

What to Feed a Gerrymander

Ben Conlee
 Abe Othman
 Chris Yetter
 Harvard University
 Cambridge, MA

Advisor: Clifford H. Taubes

Summary

Gerrymandering, the practice of dividing political districts into winding and unfair geometries, has a deleterious effect on democratic accountability and participation. Incumbent politicians have an incentive to create districts to their advantage (California in 2000, Texas in 2003); so one proposed remedy for gerrymandering is to adopt an objective, possibly computerized, methodology for districting.

We present two efficient algorithms for solving the districting problem by modeling it as a Markov decision process that rewards traditional measures of district “goodness”: equality of population, continuity, preservation of county lines, and compactness of shape. Our Multi-Seeded Growth Model simulates the creation of a fixed number of districts for an arbitrary geography by “planting seeds” for districts and specifying particular growth rules. The result of this process is refined in our Partition Optimization Model, which uses stochastic domain hill-climbing to make small changes in district lines to improve goodness. We include as an extension an optimization to minimize projected inequality in district populations between redistrictings.

As a case study, we implement our models to create an unbiased, geographically simple districting of New York using tract-level data from the 2000 Census.

What is Gerrymandering?

Gerrymandering is division of an area into political districts that give advantage to one group. Frequently, this involves concentrating “unfavorable”

The UMAP Journal 28 (3) (2007) 261–280. ©Copyright 2007 by COMAP, Inc. All rights reserved. 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. Abstracting with credit is permitted, but copyrights for components of this work owned by others than COMAP must be honored. To copy otherwise, to republish, to post on servers, or to redistribute to lists requires prior permission from COMAP.

voters in a few districts to ensure that “favorable” voters will win in many more districts. To squeeze unfavorable voters into a few districts, gerrymandering creates snaky and odd shaped regions. The eponymous label was created when politician Elbridge Gerry pioneered this technique in early 19th century and his opponents claimed that the districts resembled salamanders (**Figure 1**).



Figure 1. The original “Gerry-mander” from the *Boston Centinel* (1812). Source: Wikipedia [2007], which in turn was cropped from U.S. Department of the Interior [2007].

Basic Terminology

- **Packing:** Forcing a disproportionately high concentration of a particular group into one district to lessen their impact in nearby districts.
- **Cracking:** Spreading members of a group into several districts to reduce their impact in each of these districts.
- **Forfeit district:** A district where group *A* packs the members of group *B* so that group *B* wins this district but loses several surrounding districts that *B* might win with a different districting scheme.
- **Wasted Vote:** A vote cast by a member of group *A* in a district where *A* is already assured victory, so voting has no bearing on the result. In general, the group with more wasted votes is made worse off by a districting plan.

Why Is It So Bad?

Gerrymandering reduces electoral competition within districts, since cracking/packing makes elections uncompetitive. Further, incumbent representatives are in no real danger of losing elections, so they do not campaign vigorously, which can lead to lower voter turnout. Exacerbating the problem, incumbents’ increased advantage means that they have less incentive to govern based on their constituents’ interests, so democratic accountability and engagement mutually deteriorate.

Gerrymandering also presents the practical problem that it is difficult to explain to voters why district shapes are so labyrinthine. Some districts connect demographically-similar but geographically-distant regions using thin filaments (**Figure 2**). “Niceness” of district shape almost always takes a back seat to political and racial concerns. Example: In the 2000 California realignment, Democrats and Republicans united to design incumbent-favoring districts, which resulted in the re-election of all of the 153 incumbents in 2004. How can one argue that this is in voters’ best interests?

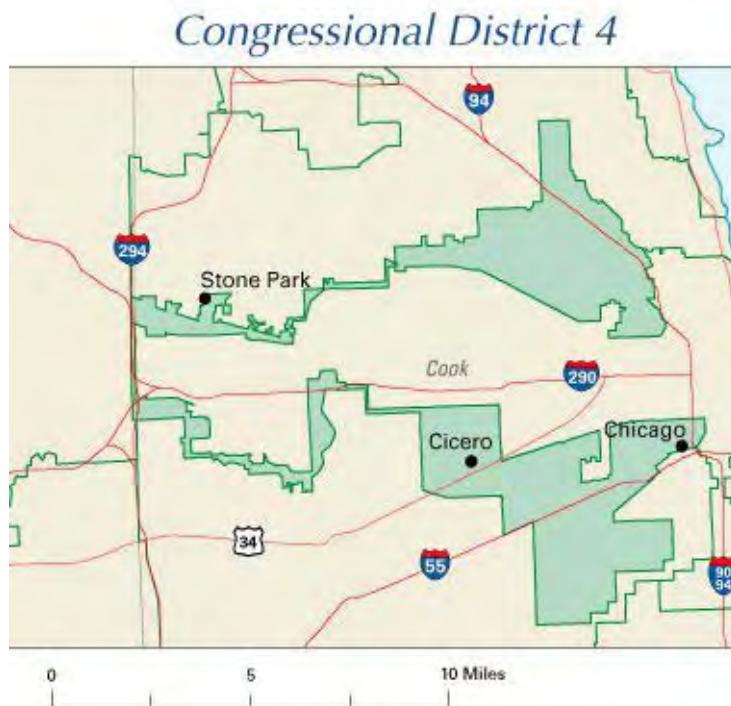


Figure 2. A present-day gerrymander, the Illinois 4th congressional district. The two “earmuffs” are connected by a narrow band along Interstate 294. Source: Wikipedia [2007], in turn cropped from U.S. Department of the Interior [2007].

