Computers & Industrial Engineering 140 (2020) 106265

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

Mathematical models of political districting for more representative governments

Hongrui Liu, Ayca Erdogan, Royce Lin, H.-S. Jacob Tsao*

Department of Industrial and Systems Engineering, San Jose State University, United States

ARTICLE INFO

Keywords:
Political districting
Math modeling
Optimization
Fairness
Competitiveness

ABSTRACT

Political Districting, the problem of partitioning a geographical region into sub-regions as electoral districts for political seat assignment, has been widely studied by researchers of both social and mathematical sciences. In the US, one Congressional representative is elected for each such district, and the Constitution requires compactness, geographical contiguity, and population balance for any such district. "Gerrymandering," i.e., misuse of re-districting once every 10 years (after each new census) to benefit a particular political party or current incumbent(s), has been practiced for decades, and political strongholds for individual parties have been formed. The vast majority, if not all, of the current literature and proposed software tools on political districting focus on how to achieve the non-political constitutional requirements. Election and re-districting are political processes; we believe that software tools are needed to achieve political purposes. In this paper, we present two mathematical optimization models to implement two new political criteria: fairness and competitiveness. Fairness aims to ensure fair allocation of seats to political parties considering the distribution of voters. Competitiveness aims to maximize the number of competitive districts to prevent districting solutions that provide clear advantage to one political party. We show that the feasibility problem associated with each of the two optimization models is NP-complete. Both models are implemented with a case study of South Carolina. The numerical results of six scenarios demonstrate their effectiveness and efficiency. A strategy is proposed to deal with the issue of computational complexity associated with a state that is partitioned into a large number of indivisible area/population units, e.g., census tracks, as building blocks of electoral districts. The two models can be extended or modified to address fairness and competitiveness issues in some political systems with three or more political parties.

1. Introduction

Political Districting (or re-districting) partitions a given geographical region into a given number of districts to which a known number of political seats are allocated. The number of seats allocated to the region is usually determined by its population. Throughout this paper, we use the US Congressional Elections as the example, where the given geographical region is a state and exactly one seat is allocated to each district. In the US, a census takes place every 10 years. A change in population size of a state may lead to a different number of political seats allocated to the state and hence lead to the necessity of re-districting of the state. Proper re-districting is a critically important problem, because different districting combinations may result in different seat assignments thus can change election outcomes.

The US Constitution and Supreme Court rulings mandate three key legal requirements for districting or re-districting: proportional equality (i.e., approximately equal population size for all districts), spatial contiguity and compactness. Each district consists of multiple indivisible areas often called "population units" or "area units" (or simply "areas" in what follows), which could be counties in some states and census tracks or others in other states. Contiguity is achieved when it is possible to go from each unit in the district to another without crossing borders to another district. Proportional population and spatial contiguity can be clearly defined. Compactness on the other hand pertains to the shape of the area, is difficult to rigorously define, and can be defined in many different ways, using geography or population. Geographical compactness may be thought of essentially requiring the shape of the area to resemble as much as possible a circle or a square, without any spikes. Population-based compactness considers the distance of the population to the center of the district. Partisan politicians and political incumbents have taken advantage of this vagueness for their own political advantages. Decades of indulgence has resulted in

* Corresponding author.
E-mail addresses: hongrui.liu@sjsu.edu (H. Liu), ayca.erdogan@sjsu.edu (A. Erdogan), jacob.tsao@sjsu.edu (H.-S.J. Tsao).

https://doi.org/10.1016/j.cie.2019.106265
Received 18 July 2019; Received in revised form 26 November 2019; Accepted 30 December 2019
Available online 31 December 2019
0360-8352/

many districts resembling a salamander, whose body of head through the tail or four limbs and their toes certainly is not compact by any measure. Part of the name of an early US politician particularly known for such indulgence was attached to part of the name of this critter, giving rise to the term Gerrymandering.

Decades of Gerrymandering has resulted in political strongholds for either of the two major political parties of the US, where (i) the voters for one party in one district outnumber those for the other so much so that campaigning by either party is waste of effort and (ii) the ratio between the number of seats won by the two parties in a State is far from the corresponding ratio between the two total numbers of voters. These are two fundamental problems of Gerrymandering: non-competitiveness and unfairness. These two serious problems motivated our research and the two mathematical models.

Re-districting was typically conducted by a commission and has been done through negotiation and consensus building, without any modern computer tools. Re-districting can be formulated and solved as a difficult purely mathematical problem, without any partisan biases; such efforts started as early as 1960s but continued intermittently and sparsely. While such commissions and their staff members continue to tinker manually with the complex combinatorial problem of re-districting, our research will assist them by developing a suite of mathematical models and their computer implementations to develop and optimize districting plans according to not only the US Constitutional mandates but also two different criteria designed for possible use by local re-districting commissions to achieve fairer and more competitive elections.

American politics has become rather polarized recently, and the "establishments" on both major political parties do not seem to represent the population well. The branch of the US federal government that represents the population most directly is the House of Representatives; we focus on election of its members. Despite its historically low approval ratings (in the teens) in recent years, the vast majority of the incumbents continued to be re-elected. A root cause of this is "gerrymandering" or partisan re-districting, i.e., using incumbents' power to redraw a state's election districts to their advantage after each US Census (held every 10 years and to be held in 2020 again). Despite rulings of some such practices by the Supreme or other Courts as unconstitutional, the situation has not improved. Gerrymandering is possible because of not only the vagueness of compactness but also the numerous ways to satisfy the other two Constitutional mandates.

In this paper, we present two mathematical models for the political redistricting problem, which attempt to deal with additional aspects that were not explicitly stated in the abovementioned three principles: fairness and competitiveness. An example fairness requirement would stipulate that the percentage of districts of a state in which a party enjoyed a clear majority in recent and similar elections must be close to or at least not too far from the percentage of voters across the entire state that voted for the party's nominees in the elections. For example, a state with 6 seats and a 60% vs. 40% voter-split should have close to a 4 vs. 2 seat-split. An example competitiveness requirement would stipulate that a given number of the districts of a state must be competitive in the sense that, in a competitive district, the relative percentage of the voters of either of the two major parties in recent and similar elections must be between 45% and 55% if possible.

Our goal is to provide an easy-to-understand and practical decision support tool that enables the neutral commissions who are responsible to conduct the redistricting to explore various options and arrive at better decisions. The two optimization models proposed in this paper can be extended or modified to address fairness and competitiveness issues in some political systems with three or more political parties.

We prove that the feasibility problem (i.e., the decision problem or the recognition version) associated each of the two optimization models is NP-Complete. Both models are implemented with a case study of South Carolina. The numerical results of six scenarios demonstrate their effectiveness and efficiency. A strategy is proposed to deal with the issue of computational complexity associated with a state that is partitioned into a large number of indivisible area/population units, e.g., census tracks, as building blocks of electoral districts.

The rest of this paper is organized as follows. Literature review is summarized in Section 2. Section 3 describes and justifies the mathematical formulations. Computational complexity is discussed in Section 4. The results of a case study and their analysis are summarized in Section 5, followed by concluding remarks of Section 6.

2. Literature review

Researchers from various areas including political, social, and computer sciences, as well as operations research have studied political redistricting. It has been suggested that political redistricting should be done by an automated and impersonal procedure to prevent gerrymandering (Vickrey, 1961).

The mathematical modeling efforts started by the work by Hess, Weaver, Siegfeldt, Whelan, and Zitlau (1965) and Garfinkel and Nemhauser (1970). Hess et al. indicated that districting problem is analogous to warehouse location problem, which selects a number of warehouses (population centers in the districting case), and allocates customers (population units) to the open warehouses, with the exception of almost equal capacity (district population) requirement. The optimal solution to the integer-programming model based on location problem does not guarantee contiguity, thus they developed a heuristic method that finds a solution that is contiguous and compact. Similar heuristics are proposed and used as computer-aided redistricting approaches by many researchers while incorporating several different criteria (George, Lamar, & Wallace, 1997; Hojati, 1996; Kaiser, 1966; Morrill, 1973, 1976; Nagel, 1964, 1972; Thoreson & Liittschwager, 1967).

Garfinkel and Nemhauser's pioneering paper on political districting introduced a new formulation that considers all possible districts and selecting an optimal subset of those districts that minimizes a cost function that considers maximum deviation of district population from overall population mean. Their model is the one that first introduced the population deviation from the mean, to prevent districts with disproportionally large or small populations. They proposed a two-phase approach: In phase one, all feasible districts based on contiguity, compactness and limited deviation from the population mean are generated. In the second phase, the set of districts which covers each population unit only once while minimizing the maximum deviation from the population mean for each district is found. A similar two-stage approach is used by Nygreen for redistricting in Wales, in which the second phase utilized a set-partitioning approach (Nygreen, 1988).

Mehrotra, Johnson, and Nemhauser (1998) presented a mathematical model that considers the three principles of districting and used a three-step procedure to reach to a final districting plan. The pre-processing phase generates districts using a clustering heuristic, second phase uses Branch and Price to find an improved solution. Finally, the post-processing step modifies the solution to ensure population equality. They claim that, since the model does not consider political data, it would be free from potential gerrymandering accusations. However, with the increase in US population both in absolute numbers and demographic complexity, such models that ignore residential political patterns may result in redistricting plants that would disregards minority groups and their distinct interests which should receive political protection (Forest, 2005).

