Graphs, Geometry, and Gerrymanders
a guide for the modern mathematician
Zachary Schutzman

University of Pennsylvania & Metric Geometry and Gerrymandering Group

Diet Graduate Seminar, U. of Toronto Mathematics

February 21st, 2019
This is based on work with...

gerrychain + trees
https://github.com/mggg/gerrychain

Daryl DeFord
Moon Duchin
Max Hully
Lorenzo Najt
The Gerrychain Team

Metagraph
https://mggg.org/metagraph

Moon Duchin
Seth Drew
Eugene Henninger-Voss
Amara Jager
Heather Newman
WHAT IS REDISTRICTING?

WHY DOES IT MATTER?
Redistricting

...is done regularly.

...occurs at a variety of levels.

...is baked into our system of government.
What is the ‘correct’ way to draw districts? Is there even a good notion of ‘correct’? Can we identify when a districting plan is ‘incorrect’?

Politicians draw political districts. Incentives do not always align.
The Definition

Gerrymandering is the intentional drawing of electoral districts to favor or disfavor some political outcome.
A brief history of Canadian redistricting

Pre-1950s: big problem
1957: Manitoba tries something new
Now: Canada uses independent commissions for federal districts

Ridings within a province can differ in population by up to 25% legally, 15% in practice.
Recent Issues


Right Now: Reducing the number of Toronto City Council wards
Where do people live?

In general, Canada’s issue is *malapportionment*, not *gerrymandering*.

Nationally, the smallest riding has about $\frac{1}{4}$ the population of the largest.
Gerry's Salamander
Gerry's Salamander
Tradition of Abuse

Draw incumbents in or out
Preserve an Incumbent
Tradition of Abuse

Draw incumbents in or out
Ohio 1880s
Ohio: Six Times in Twelve Years

12 Ninth Apportionment — 1878 to 1880
13 Tenth Apportionment — 1880 to 1882
14 Eleventh Apportionment — 1882 to 1884
15 Twelfth Apportionment — 1884 to 1886
16 Thirteenth Apportionment — 1886 to 1890
17 Fourteenth Apportionment — 1890 to 1892
Tradition of Abuse

Draw incumbents in or out
Ohio 1880s
Tuskegee
Tuskegee, Alabama

(The entire area of the square comprised the City prior to Act 140. The irregular black-bordered figure within the square represents the post-enactment city.)
Geometry is no longer sufficient
Why?

Better data
Better software
Better ad targeting

But! Public tools have largely caught up with private ones.
The abundance of data means that it’s easier to draw a hard-to-detect gerrymander. One way to demonstrate that something is a gerrymander is to show that it is an extreme outlier with respect to some measure we care about. How do we go about demonstrating this?
GRAPHS AND METAGRAPHS
A Graph Theory Problem

Districts are formed from atomic geographic units (Census blocks)
Can we get anything by using the tools of graph theory?
Constructing the dual graph
A **districting plan** (on a dual graph) is a *partition* of the vertices into $k$ disjoint, connected pieces which satisfy the criteria we care about.

- equal population
- partisan, racial metrics
- not splitting towns
Hardness

Unfortunately, no matter how you slice it, redistricting is NP-Hard.

Population constraints, racial and partisan metrics $\rightarrow$ SUBSETSUM
Geometric constraints (minimize cut edges, e.g.) $\rightarrow$ MILP
Optimization problems (find the ‘fairest’ plan ...) $\rightarrow$ $k$-KNAPSACK

However, NP-Hard doesn’t mean impossible.
The Dream

We’re working in a very particular corner of the universe of the problem.

Maybe it’s easy for our setting?

Does our problem have enough structure that we can just write down all of the plans and look at every one?
A Combinatorial Warmup

Let’s take the fictional state of Gridlandia.

How many ways are there to divide Gridlandia into 3 connected pieces of size 3?
Gridlandia Enumeration
And larger?

4x4 into 4 pieces of size 4? → 117
5x5 into 5 pieces of size 5? → 4006
6x6 into 6 pieces of size 6? → 451206
7x7 into 7 pieces of size 7? → 158753814 (10^8)
8x8 into 8 pieces of size 8? → 187497290034 (10^{11})
9x9 into 9 pieces of size 9? → 706152947468301 (10^{14})
10x10 into 10 pieces of size 10? → Open Problem!

Since we can’t write down all the plans, we need some way of sampling them.
Let’s imagine (because we definitely can’t write it down) a graph $\mathcal{M}$

There is a vertex for each (valid) districting plan.

The edges are interesting. How do we define ‘adjacent’?
Adjacency

We can do whatever we want! Let \( f : \mathbb{D} \times [0, 1] \rightarrow \mathbb{D} \) be any function satisfying the following:

\[
\text{Im}(f) = \mathbb{D}
\]

\( f \) is reversible (i.e. if \( f(D_1, \alpha) = D_2 \), then there exists \( \beta \) such that \( f(D_2, \beta) = D_1 \))

\( f \) is efficiently computable

For example, \( f \) can be the function "choose two adjacent districts and a cell in each and try to swap them"
A quick illustration

Lots more stuff at http://mggg.org/metagraph
For our function $f$, put an edge $(D_i, D_j)$ in the metagraph $\mathcal{M}$ if $f(D_i, \alpha) = D_j$ for some $\alpha$.

$\text{Im}(f) = \mathbb{D} \to \mathcal{M}$ is connected
  $\to$ each edge is equipped with a transition probability

$f$ is reversible $\to \mathcal{M}$ is bidirected

$f$ is efficiently computable $\to$ we can simulate a random walk
A quick illustration

Lots more stuff at http://mggg.org/metagraph
We can define a random walk on the metagraph by starting at some vertex $D_1$, picking a random $\alpha$, moving to $f(D_1, \alpha)$, and repeating. This is a Markov chain.
So what?

If we run this random walk long enough, the distribution of this sample converges to the **stationary distribution**!

We can pick the stationary distribution by carefully picking the transition probabilities!
How do we use this?

Graphic from DeFord, Duchin & Solomon *Report for the VA House of Delegates*
Proposals

Determining which plans are “adjacent” affects the sampling procedure. The flip walk moves very slowly through the space. The flip walk also moves towards plans which are geometrically nonsensical.

Is there a way to move around the space more quickly which avoids both of these?
Consider the following proposal function:

- pick two adjacent districts in $D_1$
- merge them into one "superdistrict"
- find a random spanning tree of the superdistrict
- cut the tree in half to make two new districts
- let $D_2$ be the new plan
A quick illustration

Graphic from DeFord, Duchin & Solomon *Report for the VA House of Delegates*
Recombination

Some properties:
  favors generating districts with lots of spanning trees
  population constraints make the choice of cut (roughly) unique
  metagraph nodes have high degree

Experiments with ReCom show that it does, in practice, improve the quality of the sample that the MCMC process generates!
There will be nation-wide redistricting at all levels in 2021
Awareness of the problem is higher than ever
How do we get it right?
THANK YOU!