However, gerrymandering can be considered appropriate in specific situations. For instance, the Arizona Legislature gerrymandered a division between the historically hostile Hopi and Navajo tribes even though the Hopi reservation is entirely surrounded by the Navajo reservation.

The Legality of Gerrymandering

Though gerrymandering is objectionable to many, it is legal. The Voting Rights Act of 1965, which eliminated poll taxes and other discriminatory voting policies, may have inadvertently increased gerrymandering. One interpretation of the Act is that it mandates nondiscriminatory election *results*, which has led to a strange reversal of vocabulary in which creating “majority-minority” districts is considered beneficial. These gerrymandered districts are packed

with minorities to guarantee minority representation in Congress.

However, in *Shaw v. Reno* (1993), and later in *Miller v. Johnson* (1995), the Supreme Court ruled that racial/ethnic gerrymanders are unconstitutional. Nevertheless, *Hunt v. Cromartie* (1999) approved of a seemingly racial gerrymandering because the motivation was mostly partisan rather than racial. The recent case *League of United Latin American Citizens v. Perry* (June 2006) held that states are free to redistrict as often as they like so long as the redistrictings are not purely racially motivated.

Assumptions and Notation

What Can We Consider When Districting?

1. Population equality between districts (legally mandated)
2. Contiguity of districts (legally mandated, excepting islands)
3. Respect for legal boundaries (counties, city limits, townships)
4. Respect for natural geographic boundaries
5. Compactness of district shapes
6. Respect for human-made boundaries (highways, parks, etc.)
7. Respect for socioeconomic similarity of constituents
8. Similarity to past district boundaries
9. Partisan political concerns
10. Desire to make districts (un)competitive
11. Racial/ethnic concerns
12. Desire to protect (or unseat) incumbent politicians

In our model, we consider only the top seven factors. The case *SC State Conference of Branches v. Riley* (1982) ruled that past districts (factor 8) are a legitimate tool for creating new districts, but we ignore past districtings, since they are heavily biased by factors 9–12, all related to political or racial concerns.

Geography and Similar Characteristics

The U.S. Census Bureau provides data on legal, natural, and human-made boundaries as well as socioeconomic similarity of regions. In each census, the United States population is divided at several levels of accuracy, the smallest of which are: *blocks* (40 people on average), *block groups* (1,500 people), and *tracts*

(4,500 people). We follow the practice in Young [1988] by districting based on tracts.

Census tract boundaries normally follow visible features, but may follow governmental unit boundaries and other non-visible features, and they always nest within counties. Census tracts are designed to be relatively homogenous units with respect to population characteristics, economic status, and living conditions at the time the users established them.

[Caliper Corporation n.d.]

For these reasons, we believe that tracts are acceptably small and homogeneous to use as a base unit. Further, a tract is completely contained a county, so we can easily check whether or not a district breaks county integrity.

Notation

Let n be the number of districts and m the number of census tracts. We denote districts by D_i , tracts by T_l and the set of all tracts by $\Gamma = \{T_l\}_{1 \leq l \leq m}$, which we call a *state*. Denote the set of all districts at a particular time by $\Delta = \{D_i\}_{1 \leq i \leq n}$; we call this a *partition* of the state.

Adjacency

Define the symmetric relation $T_p \sim T_q$ for tract pairs (T_p, T_q) that are adjacent. Let $d(T_l)$ be the district to which T_l belongs. We also naturally extend the definition of d to sets of tracts.

Define the *neighbor set* of tract T_l by $a_T(T_l) = \{T_p \in \Gamma | T_l \sim T_p\}$, all tracts neighboring T_l ; and define $a_D(T_l) = d(a_T(T_l))$ to be the set of all districts containing neighbors of T_l . Every tract borders at least one other tract, so all $a_T(T_l)$ and $a_D(T_l)$ are nonempty.

Borders

Let the *border* of district D_i be $\partial D_i = \{T_l \in D_i | a_D(T_l) \neq \{D_i\}\}$, the set of tracts in D_i adjacent to at least one district other than D_i . The *interior* of district D_i is $I_i = D_i \setminus \partial D_i$, the set of tracts in D_i whose neighbors are all in D_i . Let $m_i = |D_i|$ be the number of tracts in D_i and $b_i = |\partial D_i|$ the number of tracts bordering D_i .

The *frontier* of D_i is $F_i = (\cup_{T_l \in D_i} a_T(T_l)) \setminus D_i$, the set of tracts outside of D_i that border the boundary tracts of D_i .

Counties

We denote a county by C_j and the set of all counties by Λ . Districts can (and often do) break county boundaries, but tracts are contained entirely within counties, so a county is a set of tracts. Districts are also sets of tracts, so we interpret $D_i \cap C_j$ as the set of tracts in both district D_i and county C_j .

Population

Let the population of the state be P and let $\bar{p} = P/n$ be the optimal district size. We use the function $p(\cdot)$ to denote the population of an object; for instance, $p(T_l)$ and $p(C_j)$ are the populations of tract T_l and county C_j , respectively. We use the shorthand $p_i = p(D_i)$ for the population of districts.

Table 1 is a useful reference of these numerous definitions.

Table 1.
Variables and their meanings.