Mathematical models for redistricting listed above does not deal with contiguity directly, however it is somewhat addressed by the objective functions that minimize distance-based metrics. Thus, the solutions do not guarantee contiguity. In most redistricting literature, contiguity is handled through post-optimization processing. Many authors propose graph theoretical models that starts with a connected n-node graph, i.e. contiguity graph, and aims to find a compact partition of the graph to several connected graphs while ensuring population balance (Bodin, 1973). Others used metaheuristic methods to find solutions to districting problems. Bacao, Lobo, and Painho (2005) developed genetic algorithms and applied to the districting problem in Lisbon (Bacao et al., 2005). Forman and Yue proposed a Traveling Salesman Problem based genetic algorithm and tested it on the districting problems of several states in the US (Forman & Yue, 2003). Ricca and Simeone (2008) compared performances of four local search heuristics (simulated annealing, tabu search, a basic descent, and the Old Bachelor Acceptance Algorithm) for political redistricting as graph partitioning problem with additional criteria such as “fair apportionment” and “conforming administrative boundaries”. While enforcing contiguity constraint in graphs seems simple, that is considering edges that only exist if two nodes are geographical neighbors, ensuring this on large graphs with many vertices becomes computationally intractable. King, Jacobson, Sewell, and Cho (2012) introduce “geo-graphs”, which provides a new approach to graph partitioning, by considering each vertex as a geographical region. Using planar graph duality, the authors show that geo-graphs produce feasible partitions very efficiently independent of the number of nodes.

Bozkaya, Erkut and Laporte considered the political redistricting as a multi-criteria optimization problem (Bozkaya, Erkut, & Laporte, 2003). They considered contiguity as a hard constraint using adjacency lists, and all other criteria including population equity, compactness, socioeconomic homogeneity, similarity to existing districting plan, and integrity of communities as a weighted some of related metrics in the objective function. They proposed a tabu-search based heuristic that produces high feasible and quality solutions. For other local search methods applied to political districting problem, readers are referred to the work by Ricca, Scozzari, and Simeone (2013). For a general review of territorial design problems, and solutions algorithms, the readers are referred to Kalcsics, Nickel, and Schröder (2005).

Despite the many attempts to develop mathematical models of redistricting that can address the three principals of equal population, compactness and contiguity, without considering the information of registered voters, the resulting new redistricting plan may lead to disproportional representation of parties. In this paper, we present two models that aim to provide districting plans that consider proportionality of the voters for more fairness and higher competitiveness on top of the traditional three principles mentioned above.

3. Mathematical formulations

In this section, we propose mathematical formulations that address fairness and competitiveness, in addition to the other legal requirements described in Section 1 for the political districting/redistricting problem. We start by describing our general approach and notations. We then illustrate the mathematical formulation of the political districting/redistricting problem for a state in the United States to address fairness and competitiveness.

3.1. General approach and notation

We employ a hubbing approach to formulate a Binary Linear Programming (BLP) problem for the political districting/redistricting problem. The hubbing approach has been used in telecommunication and transportation industries for decades to deal with the problem of partitioning a large geographical area into small area units and grouping the small area units into districts, each of which has a “central office” or a “hub”, respectively. In hubbing, the contiguity and compactness are handled indirectly by minimizing the objective function of total (across all districts) weighted sum of distances from a hub to the area units it serves, weighted by the amount of demand. In addition, the distances can be raised to any arbitrarily high power to ensure continuity and compactness. The binary decision variables include (a) one yes-or-no decision for each area unit j to be designated as the hub of a district and (b) one yes-or-no decision for whether an area unit i is assigned to a hub area-unit j and hence the corresponding district.

For the political districting/redistricting problem, a given state consists of n indivisible areas and we simply call them “areas”, which could be counties in some states, census tracks or others in other states. The n areas must be partitioned into a given number of m districts. Each district is indexed by one and only one of its areas, and the indexed area is called and treated as the hub of the district. Those areas that are assigned to a common hub constitute a district. In other words, a hub, together with all its spokes, constitutes a district. Each area (of the state) is assigned to one and only one hub. Unlike the hubbing problems facing the telecommunication and transportation industries, our two optimization problems do not involve any flow between any pair of hubs and resemble the facility location problem.

Notations:

n: a given number of areas (i.e., area units) in a state;
m: a required and given number of hubs/districts in a state;
$i = 1, 2, \ldots, n$, is the area index.

For each area i, $i = 1, 2, \ldots, n$:

P_i = the given population size of area i; $\underline{P} = (P_1, P_2, \ldots P_n)$
R_i = the given number of Republican voters of area i; $\underline{R} = (R_1, R_2, \ldots R_n)$
D_i = the given number of Democratic voters of area i; $\underline{D} = (D_1, D_2, \ldots D_n)$
U_i = the given number of all other voters (undecided affiliation) of area i; $\underline{U} = (U_1, U_2, \ldots U_n)$
$V_i = R_i + D_i + U_i$ = the given total number of voters of area i.

For a state,

$P = \sum_{i=1}^{n} P_i$ = population size of the entire state;
$R \equiv \sum_{i=1}^{n} R_i$ = the number of Republican voters across the entire state;
$D \equiv \sum_{i=1}^{n} D_i$ = the number of Democratic voters across the entire state;
$V \equiv \sum_{i=1}^{n} V_i$ = the number of voters across the entire state.
D_{ij} = the given distance between the center of area i and the center area j; $\mathbf{D} = [D_{ij}]$
δ = a given tolerated percentage deviation from the target population P/m in any one district;
α = a given threshold parameter to restrict the distance from an area to its hub;
β_1 = a given lower bound for the number of Republic districts, i.e., districts in which, according to voter estimates, there are more Republican voters than Democratic voters;
β_2 = a given upper bound for the number of Republic districts;
T_R= a given target number of Republican districts;
γ = a given required minimum number of competitive districts;
σ = a given competitiveness parameter; if the (relative) percentage of Republican voters (with respective the total number of Republican and Democratic voters) in a district is within the range

of $[0.5 - \sigma, 0.5 + \sigma]$, then the district's election is considered competitive.

As mentioned earlier, we present two models in this paper. In the first model, we deal with fairness requirement in addition to the other three legal requirements, namely equal population, contiguity and compactness. We address the competitiveness requirement in the second model.

3.2. The fairness model

Political representation is fair if the proportions of hubs/districts favoring the two parties, are close to the corresponding proportions of the voters of the entire state. For example, if 30% of the voters of California are Republican, 60% of them are Democratic and 10% of them are neither, then it would be fair if the proportion of hubs/districts in which there are more Republican voters than Democratic voters is close 1/3 = 30%/(30% + 60%) and the proportion of hubs/districts in which there are less Republican voters than Democratic voters is close to 2/3. We acknowledge that such fairness may not be achievable. For example, the voter splits in all unit areas are close to this state-voter split, then all representatives of California would be Democrats. The complete formulation of the Political Districting (or re-districting) problem is described in this section. The geographical "contiguity" and "compactness" is achieved by optimizing a distance-based (D_{ij}) and voter-size-based (V_i) objective function, and the equal-population and the fairness requirement are modeled as constraints.

Decision Variables:

Binary hub decision variables: $y_j = 1$ if area j is a hub; $y_j = 0$ otherwise; $j = 1, 2, \ldots, n$. $\underline{y} = (y_1, y_2, \ldots, y_n)$

Area-to-hub assignment: $x_{ij} = 1$ if area i is assigned to area j as its hub; $x_{ij} = 0$ otherwise. $\mathbf{X} = [x_{ij}]$

Republic Hub/District indicator: $z_j = 1$ if area j is a hub/district (i.e., if $y_j = 1$) and there are more Republican voters than Democratic voters in the district. $\underline{z} = (z_1, z_2, \ldots, z_n)$

Objective Function:

$$\text{Min} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} V_i D_{ij} x_{ij} \tag{1}$$

To ensure "contiguity" and "compactness" with distance-oriented spoke weights, we can use distances directly or some other measures. For example, using squared hub-to-spoke distances may lead to a higher degree of "contiguity" and "compactness." We can try this if use of the distances themselves directly does not lead to sufficient degree of "contiguity" and "compactness."

Constraints:

Hubbing Constraints:

$$x_{ij} \le y_j, \quad \forall\, i = 1, 2, \cdots, n; \ , \quad \forall\, j = 1, 2, \cdots, n, \tag{2}$$

$$D_{ij} x_{ij} \le \alpha y_j, \quad \forall\, i = 1, 2, \cdots, n; \ , \quad \forall\, j = 1, 2, \cdots, n, \tag{3}$$

$$\sum_{j=1}^{n} x_{ij} = 1, \quad \forall\, i = 1, 2, \cdots, n, \tag{4}$$

$$\sum_{j=1}^{n} y_j = m. \tag{5}$$

The constraints in Eq. (2) ensure that an area i can only be assigned to an area j if is area j is selected as a hub. The constraints in Eq. (3) ensure that if an area j is selected as a hub and if area i is assigned to area j, then the distance from area i to area j cannot exceed a threshold α. Eq. (3) helps to achieve adjacency. We use Eq. (4) to ensure that one area can only be assigned to exactly one hub and Eq. (5) to ensure the total number of hubs is equal to the total number of districts.

