Mathematics of Gerrymandering

Sarah Ahmad Leila Dahlia Erik Raaum Evan Rogers Emma Todd

University of Illinois at Chicago

MSCS Undergraduate Research Laboratory

UIC

Introduction

Gerrymandering is a political phenomenon in which electoral district boundaries are drawn to favor a certain political party or demographic group. Its applications dramatically affect the results of elections and remain contested in various courts across the United States.
Due to the impacts of gerrymandering on electoral outcomes and fair representation, governments have turned to various non-partisan or cross-party organizations for the creation of district lines.
In addition, researchers from various disciplines have developed statistical methods to evaluate electoral districts and create a more formal process to examine gerrymandering.
To approach these questions mathematically the problem is translated into graph theory. The graph takes the smallest blocks of voters, typically census tracts, as the vertices and uses edges to encode whether two tracts touch. A map dividing the populating into k districts is a decomposition of this graph into k equal size connected pieces.
The number of possible maps is too large to possibly explore them all. So, to get a sense of what a "typical" map looks like, one produces a large number of maps at random and look at the statistics of whatever quantities one is studying. If a map in real life is an extreme outlier compared to the random maps then that is evidence it was produced by some non-random process.
The random maps are generated by a random walk[1] among possible maps. Each step of the walk takes two current districts that touch and combines them into one larger district and then splits this at random into two new districts. Our project explores how this key step behaves in some simple examples of $n \times n$ grids and investigates the shapes of the pieces when the algorithm is applied to divide the grid into two connected halves.

Definitions and Key Ingredients

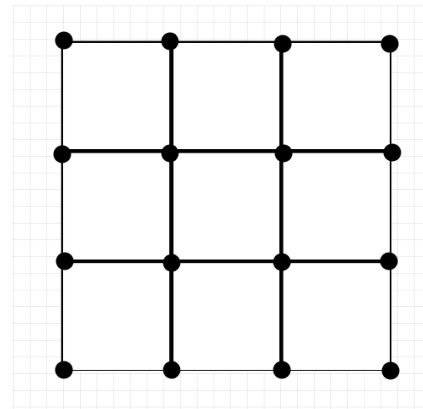
Definition 1: Spanning Tree A spanning tree is a connected subset of a graph G which includes all vertices of G with the fewest number of edges.

Definition 2: Half-edge An edge of a spanning tree is a half-edge if there are an equal number of vertices on either side of the edge.

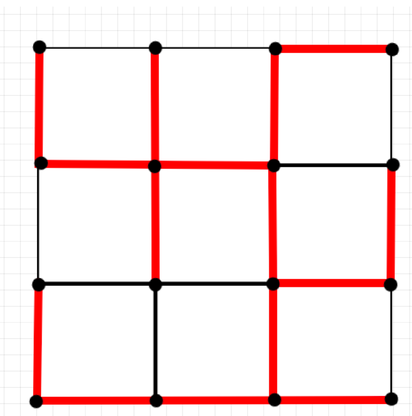
Wilson's Algorithm Wilson's Algorithm[2] is a method to produce a random spanning tree for a graph. It proceeds recursively, where at each stage a random vertex not currently in the tree is selected and connected to the tree by a path built via a random walk from which all loops are erased. The key property is that this algorithm produces all spanning trees with equal probability.

Gerrymandering and Graph Theory

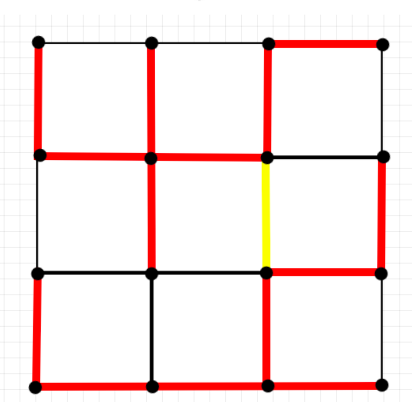
The Metric Geometry and Gerrymandering Group at Tufts University abstracted voting districts into graphs, where each vertex represents a single voting unit and each edge represents a boundary between them.