Variable	Definition
n	Number of congressional districts
D_i	The i th district ($1 \leq i \leq n$)
Δ	Set of all districts in a state, a <i>partition</i>
m	Number of census tracts
T_l	The l th tract in ($1 \leq l \leq m$)
Γ	Set of all tracts in a state
$d(T_l)$	District to which tract T_l belongs
$T_p \sim T_q$	Tracts T_p and T_q are adjacent
$a_T(T_l)$	Set of tracts adjacent to tract T_l
$a_D(T_l)$	Set of districts containing tracts neighboring T_l
∂D_i	Border of D_i , tracts that neighbor another district
I_i	Interior of D_i , tracts that do not neighbor another district
m_i	Number of tracts in D_i
b_i	Number of tracts in ∂D_i
F_i	Set of all tracts outside of D_i that border ∂D_i
C_j	The j th county
$c(T_l)$	The county to which tract T_l belongs
$c(D_i)$	The set of counties containing district D_i
P	Total population of the state
\bar{p}	Average population of a district
$p(\cdot)$	Population of an arbitrary object
p_i	Population of district D_i

Past Models

Cirincione et al. [2000] judge the quality of a districting plan based on equal population, preservation of county integrity, and district area compactness. They require that district populations differ by no more than 1% from exact equality of number of constituents and point contiguity of a district. They construct districts by picking a random block group, then adding additional block groups to the new district until the population reaches \bar{p} . At this point, they repeat the process starting with a new random block group. Compactness is based on minimum bounding rectangles, and county integrity is encouraged by “randomly” selecting new block groups with a preference for block groups in counties already in the emerging district.

Mehrotra et al. [1998] and Garfinkel and Nemhauser [1970] implement a “branch-and-price” method in the optimization step. They first obtain a dis-

tricting and then optimize over constraints such that population sizes are allowed to vary. In a final step, they split up population units to ensure population equality. They define compactness in a graph-theoretical manner, where connected nodes are adjacent tracts. They define the “center” of a district to be the tract with the smallest maximum distance to another other tract. They consider a district compact when sum of distances from each node to the center is small.

We do not use their measure, since it does not uniquely define the center of a graph, and (contrary to their claims) does allow for oddly-shaped districts, such as a district whose graph is a star-shaped tree with one tract in the center and many noncontiguous paths emanating from it. Such a tree structure is one a salient feature of gerrymandering.

We also do not use a “branch-and-price” method of optimization. Following suggestions of Nagel [1965] and Kaiser [1966], we employ a local search algorithm in which tracts are swapped between existing districts to maximize the objective function.

Measuring Compactness

The notion of compactness of a planar region has no uniformly accepted definition. Young [1988] suggests that any reasonable measure of compactness should consider population units (census tracts in our case study) as indivisible but laments that no one measure seems to work well for all geographic configurations.

Young’s measures include the maximum total perimeter, the relative height and width, and the moment of inertia of the district. All these fail to consider both perimeter and area simultaneously.

The Isoperimetric Theorem states that the quantity A/P^2 , the ratio of the area A of a planar region (not necessarily contiguous) to the square of its perimeter, is maximized at $1/4\pi$ when the region is circular. We define *compactness* of a region as the ratio $4\pi A/P^2$. This ratio is 1 for the circle, with higher values indicating greater compactness. The compactness of a square is $4\pi/16 \approx 0.785$, an upper bound for compactness of any rectangle.

This ratio is a good measure of “regularity” of a region. Specifically, any shear of factor s applied to a circle decreases the compactness by a factor of s , and any concave region has lower compactness than its convex hull. In fact, the convex hull of a concave region has greater area *and* smaller perimeter.

The Multi-Seeded Growth Model

We take a two-stage approach to finding the best districting. In the *Multi-Seeded Growth Model* (MSGM), we find an initial allocation of n districts so that the partition has modest levels of population equality and county preservation.

Our *Partition Optimization Model* (*POM*) edits and improves the rough sketch from *MSGM*.

The reason that we use two phases is speed. Our initial inclination was to allocate tracts randomly to the n districts and then optimize by swapping tracts to improve some objective function. However, a random initial configuration is so far from the global maximum that the search might take years.

The *MSGM* generates a very crude districting that ensures district contiguity and tries for population equality and county preservation. Its districts are unacceptable for an actual plan but save enormous amounts of computing time.

How It Works

We grow the n districts simultaneously until they cover the state.

We start by allocating the entire state to a dummy district D_0 , and then allocate n tracts that serve as the initial “seeds” for the final districts, such that each D_i begins as only a single tract. While $|D_0| > 0$, we consider the set S of all possible moves that involve taking a district from D_0 while preserving contiguity. That is:

$$S(D_0; D_1, \dots, D_n) = \bigcup_{i=1}^n \bigcup_{T_l \in F_i} M(T_l, D_0, D_i),$$

where $M(T_l, D_i, D_j)$ represents a *move of tract* T_l from D_i to D_j and F_i is the set of tracts that border T_l . Each move is scored by desirability of the prospective partition according to the score if we were to accept only that move. We perform the top 3% of moves. This method preserves contiguity, because by definition any $T_l \in F_i$ must be contiguous with D_i , and thus the D_i are contiguous at each step.