Constraints defining Republican District Indicator z_j:

For every $j = 1, 2, \ldots, n$,

$$(z_j - 1)M + y_j \le \sum_{i=1}^{n} (R_i - D_i) x_{ij} \le M z_j, \tag{6}$$

$$z_j \le y_j, \tag{7}$$

where M is a very big number. In Eq. (6), if j is a hub (i.e., if $y_j = 1$), then $\Delta_j \equiv \sum_{i=1}^{n} (R_i - D_i) x_{ij}$ is the number of Republican voters minus the Democratic voters in that corresponding district. If $\Delta_j > 0$, the second inequality in (6) forces $z_j = 1$. If $\Delta_j < 0$, the first inequality in (6) forces $z_j = 0$. In addition, if $\Delta_j = 0$, the first inequality in (6) forces $z_j = 0$, because $z_j = 1$ leads to the contradiction of $y_j = 1 \le 0$. In short, If j is a hub (i.e., if $y_j = 1$), then $z_j = 1$ if and only if there are strictly more Republican voters than Democratic voters in district j, i.e., $\Delta_j > 0$. Moreover, if j is not a hub, i.e., if $y_j = 0$, this constraint does not impose any restriction on any of the other decision variables. Note that the presence of y_j in this constraint may not be necessary, although it does rule out the rare possibility that the total number of republican voters in a district is exactly equal to its democratic counterpart. The constraints in Eq. (7) ensure z_j can only be 1 if area j is a hub/district (i.e., if $y_j = 1$).

Fairness may be defined in multiple ways. The most straightforward way is to set a lower bound on the number of districts favoring each party. For example,

$$\sum_{j=1}^{n} z_j \ge \beta_1 \tag{8.a}$$

$$\sum\nolimits_{j=1}^{n} z_j \le \beta_2 \tag{8.b}$$

where β_1 is a given lower bound on the number of Republic districts and β_2 is a given upper bound on the number of Republic districts. Particular choices of β_1 and β_2 include $[m\frac{R}{R+D}]$ and $\left[m\frac{R}{R+D}\right] + 1$, *where* $[\ * \ ]$ denotes the largest integer that is less than or equal to the real number $*$. This choice of β_1 stipulate that the number of Republican districts cannot be less than $[m\frac{R}{R+D}]$.This choice of β_2 stipulates that the number of Democratic districts cannot be less than $[m\frac{D}{R+D}]$.

Another way is to require a specific split of the total number of districts between the two parties according to the corresponding split of the total number of voters of a State. More precisely, we formulate the fairness requirement in Eq. (8.c) below by making the total number of Republican districts (where there are more Republican voters than Democratic voters in the district) equal to a given target number T_R of Republican districts. In this case, the fairness constraint is simply:

$$\sum_{j=1}^{n} z_j = T_R. \tag{8′}$$

Particular choices for T_R include $\frac{R}{R+D}$ or its closest integer.

The equal-population constraints are formulated as:

$$\frac{P}{m}(1 - \delta) y_j \le \sum_{i=1}^{n} P_i x_{ij} \le \frac{P}{m}(1 + \delta) y_j, \quad j = 1, 2, \cdots, n. \tag{9}$$

The constraints in Eq. (9) ensure that population in any one district cannot deviate more than the tolerated percentage δ% from the target population P/m.

3.3. The competitiveness model

A district is said to be competitive if the percentage of Republican voters (or, equivalently, the percentage of Democratic voters) with

Table 1
(Scenario 1): Basic with Regular Distance (D_{ij}) in Objective Function and without Fairness and Competitiveness.

District 1: $y_{10} = 1$, Hub = Area 10			District 2: $y_{27} = 1$, Hub = Area 27			District 3: $y_{37} = 1$, Hub = Area 37		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
7	Beaufort	$x_{7\_10} = 1$	1	Abbeville	$x_{1\_27} = 1$	2	Aiken	$x_{2\_37} = 1$
8	Berkeley	$x_{8\_10} = 1$	4	Anderson	$x_{4\_27} = 1$	3	Allendale	$x_{3\_37} = 1$
10	Charleston1	$x_{10\_10} = 1$	26	Greenville2	$x_{26\_27} = 1$	5	Bamberg	$x_{5\_37} = 1$
11	Charleston2	$x_{11\_10} = 1$	27	Greenville3	$x_{27\_27} = 1$	6	Barnwell	$x_{6\_37} = 1$
12	Charleston3	$x_{12\_10} = 1$	28	Greenwood	$x_{28\_27} = 1$	20	Dorchester	$x_{20\_37} = 1$
17	Colleton	$x_{17\_10} = 1$	34	Laurens	$x_{34\_27} = 1$	21	Edgefield	$x_{21\_37} = 1$
29	Hampton	$x_{29\_10} = 1$	41	Oconee	$x_{41\_27} = 1$	36	Lexington1	$x_{36\_37} = 1$
31	Jasper	$x_{31\_10} = 1$				37	Lexington2	$x_{37\_37} = 1$
						40	Newberry	$x_{40\_37} = 1$
						42	Orangeburg	$x_{42\_37} = 1$
						45	Saluda	$x_{45\_37} = 1$
District 4: $y_{38} = 1$, Hub = Area 38			District 5: $y_{46} = 1$, Hub = Area 46			District 6: $y_{44} = 1$, Hub = Area 44		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
15	Chesterfield	$x_{15\_38} = 1$	13	Cherokee	$x_{13\_46} = 1$	9	Calhoun	$x_{9\_44} = 1$
18	Darlington	$x_{18\_38} = 1$	25	Greenville1	$x_{25\_46} = 1$	14	Chester	$x_{14\_44} = 1$
19	Dillon	$x_{19\_38} = 1$	46	Spartanburg1	$x_{46\_46} = 1$	16	Clarendon	$x_{16\_44} = 1$
23	Florence	$x_{23\_38} = 1$	47	Spartanburg2	$x_{47\_46} = 1$	22	Fairfield	$x_{22\_44} = 1$
24	Georgetown	$x_{24\_38} = 1$	49	Union	$x_{48\_46} = 1$	32	Kershaw	$x_{32\_44} = 1$
30	Horry	$x_{30\_38} = 1$	51	York	$x_{51\_46} = 1$	33	Lancaster	$x_{33\_44} = 1$
38	Marion	$x_{38\_38} = 1$				35	Lee	$x_{35\_44} = 1$
39	Marlboro	$x_{39\_38} = 1$				43	Richland1	$x_{43\_44} = 1$
50	Williamsburg	$x_{50\_38} = 1$				44	Richland2	$x_{44\_44} = 1$
						48	Sumter	$x_{48\_44} = 1$

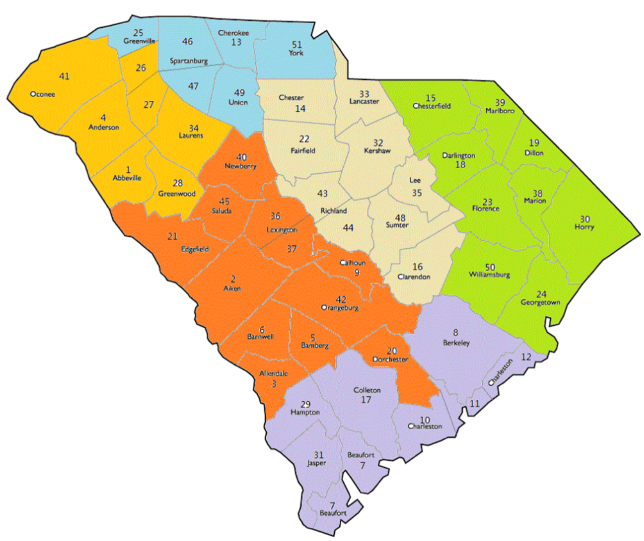

Fig. 1. (Scenario 1): Basic with Regular Distance (D_{ij})and without Fairness and Competitiveness.

respect to the total number of Republican or Democratic voters in a district is within a given range centered at 50%, e.g., a range of 5% below and 5% above 50% (i.e., between 45% and 55%). We acknowledge that the competitiveness may not always be achievable. For example, the competitiveness cannot be achieved for a state in which each of the unit areas is dominated by one of the two parties; or requiring competitiveness may make the "contiguity" and "compactness" unachievable. The formulation is specified as follows.

Additional Decision Variables:

Additional variables need to be defined in this problem to address this new requirement of competitiveness.

The "greater-than-lower-bound" indicator variables:

$u_j = 1$ if area j is a hub/district (i.e., if $y_j = 1$) and the relative percentage of Republican voters with respect to the total number of Republican and Democratic voters in the district is greater than or equal to the lower bound of the competitiveness range, i.e., $0.5 - \sigma$, where σ is the threshold parameter to express the required tightness of competition for it to be considered as competitive. The default value of σ is 0.05; $u_j = 0$ otherwise.

The "less-than-upper-bound" indicator variables:

$v_j = 1$ if area j is a hub/district (i.e., if $y_j = 1$) and the relative percentage of Republican voters with respect to the total number of Republican and Democratic voters in the district is less than or equal to the upper bound of the competitiveness range, i.e., $0.5 + \sigma$, where σ is the threshold parameter to express the required tightness of competition for it to be considered as competitive. The default value of σ is 0.05; $v_j = 0$ otherwise.

$w = \sum_{j=1}^{n} (u_j + v_j) - m$ is the total number of competitive districts. This is valid because if and only if the Republican percentage of a district is both greater than or equal to the lower bound and less than or equal to the upper bound, $u_j + v_j = 2$. Otherwise, $u_j + v_j = 1$.

Objective Function:

The objective function is identical to that of the Fairness Model.

$$\text{Min} \quad \sum_{i=1}^{n} \sum_{j=1}^{n} V_i D_{ij} x_{ij}$$

Constraints:

The constraints needed to define a measure of competitiveness and the total number of competitive districts are as follows:

$$u_j \le y_j, \quad \forall\, j = 1, 2, \cdots, n, \tag{10}$$

$$v_j \le y_j, \quad \forall\, j = 1, 2, \cdots, n, \tag{11}$$

Table 2
(Scenario 2): Basic; Squared Distance ($\boldsymbol{D}_{ij}^2$) in Objective Function and without Fairness and Competitiveness.

