senate
- School Districts
- Municipal Representatives
- ...

Representative Democracy

Example (What types of districts?)

- Congressional Representatives
- State House and Senate
- School Districts
- Municipal Representatives
- ...

- Data
  - Decennial census
  - Fundamental subunits
  - Voting data

Representative Democracy

Example (What types of districts?)

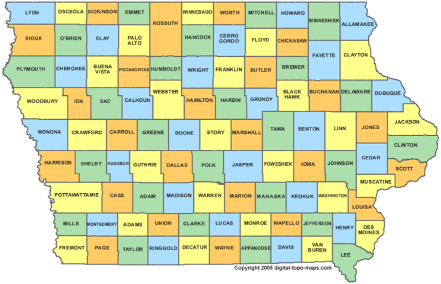
- Congressional Representatives
- State House and Senate
- School Districts
- Municipal Representatives
- ...

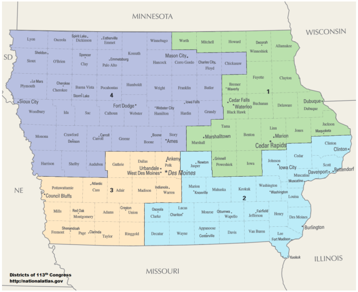
- Data
  - Decennial census
  - Fundamental subunits
  - Voting data
- Partisan Considerations
  - Number of seats
  - Ease of victory
  - Incumbency
  - ...

State Level Differences

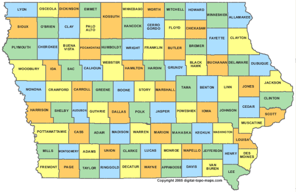
- Number and type of districts
- Constraints on districts
- Map drawers
- Map approval process
- Political geography
- ...

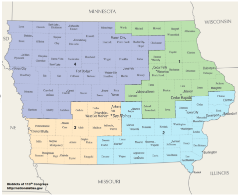
Example: Iowa





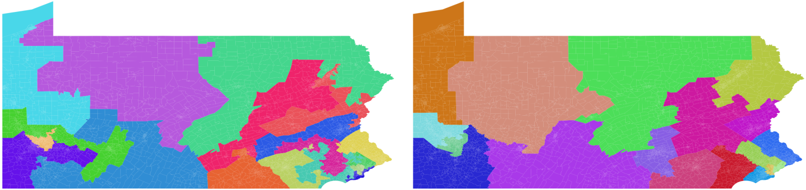
Example: Iowa





- 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!

Common Rules

Example (What are some standard requirements?)

- Population Balance

Common Rules

Example (What are some standard requirements?)

- Population Balance
- Contiguity
- Compactness
- Communities of Interest
- VRA Compliance
- Municipal Boundaries
- ...

What is, not what might be...

- Why don't we change the law?
- Why focus on these specific rules?
- Why not just let computers draw the lines?
- Why involve mathematicians at all?
- ...

Recent Legislation

- Very Process Focused
- Too specific
- Not specific enough
- Exploitable

Recent Legislation

- Very Process Focused
- Too specific
- Not specific enough
- Exploitable

	Michigan	Utah	Colorado	Missouri
1	VRA + POP	VRA + POP	POP ($<$ 5%) + Contiguous	POP
2	Contiguous	Municipalities	VRA	VRA
3	Communities of interest	Compactness	COI + Municipal	Representatives of choice
4	Partisan fairness	Contiguity	Compactness	Partisan fairness + competitiveness
5	No incumbency protection	Communities of interest	Competitiveness	Contiguity
6	Municipal boundaries	Natural boundaries		Municipal boundaries
7	Compactness	Align boundaries		Compactness

Example: Competitiveness

Example (What is a competitive district/plan?)

Example: Competitiveness

Example (What is a competitive district/plan?)

?

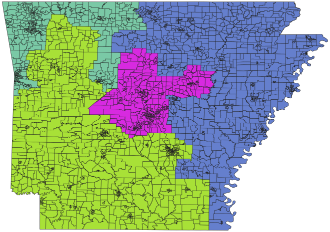
Example: Competitiveness

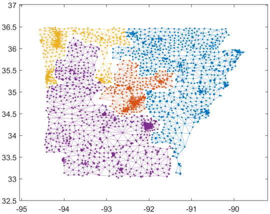
Example (What is a competitive district/plan?)

?

- Arizona
- Missouri
- New Jersey (almost)
- ...