Even though in the *MSGM* we do not consider moves between two “true” districts (rather, we consider only moves between a true district and the dummy district), the score of a move does not exist in isolation. Consider two adjacent districts D_i and D_j , a shared frontier tract $T_l \in F_i \cap F_j$, and an unshared frontier tract $T_k \in F_i \cap F_j^c$. The acceptance of $M(T_k, D_0, D_i)$ alters the heuristic value of every move associated with F_i , which could potentially affect the optimality of further moves with D_i , such as the acceptance of $M(T_l, D_0, D_i)$ rather than $M(T_l, D_0, D_j)$. Furthermore, the acceptance of $M(T_l, D_0, D_i)$ likely expands the size of F_i . Perhaps there is an optimal move opened up in this new frontier that we do not even consider, because we have not even calculated its value.

It would be better to perform only the best move, but such a strategy is too computationally intensive. We compromise by taking in each step an elite fraction of the moves before recalculating S and the values of its associated moves. In this respect, our approach is analogous to the strategy of *modified policy iteration* for solving a Markov decision problem, in which a fixed number of rounds of value iteration are made between policy iterations. The tradeoff

of possible inefficiency is more than compensated for by speed gain, especially considering that the solution obtained by MSGM will be further refined by POM.

The MSGM scheme uses a variable number of moves between recalculating the value of the frontier. Our scheme causes us to be delicate in our selections of tract allocations, making moves virtually one at a time at the beginning and end of the MSGM. By focusing on the beginning and end of the problem, we attempt to avoid having a single district grow too large through inefficient allocation.

Unlike Cirincione et al. [2000], we use random initial seeds weighted by population rather than seeds equally spaced around the state. The process works as follows: While there are still random seeds to be selected, we find a candidate initial seed tract T_l in D_0 . We accept T_l as an initial seed with probability $p(T_l)/\hat{p}$ so that tract selection is proportional to population. The MSGM algorithm produces the best initial results when all the districts have the same population rather than the same number of tracts. The geographically optimal placement of five (or fewer) starting seeds in the NYC Metropolitan area and Long Island evinces the fallibility of the equidistant initial-seed method.

The heuristic by which we rank candidate moves has two components: a population score and a county score.

Population Score

We want to minimize egregious disparities in population between districts. The population component of our heuristic should give the highest score to a district when $p_i = \bar{p}$. Additionally, we want to penalize large deviations from the optimal population, so our function should be concave down.

Let $f(p_i)$ be the population heuristic score for a district with population D_i . We use a piecewise definition for f :

$$f(p_i) = \begin{cases} M \sqrt{\frac{p_i}{\bar{p}}}, & \text{if } p_i \leq \bar{p}; \\ M - \frac{4M}{p_i^2} (p_i - \bar{p})^2, & \text{if } p_i > \bar{p}. \end{cases}$$

Notice that f is steeper for values $p_i > \bar{p}$ because we do not want growing districts to engulf too much population; we penalize deviations above \bar{p} worse than deviations below \bar{p} . **Figure 3** shows the function f .

County Preservation Score

We measure a district's county preservation score in terms of the percentage of counties that it completes on a population basis. To encourage growing districts to add remaining tracts in nearly complete counties, the marginal value of

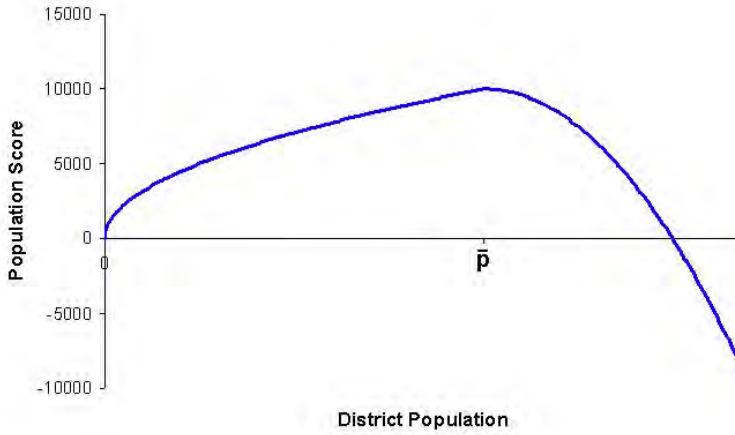


Figure 3. MSGM heuristic for population.

adding these should increase with the fraction of the population already contained in that district. To accomplish this, we use the square of the proportion contained in a county. The county score g for a district D_i is:

$$g(D_i) = \sum_{C_j \in \Lambda} \left(\frac{\sum_{T_l \in D_i \cap C_j} p(T_l)}{p(C_j)} \right)^2. \quad (1)$$

For instance, if a district completely contains one county and contains 30% of each of two other counties' populations, its score would be $(1^2 + 0.3^2 + 0.3^2) = 1.18$. **Figure 4** shows a plot of the county score that a district receives based on what percentage of a county's population said district contains.

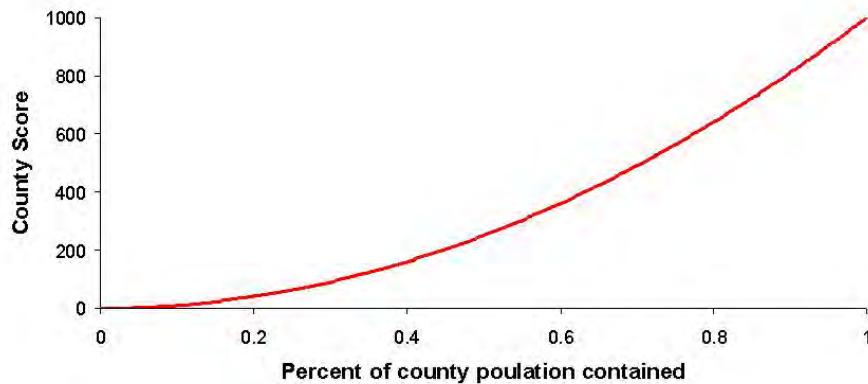


Figure 4. MSGM heuristic for county completeness.