District 1: $y_{10} = 1$, Hub = Area 10			District 2: $y_{27} = 1$, Hub = Area 27			District 3: $y_{42} = 1$, Hub = Area 42		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
7	Beaufort	$x_{7\_10} = 1$	1	Abbeville	$x_{1\_27} = 1$	2	Aiken	$x_{2\_42} = 1$
8	Berkeley	$x_{8\_10} = 1$	4	Anderson	$x_{4\_27} = 1$	3	Allendale	$x_{3\_42} = 1$
10	Charleston1	$x_{10\_10} = 1$	26	Greenville2	$x_{26\_27} = 1$	5	Bamberg	$x_{5\_42} = 1$
11	Charleston2	$x_{11\_10} = 1$	27	Greenville3	$x_{27\_27} = 1$	6	Barnwell	$x_{6\_42} = 1$
12	Charleston3	$x_{12\_10} = 1$	28	Greenwood	$x_{28\_27} = 1$	9	Calhoun	$x_{9\_42} = 1$
17	Colleton	$x_{17\_10} = 1$	34	Laurens	$x_{34\_27} = 1$	20	Dorchester	$x_{20\_42} = 1$
31	Jasper	$x_{31\_10} = 1$	41	Oconee	$x_{41\_27} = 1$	21	Edgefield	$x_{21\_42} = 1$
						29	Hampton	$x_{29\_42} = 1$
						36	Lexington1	$x_{36\_42} = 1$
						37	Lexington2	$x_{37\_42} = 1$
						42	Orangeburg	$x_{42\_42} = 1$
						45	Saluda	$x_{45\_42} = 1$

District 4: $y_{38} = 1$, Hub = Area 38			District 5: $y_{46} = 1$, Hub = Area 46			District 6: $y_{44} = 1$, Hub = Area 44		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
15	Chesterfield	$x_{15\_38} = 1$	13	Cherokee	$x_{13\_46} = 1$	14	Chester	$x_{14\_44} = 1$
18	Darlington	$x_{18\_38} = 1$	25	Greenville1	$x_{25\_46} = 1$	16	Clarendon	$x_{16\_44} = 1$
19	Dillon	$x_{19\_38} = 1$	40	Newberry	$x_{40\_46} = 1$	22	Fairfield	$x_{22\_44} = 1$
23	Florence	$x_{23\_38} = 1$	46	Spartanburg1	$x_{46\_46} = 1$	32	Kershaw	$x_{32\_44} = 1$
24	Georgetown	$x_{24\_38} = 1$	47	Spartanburg2	$x_{47\_46} = 1$	33	Lancaster	$x_{33\_44} = 1$
30	Horry	$x_{30\_38} = 1$	49	Union	$x_{48\_46} = 1$	35	Lee	$x_{35\_44} = 1$
38	Marion	$x_{38\_38} = 1$	51	York	$x_{51\_46} = 1$	43	Richland1	$x_{43\_44} = 1$
39	Marlboro	$x_{39\_38} = 1$				44	Richland2	$x_{44\_44} = 1$
50	Williamsburg	$x_{50\_38} = 1$				48	Sumter	$x_{48\_44} = 1$

Fig. 2. (Scenario 2): Basic with Squared Distance (D_{ij}^2) and without Fairness and Competitiveness.

$$(u_j - 1)M \le \sum_{i=1}^{n} R_i x_{ij} - (0.5 - \sigma) \sum_{i=1}^{n} (R_i + D_i) x_{ij} \le M u_j, \quad \forall\, j = 1, 2, \cdots, n \tag{12}$$

$$(v_j - 1)M \le (0.5 + \sigma) \sum_{i=1}^{n} (R_i + D_i) x_{ij} - \sum_{i=1}^{n} R_i x_{ij} \le M v_j, \quad \forall\, j = 1, 2, \cdots, n \tag{13}$$

$$w = \sum_{j=1}^{n} (u_j + v_j) - m. \tag{14}$$

where M is a very big number. The constraints in Eqs. (10) and (11) ensure u_j and $v_j$can only be 1 if area j is a hub/district (i.e., if $y_j = 1$). We use the linear function $\sum_{i=1}^{n} R_i x_{ij} - (0.5 - \sigma) \sum_{i=1}^{n} (R_i + D_i) x_{ij}$ in Eq. (12) to represent the number of Republican voters in excess of $(0.5 - \sigma)$ times the combined number of Republican and Democratic voters. Similarly, we use the linear function $(0.5 + \sigma) \sum_{i=1}^{n} (R_i + D_i) x_{ij} - \sum_{i=1}^{n} R_i x_{ij}$ to represent the shortage of the Republican voters with respect to $(0.5 + \sigma)$ times the combined number. If j is a hub (i.e., if $y_j = 1$), and if the percentage of Republican voters with respect to the total Republican and Democratic voters in the corresponding district is greater than the lower bound of the competitiveness range 0.5 -σ, the second inequality in (12) forces u_j to be 1. Otherwise, the first inequality in (12) forces u_j to be 0. If j is a hub (i.e., if $y_j = 1$), and if the percentage of Republican voters with respect to the total number of Republican and Democratic voters in the corresponding district is less than or equal to the upper bound of the competitiveness range $0.5 + \sigma$, the second inequality in (13) forces v_j to be 1. Otherwise, the first inequality in (13) forces v_j to be 0. The total number of competitive district τ is thus defined in Eq. (14). With this definition, actual fairness constraints can be easily imposed. In particular, w can be required to be no less than a lower bound γ, with the following constraint:

$$w \ge \gamma \tag{15}$$

Another possibility is to require that w be equal to a given desired number.

4. Computational complexity

In this section, we first discuss the computational complexity of fair districting. We then prove that the feasibility problem associated with each of the two optimization problems is NP-complete and hence establish that the two optimization problems are NP-hard.

The computational complexity of fair redistricting problems has received some attention very recently but in the literature of theoretical computer science. Kueng, Mixon, and Villar (2019) studied the computational complexity of a fair redistricting problem that is related to the fairness model proposed in this paper. However, there are major

Table 3
(Scenario 3): Squared Distance ($\boldsymbol{D}_{ij}^2$) in Objective Function with Fairness.

District 1: $y_{24} = 1$, Hub = Area 24			District 2: $y_{27} = 1$, Hub = Area 27			District 3: $y_{37} = 1$, Hub = Area 37		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
8	Berkeley	$x_{8\_24} = 1$	1	Abbeville	$x_{1\_27} = 1$	2	Aiken	$x_{2\_37} = 1$
12	Charleston3	$x_{12\_24} = 1$	4	Anderson	$x_{4\_27} = 1$	9	Calhoun	$x_{9\_37} = 1$
23	Florence	$x_{23\_24} = 1$	26	Greenville2	$x_{26\_27} = 1$	20	Dorchester	$x_{20\_37} = 1$
24	Georgetown	$x_{7\_24} = 1$	27	Greenville3	$x_{27\_27} = 1$	21	Edgefield	$x_{21\_37} = 1$
30	Horry	$x_{30\_24} = 1$	28	Greenwood	$x_{28\_27} = 1$	36	Lexington1	$x_{36\_37} = 1$
38	Marion	$x_{38\_24} = 1$	34	Laurens	$x_{34\_27} = 1$	37	Lexington2	$x_{37\_37} = 1$
			41	Oconee	$x_{41\_27} = 1$	43	Richland1	$x_{40\_37} = 1$
						45	Saluda	$x_{45\_37} = 1$

District 4: $y_{46} = 1$, Hub = Area 46			District 5: $y_{17} = 1$, Hub = Area 17			District 6: $y_{32} = 1$, Hub = Area 32		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
13	Cherokee	$x_{13\_46} = 1$	3	Allendale	$x_{3\_17} = 1$	14	Chester	$x_{14\_32} = 1$
25	Greenville1	$x_{25\_46} = 1$	5	Bamberg	$x_{5\_17} = 1$	15	Chesterfield	$x_{15\_32} = 1$
40	Newberry	$x_{40\_46} = 1$	6	Barnwell	$x_{6\_17} = 1$	18	Darlington	$x_{18\_32} = 1$
46	Spartanburg1	$x_{46\_46} = 1$	7	Beaufort	$x_{7\_17} = 1$	19	Dillon	$x_{19\_32} = 1$
47	Spartanburg2	$x_{47\_46} = 1$	10	Charleston1	$x_{10\_17} = 1$	22	Fairfield	$x_{22\_32} = 1$
49	Union	$x_{49\_46} = 1$	11	Charleston2	$x_{11\_17} = 1$	32	Kershaw	$x_{32\_32} = 1$
51	York	$x_{51\_46} = 1$	16	Clarendon	$x_{16\_17} = 1$	33	Lancaster	$x_{33\_32} = 1$
			17	Colleton	$x_{17\_17} = 1$	35	Lee	$x_{35\_32} = 1$
			29	Hampton	$x_{29\_17} = 1$	39	Marlboro	$x_{39\_32} = 1$
			31	Jasper	$x_{31\_17} = 1$	44	Richland2	$x_{44\_32} = 1$
			42	Orangeburg	$x_{42\_17} = 1$	48	Sumter	$x_{48\_32} = 1$
			50	Williamsburg	$x_{50\_17} = 1$			

