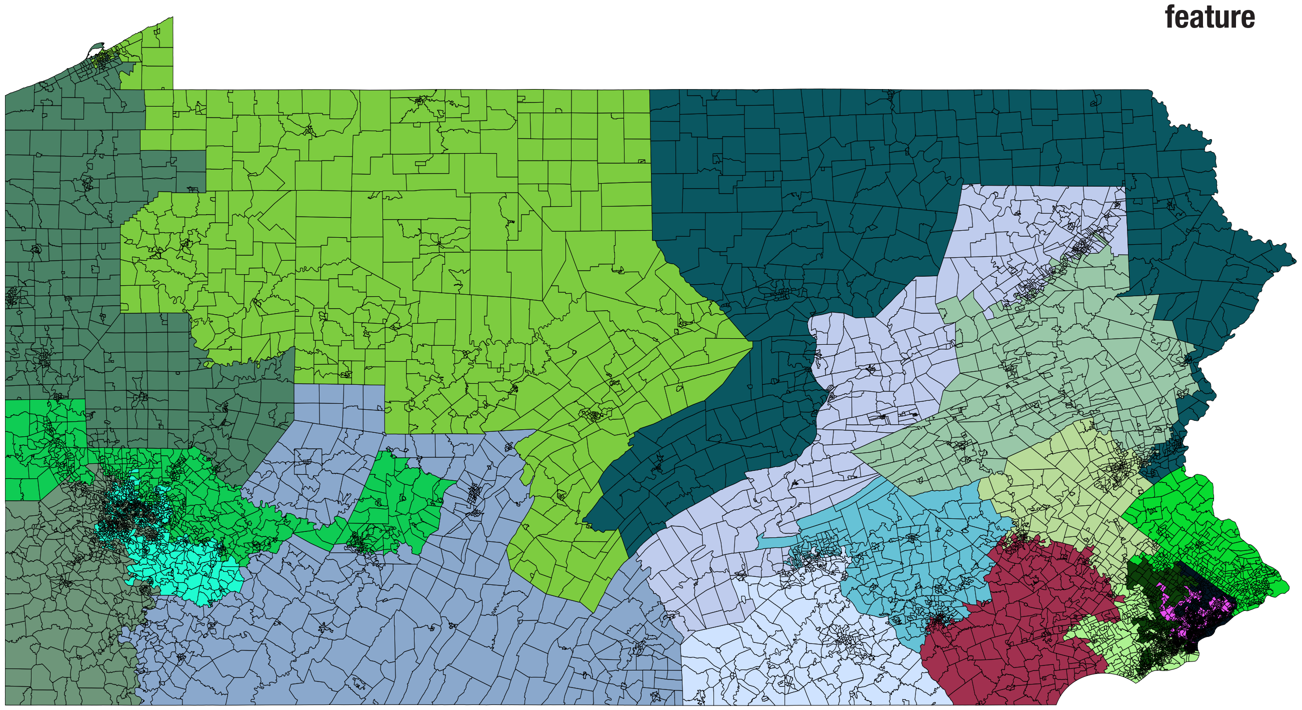
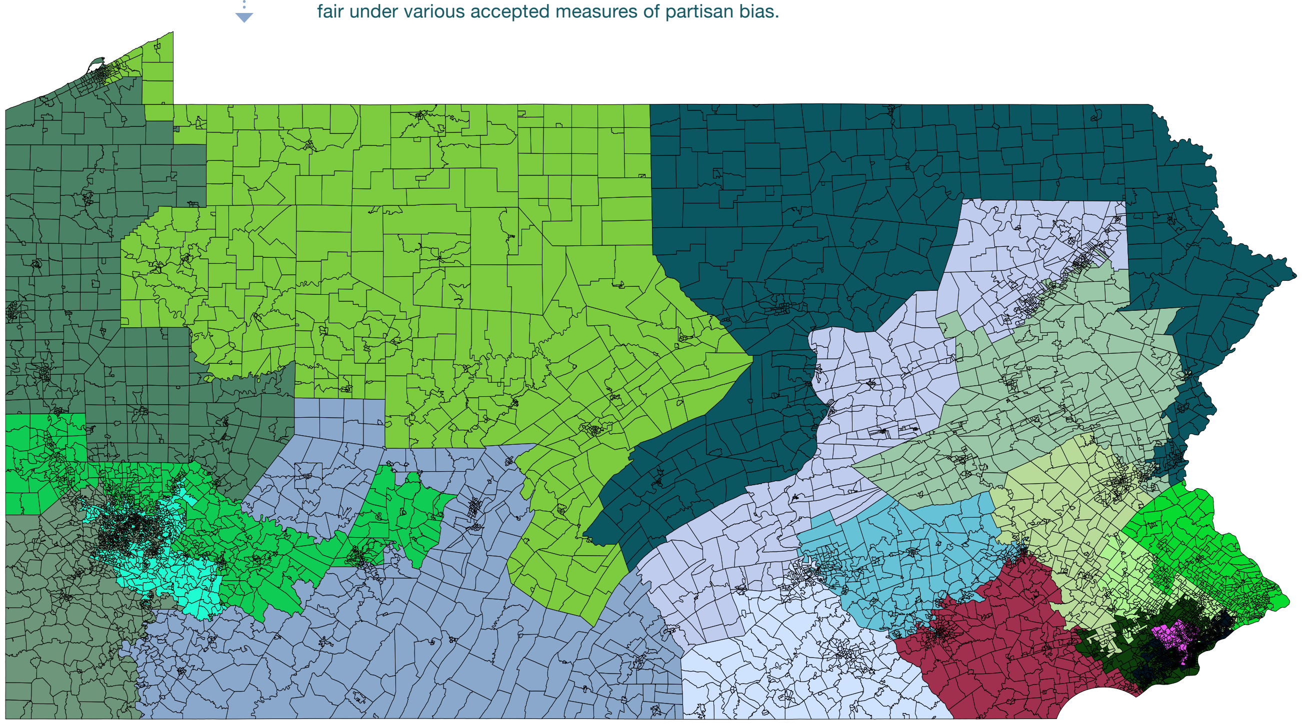
Can Math Solve the Gerrymandering Problem?