Partition Optimization Model

We refine the MSGM solution through local search.

The Objective Function

The only characteristics of the district and the county that we use are the populations $p(P) = \{p_i\}_{1 \leq i \leq n}$, the compactness measures $c(P) = \{c_i\}_{1 \leq i \leq n}$, and the fractions $\rho(P) = \{\rho_{i,r} | 1 \leq i \leq n, 1 \leq r \leq c\}$ of the population of county r that is contained in district i . We would like our score function $s(P) = s(p(P), c(P), \rho(P))$ to have the following properties:

1. The score function should be unimodal as a function of p_i , with mode at $p_i = \bar{p}$.
2. The score should increase more by adding tracts that lie in $\chi(D_i)$, so that we prefer having as few districts as possible in a given county.
3. The score should increase more by adding tracts that increase the sum of all compactness measures by the greatest amount.

These desiderata suggest that we consider the three vectors $p(P)$, $c(P)$, $\rho(P)$ independently of one another in the score function. In other words, we would like our score function to be a separable function of these three vectors, that is, have the form

$$s(P) = f(p(P)) + g(c(P)) + h(\rho(P)),$$

where f, g, h are functions.

One (Wo)man, One Vote

The state has total population P and average population of $\bar{p} = P/n$ per district. Letting p_i be the population in district i , we consider three potential metrics for the population variance between districts:

1. Variance: $\text{Var}(p_1, p_2, \dots, p_n)$
2. Maximum deviation: $\max\{|p_i - \bar{p}|\}$
3. Maximum difference: $\max\{p_i\} - \min\{p_i\}$

For each measure, lower values are preferable and the minimum is 0. We submit that variance is the best alternative. To see why, consider two possible population distributions between districts:

- *Situation A:* One district has a population of $1.05\bar{p}$, one has $0.95\bar{p}$, and all of the others have \bar{p} .
- *Situation B:* Half of the districts have population $1.05\bar{p}$ and half have $0.95\bar{p}$ (any left-over odd district has \bar{p}).

In Situation A, only two districts are different from the ideal population level \bar{p} ; but in Situation B, very few districts have population \bar{p} . So a good metric should rank *B* worse than *A*. Clearly, the variance of populations is higher in *B* than in *A*, so variance passes this test. The maximum deviation test gives $0.05\bar{p}$ for both *A* and *B*, and the maximum difference gives $0.1\bar{p}$ for both.

We see that variance is the best measure of similarity, since it factors in the pairwise difference in all district populations.

By penalizing extreme variation away from \bar{p} , MSGM creates districts with approximate population equality. However, in one typical run, the final populations of districts vary from 600,000 to 700,000, an unacceptable difference.

Compactness

To measure the compactness of a district, we would ideally use our compactness measure:

$$c_i = \frac{\text{Area}(D_i)}{[\text{Perimeter}(D_i)]^2},$$

such that:

$$g(c(P)) = \beta \sum_{i=1}^n c_i,$$

where β is some constant.

Unfortunately, we could not calculate the perimeter of an arbitrary tract (the C++ library that we used to interact with our census data shapefiles featured massive memory leaks for large-scale union operations, questionable accuracy for pairwise unions, and seemingly arbitrary calculations of intersection length).

Yet it is a poor craftsman who blames the tools, so we adopt a different measure of compactness. The *clustering coefficient* provides a rough approximation for compactness:

$$cc(D_i) = \frac{\sum_{T_l \in D_i} |\{T_k \in D_i | T_k \sim T_l\}|}{\binom{m_i}{2}},$$

such that

$$g(c(P)) = \beta \sum_{i=1}^n cc(D_i),$$

where β is some constant. The clustering coefficient provides a ratio of the total number of interdistrict boundaries to the maximum possible number of interdistrict boundaries. If all tracts were uniformly shaped, this measure would

prize square- and circle-shaped districts, while winding single-tract-width districts would be penalized. However, given the asymmetry of tract shapes, this measure does little to reflect negatively upon district shapes such as the dumbbell (two circular clusters of tracts connected by a narrow band of tracts). In general however, the clustering coefficient values adding to districts tracts that are “close” and removing from districts those tracts that are auxiliary.

County Preservation

We adopt the same county preservation measure (1) used in the MSGM, with the option of adding a scaling factor to the entire function to refine empirical performance.

Search Method and Neighborhood Function

To refine our solution from MSGM, we must move tracts between districts. Yet the space of all possible contiguous moves is too large. We consider a range of possible moves with respect to only one district and perform the best move on this dramatically reduced state space.

By selecting our target district at random in each iteration, our strategy is best described as *stochastic domain hill climbing*, a method that combines the best aspects of both random and deterministic local search methods. We perform optimal moves while avoiding getting stuck trying to increase the score of only a single district. Simple first-order moves on the district level, that is, adding or removing individual tracts, cannot reduce the variance metric to the extremely low standard that we demand, so we include second-order moves, that is, “swaps”—both an add and a remove within a single operation.

