# 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 Part II: MCMC and GerryChain

Daryl DeFord

CSAIL – GDP Group

IAP 2019
Massachusetts Institute of Technology
January 10, 2019

# MORAL:

Computational Redistricting is NOT a solved problem!

## 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. IAP Info:
   - Resources: people.csail.mit.edu/ddeford/CAPR
   - Today 12-1 MCMC and GerryChain
   - Monday 1/14 GerryChain drop in 12-3
   - 1/22 12-1 Graph Partitions
   - 1/29 12-1 In–depth state examples

# Outline

# Geometric Probability

**Question**

*What is the expected distance between two random points on* $[0, 1]$*?*

# Geometric Probability

## Question

*What is the expected distance between two random points on* $[0,1]$*?*

## Answer

$$\int_0^1 \int_0^1 |x-y| dx dy = \frac{1}{3}$$

# Geometric Probability

## Question

*What is the expected distance between two random points on $[0,1]$?*

## Answer

$$\int_0^1 \int_0^1 |x-y| dx dy = \frac{1}{3}$$

## Question

*What is the expected distance between two random points on $[0,1]^n$?*

## Answer

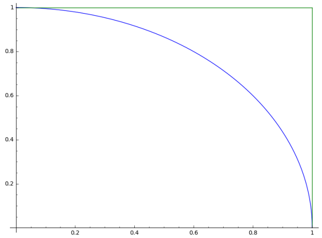
$$\int_0^1 \cdots \int_0^1 \sqrt{\sum_{j=1}^{n} (x_j - y_j)^2 dx_1 \cdots dx_n dy_1 \cdots dy_n} = \quad :($$

# Numerical Integration

## Question

*What is the area "under" the curve?*

$$\int_0^1 \int_0^{\sqrt{1-x^2}} 1 dx dy$$



# Properties of Monte Carlo Methods

- Draw (independent) samples from a random distribution
- Compute some measure for each draw
- Repeat lots and lots of times
- Average/aggregate the derived data

# Properties of Monte Carlo Methods

- Draw (independent) samples from a random distribution
- Compute some measure for each draw
- Repeat lots and lots of times
- Average/aggregate the derived data

Today, we tend to take access to random numbers for granted but some of the first applications of Monte Carlo were physical systems for generating random numbers (and the modern version was invented to analyze solitaire).

# What is a Markov chain?

**Definition (Markov Chain)**

A sequence of random variables $X_1, X_2, \ldots,$ is called a Markov Chain if

$$\mathbb{P}(X_n = x_n : X_1 = x_1, \ldots, X_{n-1} = x_{n-1}) = \mathbb{P}(X_n = x_n : X_{n-1} = x_{n-1}).$$

# Markov Formalism

Given a finite state space $X = x_1, x_2, \ldots, x_n$ we can specify a Markov chain over $X$ with transition probabilities $p_{i,j} = \mathbb{P}(X_m = i : X_{m-1} = j)$ and associated transition matrix $P = [p_{i,j}]$.

# Markov Formalism

Given a finite state space $X = x_1, x_2, \ldots, x_n$ we can specify a Markov chain over $X$ with transition probabilities $p_{i,j} = \mathbb{P}(X_m = i : X_{m-1} = j)$ and associated transition matrix $P = [p_{i,j}]$.

**Definition (Random Walk)**

We can also view a Markov chain as random walk on a directed, weighted graph, with weights given by the $p_{i,j}$.

# Desirable Adjectives

- Irreducible: A chain is irreducible if each state is (eventually) reachable from every other state.
- Aperiodic: A chain is aperiodic if for each state, the GCD of the lengths of the loops, starting and ending at that state is equal to 1.
- Steady State Distribution: A distribution $\pi$ is said to be a steady state of the chain if $\pi = \pi P$.
- Reversible: A chain with steady state $\pi$ is called reversible if it satisfies the detailed balance condition:

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

# Key Theorem

If the chain is irreducible and periodic then $\lim_{m\to\infty} P^m = 1\pi$ for a unique $\pi$. Even better, if $y_1, y_2, \ldots, y_m$ are samples from $\pi$ then,

$$\lim_{m\to\infty} \frac{1}{m} \sum_{i=1}^{m} f(y_i) = \mathbb{E}[f]$$

The key idea of MCMC is to create an irreducible, aperiodic Markov chain whose steady state distribution $\pi$ is the distribution we are trying to sample from.

