

The impact of highly compact algorithmic redistricting on the rural-versus-urban balance

Archer Wheeler
Brown University

Philip N. Klein
Brown University

ABSTRACT

It is commonly believed that, in congressional and state legislature elections in the United States, rural voters have an inherent political advantage over urban voters. We study this hypothesis using an idealized redistricting method, *balanced centroidal power diagrams*, that achieves essentially perfect population balance while optimizing a principled measure of compactness. We find that, using this method, the degree to which rural or urban voters have a political advantage depends on the number of districts and the population density of urban areas. Moreover, we find that the political advantage in any case tends to be dramatically less than that afforded by district plans used in the real world, including district plans drawn by presumably neutral parties such as the courts. One possible explanation is suggested by the following discovery: modifying centroidal power diagrams to prefer placing boundaries along city boundaries significantly increases the advantage rural voters have over urban voters.

CCS CONCEPTS

- **Theory of computation** → Design and analysis of algorithms;
- **Computing methodologies** Model development and analysis;
- **Applied computing** Law, social and behavioral sciences.

KEYWORDS

redistricting, rural-versus-urban, balanced centroidal power diagrams

ACM Reference Format:

Archer Wheeler and Philip N. Klein. 2020. The impact of highly compact algorithmic redistricting on the rural-versus-urban balance. In *28th International Conference on Advances in Geographic Information Systems (SIGSPATIAL '20)*, November 3–6, 2020, Seattle, WA, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3397536.3422249>

1 INTRODUCTION

Representatives to the U.S. House of Representatives and to many state legislative bodies are selected by winner-take-all elections across districts in states. A *district plan* for a state is a partition of the state's map into regions, called *districts*. A state's districts should be close to equal in population. Moreover, districts are expected to be *compact* and *contiguous* (notions that are not formally defined in the law). It is well known that district plans have been engineered

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SIGSPATIAL '20, November 3–6, 2020, Seattle, WA, USA

© 2020 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8019-5/20/11.

<https://doi.org/10.1145/3397536.3422249>

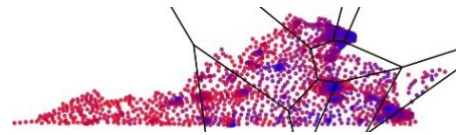


Figure 1: Algorithmic redistricting of Virginia using balanced power diagrams and populations from the 2010 census. Each dot represents the results of a precinct in the 2016 election with a color gradient corresponding to the outcome.

to provide advantage to individual candidates or to parties (this is called *gerrymandering*) [1, 8, 17]. Gerrymandered districts can lead to the advantaged person or party being less responsive to voter preferences.

Voters in rural areas and voters in urban areas tend to vote for opposing parties, in the US and elsewhere [15]. It is considered well-established that geography—what parts of the map are urban and what parts are rural, and how many people live in each—has a major impact on the relative electoral success of rural voters versus urban voters. Rodden [15] has written the definitive work on the phenomenon, addressing its historical origins and its implications for the present. While he clearly acknowledges the role of gerrymandering, he convincingly argues that the rural-voter advantage is inherent in the geography—the dense packing of left-leaning voters into urban areas, and the dispersion of right-leaning voters through the larger rural areas. Rodden suggests that “a party-blind process that produces geometrically compact districts” would simply benefit the rural party, but we find the phenomenon is more nuanced.

In this paper we explore the hypothesis that rural-voter advantage is inherent in the geography, rather than being a consequence of features of specific district plans. For this exploration, we use the sort of party-blind redistricting algorithm for optimizing compactness that Rodden cautioned against. Fryer and Holden [10] state three properties that they argue any measure of compactness should satisfy, and propose a measure, RPI, that uniquely satisfies these properties. In this paper, we use a method [4] that we believe tends to find district plans that are nearly optimal with respect to RPI. We analyzed these district plans as follows: we simulated elections in a subset of U.S. states for which there are high-quality 2016 precinct presidential election results, and calculated the likelihood of electoral outcomes.