If the maximum connectedness of any tract on the graph is k , checking for all adds and removes separately for district D_i involves considering

$$\mathcal{O}(k \cdot |\partial D_i| + |F_i|) = \mathcal{O}(km_i)$$

moves, while looking at all swaps involves considering $\mathcal{O}(k \cdot |\partial D_i| \cdot |F_i|) = \mathcal{O}(km_i^2)$ moves. One might contend, then, that the operation of checking *every* district for first-order moves might be a better algorithm, since it would take $\mathcal{O}(\sum_{i=1}^n km_i) = \mathcal{O}(nkm_i)$ heuristic evaluations. One could even supplement such an algorithm with a degree of randomness, to avoid being caught in a loop of futility, by employing simulated annealing, stochastic hill climbing, or tabu search on the resulting list of possible future states. We found, however, that checking for second-order moves provides far better empirical results with acceptable time performance, while an algorithm enumerating all the possible second-order states, requiring $\mathcal{O}(\sum_{i=1}^n km_i^2) = \mathcal{O}(nkm_i^2)$ heuristic evaluations, was too slow to be effective.

The heart of POM is **Algorithm 1**. For simplicity and readability:

- $M_{\text{add}}(D_i)$ is the set of all moves in which we add a frontier tract to D_i ,

- $M_{\text{remove}}(D_i)$ is the set of all moves in which we remove a border tract from D_i , and
- M^{-1} is the move inverse to M , such that applying both M and M^{-1} in turn has no effect.

Input: Iteration count $iter$, initial partition P .

Output: Final partition P .

```

count ← 0
while count < iter do
    curscore ← s( $P$ )
     $D \leftarrow randomDistrict()$ 
    bestscore ← curscore
    foreach  $M_a \in \{\emptyset \cup M_{\text{add}}(D)\}$  do
        foreach  $M_r \in \{\emptyset \cup M_{\text{remove}}(D)\}$  do
            performMove( $M_a$ )
            performMove( $M_r$ )
            if isContiguous( $P$ ) then
                tmpscore ← s( $P$ )
                if tmpscore > bestscore then
                    bestscore ← tmpscore
                    bestadd ←  $M_a$ 
                    bestremove ←  $M_r$ 
                end
            end
            performMove( $M_a^{-1}$ )
            performMove( $M_r^{-1}$ )
        end
    end
    if bestscore > curscore then
        performMove(bestadd)
        performMove(bestremove)
        count ← count + 1
    end
    return  $P$ 
end
```

Algorithm 1: Stochastic domain hill-climbing algorithm for districting.

We guarantee that the solution will be contiguous by not even considering moves that would break contiguity, and that we perform a move only if it increases the score of our current state.

Achieving Absolute Equality

U.S. law mandates that the populations of each district in a state be equal *to within one person* according to the census data [*Karcher v. Daggett* (1983)]. We deal with entire census tracts, so our algorithm cannot meet that standard. This last step must be implemented by splitting tracts between districts.

To our knowledge, this problem beyond population unit level (no smaller than block groups) has not been addressed in the literature. Clearly, the simplest way to do this is to split one of the border tracts. While we do not implement this, we describe a methodology for it.

Let G be a graph whose vertices are the districts and whose edges are the pairs of bordering districts. If we can find a pair of districts such that splitting a border tract between them gives both districts populations within one person of the mean population, then we would optimally do so and ignore those two districts for the remainder of the algorithm. However, to guarantee that the algorithm finishes, we require that the graph G remain connected (otherwise, G may divide into two or more connected components, such that the constituent districts cannot attain populations equal to the overall mean). Taking out two districts at a time by splitting only a single tract splits the fewest possible tracts.

We search for an edge of G such that removal of its two vertices and all edges emanating from them leaves a new graph $G_1 \subset G$ that is connected. We call the deletion of a single vertex from a graph that leaves the graph connected a *paring*. If these two vertices have some special properties, we perform the double paring and then perform the algorithm on G_1 , and continue until all districts have equal population. If no such pair of districts exists, we then perform a single paring and ensure that the removed district has population \bar{p} before removing it. Define *tract splitting* to be the process of splitting a border tract into two disjoint areas and two disjoint populations allocated between two bordering districts.

There always exists an edge on a connected graph G that permits a double paring of G , except for a very specialized set of connected graphs. However:

Theorem. *Every connected graph permits a paring.*

A proof of this theorem is given in the [Appendix](#).

We recursively update the districts to get population equality. We iteratively pare the graph G of districts such that each time we pare a district or pair of districts, those districts have populations which equal the population mean. By the theorem, this process always ends with all districts having equal population.

The algorithm removes at least one vertex from G at each step, and the whole algorithm can therefore be performed with $(n - d)$ tract splittings, where n is the number of districts and d is the number of double parings performed.

Case Study: New York

The Data

The 2000 census for New York State contains 4,907 tracts, some with no population [Empire State Development 2007]. These empty districts are the “holes” on our maps. Trimming these tracts leaves 4,827 tracts to examine.

Results

Running the MSGM on our initial allocation gives 29 haggard districts, varying from 281,000 to 970,000 population. We use this solution as a starting point. Though our algorithmic process of refinement is stochastic, generally more than 90% of the moves in any run involve swaps; this is particularly true at the very end of a run, where population differences between districts are minute. As a result, swapping provides a way to adjust population smoothly and also “cleans up” tattered fringes of districts, increasing their compactness even with vigorous population changes. After refinement, district populations ranged from 652,561 to 655,760.