# Aladdin k–grams

1. ' k,"be prw henghine "wd." alwe amad En mofed in tutheofurg Chithigerex d "I kisojep, blyimator, thr '
2. ' "I warder vizie the broom the pultail his pier, boom. Thearts and of firstrand dozen ise jewels andly chough who pull they two," Thind here mandsorrieve if goned to she som his enormed the gave firs. '
3. The Sultansported threat all him all he return to see Fatima, sent a certain the lamp. "Build in first, terried from his son if her forted the
4. He them to that this lamp I left her at she was not having gold and seeing each of the the city, showed to me, and her mothere which sparkled and garden, which would only two little cotton, forgot to the vizier and tell her next day the palace wherefore

# What is MCMC?

In our Monte Carlo methods we just required that we sample from our space uniformly but this isn't always easy to do. MCMC gives us a way to sample from a desired pre–defined distribution by forming a related Markov chain (or walk) over our state space, with transition probabilities determined by a multiple of the distribution that we are trying to sample from.

# Proportional to a distribution!?!

A common way this arises is when we have a score function or a ranking on our state space and want to draw proportionally to these scores. Given a score $s : X \to \mathbb{R}$ we want to sample from $X$ with probabilities

$$\mathbb{P}(X_i) = \frac{s(X_i)}{\sum_j S_j}$$

When $|X|$ is enormous, we don't want to/can't compute the denominator directly. Also, uniform sampling over–prioritizes low score spaces. This is also an advantage to local methods.

# Terminology

**State Space:** The underlying set of elements we are allowed to sample.

# Terminology

**State Space:** The underlying set of elements we are allowed to sample.

**Score:** A function $s : X \to \mathbb{R}_{\geq 0}$ that determines our target distribution.

# Terminology

**State Space:** The underlying set of elements we are allowed to sample.

**Score:** A function $s : X \to \mathbb{R}_{\geq 0}$ that determines our target distribution.

**Proposal Distribution:** A Markov chain $Q$ over $X$.

# Terminology

**State Space:** The underlying set of elements we are allowed to sample.

**Score:** A function $s : X \to \mathbb{R}_{\geq 0}$ that determines our target distribution.

**Proposal Distribution:** A Markov chain $Q$ over $X$.

**Metric:** Another function $f : X \to \mathbb{R}$ that is our quantity of interest for the distribution.

# Metropolis Procedure

Given that we have a given score function, proposal distribution, metric, and initial graph $g_0$ we generate new graphs $g_n$ by:

1. Generating $\hat{g}$ according to the proposal distribution $Q(\hat{g} : g_i)$.
2. Compute the acceptance probability: $\alpha = \min\left(1, \frac{s(\hat{g})}{s(g_i)} \frac{Q(g_i|\hat{g})}{Q(\hat{g}|g_i)}\right)$
3. Pick a number $\gamma$ uniformly on $[0, 1]$
4. Set

$$g_{i+1} = \begin{cases} \hat{g} & \text{if } \gamma < \alpha \\ g_i & \text{otherwise/} \end{cases}$$

# Example: Scrabble scores from uniform

- State space: 5 letter words
- Proposal: Change a uniform letter uniformly
- Score: Scrabble Score
- Example:
  1. Start at 'aaaaa'
  2. Propose 'aaaza'
  3. Compute:
  $$\alpha = \min(1, \frac{14}{5})$$
  4. Accept!
  5. Propose 'aaaya'
  6. Compute:
  $$\alpha = \min(1, \frac{8}{14})$$
  7. Reject!
  8. etc.

# Scrabble Scores from Text

- State space: Letters
- Proposal: Aladdin text
- Score: Scrabble Score
    1. Start at $m$
    2. Propose $a$
    3. $a \to m$ is $3.96\%$
    4. $m \to a$ is $18.41\%$
    5. Compute:

$$\left(\frac{1}{2}\right)\left(\frac{.0396}{.1841}\right) \approx .1$$

- Accept?

# 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

# Election Counterfactuals

We are actually evaluating the question: What might have happened if the districts looked like this?