CMU mathematicians' work on detecting bias in congressional district maps makes its debut before the U.S. Supreme Court



2^{40}

Our method demonstrates the current districting of Pennsylvania is an outlier, in the sense that it is more biased than the overwhelming majority of geometrically similar districtings. The districting above is the current Congressional districting of Pennsylvania, and the districting below is an example of an alternative districting of Pennsylvania preserving the same counties as the present districting, and with similar overall district geometry (as measured by total district perimeter). This example was produced by roughly 750 billion steps of our Markov Chain. Furthermore, the districting below is more fair under various accepted measures of partisan bias.



In early October, the U.S. Supreme Court heard oral arguments in Gill v. Whitford, a case about partisan gerrymandering in Wisconsin. The highly anticipated case could transform the way congressional and legislative district lines are drawn. And mathematicians — including Carnegie Mellon's Alan Frieze and Wesley Pegden — are weighing in. Frieze and Pegden's work is cited in an amicus brief filed on behalf of the plaintiffs that addresses a question at the heart of the case: Is there a way to determine, in an unbiased, practical way, that extreme partisan gerrymandering has actually occurred?

Politicians have been gerrymandering — drawing congressional districts to favor one party or candidate over another — since the early 1800s, and the Supreme Court has considered its constitutionality many times. In 2004, the high court upheld a lower court's ruling that the partisan gerrymander in question was not unconstitutional, in part because no standards existed for adjudicating partisan gerrymandering claims. But Justice Anthony Kennedy posited that such a standard might one day exist. Frieze and Pegden's work could provide that standard.

In a paper published in the *Proceedings of the National Academy of Sciences*, Frieze, Pegden and the University of Pittsburgh's Maria Chikina used a Markov chain (see page 18) to rigorously demonstrate bias in the congressional districting maps of the state of Pennsylvania.

"The idea is, if you want to evaluate whether a particular division of a state is fair, a natural thing to try to do is to compare it to a typical districting — in other words, a random one," Pegden said. "Our method takes the actual layout of where people live, their political affiliation, and tests in a rigorous way whether the current districting is much worse than other typical districtings of the same state."