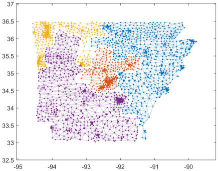
Mathematical Formalism

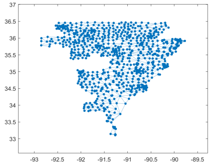


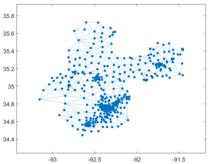


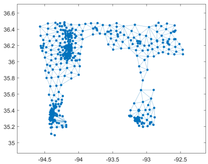
In order to study this problem mathematically we need to abstract the process of districting into the realm of mathematical objects. The first step is to discretize!

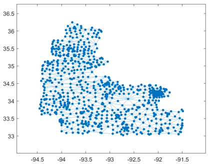
Graph Partitioning











Example: Population Equality

Example (How should we measure population imbalance?)

- Largest deviation from average

Example: Population Equality

Example (How should we measure population imbalance?)

- Largest deviation from average
- Largest difference between districts
- Euclidean distance to ideal
- Sum of individual deviations
- Sum of pairwise deviations
- ...

Other Measures

- Choice of geographic units
- Population Balance
- Contiguity
- Compactness
- Communities of Interest
- VRA Compliance
- Municipal Boundaries
- ...

MORAL:

Computational Redistricting is NOT a solved problem!

What is gerrymandering?

Example (What is gerrymandering?)

What is gerrymandering?

Example (What is gerrymandering?)

?

What is gerrymandering?

Example (What is gerrymandering?)

?

Example (Who is harmed by gerrymandering?)

What is gerrymandering?

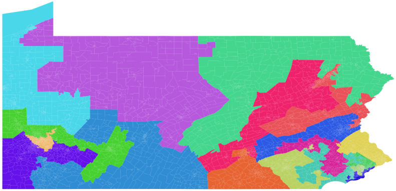
Example (What is gerrymandering?)

?

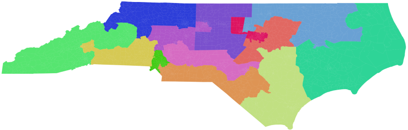
Example (Who is harmed by gerrymandering?)

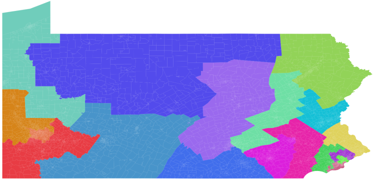
- Partisan
- Racial
- Incumbents
- ...

Ugly Shapes



Partisan Imbalance





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

Compactness and Partisan Measures

- Compactness
  - Polsby–Popper
  - Reock
  - Total perimeter
  - Convex hull
  - Discrete metrics
  - ...
- Partisan Imbalance
  - Mean–Median
  - Partisan Bias
  - Efficiency Gap
  - Proportionality
  - ...

MORAL:

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

All hope is not lost...

All hope is not lost...

- 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

All hope is not lost...

- 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.

All hope is not lost...

- 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.

Arkansas Outlier Example

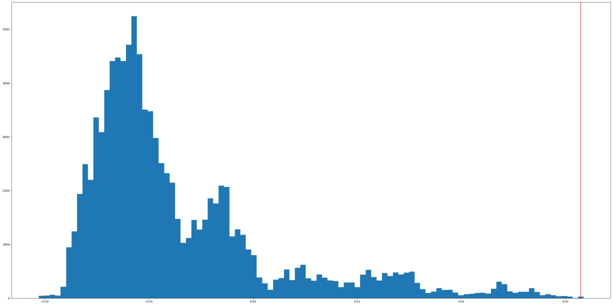
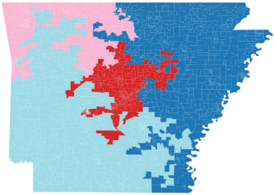
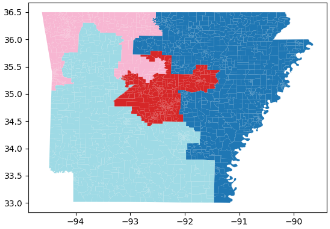


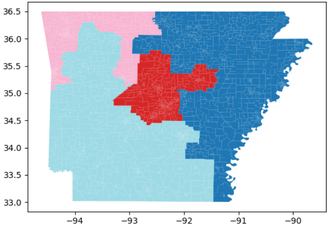
Figure: Mean–Median score using senate 2016 election data on 1,000,000 plans.