We find that using these compact district plans leads to elections that are significantly more competitive and exhibit less partisan advantage than existing district plans. This is true even in the case of Virginia where the existing district plan was redrawn by a court. Moreover, the partisan advantage does not consistently belong to the rural party. One possible explanation for this result is that the algorithmically generated district plans pay no attention to

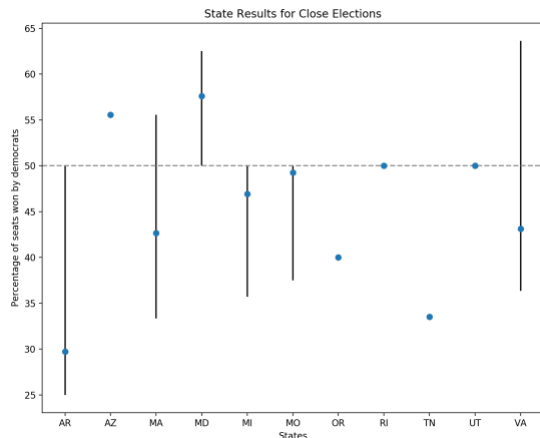


Figure 2: Percentage of seats won by Democrats for all available states given a hypothetical close election within 1% of the popular vote. Results are extrapolated from the 2016 election using a uniform swing to achieve a close popular vote. Error bars show the 95% confidence interval of possible outcomes.

municipal and county lines. We found that modifying the algorithm to prefer to locate district boundaries on or near such administrative lines significantly increases the electoral advantage of rural voters. In many states, it is expected that district plans take into account these administrative lines.

To understand how election results depend on the parameters of rural-versus-urban geography, we apply the same analysis to synthetic data. This enables us to understand the effects of three factors: number of districts, urban density, and party preference distribution. We find that the most important factors are the number of districts and the population density distribution (as opposed to the party preference distribution). When the number of districts is below a threshold (around five), the rural party has an advantage regardless of other parameters. Above that threshold, contrary to what one might expect, increasing urban population density *advantages* the urban party

2 METHODS

We use a measure of compactness, RPI, proposed by Fryer and Holden [10] based on the locations of residences within a state. Under this measure, an optimally compact district plan is one that minimizes the sum of mean squared distances between voters in each district.

There have been many proposed quantitative measures of compactness [6, 11]. We find that techniques which measure border length are highly sensitive to geographic features such as rivers or state boundaries. Since our goal is to measure rural-versus-urban advantage, we think our measure of compactness should be defined by resident locations rather than the shape of resulting districts. Fryer and Holden argue that RPI is equivalent to any measure based on resident locations which maintain three desired properties [10]. To solve this problem we build on earlier work by Cohen-Addad,

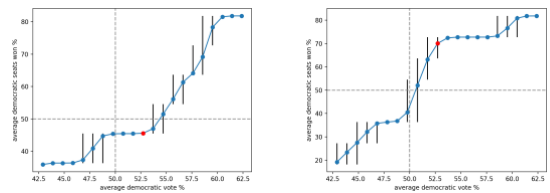


Figure 3: Simulated election results in Virginia for different district plans. Left: Existing district plan. Right: Algorithmically generated plan. Each graph shows likely outcomes given a statewide popular vote. Error bars show the 95% confidence interval of outcomes. The point marked in red would be the result during the 2016 election.

Klein and Young which presents a modified version of Lloyd’s algorithm using balanced power diagrams [5]. The resulting district plans are similar to weighted voronoi diagrams. See Figure 1.

2.1 District Plan Construction

We use 2010 census block data to build United States House of Representative districts using the power diagram capacitated k-means algorithm. Many previous works use significantly more coarse datasets for district building such as precincts [2, 7, 9]. We allow for some census blocks lying on the border of districts to split their population. However, Cohen-Addad, Klein & Young show that this can be fixed in practice with small perturbations in blocks along the border while still achieving perfect population balance [5].