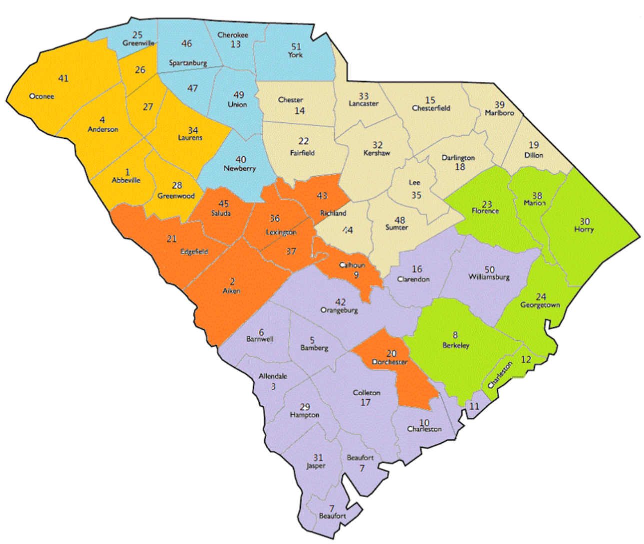

Fig. 3. (Scenario 3): Squared Distance (D_{ij}^2) with Fairness.

differences. First of all, they focused on proof of NP-completeness and did not discuss any solution algorithm. In developing districts, they considered individual voters and their physical locations; there is no concept of population or area unit, which is a fundamental construct in this paper. Their districts are formed as the convex hulls of the people assigned to the respective districts. As for district compactness, they impose a constraint on the maximum possible distance between any two persons in the same district. Their focus was on the feasibility problem (i.e., the decision or recognition problem), and hence no objective function was needed. Puppe and Tasnádi (2008) studied a similar problem with a different set of assumptions. Like the work of Kueng et al. (2019), they focused on proof of NP-completeness and did not discuss any solution algorithm. They also considered individual voters without the concept of area unit. A major difference is as follows. With n denoting the total number of voters and d denoting the total number of districts, they assumed, for simplicity, that there is a positive integer k such that $n = d(2k + 1)$. They considered a mathematical concept of a "geography" but defined it as a given family of subsets of $2k + 1$ individual voters such that the family contains at least d such subsets that partition the entire voter population.

In the rest of this section, we prove that the feasibility problem associated with each of the two optimization models is NP-complete. We first define our feasibility problems and then prove their NP-completeness.

Two variations of our fair districting problems have been defined in Section 3. The one with Eq. (8′) as the fairness constraint is more stringent and hence more difficult to solve, therefore we focus on the one with Eqs. (8.a) and (8.b) and prove its NP-completeness.

Definition 1. Given a set of input data of $n, m, \underline{P}, \underline{D}, \underline{U}, \mathbf{D}, \delta, \alpha, \beta_1$ and β_2, the feasibility problem of determining whether there exists a fair districting plan $\mathbf{X}$, $\underline{y}$ and $\underline{z}$ that satisfy Eqs. (2)–(9) is an instance of the fair districting problem.

Theorem 1. *The fair districting problem is NP-complete.*

Proof. Given a set of input data $\mathbf{X}$, $\underline{y}$ and $\underline{z}$, checking its feasibility requires verifying the $2n^2 + 4n + 3$ constraints of Eqs. (2)–(9). Therefore, the verification can be completed in polynomial time with respect to the size of the input, and the fair districting problem is NP. The uncapacitated facility location (UFL) problem has several variations; they have been shown to be NP-complete (Garey & Johnson, 1979; Karp, 1972). We next show that the UFL problem is a special case of the fair districting problem. Note that Eqs. (2), (4) and (5) constitute a UFL. They, unlike the rest of the constraints of this problem, involve only n and m and do not involve any other problem parameters. Eqs. (6) and (7) are "definitional" constraints and do not infringe on the feasibility of any given set of $\mathbf{X}$, $\underline{y}$ and $\underline{z}$. They are included to recognize and count the number of Republican districts. Eqs. (3), (8.a), (8.b) and (9) are real constraints. However, if α is set to a value α^*that is larger than the maximum distance between any two points of the state, then Eq. (3) does not pose any real constraint. In addition, if $\beta_1$is set to 0 (or any smaller number) and β_2 is set to m (or

Table 4
(Scenario 4): Cubed Distance ($\boldsymbol{D}_{ij}^3$) in Objective Function with Fairness and $\delta = 20\%$

District 1: $y_{10} = 1$, Hub = Area 10			District 2: $y_{27} = 1$, Hub = Area 27			District 3: $y_{38} = 1$, Hub = Area 38		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
7	Beaufort	$x_{7\_10} = 1$	1	Abbeville	$x_{1\_27} = 1$	18	Darlington	$x_{18\_38} = 1$
8	Berkeley	$x_{8\_10} = 1$	4	Anderson	$x_{4\_27} = 1$	19	Dillon	$x_{19\_38} = 1$
10	Charleston1	$x_{10\_10} = 1$	25	Greenville1	$x_{25\_27} = 1$	23	Florence	$x_{23\_38} = 1$
11	Charleston2	$x_{11\_10} = 1$	26	Greenville2	$x_{26\_27} = 1$	24	Georgetown	$x_{24\_38} = 1$
12	Charleston3	$x_{12\_10} = 1$	27	Greenville3	$x_{27\_27} = 1$	30	Horry	$x_{30\_38} = 1$
17	Colleton	$x_{17\_10} = 1$	28	Greenwood	$x_{28\_27} = 1$	38	Marion	$x_{38\_38} = 1$
20	Dorchester	$x_{20\_10} = 1$	34	Laurens	$x_{34\_27} = 1$	50	Williamsburg	$x_{50\_38} = 1$
31	Jasper	$x_{31\_10} = 1$	41	Oconee	$x_{41\_27} = 1$			
District 4: $y_{49} = 1$, Hub = Area 49			District 5: $y_{32} = 1$, Hub = Area 32			District 6: $y_{42} = 1$, Hub = Area 42		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
13	Cherokee	$x_{13\_49} = 1$	14	Chester	$x_{14\_32} = 1$	2	Aiken	$x_{2\_42} = 1$
36	Lexington1	$x_{36\_49} = 1$	15	Chesterfield	$x_{15\_32} = 1$	3	Allendale	$x_{3\_42} = 1$
37	Lexington2	$x_{37\_49} = 1$	22	Fairfield	$x_{22\_32} = 1$	5	Bamberg	$x_{5\_42} = 1$
40	Newberry	$x_{40\_49} = 1$	32	Kershaw	$x_{32\_32} = 1$	6	Barnwell	$x_{6\_42} = 1$
45	Saluda	$x_{45\_49} = 1$	33	Lancaster	$x_{33\_32} = 1$	9	Calhoun	$x_{9\_42} = 1$
46	Spartanburg1	$x_{46\_49} = 1$	35	Lee	$x_{35\_32} = 1$	16	Clarendon	$x_{16\_42} = 1$
47	Spartanburg2	$x_{47\_49} = 1$	39	Marlboro	$x_{39\_32} = 1$	21	Edgefield	$x_{21\_42} = 1$
49	Union	$x_{49\_49} = 1$	43	Richland1	$x_{43\_32} = 1$	29	Hampton	$x_{20\_42} = 1$
51	York	$x_{51\_49} = 1$	48	Sumter	$x_{48\_32} = 1$	42	Orangeburg	$x_{42\_42} = 1$
						44	Richland2	$x_{44\_42} = 1$

Table 5
(Scenario 5): Squared Distance ($\boldsymbol{D}_{ij}^2$) in Objective Function with Competitiveness but without Fairness.

District 1: $y_{10} = 1$, Hub = Area 10			District 2: $y_{27} = 1$, Hub = Area 27			District 3: $y_{42} = 1$, Hub = Area 42		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
7	Beaufort	$x_{7\_10} = 1$	1	Abbeville	$x_{1\_27} = 1$	2	Aiken	$x_{2\_42} = 1$
8	Berkeley	$x_{8\_10} = 1$	4	Anderson	$x_{4\_27} = 1$	3	Allendale	$x_{3\_42} = 1$
10	Charleston1	$x_{10\_10} = 1$	26	Greenville2	$x_{26\_27} = 1$	5	Bamberg	$x_{5\_42} = 1$
11	Charleston2	$x_{11\_10} = 1$	27	Greenville3	$x_{27\_27} = 1$	6	Barnwell	$x_{6\_42} = 1$
12	Charleston3	$x_{12\_10} = 1$	28	Greenwood	$x_{28\_27} = 1$	9	Calhoun	$x_{9\_42} = 1$
17	Colleton	$x_{17\_10} = 1$	34	Laurens	$x_{34\_27} = 1$	20	Dorchester	$x_{20\_42} = 1$
29	Hampton	$x_{29\_10} = 1$	41	Oconee	$x_{41\_27} = 1$	21	Edgefield	$x_{21\_42} = 1$
31	Jasper	$x_{31\_10} = 1$				36	Lexington1	$x_{36\_42} = 1$
						37	Lexington2	$x_{37\_42} = 1$
						42	Orangeburg	$x_{42\_42} = 1$
District 4: $y_{38} = 1$, Hub = Area 38			District 5: $y_{46} = 1$, Hub = Area 46			District 6: $y_{44} = 1$, Hub = Area 44		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
15	Chesterfield	$x_{15\_38} = 1$	13	Cherokee	$x_{13\_46} = 1$	14	Chester	$x_{14\_44} = 1$
18	Darlington	$x_{18\_38} = 1$	25	Greenville1	$x_{25\_46} = 1$	16	Clarendon	$x_{16\_44} = 1$
19	Dillon	$x_{19\_38} = 1$	40	Newberry	$x_{40\_46} = 1$	22	Fairfield	$x_{22\_44} = 1$
23	Florence	$x_{23\_38} = 1$	46	Spartanburg1	$x_{46\_46} = 1$	32	Kershaw	$x_{32\_44} = 1$
24	Georgetown	$x_{24\_38} = 1$	47	Spartanburg2	$x_{47\_46} = 1$	33	Lancaster	$x_{33\_44} = 1$
30	Horry	$x_{30\_38} = 1$	49	Union	$x_{48\_46} = 1$	35	Lee	$x_{35\_44} = 1$
38	Marion	$x_{38\_38} = 1$	51	York	$x_{51\_46} = 1$	43	Richland1	$x_{43\_44} = 1$
39	Marlboro	$x_{39\_38} = 1$				44	Richland2	$x_{44\_44} = 1$
50	Williamsburg	$x_{50\_38} = 1$				45	Saluda	$x_{45\_44} = 1$
						48	Sumter	$x_{48\_44} = 1$

any larger number), then Eqs. (8.a) and (8.b) do not pose any real constraint. Finally, if σ is set to $m-1$ (or any larger number), then Eq. (9) does not pose any real constraint. With these choices of parameter values, the fair districting problem is simply the UFL. Therefore, the UFL is a special case of the fair districting problem. The fact that UFL is NP-complete and the fact that the fair districting problem is NP imply that the fair districting problem is NP-complete as well.□