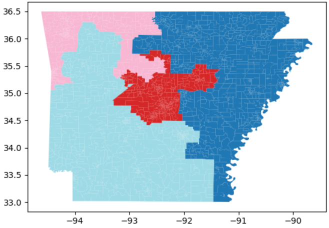
Which ensembles?



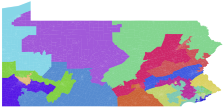


Arkansas Tree Ensembles



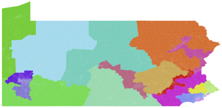


Pennsylvania Landscapes

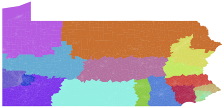


2011

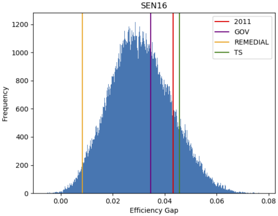
538 GOP

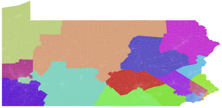


538 Dem



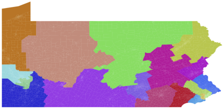
8th Grade



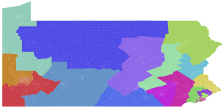


538 Compact

Gov

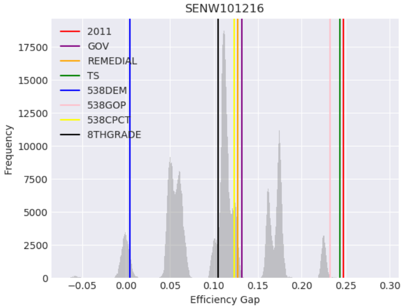


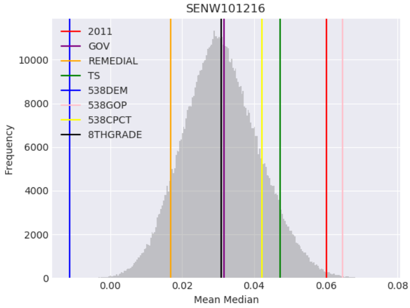
Remedial



TS

Pennsylvania Landscapes





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

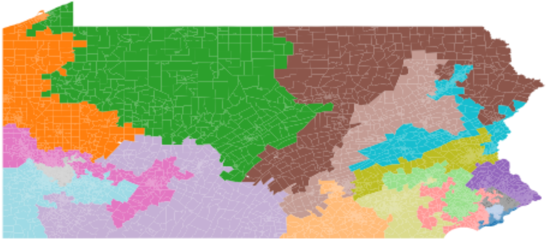
MCMC on partitions

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

Pennsylvania Single Edge Flip



Pennsylvania Recombination Steps

Necessary Inputs

- Geographic Units
  - Census vs. Political
  - Size of problem
- Demographic Data
  - Population
  - Voting Population
  - Race (for VRA)
- Initial Districting Plan
  - Discretization
- Voting Data
  - Aggregated in unusual units

Data Availability

Example (What adjective best describes US Electoral data?)

Data Availability

Example (What adjective best describes US Electoral data?)

Abominable*

Data Availability

Example (What adjective best describes US Electoral data?)

Abominable*

* Alternatively, any adjective from "Your a mean one, Mr. Grinch."

github.com/mggg-states

- Currently complete data for 7 states
- Varying levels of confidence in data fidelity
- Working to collect data for all states
- More importantly, constructing a pipeline for processing 2020 census data

Other Projects:

- OpenElections Project
- NvKelso
- Princeton Gerrymandering Project
- Redistricting Reform Project

MGGG Software

1. github.com/mggg
    - Districtr
    - GerryChain
2. github.com/gerrymandr
    - Compactness measures
    - Segregation Measures
    - State Specific Analyses
    - Jupyter Notebooks
    - ...
3. mggg.org
    - GridLandia
    - MetaGraph Sizes
    - EcologicalInference Apps

MGGG Software

1. github.com/mggg
   - **Districtr**
   - **GerryChain**
2. github.com/gerrymandr
   - Compactness measures
   - Segregation Measures
   - State Specific Analyses
   - Jupyter Notebooks
   - ...
3. mggg.org
   - GridLandia
   - MetaGraph Sizes
   - EcologicalInference Apps

MORAL:

Computational Redistricting is NOT a solved problem!

The End

Thanks!