They began with a current map of Pennsylvania's congressional districts, and then ran the chain, which changed the map in random steps — wiggling little municipalities here and moving little groups of people across boundaries there. The shapes of the districts slowly changed while keeping a roughly equal population in each. The mathematicians observed that their simulation's randomly "redrawn" districts are much fairer that the initial district map they started with.

"Out of all the trillion districtings that we saw in this sequence of random steps, none of them were as bad as the very first one. What that shows is that this initial districting is much worse than other random districtings," Pegden said, adding, "There is no way that this map could have been produced by an unbiased process."

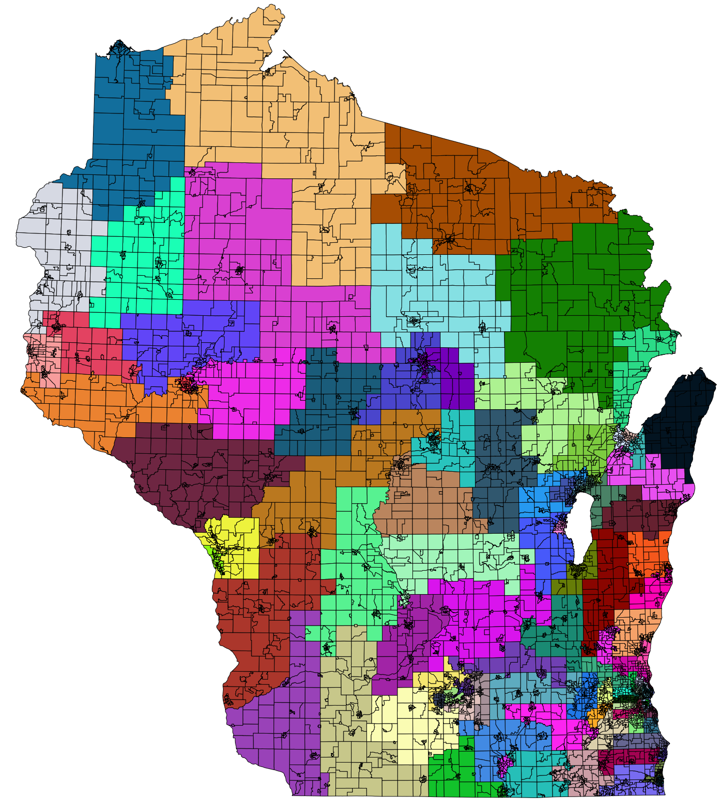
In June, Frieze and Pegden's method became part of a lawsuit filed by the League of Women Voters of Pennsylvania. The lawsuit is asking that the state's congressional district map be thrown out.

Shortly after that, Pegden heard from Nicholas Stephanopoulos, one of the attorneys for the plaintiff in the Wisconsin case. Stephanopoulos was interested in seeing what their analysis could say about Wisconsin. Pegden got to work. "To be doing some math on a deadline because the Supreme Court is going to hear a case? You don't usually have that," he joked.

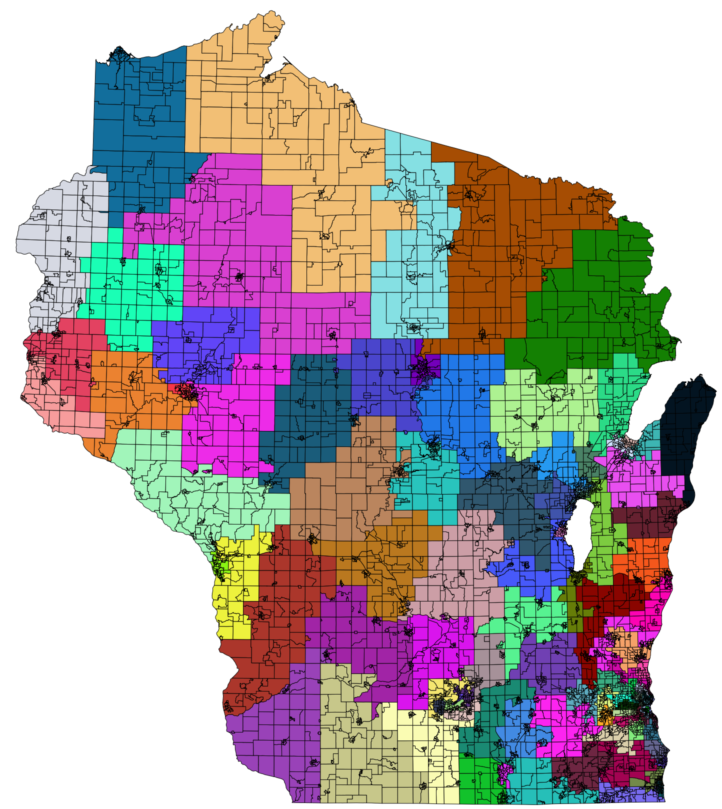
Pegden downloaded the Wisconsin voter data and ran the simulation. It turned out that the Wisconsin result was, in some sense, even cleaner than Pennsylvania's. "In Wisconsin, they did an amazing job gerrymandering. It's a beauty of extreme gerrymandering. It's really unbelievable."