- Data
- No time travel
- Hypothesis testing
- Landscapes

# MCMC on partitions

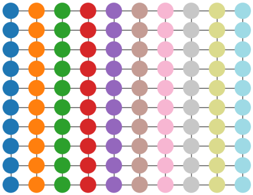
1. Set constraints to define the state space
2. Start with an initial plan
3. **Propose a modification**
4. Verify that the modification satisfies the constraints
5. Accept using MH criterion
6. Repeat

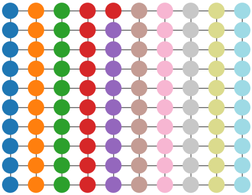
## Why?

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

# Single Edge Flip Proposals

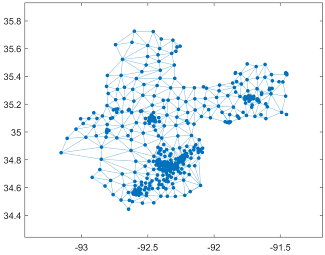
1. Uniformly choose an edge between districts
2. Change one of the incident node assignments to match the other



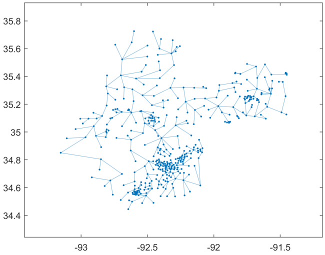


- Mattingly et al. (2017, 2018) Court cases in NC and WI.
- Pegden et al. Assessing significance in a Markov chain without mixing, PNAS, (2017). Court case in PA.