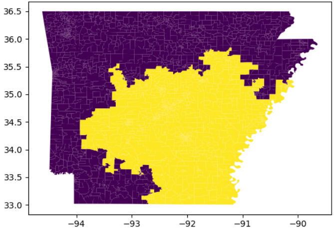
Advertisements

1. VRDI – 6 week summer program for graduate and undergraduate students (Deadline 2/1)
   - Application: tinyurl.com/apply-vrdi-2
   - Information: gerrydata.org
2. Contact:
   - Email: ddeford at mit.edu
   - Website: mggg.org
   - Slack channel: GerryChat.slack.com
3. Research Projects
   - Math Problems: tinyurl.com/gerryprojects
   - Data Problems: tinyurl.com/GerryChainProjects
4. More IAP sessions:
   - Resources: people.csail.mit.edu/ddeford/CAtPR
   - Today 12-1 More emphasis on building state level data sets
   - Thursday 12-1 MCMC and GerryChain
   - 1/22 12-1 Graph Partitions
   - 1/29 12-1 In–depth state examples

Agglomerative

Method

- *Start with each node in own component*
- *Select an arbitrary edge between two components*
    - *Merge clusters if population allows and doesn't disconnect the complement*
    - *If population doesn't allow, delete edge*
    - *If merging would disconnect the graph, merge the smallest population component*
- *Repeat until only 2 clusters*



Flood Fill

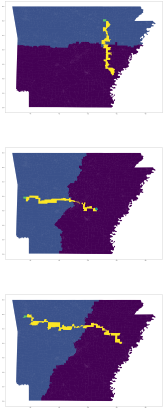
Method

- *Select a node at random*
- *Select a random neighbor of the current cluster*
- *Add if population allows and doesn't disconnect the complement*
- *Repeat until population balanced*

Min Cut

Method

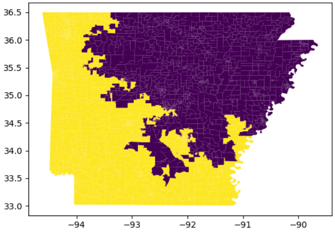
- *Select random source and sink nodes*
- *Weight the edges in the graph by* $10^{min\ distance - 3}$
- *Compute the min cut*
- *Repeat until population balanced*



Path Fill

Method

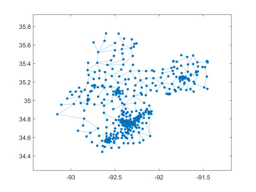
- *Start with an arbitrary node*
- *Select a node not in the district*
- *Add all the nodes on a shortest path from the new node to the district if it doesn't disconnect the complement or add too much to the population*
- *Repeat until population balanced*

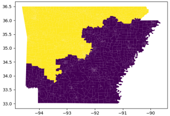


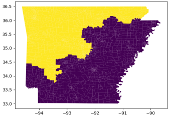
Tree Partitions

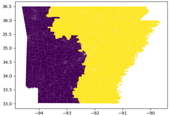
Method

- *Generate a uniform spanning tree*
- *Cut an edge that leaves population balanced components*



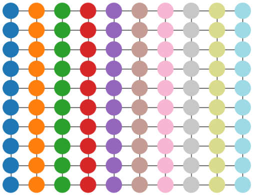


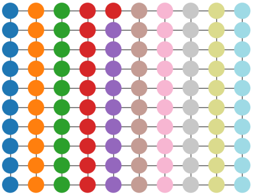




Single Edge Flip Proposals

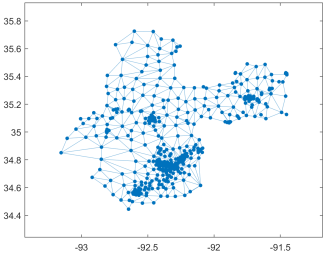
1. Uniformly choose an edge between districts
2. Change one of the incident node assignments to match the other



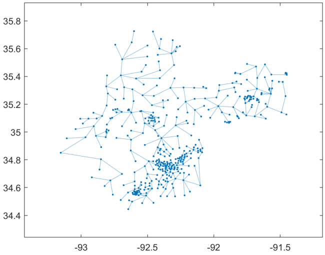


- Mattingly et al. (2017, 2018) Court cases in NC and WI.
- Pegden et al. Assessing significance in a Markov chain without mixing, PNAS, (2017). Court case in PA.