Definition 2. Given a set of input data of n, m, $\underline{P}$, $\underline{D}$, $\underline{U}$, $\mathbf{D}$, δ, α, γ, and σ, the feasibility problem of determining whether there exists a competitive districting plan $\mathbf{X}$, y, u, v and w that satisfy Eqs. (2)–(5) and (10)–(15) is an instance of the competitive districting problem.

Theorem 2. *The competitive districting problem is NP-complete.*

Proof. Given a set of input data $\mathbf{X}$, y, u, v and w, checking its feasibility

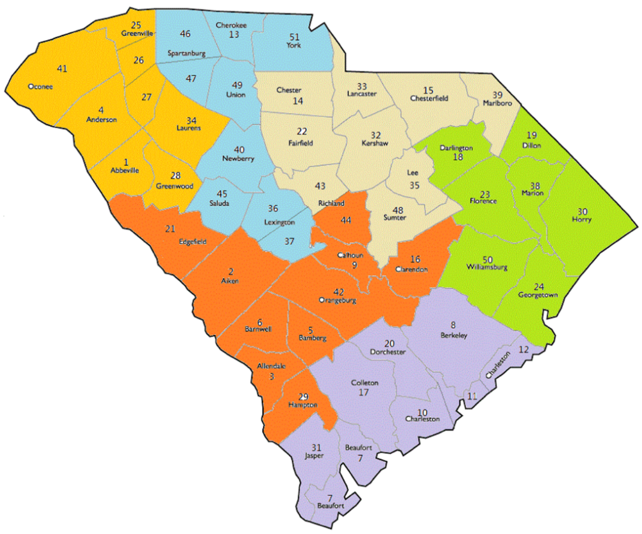

Fig. 4. (Scenario 4): Cubed Distance (D_{ij}^3) with Fairness and $\delta = 20\%$.

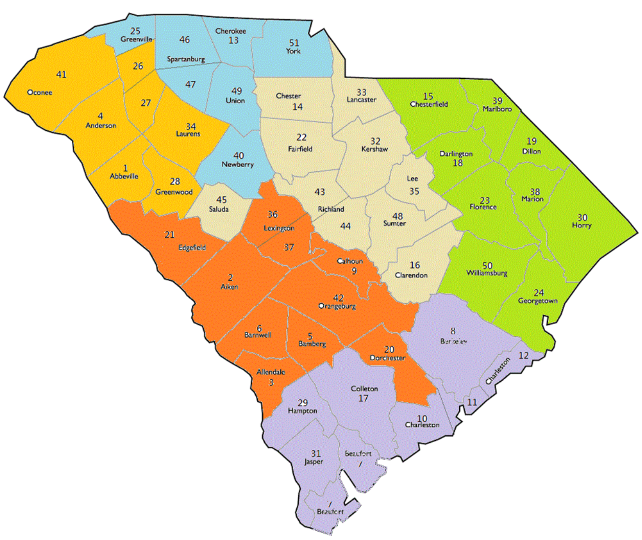

Fig. 5. (Scenario 5): Squared Distance (D_{ij}^2) with Competitiveness.

requires verifying the $2n^2 + 5n + 3$ constraints of Eqs. (2)–(9) and (10)–(15). Therefore, the verification can be completed in polynomial time with respect to the size of the input, and the competitive districting problem is NP. As in Theorem 1, we next show that the UFL problem is a special case of the competitive districting problem. There are several ways to make Eqs. (10)–(15) pose no real constraints on any given set of $\mathbf{X}$, y, u, v and w. Note that Eqs. (10)–(13) are "definitional" constraints used to recognize a competitive district and do not pose any real constraint. Eq. (14) is used to count the total number of competitive districts and does not pose any real constraint either. Setting γ to 0 in Eq. (15) results in the set of Eqs. (10)–(15) posing no real constraint all together; doing so effectively reduces the constraint set to Eqs. (2)–(5), which is equivalent to the UFL problem. With this choice of γ value, the competitive districting problem is simply the UFL. Therefore, the UFL is a special case of the competitive districting problem. The fact that UFL is NP-complete and the fact that the competitive districting problem is NP imply that the competitive districting problem is NP-complete as well.□

5. A case study: six scenarios

In this section, we use South Carolina (SC) to demonstrate the models. There are 46 counties (unit areas) in SC, and the State must be partitioned into six districts. Some populations of some counties are already larger than 1/6 of the State population and hence must be divided. We adopted the pre-processing done in (Mehrotra et al., 1998) for South Carolina. A goal of the preprocessing is to have the population of each unit area be between 2% and 25% of 1/6 of the state population. The preprocessing results in 51 unit areas. The 51 unit areas will be partitioned into 6 districts. We used the population data of the 2000 US Census (US Department of Commerce, 2019). We used the South Carolina result of the 2000 Presidential Election (Wikipedia, 2019) as data source for the he numbers of Democratic, Republican, and other voters. More precisely, voters casting their votes for the Democratic candidate Al Gore are considered Democratic voters; voters casting their votes for the Republican candidate George W. Bush are considered Republican voters. The 51 unit areas, their populations, the total numbers of people casting their votes, the numbers of Republican, Democratic, and other voters, and their geographic location are provided in Table A1 in the Appendix. For each of those counties that had a large population and were partitioned into smaller sub-county unit areas, we simply divide the total numbers by the number of sub-county areas. We also retain the identical geographical locations for simplicity. The percentage of the Republican voters is 56.5%, the percentage of the Democratic voters is 41.5% and the percentage of other voters is 2% as shown in Table A1 in the Appendix. The distance between each pair of the areas i and j is calculated with the formula $D_{ij} = \sqrt{(h_i - h_j)^2 + (v_i - v_j)^2}$, where (h_i, v_i) is the geographical center location of area i.

We illustrate applications of these two models with six scenarios. We study three sets of two scenarios each. The first set of two scenarios focuses on the three congressional mandates. The next two sets go beyond the three mandates and consider one more criterion as well. The second set considers fairness while the third set considers competitiveness.

For their software implementations, we wrote Python programs to create the corresponding BLP-specification and data files that can be read by IBM-CPLEX. For all the scenarios, we adopted all default CPLEX parameters for solving integer linear programs, including the MIP tolerance mipgap of 0.0001. The hardware used for the implementations is a Lenovo ThinkPad X1 Carbon 5th Signature Ed., with Intel Core i7-7600U at 2.8 GHz and with a 16-GB RAM. The results of these scenarios and the run time of each scenario are summarized in Table 7.

Scenario 1: We first solve the problem without considering the fairness and competitiveness constraints. The model is reduced to simply satisfying the three basic requirements: proportional equality (i.e., approximately equal population size for all districts), spatial contiguity and compactness. The proportional equality requirement is enforced through constraint (9) with the default percentage deviation from the target population $\delta = 5\%$. The spatial contiguity and compactness requirements are enforced through the objective function (1) to minimize the weighted sum of distances from a hub to the area units it serves, weighted by the amount of demand (number of voters). The spatial contiguity and compactness requirements can further be enforced through the threshold constraint (4) through a specified α value. A large α value will basically disable the constraint, while a feasible solution may not be obtained with a small α value. For this case study, we use a large α value (i.e. 100) so that the spatial contiguity and compactness requirements are achieved purely through minimizing the objective function. The solution to scenario 1 is shown in Table 1 and the geographic district map is illustrated in Fig. 1. The objective value is 3.1635112200e+07. The computation took 0.39 s. The map shows the

Table 6
(Scenario 6): Cubed Distance ($\boldsymbol{D}_{ij}^{3}$) in Objective Function with Competitiveness but without Fairness.