# Tree based methods

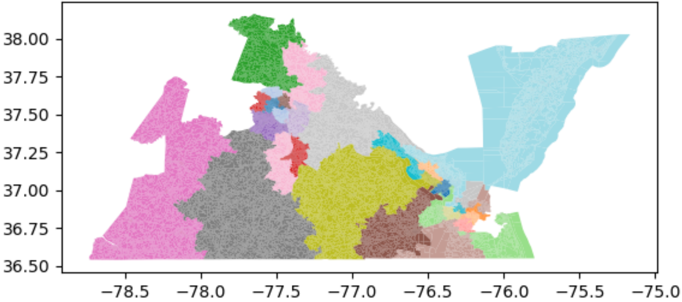



(a) District




(b) Spanning Tree

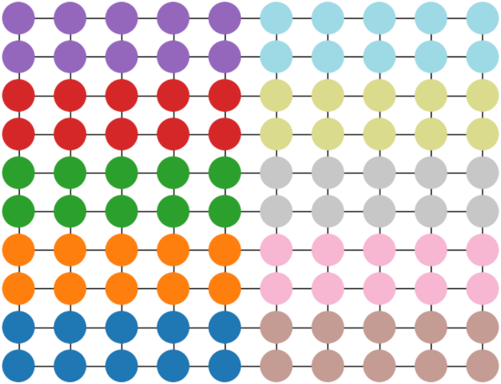
# Tree Seeds Ensemble



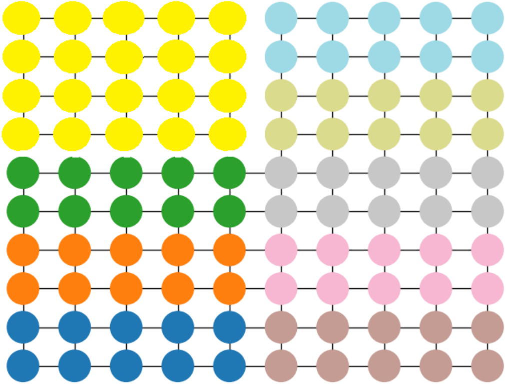
# Recombination Steps

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

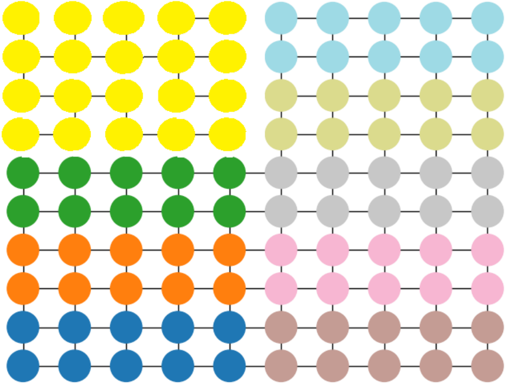
# Recombination Step Example



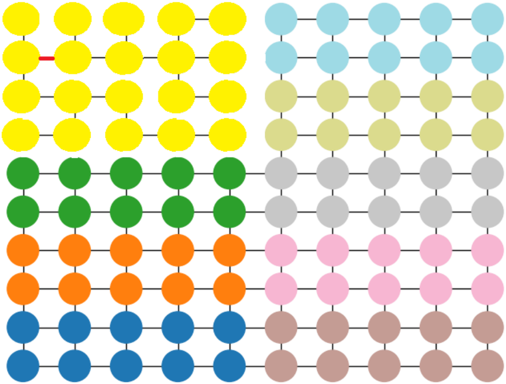
# Recombination Step Example



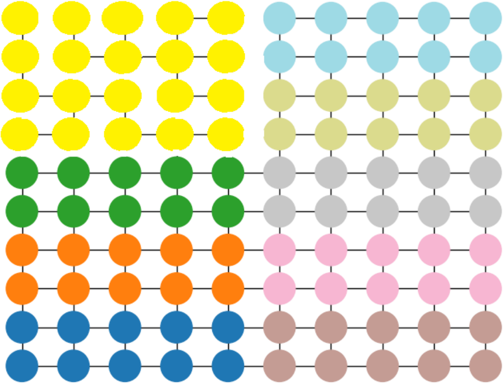
# Recombination Step Example



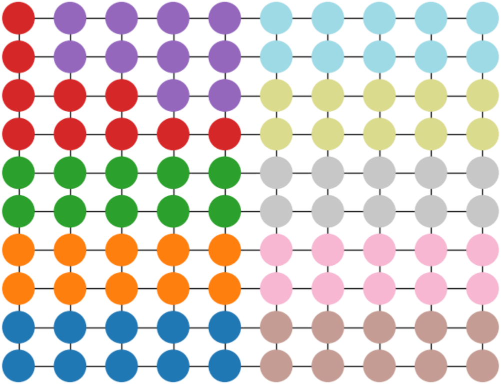
# Recombination Step Example



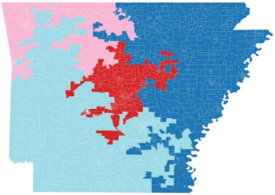
# Recombination Step Example

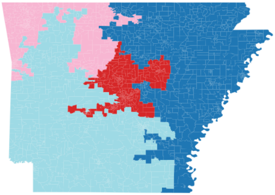


# Recombination Step Example

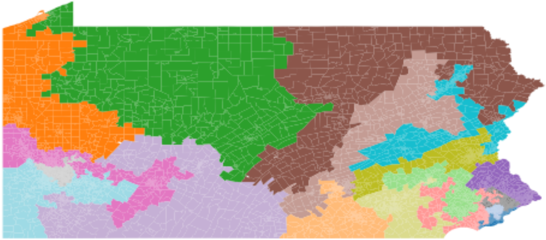


# Single Edge Ensembles





# Pennsylvania Single Edge Flip



# Pennsylvania Recombination Steps

# Energy Functions

- Weight plans proportional to $e^{-\beta \sum w_i \mathrm{scores}(D_i)}$
- Varying $\beta$ controls the strictness of the constraints
- Varying $w_i$ changes the relative strengths of the scores
- Exploit vs. Explore

## Examples

*If our two considerations are population balance and compactness we might use something like:*

$$e^{-\beta(17 \cdot \textit{Population Deviation} + 132 \cdot \textit{Total Perimeter})}$$

# Winnowing (Individual Districts)

- Strict 1% population bound
- Strict compactness bound
- Strict VRA bounds
- No triply split counties

# Creativity!

- Scores
  - Compactness
  - Aggregating measures
  - Hard vs. Soft constraints
  - ...
- Proposals
  - Boundary Flip
  - Tree Methods
  - Your favorite graph method here
  - ...
- Metrics
  - Sorted vote percentage vector
  - Partisan Metrics
  - Competitiveness?
  - ...

# MORAL:

Computational Redistricting is NOT a solved problem!

# GerryChain Components

- Inputs:
  - Dual Graph
  - Population Data
  - Vote Data
- Initialize
  - Updaters
  - Scores
  - Initial Plan
- Proposals
  - Input: Current state
  - Output: New state
- Validators
  - Input: Proposed state
  - Output: Binary – Pass/Fail
- MH Acceptance function
  - Input: Proposed state
  - Outpit: Binary – Accept/Reeject
- Output
  - Partitions

# The End

Thanks!