Will the Supreme Court justices agree? We'll find out next spring.

At right, above is the present Assembly districting of Wisconsin, and below that is an alternative map produced by roughly 1 trillion steps of our Markov chain. Note that superficially, the districtings are similar; our test works by showing that the current districting is politically unusual among the districtings with similar geometric properties.



"In Wisconsin ... It's a beauty of extreme gerrymandering."



Detecting Gerrymandering with Markov Chains

by Wes Pegden

In past and ongoing legal cases concerning the constitutionality of gerrymandering political districtings, a key question for courts has been the extent to which it is possible to rigorously evaluate claims of gerrymandering in a practical way.

Take, for example, the 2012 elections for congressional seats in Pennsylvania. Democrats won more than 50 percent of the total vote for Congressional seats in Pennsylvania, but won only five out of the 18 seats at stake. To Democrats, this is clear grounds to call foul. But wait! Republicans can (and do) point out that the distribution of Democratic voters itself is likely to cost the Democrats Congressional seats, without any nefarious districting efforts by Republicans. Cities tend to be strong bastions of support for the Democratic party: the current Congressman for Pennsylvania's 2nd district (in Philadelphia) won his general election with more than 90 percent of the vote. On the other hand, no Congressional district in the country is so solidly Republican. Democratic voters really are more concentrated than Republican voters are, which really does create a greater potential for Democrats to waste more votes in districts where they have strong supermajorities, and thus win fewer seats in Congress than Republicans would at the same overall level of support.

The question, then, is how to tell when the districting of a state is responsible for bias toward one political party, rather than just the state's existing political geography. How can we tell when the bias of a particular political districting is really atypical among the space of valid districtings of the same state? To rephrase this mathematically, how can we tell that a districting is more biased than a random valid districting of the same state?

First, we need to have a notion of what constitutes a valid districting of the state. Certainly, districts should be contiguous, and nearly equal in population. There should also be geometric constraints on the districts (for example, ensuring that the square of the perimeter of a district is not much larger than its area) to forbid fractal-like districts. Other requirements for valid districtings may include not splitting many counties, and respecting other natural requirements considered by legislatures.

Let's imagine that we have a bag full of hypothetical "valid" districtings of a state satisfying all of our requirements. Now we just need to pull a lot of random districtings out of our bag and compare them to the actual districting of our state. We would conclude that our districting is gerrymandered if it is more biased toward one political party than the vast majority of the random samples. (For you statisticians: if it's more biased than 99.9999995% of random districtings, it would be gerrymandered at a p-value of .000000005.)

The problem is: How do we pull things out of this bag of districts? In other words, how do we generate random geometric partitions of a region (a state) satisfying various constraints? One natural way to try to draw samples from a weird distribution like this is to use Markov chains, which are used for sampling in areas as disparate as protein folding and statistical mechanics.