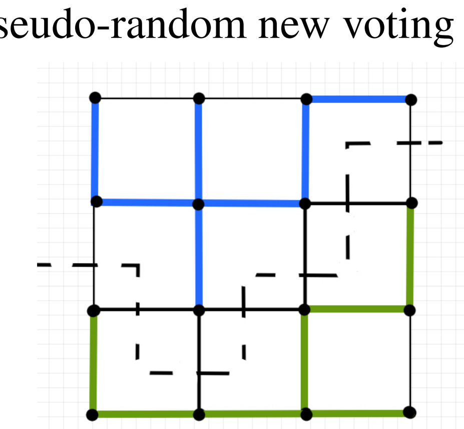
They divide this graph in half by first using Wilson's Algorithm to pick a uniform random spanning tree of this new graph.



They then check if this spanning tree has a half-edge, an edge that divides the graph into two graphs with $\frac{n}{2}$ vertices. If it does not, the algorithm is repeated again. If it does, they take each of the new graphs as separate voting districts.



By repeating this algorithm many times, the team creates a pseudo-random new voting block.



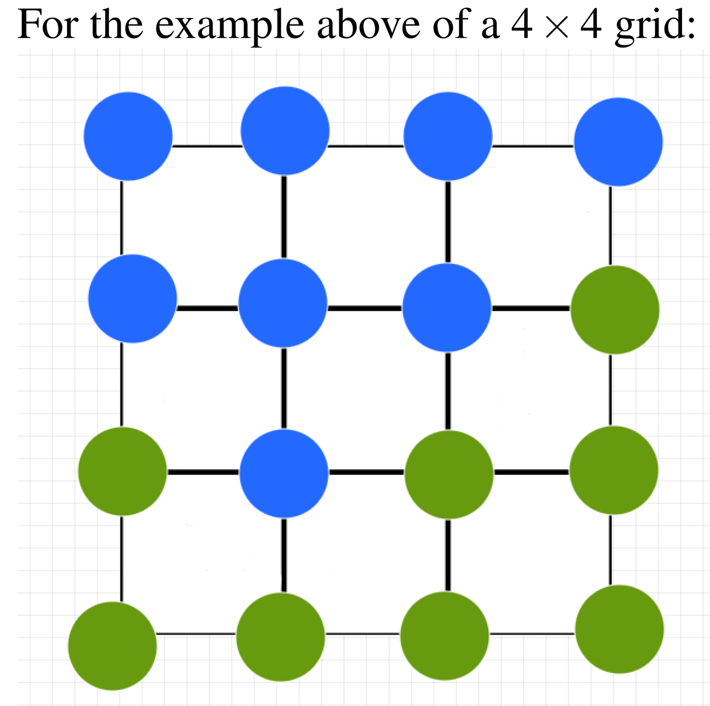
An important measure of compactness is the boundary measure. In our case, boundary is the count of edges between the two new graphs. In the 4×4 case above, we see by the dotted line that seven edges had to be cut to form the resulting graphs.

Likelihood of a Division

To determine how likely any particular division is, we need to know how many different spanning trees there are that produce the given splitting. If the two regions are graphs A and B we can calculate this as:

#{ Spanning trees in A} $\cdot$ #{ Spanning trees in B } $\cdot$ of edges connecting A to B

For the example above of a 4×4 grid:



the blue graph has 15 spanning trees, the green region has 4, and there are 7 edges connecting the two, so there are 420 different trees that would produce this. For contrast, if we divide the 4×4 grid into two 2×4 grids then each side has 56 spanning trees and their are 4 edges joining them. So this division can arise from $56 \cdot 56 \cdot 4 = 12544$. So this simple division is almost 30 times more likely to be produced. In general it seems that more compact shapes are more likely to occur.