Tree based methods

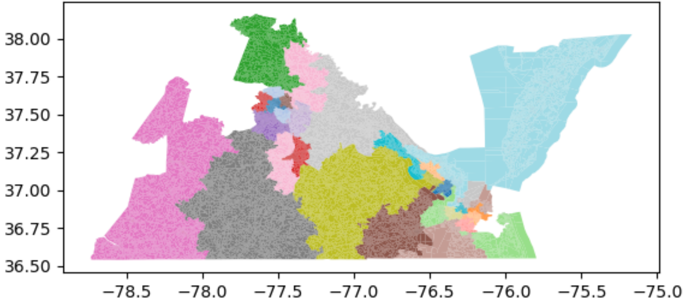


(a) District



(b) Spanning Tree

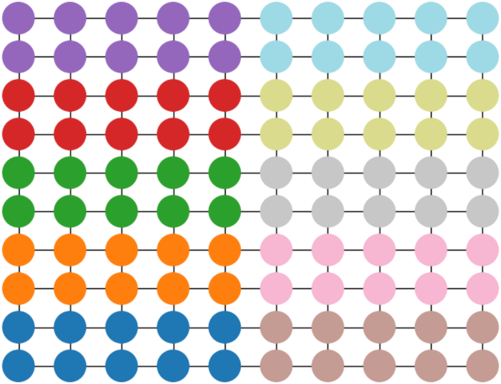
Tree Seeds Ensemble



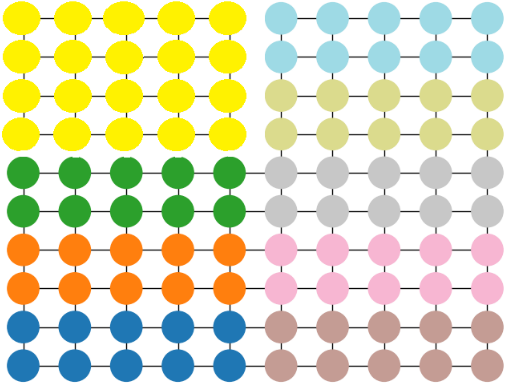
Recombination Steps

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

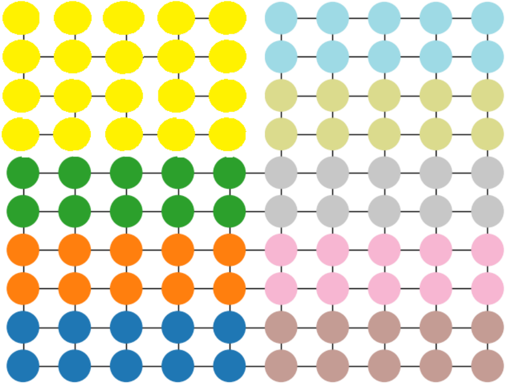
Recombination Step Example



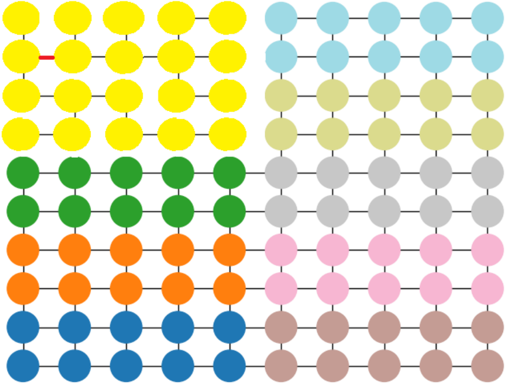
Recombination Step Example



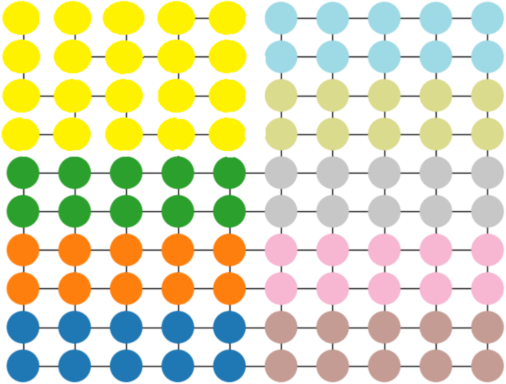
Recombination Step Example



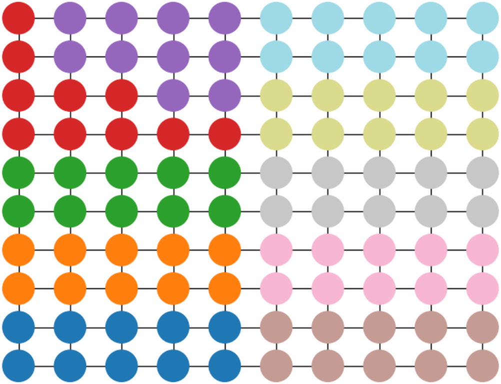
Recombination Step Example



Recombination Step Example



Recombination Step Example



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.