District 1: $y_{10} = 1$, Hub = Area 10			District 2: $y_{27} = 1$, Hub = Area 27			District 3: $y_{42} = 1$, Hub = Area 42		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
7	Beaufort	$x_{7\_10} = 1$	1	Abbeville	$x_{1\_27} = 1$	2	Aiken	$x_{2\_42} = 1$
8	Berkeley	$x_{8\_10} = 1$	4	Anderson	$x_{4\_27} = 1$	3	Allendale	$x_{3\_42} = 1$
10	Charleston1	$x_{10\_10} = 1$	25	Greenville1	$x_{25\_27} = 1$	5	Bamberg	$x_{5\_42} = 1$
11	Charleston2	$x_{11\_10} = 1$	26	Greenville2	$x_{26\_27} = 1$	6	Barnwell	$x_{6\_42} = 1$
12	Charleston3	$x_{12\_10} = 1$	27	Greenville3	$x_{27\_27} = 1$	9	Calhoun	$x_{9\_42} = 1$
17	Colleton	$x_{17\_10} = 1$	41	Oconee	$x_{41\_27} = 1$	20	Dorchester	$x_{20\_42} = 1$
29	Hampton	$x_{29\_10} = 1$				21	Edgefield	$x_{21\_42} = 1$
31	Jasper	$x_{31\_10} = 1$				36	Lexington1	$x_{36\_42} = 1$
						37	Lexington2	$x_{37\_42} = 1$
						42	Orangeburg	$x_{42\_42} = 1$
						45	Saluda	$x_{45\_42} = 1$
District 4: $y_{38} = 1$, Hub = Area 38			District 5: $y_{49} = 1$, Hub = Area 49			District 6: $y_{32} = 1$, Hub = Area 32		
Area	County	x_{ij}	Area	County	x_{ij}	Area	County	x_{ij}
16	Clarendon	$x_{16\_38} = 1$	13	Cherokee	$x_{13\_49} = 1$	14	Chester	$x_{14\_32} = 1$
18	Darlington	$x_{18\_38} = 1$	28	Greenwood	$x_{28\_49} = 1$	15	Chesterfield	$x_{15\_32} = 1$
19	Dillon	$x_{19\_38} = 1$	34	Laurens	$x_{34\_49} = 1$	22	Fairfield	$x_{22\_32} = 1$
23	Florence	$x_{23\_38} = 1$	40	Newberry	$x_{40\_49} = 1$	32	Kershaw	$x_{32\_32} = 1$
24	Georgetown	$x_{24\_38} = 1$	46	Spartanburg1	$x_{46\_49} = 1$	33	Lancaster	$x_{33\_32} = 1$
30	Horry	$x_{30\_38} = 1$	47	Spartanburg2	$x_{47\_49} = 1$	43	Richland1	$x_{43\_32} = 1$
35	Lee	$x_{35\_38} = 1$	49	Union	$x_{48\_49} = 1$	44	Richland2	$x_{44\_32} = 1$
38	Marion	$x_{38\_38} = 1$	51	York	$x_{51\_49} = 1$	48	Sumter	$x_{48\_32} = 1$
39	Marlboro	$x_{39\_38} = 1$						
50	Williamsburg	$x_{50\_38} = 1$						

contiguity and compactness of all six districts.

The detailed population composition data of each district is summarized in Table 5 in which one can find the total population P_i, the number of Republican voters R_i, the number of Democratic voters D_i, the percentage deviation of the population from the target population P/m and the ratio of the Republican voters to the total number of Republican and Democratic voters for each district.

In scenario 1, only the proportional equality requirement constraints are enforced, i.e., the population size in any one district cannot deviate more than the tolerated 5% from the target population P/m. We verified from Table 7 that all the six districts satisfy the requirement. The fairness constraints are not enforced and the competitiveness is not addressed in this scenario. Out of the six districts, five (districts 1, 2, 3, 4 and 5 as shown in Table 7) are Republican districts, i.e. the number of the Republican voters is larger than that of the Democratic voters in each of the five districts; one (district 6) is a Democratic district. There are three competitive districts, i.e. the percentage of Republican voters (or, equivalently, the percentage of Democratic voters) with respect to the total number of Republican and Democratic voters in a district is in the range of [0.45, 0.55] with $\sigma = 0.05$, in scenario 1.

Scenario 2: In this scenario, we raise the importance of the distance by using a squared distance, i.e., D_{ij}^{2}, in the objective function to demonstrate the ease of ensuring a higher degree of continuity and compactness. (This is the only difference from Scenario 1.) The solution to scenario 2 is shown in Table 2 and the geographic district map is illustrated in Fig. 2. The objective value is 1.0873385059e+09. The districting results from the scenarios 1 and 2 are very similar for South Carolina as shown in Figs. 1 and 2. The map of scenario 2 shows the contiguity and compactness of all six districts, not surprisingly.

The detailed population composition data of each district is summarized in Table 7 together with the other scenarios. We verified that all the six districts satisfy the population-equality requirement. Out of the six districts, five (districts 1, 2, 3, 4 and 5 as shown in Table 7) are Republican districts, i.e., the number of the Republican voters being larger than that of the Democratic voters in that district; one (district 6) is a Democratic district. There are two competitive districts in scenario 2. The computation took 0.67 s.

We use Scenario 2 as a basis to study of the impact of fairness requirements and that of competitiveness requirements.

Scenario 3: Starting from Scenario 2, the fairness constraints (6), (7) and (8′) are enforced in this scenario with M being a very large number 99999. The fairness requirement is that the target number T_R of Republican district be equal to the integer closest to the proportion $\frac{R}{R+D}$ of the Republican voters among voters of the two parties. The percentage of the Republican voters is 56.5% and the percentage of the Democratic voters is 41.5% in South Carolina. A fair number of Republican districts is calculated as $m \times 56.5\%/(56.5\% + 41.5)$. With proper rounding, it can be set to four, resulting in two as a fair number of Democratic districts. The solution to scenario 3 is shown in Table 3 and the geographic district map is illustrated in Fig. 3.

The detailed population composition data of each district is summarized in Table 7. We can verify that all the six districts satisfy the population-equality requirement. Out of the six districts, four (districts 1, 2, 3 and 4 as shown in Table 7) are Republican districts and two (districts 5 and 6) are Democratic districts, as expected. There are 2 competitive districts in scenario 3. The objective value is 1.5518769245e+09, which is higher than the solution in Scenario 2. The continuity and compactness is compromised to achieve the required fairness. Further raising the power of distance to three would not produce fair districts that are also contiguous. This is indicative of the difficulty for the current data to satisfy all the constraints. The computation took 50.5 s. In the next scenario, we use the same data to

Table 7
Result Summary.

		PopulationP_i	Rep.R_i	Dem.D_i	Equal Population	Fairness		Competitiveness	
					|% Dev.| From Target	Rep.	Dem.	$R_i/(R_i+D_i)$	Competitive
Scenario 1 (Basic) Run time: 0.39 sec.	Dist. 1	653,345	120,565	99,705	0.75%	Yes		0.547351	Yes
	Dist. 2	647037.3	141745.3	76208.67	0.22%	Yes		0.650345	No
	Dist. 3	677,792	148,574	91,719	4.52%	Yes		0.618303	No
	Dist. 4	620,622	107,954	96,150	4.29%	Yes		0.528917	Yes
	Dist. 5	627361.7	132462.7	74213.33	3.26%	Yes		0.64092	No
	Dist. 6	664,649	108,740	117,220	2.50%		Yes	0.481236	Yes
Scenario 2 (Basic; D_{ij}^2) Run time: 0.67 sec.	Dist. 1	631,959	117,767	94,809	2.55%	Yes		0.554	No
	Dist. 2	647037.3	141745.3	76208.67	0.22%	Yes		0.650345	No
	Dist. 3	678,255	147,096	95,250	4.59%	Yes		0.606967	No
	Dist. 4	620,622	107,954	96,150	4.29%	Yes		0.528917	Yes
	Dist. 5	663469.7	139954.7	78641.33	2.31%	Yes		0.640244	No
	Dist. 6	649,464	105,524	114,157	0.15%		Yes	0.480351	Yes
Scenario 3 (D_{ij}^2; Fair) Run time: 50.5 sec	Dist. 1	659,627	123405.7	97286.67	1.72%	Yes		0.559175	No
	Dist. 2	647037.3	141745.3	76208.67	0.22%	Yes		0.650345	No
	Dist. 3	674278.5	149,188	92691.5	3.98%	Yes		0.616786	No
	Dist. 4	663469.7	139954.7	78641.33	2.31%	Yes		0.640244	No
	Dist. 5	620,019	106261.3	107465.3	4.39%		Yes	0.497183	Yes
	Dist. 6	626375.5	99,486	102922.5	3.41%		Yes	0.491511	Yes
Scenario 4 (D_{ij}^3; Fair; $\delta = 20\%$) Run time: 13.34 sec	Dist. 1	728,372	138,501	106,977	12%	Yes		0.564209	No
	Dist. 2	773,576	172,650	90,812	19%	Yes		0.655313	No
	Dist. 3	549,036	98,989	84,979	15%	Yes		0.538077	Yes
	Dist. 4	772,126	171,243	89,550	19%	Yes		0.656624	No
	Dist. 5	528209.5	84,221	87739.5	18%		Yes	0.489769	Yes
	Dist. 6	539487.5	94,437	95158.5	16%		Yes	0.498097	Yes
Scenario 5 (D_{ij}^2; Comp.) Run time: 1.03 sec	Dist. 1	653,345	120,565	99,705	0.75%	Yes		0.547351	Yes
	Dist. 2	647037.3	141745.3	76208.67	0.22%	Yes		0.650345	No
	Dist. 3	637,688	140,200	87,672	1.66%	Yes		0.615258	No
	Dist. 4	620,622	107,954	96,150	4.29%	Yes		0.528917	Yes
	Dist. 5	663469.7	139954.7	78641.33	2.31%	Yes		0.640243	No
	Dist. 6	668,645	109,622	116,839	3.11%		Yes	0.484066	Yes
Scenario 6 (D_{ij}^3; Comp.) Run time: 0.76 sec.	Dist. 1	653,345	120,565	99,705	0.75%	Yes		0.547351	Yes
	Dist. 2	637,738	148,355	74,753	1.65%	Yes		0.664947	No
	Dist. 3	656,869	144,298	90,354	1.30%	Yes		0.614945	No
	Dist. 4	630,475	109,549	99,937	2.78%	Yes		0.522942	Yes
	Dist. 5	672,769	133,345	80,097	3.75%	Yes		0.624736	No
	Dist. 6	639,611	103,929	110,370	1.37%		Yes	0.484972	Yes