While our balanced power diagram technique does use a randomized start, it consistently finds the same result. For instance, over a hundred runs on Virginia our algorithm found 99 identical results and one result with slightly different weights, but differed only in handling a tiny number of census blocks along the border. This property is consistent across different states. Because of this, we suspect that the algorithm is finding a nearly optimal solution.

2.2 Simulating Elections

Voting results, however, are not reported by census block. For historical voting outcomes we use precinct level results since they are the most geographically fine grained data available. We used data from openprecincts.org which is rigorously compiled and validated [14]. We take these historical precinct results and place all voters within the district where the centroid of the precinct lies. We convert historical results into a normal probability distribution and then average results over a thousand randomized elections. We additionally compute hypothetical elections using a uniform popular vote swing across all precincts.

3 STATE RESULTS

Figure 2 shows likely election outcomes in a close election for all states with sufficiently high quality data and at least two house districts. Since the United States uses a “first pass the post” system one should expect a nonlinear relationship between votes a party received and the number of seats won. However, there is little justification for winning a majority of districts given a minority of

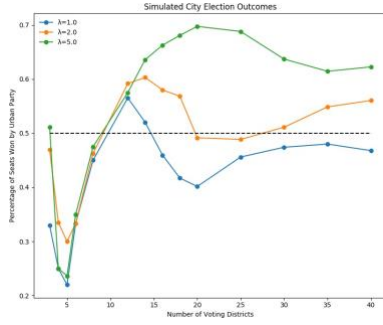


Figure 4: Average election outcomes with a 50/50 popular vote split given the number of districts drawn. The parameter λ controls the population distribution with a higher λ representing a denser population.

the popular vote. Because of this, we present in Figure 2 results for close elections (within 1% of the vote) rather than elections close to a state’s historical results. These results show political advantage of the district plans not likely outcomes. The importance of this kind of measure is argued and formalized by the legal scholars Grofman and King [12].

Figure 3 shows a more detailed look at results in Virginia. It is important to note that this current map was redrawn by a court in 2016 [13]. There are two important points to consider in these charts: the results given an equal 50/50 percent split between voters, and the percentage of overall votes needed before one party is expected to win a majority of districts. This breaks election outcomes into four quadrants. Results in the upper left or lower right quadrants indicate hypothetical elections where a party wins a majority of the seats with a minority of the vote. For more detailed results of all states shown in Figure 2 see redistrictingproject.com.

Our results show that there is perhaps a small advantage for the more rural Republican party when looking at the states in aggregate. However, most states are competitive for both parties during a close election. Additionally, these results do not show the consistent rural advantage that Rodden argues is inherent to geography [15].

4 WHEN CITIES LOSE

4.1 Simulations

In order to answer which factors of population density advantage urban or rural voters we devise a model of synthetic voter distributions. Since in the United States political affiliation is highly correlated to urban density [15], we focus our attention to outcomes around that of a hypothetical city. To do this, we fix a model and examine population density’s effect on the outcome of election results. We use a power law exponential distribution to model population density’s falloff from an urban center [3]. The density of this distribution is determined by a parameter λ with

$$p(x) = \lambda e^{-\lambda x}$$

This distribution is used to place voters’ location *independently of how they will vote*.

We also need a second distribution determining the probability that a voter will vote for each party as a function of their distance

from the city center. Once voter locations are fixed, we select urban party voters without replacement proportionally to $\frac{1}{1 + \alpha \|x\|}$ with

parameter α and distance from the city center $\|x\|$. We set voting behavior in this manner to ensure an exactly even split between the two parties. The last parameter k is the number of districts drawn.