The results in **Figures 5–7** demonstrate a partitioning into contiguous, compact, and reasonable districts. Furthermore, the simulations that produced these visually pleasing results also achieve extremely high degrees of population equality and county preservation.

Analysis of the Models

Solving the Problem

By combining the Multi-Seeded Growth Model with the Partition Optimization Model, we create fair and geometrically compact districts. The districts conform to well-accepted measures of goodness: population equality, contiguity, preservation of county boundaries, and compactness of shape.

The districts produced are both simple and fair. Geometric simplicity is measured by compactness, as determined by how close the members of a districts live relative to one other. Additionally, our method penalizes splitting counties between several districts, so that neighboring citizens, who have similar concerns, have the same representative. Fairness of our methodology is evident in its indifference to partisan politics, incumbent protection, and race/ethnicity.

Strengths

The model successfully generates district partitions that simultaneously excel against the standard metrics of county integrity, compactness, and popu-

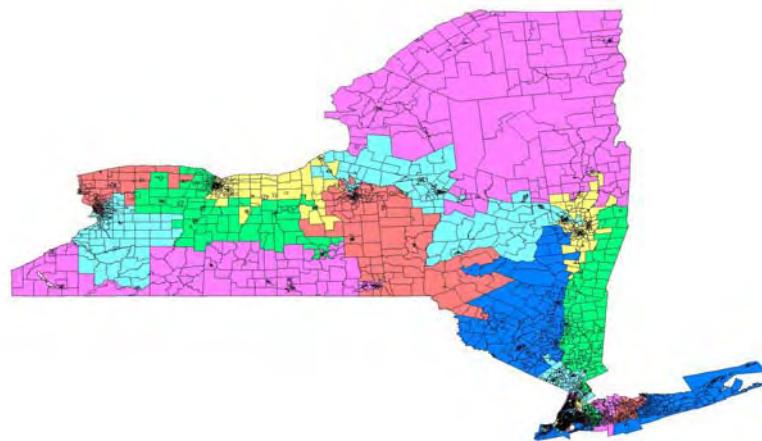


Figure 5. New York congressional districts from the POM.

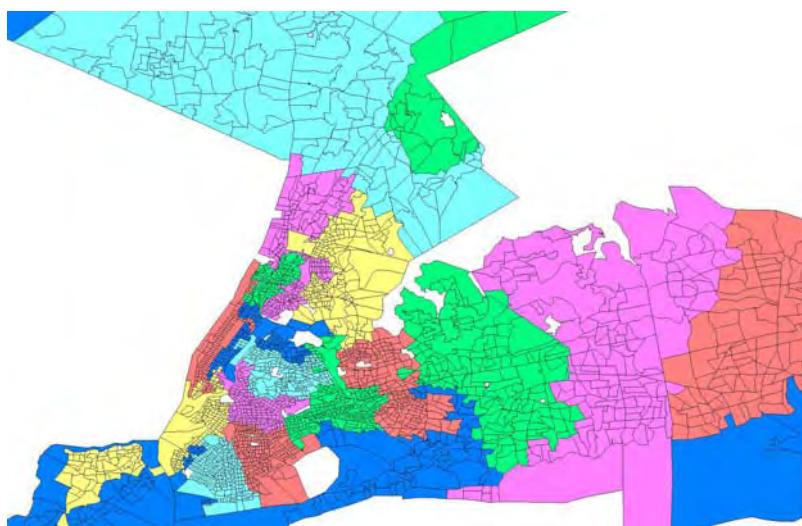


Figure 6. NYC metro-area POM.

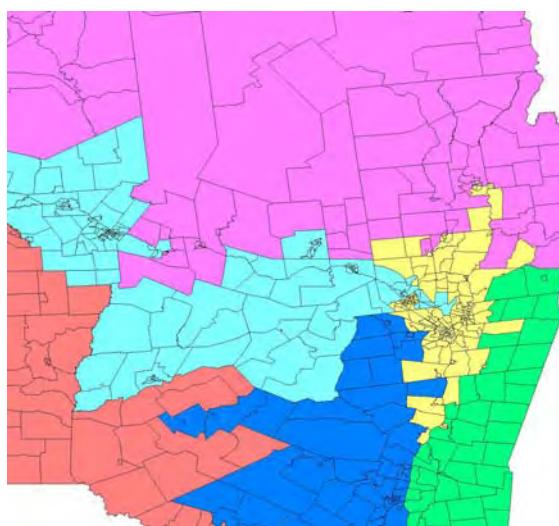


Figure 7. Close-up of the Albany area POM.

lation equality. Unlike other models in the literature, we provide an algorithm for reducing population differences to at most one person by breaking up a minimal number of tracts.

The model runs independently of the distribution of population, and works well both in low- and high-density locales, and with regular and oddly-shaped census tracts. This is evidenced by the successful districtings that our model produces in rural, small city, and large metropolitan areas (**Figures 5–7**).

The algorithm can generate districts for a large state in less than an hour.

Weaknesses

The model assumes contiguity of the entire state; so in cases where contiguity cannot be forced, such as Hawaii or Michigan, we must change the algorithm slightly. One solution could be to divide the state into several regions and run the model separately on each region, allocating the proportionally correct number of representatives to each region based on population.