Counting Spanning Trees

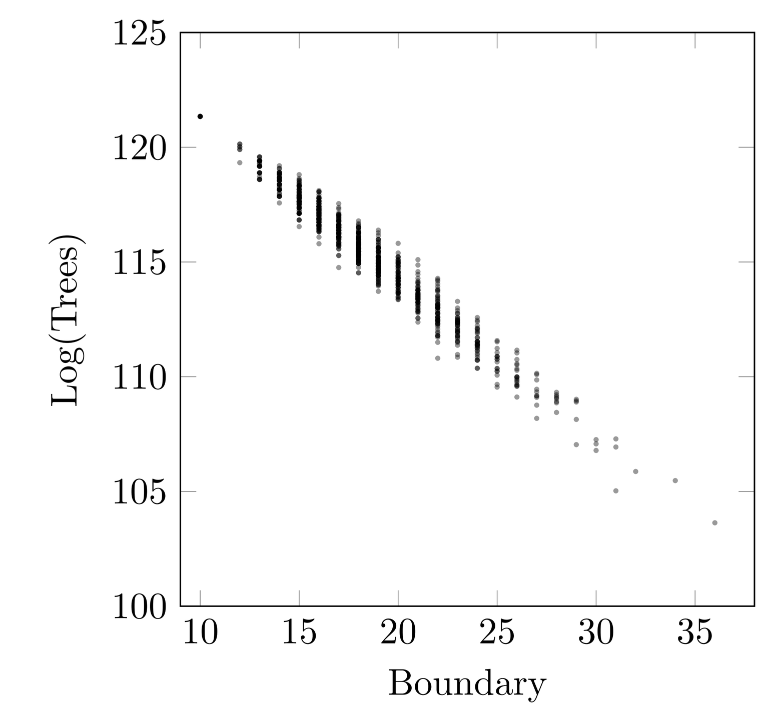
To compute the number of spanning trees in a connected graph we use the Laplacian Matrix of the graph. For a graph with n vertices this is the $n \times n$ matrix L where the diagonal entry l_{ii} is defined as the degree of the i^{th} vertex, and for $i \neq j$ $l_{ij} = -1$ if vertex i and vertex j are adjacent and 0 otherwise. Kirchoff's Matrix Theorem[3] tells us that this matrix has $n-1$ non-zero eigenvalues, $\lambda_1, \lambda_2, \cdots, \lambda_{n-1}$ and the number of spanning trees of the graph is:

$$\frac{1}{n}\prod_{i=1}^{n-1}\lambda_i$$

This can be computed as the determinant of a reduced Laplacian matrix. As these numbers can get quite large our code actually computes the log of this determinant instead.

The 10×10 Grid

We ran this algorithm to produce 1000 divisions of a 10×10 grid into two districts, and calculated the the boundary and the number of spanning trees that would produce it. The scatter plot below shows the results, which suggest that the log of the number of trees is approximately linear in the boundary. This would mean the likelihood of any given division would decrease exponentially with the boundary size.



Further Questions

The boundary is one measure of the shape of a division, and we see that it correlates well with likelihood. However it does not completely predict it, so other aspects of the shapes must also play a role. One possibility is to explore the frequency with which the edges of the graph occur in the boundaries of splitting and then weight the measure of the boundary to take into account that some edges are less likely than other. There are known formulas for the frequency with which edges occur in spanning trees, but for this purpose one would need to be able to find the frequency for only those trees with half-edges. We were able to do this for small grids but did not in general.
Another interesting question is how likely a random spanning tree is to contain a half-edge. In our calculations this probability appeared to decrease roughly like $\frac{1}{n}$ for the $n \times n$ grid, it would be very interesting to establish rigorous estimates.

References

[1] Daryl DeFord, Moon Duchin, and Justin Solomon.
Recombination: A family of markov chains for redistricting.
10 2019.

[2] James Gary Propp and David Bruce Wilson.
Exact sampling with coupled markov chains and applications to statistical mechanics, 1996.

[3] Russell Lyons and Y. Peres.
Probability on trees and Networks.
Cambridge University Press, 2016.

[4] Study of reform proposals for chicago city council.

[5] Tara Abrishami, Nestor Guillen, Parker Rule, Zachary Schutzman, Justin Solomon, Thomas Weighill, and Si Wu.
Geometry of graph partitions via optimal transport.
10 2019.

[6] Moon Duchin and Bridget Eileen Tenner.
Discrete geometry for electoral geography.
8 2018.

[7] Richard Barnes and Justin Solomon.
Gerrymandering and compactness: Implementation flexibility and abuse.
Political Analysis, 29:448–466, 10 2021.

[8] Elle Najt, Daryl DeFord, and Justin Solomon.
Complexity and geometry of sampling connected graph partitions.
8 2019.

[9] Gregory F Lawler.
A self-avoiding random walk, 1980.

[10] Moon Duchin, Tom Needham, and Thomas Weighill.
The (homological) persistence of gerrymandering.
7 2020.

[11] Daryl DeFord, Natasha Dhamankar, Moon Duchin, Varun Gupta, Mackenzie McPike, Gabe Schoenbach, and Ki Wan Sim.
Implementing partisan symmetry: Problems and paradoxes.
8 2020.