Essentially, a Markov chain is a process that takes a random walk along a set of possibilities by making a sequence of small changes. Formally speaking, it is a sequence of random variables

$$X_0, X_1, X_2, X_3, \ldots$$

taking values on some state-space Σ, such that, for each i, and for any states $\sigma_0, \ldots, \sigma_i \in \Sigma$, the conditional probability

$$\Pr(X_i = \sigma_i \mid X_0 = \sigma_0, X_1 = \sigma_1, \ldots, X_{i-1} = \sigma_{i-1})$$

depends only on the choice of i, σ_i, and σ_{i-1}. In other words, the process is a random walk in the sense that it has no memory; where it ends up at step i depends only on where it was at step $i-1$.

In the case of political districtings, a state is divided into thousands of little geographic regions (precincts) used by the census. We can make a small random change to a districting by selecting a precinct on the boundary of two districts, and switching which district it belongs to, so long as the result still satisfies our requirements on valid districtings. The magic of Markov chains is that, under relatively mild assumptions, there are nice theorems that tell you that no matter where you start a Markov chain, after you run it for long enough, it will essentially be at a random configuration, drawn from a *stationary distribution* of the Markov chain; this is a distribution π such that

$$X_0 \sim \pi \Longrightarrow (\forall i)\ X_i \sim \pi$$

In other words, the stationary distribution π has the property that drawing a random sample according to π and then running the Markov chain for any number of steps still results in a sample drawn according to π.

In the case of political districtings, this means that we just need to run the chain long enough, and it will eventually spit out random valid districtings of our state according to the stationary distribution of the chain, which is just what we need to make our statistical claims of gerrymandering! We should note that it is possible to carefully construct our chain so that π is uniform over all valid districtings of our state.

The problem is that, though these nice theorems do tell us that our Markov chain will generate random districtings eventually, they unfortunately tell us nothing about how long we have to run it before this is the case. Thus, we can't actually run our chain for a long time and then claim rigorously to have shown that a districting is gerrymandered. For all we know, we may just not have run the Markov chain long enough to see truly random districtings for comparison. This problem actually pervades many scientific applications of Markov chains. Frequently, they are used to generate random samples when there is not enough theory to guarantee that samples being generated are actually random; this can and does lead to disagreement among different research groups about when a result is valid and when it's not. And there have been situations (e.g., simulations of the Potts model from statistical physics) in which practitioners have developed modified Markov chains that they believed, based on evidence from simulations, achieved faster convergence, but later found that the chains still had exponential mixing times in many settings.

In our work with Maria Chikina and Alan Frieze, we proved a new general theorem about Markov chains that avoids this problem when the reason we are using a Markov chain to generate random samples is (as in our application) to demonstrate that a specific element of our space is an outlier. Specifically, we proved the following:

Theorem 1. Let $M = X_0, X_1, \ldots$ be a reversible Markov chain with a stationary distribution π, and suppose the states of M have real-valued labels. If $X_0 \sim \pi$, then for any fixed k, the probability that the label of X_0 is an *ε-outlier* from among the list of labels observed in the trajectory $X_0, X_1, X_2, \ldots, X_k$ is at most $\sqrt{2\varepsilon}$.

Here, being an *ε-outlier* means that the label of the state X_0 is smaller (or larger) than all but an ε fraction of the labels seen on the trajectory; in our application to redistricting, the label of a districting can be the number of seats Republicans would have won in a hypothetical election with the districting in question. The requirement that the Markov chain is reversible essentially means that the random changes the Markov chain makes when taking its random walk can be undone just as easily as they can be done; this is a common feature of Markov chains used for sampling.

Theorem 1 leads us to the following new statistical test for reversible Markov chains:

The $\sqrt{\varepsilon}$ test: Observe a trajectory $\sigma_0, \sigma_1, \sigma_2 \ldots, \sigma_k$ from the state σ_0, for any fixed k. The event that the label of σ_0 is an ε-outlier among the labels of $\sigma_0, \ldots, \sigma_k$ is significant at $p = \sqrt{2\varepsilon}$, under the null-hypothesis that $\sigma_0 \sim \pi$.

For our redistricting problem, this means that if we start our Markov chain from the current districting of our state, run it for as long as we want (since k is arbitrary here) and then observe that the current districting of our state is more biased than 99.9999995% of the districtings encountered by the Markov chain along the way, then this is significant at

$$p = \sqrt{2 \cdot .000000005} = .0001$$

which, while not as good as significance at $p = .000000005$, is still good enough for government work.