The model appears to tend toward creating districts that are either very low- or high-density, instead of splitting smaller population centers into a number of districts. Since political affiliation and race are likely correlated with population density, the algorithm may inadvertently generate districts that separate various demographic groups into separate districts, which could be viewed as gerrymandering. Yet, another camp would argue that it is appropriate to divide urban, suburban, and rural areas into separate districts, since their residents have different concerns.

Conclusion

Since the 19th century, Elbridge Gerry's lizard has grown into a terrible, twisting serpent, eating away at our democracy. It is time to put Gerrymanders on a healthier diet.

References

- Barkan, J.D., P.J. Densham, and G. Rushton. 2006. Space matters: Designing better electoral systems for emerging democracies. *American Journal of Political Science* 50 (4): 926–939.
- Bong, C., and Y. Wang. 2006. A multi-objective hybrid metaheuristic for zone definition procedure. *International Journal of Services Operations and Informatics* 1 (1/2): 146–164.
- Caliper Corporation. n.d. About census summary levels. <http://www.caliper.com/Maptitude/Census2000Data/SummaryLevels.htm>.

- Cirincione, C., T.A. Darling, and T.G. O'Rourke. 2000. Assessing South Carolina's 1990s Congressional redistricting. *Political Geography* 19: 189–211.
- Empire State Development. 2007. New York State Data Center. Census 2000. <http://www.empire.state.ny.us/nysdc/popandhous/Census2000.asp>.
- ePodunk Inc. 2007. New York: Population change, 2000 to 2003. <http://www.epodunk.com/top10/countyPop/coPop33.html>.
- Garfinkel, R.S., and G.L. Nemhauser. 1970. Optimal political districting by implicit enumeration techniques. *Management Science* 16 (4): B495–B508.
- Hunt v. Cromartie*, 526 U.S. 541 (1999).
- Karcher v. Daggett*, 462 U.S. 725 (1983).
- Kaiser, H. 1966. An objective method for establishing legislative districts. *Midwest Journal of Political Science* 10 (2) (May 1966): 210–213.
- League of United Latin American Citizens v. Perry*, 548 U.S. (2006).
- Luttinger, J.M. 1973. Generalized isoperimetric inequalities. *Proceedings of the National Academy of Sciences of the United States of America* 70: 1005–1006.
- Macmillan, W. 2001. Redistricting in a GIS environment: An optimisation algorithm using switching-points. *Journal of Geographical Systems* 3 (2): 167–180.
- _____, and T. Pierce. 1994. Optimization modeling in a GIS framework: The problem of political districting. In *Spatial Analysis and GIS*, edited by S. Fotheringham and P. Rogerson. Bristol, UK: Taylor and Francis.
- Mehrotra, A., E.L. Johnson, and G.L. Nemhauser. 1998. An optimization based heuristic for political districting. *Management Science* 44 (8): 1100–1114.
- Miller v. Johnson*, 515 U.S. 900 (1995).
- Nagel, S. 1965. Simplified bipartisan computer redistricting. *Stanford Law Review* 17: 863–899.
- Shaw v. Reno*, 509 U.S. 630 (1993).
- SC State Conference of Branches, etc. v. Riley*, (1982). 533 F. Supp. 1178 (DSC). Affirmed 459 US 1025.
- U.S. Department of the Interior. 2007. Printable maps. <http://nationalatlas.gov/printable/congress.html#list>.
- Weaver, J.B., and S.W. Hess. 1963. A procedure for nonpartisan districting: Development of computer techniques. *Yale Law Journal* 73 (1): 287–308.
- Wikipedia. 2007. Gerrymandering. <http://en.wikipedia.org/wiki/Gerrymandering>.
- Young, H.P. 1988. Measuring the compactness of legislative districts. *Legislative Studies Quarterly* 13: 105–115.

Appendix: Proof of Theorem

Theorem. Every connected graph permits a paring.

Proof: We proceed by induction on the number y of vertices. We prove a stronger statement, namely that for any connected graph G with at least two vertices, there exist at least two pairings. The claim clearly holds for $y = 2$.

Suppose that the claim holds for $y = k$, where $k \geq 2$, and that it does not hold for $y = k + 1$. Then, since $y \geq 3$, take any vertex v of G such that removal of v leaves G unconnected, and consider the two disjoint subgraphs G_1, G_2 into which G is divided upon removal of this vertex. By the induction hypothesis, there exist vertices v_1, v_2 of G_1 such that removal of either one leaves G_1 connected.

We claim that removal of one of v_1, v_2 from the original graph G leaves G connected. To see this, note that neither v_1 nor v_2 is adjacent to any vertex in G_2 , as G_1, G_2 have no common edges. If both v_1, v_2 are adjacent to v , then removal of v_1 leaves G connected. This is because if we let $G' = G - \{v_1\}$ and $G'_1 = G_1 - \{v_1\}$, then G' consists of $G'_1 \cup \{v\}$ and G_2 , which are both connected and connected to each other, as v is necessarily connected to G_2 . This means that $G - \{v_1\}$ is connected.

If one of v_1, v_2 is not adjacent to v , without loss of generality assume that it is v_1 . Then removing v_1 from G leaves the graph connected, as $G'_1 \cup \{v\}$ is connected, as is G_2 , and they are connected to each other. Some such vertex which admits a paring also exists in G_2 , yielding two vertices which permit a paring. This proves the result by induction.



Abe Othman, Chris Yetter, and Ben Conlee.