4.2 Results

Given our model for cities, we graph the results in Figure 4. Values for $\alpha > 1$ had little impact on the results, therefore we present results in Figure 4 for a fixed α with $\alpha = 2$. It is important to stress that the outcomes in Figure 4 differ solely on the population density parameter λ which is independent of voter political preference. We maintain equal votes for each party while increasing population density since voter preference is selected after voter locations are determined.

Additionally, this general trend is repeatable for different types of models. We found similar results with added noise, radially asymmetric population density and voter preference modeled using a power law distribution. Interestingly, asymmetry appears to help the urban party, whereas the amount of noise introduced had varied results, but maintained the same general trend. The urban party consistently performed better by increasing the density parameter λ . Additionally, when k was around five the urban party consistently underperformed. We suspect that this results at $k = 5$ is a result of examining a single city and this threshold would be different for a larger region.

It is also possible that the sharp dip before $k = 5$ is an artifact of synthetic data. At low values of k the algorithm tends to place a single district entirely within the city. All other districts look like long wedges cutting into the city. As k increases, however, the packed urban centers are balanced by entirely rural districts outside of the city. Districts drawn on real data, either on a city or a state level, more closely resemble synthetic redistricting with larger k . This effect could be reduced by using more realistic asymmetric population distributions which more closely resembles the real world.

4.3 County Lines

Many states require or expect that district plans largely preserve county and municipal lines [11]. In order to measure what effect

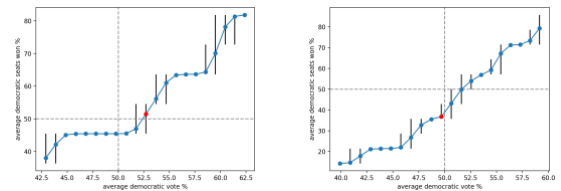


Figure 5: Algorithmic redistricting results for Virginia (left) and Michigan (right) when respecting administrative lines. Each graph shows likely outcomes given a statewide popular vote. Error bars show the 95% confidence interval of outcomes. The point marked in red would be the result during the 2016 election.

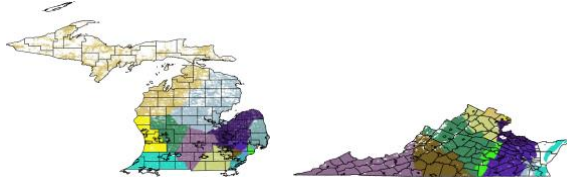


Figure 6: Algorithmically generated district plans for Virginia and Michigan when incorporating boundaries of counties and urban areas.

preserving county lines have on redistricting results, we propose an adapted power diagram technique. Instead of using a euclidian metric space, we modify distances between points by introducing a penalty for crossing administrative lines. This change increases the distance between census blocks separated by a county line. Because of this, there is more room for power diagram borders to fall between counties, and borders are less likely to split counties. Intuitively, this measure is equivalent to adding a “wall” to county boundaries. A line which crosses any wall is longer as it has to travel up and down each side.

In order to ensure convergence in this non-euclidean metric space, the algorithm must consider the centroid when the points are projected onto the county lines. It then selects either the general centroid or the projected centroid to minimize total mean squared distance. The number of counties which are split is reduced by about 10-15% when the penalty is introduced. We believe this technique is better than using a “second stage” to fit census blocks to counties as that approach could only consider local optimums nearby the first incorrect solution.

We find that incorporating county and urban area boundaries results in simulated elections that significantly advantage rural voters. We use urban areas as defined by the 2010 census [16]. As a case study we show detailed results for Virginia and Michigan in Figure 5. These results suggest that incorporating existing administrative lines into district plans could introduce unintended political bias. Even if city and county lines are not politically motivated, using them could result in politically biased districts.

5 CONCLUSION

Likely one of the reasons our maps do not heavily favor rural voters is that they tend to split dense urban areas across multiple districts. Our approach generates competitive districts that cut across urban, suburban and rural areas. In previous work by Chen and Rodden [2] exploring the electoral outcomes of algorithmically generated districts, they find split cities to be rare in their model. Their work, and related results using MCMC, builds districts by randomly combining precincts [2, 7, 9]. Chen and Rodden [2], for instance, argue their maps are compact since they combine precincts that are nearest to each other to generate district plans. However, we suspect this model is highly sensitive to local features. Since urban areas are dense, urban precincts will always be combined with other urban precincts. This process likely results in first placing cities into their own district and then building rural districts from

what is left over. While their notion of compactness makes sense from the perspective of a single district, there is no optimization which balances compactness across multiple districts.

Our approach, however, reduces voter dispersion across all districts balancing compactness from both local and global features. This means that both urban *and* rural districts are equally optimized for compactness. It is this objective to maximize global compactness that makes our maps significantly different from human drawn maps that only appear compact.

Our results suggest that contrary to assumption, population density may actually advantage the urban party in certain cases. We expect that these results differ from previous studies for two reason: we ignore administrative lines and our algorithm tends to split cities across multiple districts. When we modify our approach to respect administrative lines, we see an advantage for the rural party. We hope to counter the assumption that compactness on its own inherently favors rural voters. We suggest instead that it is attempts to respect municipal and county lines which introduce bias.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation grant CCF-1841954.

REFERENCES

- [1] Allan Borodin, Omer Lev, Nisarg Shah, and Tyrone Strangway. Big city vs. the great outdoors: Voter distribution and how it affects gerrymandering. In *IJCAI*, pages 98–104, 2018.
- [2] Jowei Chen, Jonathan Rodden, et al. Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8(3):239–269, 2013.
- [3] Colin Clark. Urban population densities. *Journal of the Royal Statistical Society. Series A (General)*, 114(4):490–496, 1951.
- [4] Vincent Cohen-Addad, Philip N. Klein, and Neal E. Young. Balanced centroidal power diagrams for redistricting. In *SIGSPATIAL/GIS*, 2018.
- [5] Vincent Cohen-Addad, Philip N Klein, and Neal E Young. Balanced centroidal power diagrams for redistricting. In *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 389–396, 2018.
- [6] Moon Duchin. Geometry versus gerrymandering. *Scientific American*, 2018.
- [7] Moon Duchin, Taissa Gladkova, Eugene Henninger-Voss, Ben Klingensmith, Heather Newman, and Hannah Wheelen. Locating the representational baseline: Republicans in massachusetts. *arXiv preprint arXiv:1810.09051*, 2018.
- [8] Jordan Ellenberg. How computers turned gerrymandering into a science. *The New York Times*, 2017.
- [9] Benjamin Fifield, Michael Higgins, Kosuke Imai, and Alexander Tarr. A new automated redistricting simulator using markov chain monte carlo. *Work. Pap., Princeton Univ., Princeton, NJ*, 2015.
- [10] Roland G Fryer Jr and Richard Holden. Measuring the compactness of political districting plans. *The Journal of Law and Economics*, 54(3):493–535, 2011.
- [11] Justin Levi. All about redistricting: Professor Justin Levi’s guide to drawing the electoral lines, 2012.
- [12] Michael D McDonald and Robin E Best. Unfair partisan gerrymanders in politics and law: A diagnostic applied to six cases. *Election Law Journal*, 14(4):312–330, 2015.
- [13] Associated Press. Virginia: Court again redraws voting districts. <https://www.nytimes.com/2016/01/08/us/virginia-court-again-redraws-voting-districts.html>, Jan 2016.
- [14] Princeton Gerrymandering Project. Openprecincts. openprecincts.org.
- [15] Jonathan Rodden. *Why cities lose: the deep roots of the urban-rural political divide*. Basic Books, 2019.
- [16] U.S. Census Bureau. 2010 census. U.S. Department of Commerce, February 2011.
- [17] Michael Wines. What is gerrymandering? What if the supreme court bans it? *The New York Times*, Mar 2019.