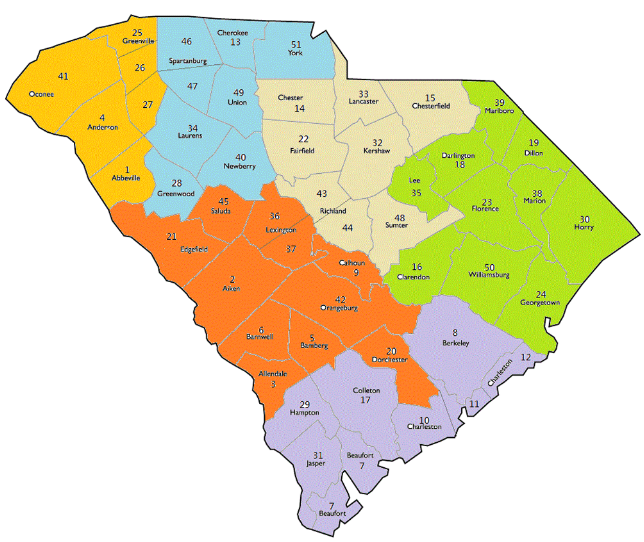

Fig. 6. (Scenario 6): Cubed Distance (D_{ij}^3) with Competitiveness.

achieve a fair solution but with a different set of parameters.

Scenario 4: In this scenario (and only in this scenario), we relax the population-equality constraints by a large deviation tolerance δ of 20% to demonstrate that the fairness model works. Staff users of a re-districting commission can refine the unit areas, with real-world boundaries of counties and townships and their population distributions taken into consideration, so that the resulting data allow fair districts that satisfy the three Constitutional mandates. To illustrate the flexibility of the models, we raise the power of the distances in the objective function to 3. Instead of requiring 4 Republican districts and 2 Democratic districts, we require that, out of the $m = 6$ districts, the number of Republican districts be greater than or equal to $[m\times56.5\%/(56.5\% + 41.5\%)] = 3$, and the number of Democratic districts be greater than or equal to $[m\times41.5\%/(56.5\% + 41.5\%)] = 2$. This is an implementation of Eqs. (8.a) and (8.b). The solution to scenario 4 is shown in Table 4 and the geographic district map is illustrated in Fig. 4. We were able to achieve contiguity and compactness by increasing the tolerance of population-equality constraints and the importance of the distance in the objective function as shown in the map.

The detailed population composition data of each district is summarized in Table 7. We can verify that all the six districts satisfy the population-equality requirement with $\delta = 20\%$. Out of the six districts, four (districts 1, 2, 3 and 4 as shown in Table 7) are Republican districts and two (districts 5 and 6) are Democratic districts. There are 3 competitive districts in scenario 4. The objective value is 4.9153041448e +10. The computation took 13.34 s.

Scenario 5: In this scenario, we consider competitiveness constraints (10), (11), (12), (13), (14) and (15) beyond the three constitutional mandates, without adding the fairness constraints. We know from the base case (Scenario 2), at least 2 competitive districts can be obtained without breaking the contiguity and compactness. In this scenario, we address the competitiveness by specifying the number of competitive districts (w) we want to achieve. We require $w = 3$. We keep the power of the distances at 2, as in scenario 2. The solution of this scenario can be found in Tables 5 and 7. The geographic district map is illustrated in Fig. 5. The objective value is 1.0891581466e + 09. The computation took 1.03 s. In this scenario, one area (area 45) is not connected to its district.

Scenario 6: In this scenario, we raise the importance of the distance by using a cubed distance i.e. D_{ij}^3, in the objective function to ensure better continuity and compactness. (This is the only difference from scenario 5.) The solution of this scenario can be found in Tables 6 and 7. The geographic district map is illustrated in Fig. 6. The objective value is 3.9985403572e + 10. The computation took 0.76 s. The continuity and compactness is achieved with 3 competitive districts in this scenario.

In the rest of this section, we analyze our numerical results. The number of binary variables is on the order of n^2, where n is the number of area units. The run times for all six scenarios of the case study are short. Four of the six scenarios were solved with approximately one second or shorter; they pertain to the basic model or the basic model plus competitiveness. The two other scenarios took much longer but less than one minute; both of them pertain to the basic model plus fairness. More extensive numerical experiments on more data sets are needed to help predict the maximum number of area units that allows numerical solutions of these models in a reasonable amount of computational time.

The primary objective of this paper is to develop a practical method conducive for development of a software tool for use by the staff of redistricting commissions. When the number of area units of a state becomes too large and the two optimization problems associated with the two models become computationally intractable, the staff can use their judgment to partition the states into regions and hence partition the problems into sub-problems. Such partitioning can be based on state geography, patterns of population concentration, and other insights, with a goal to arrive at solutions to these sub-problems whose aggregation would approximate well the solution to original problem.

6. Conclusion

Gerrymandering has been practiced by partisans for decades and has resulted in political strongholds, where one party dominates in an election district so much so that no parties would waste effort in campaigning. It has also been practiced by the incumbents so much so that, despite the low approval ratings of even less than 10% of the US Congress, high percentages of even 90% of the incumbents are elected. Election is a political process, so is redistricting. The vast majority of the past efforts on use of mathematical programming for political districting does not consider the current political realities of polarization and does not consider future political goals either. In this paper, we explicit consider fairness and competitiveness in our mathematical-programming formulation.

We also recognize that political redistricting is conducted by restricting-commission staff, who most likely are not experts on mathematical programming and do need to have software tools to support their work. The hubbing approach to achieving contiguity and compactness is a straightforward concept, and the parameters of our models are easy to understand as well. When preprocessing is needed, the goals to achieve are clear. The primary goal is to partition of an area unit with a large population to achieve an area-unit population size of 2–25% of the total population of the State or region. Another goal is to achieve balance for other political considerations. Although the commissioners, who may be partisans or incumbents, will make the final decisions, the tools like the one we built enable the commission staff to explore options not possible before. Perhaps more importantly, the simple concepts proposed in this paper enable transparency of the redistricting process, and the voters can more directly influence the commission's decisions.

Our numerical experiments demonstrate the effectiveness and the efficiency of our models, in terms of both the mathematical solutions and real-world implementations by election-commission staff.

Further political considerations can be explored, and more numerical implementations can be studied. When the number of variables, particularly binary variables, becomes too large, the NP-completeness of this problem may lead to excessive run time. Different techniques can be used to reduce it. Partitioning a state based on real-world considerations like geography and population distribution can be an effective solution, in terms of not only reducing the run time but also commission-staff's ability to produce transparent re-districting plans that can be easily understood and supported by the general public.

The two formulations proposed in this paper can be extended or modified to address fairness and competitiveness issues in some political systems with three or more political parties. A key to addressing fairness is to represent the number of districts in which the voters for a particular party outnumber those of any other parties. In the current formulation, z_j, $j = 1, 2, \ldots, n$, is the indicator variable to represent whether area unit j is the hub of a district and, if so, the Republican voters outnumber those of the Democratic Party for jth district. For elections or nations where L political parties jockey for the one political seat of a district, the corresponding number of L such indicator variables are needed. As for constraints, $\binom{L}{2}$, i.e., L Choose 2, inequality constraints analogous to Eq. (6) are needed; each of such constraint helps determine if the voters of a political party outnumber those of one other party. In addition, L inequality constraints analogous to Eq. (7) are needed; each such constraint ensures that area unit j is indeed selected as the hub of a district. With these constraints, the number of districts in which voters of each party outnumber those of all other parties can be obtained, and fairness constraints can be imposed according to what the user of the software tool as fair.

Competitiveness in a political system with three or more parties is a more complex issue. One way to define competitiveness in a district is to ensure that the leading party does not have dominance over all the other parties combined, so that the other parties can form an alliance to compete with the leading party. With such a definition, the competitiveness issue is essentially the same as the one addressed in this paper because competition is between one party and one "composite party." In situations where there are two competing alliances, the competitiveness issue remains essentially the same. In both cases, the current formulation can be readily applied. In situations where there are three major parties or where there are at least two major political parties and the total voter count for all the other parties combined is sufficiently high that an alliance can be formed to compete with the two major parties, a new definition of competitiveness and a companion set of measures of competitiveness must be developed first. These are interesting and worthy subjects of future research, some of which are being pursued by the authors. Districting is an important issue not just in government affairs but also in business and industry. The models developed in this paper can be applied to determine school districts, police districts, geo-marketing districts, electrical-power districts, etc. where the population or users can be grouped in two or more categories.

CRediT authorship contribution statement

Hongrui Liu: Methodology, Writing - original draft, Software, Formal analysis, Validation, Visualization, Writing - review & editing. **Ayca Erdogan:** Methodology, Writing - original draft, Validation, Writing - review & editing. **Royce Lin:** Software, Data curation, Validation, Visualization. **Jacob Tsao:** Conceptualization, Methodology, Writing - original draft, Validation, Writing - review & editing, Project administration.

Appendix

Table A1
South Carolina 2000 census data.

Area	County	Population P_i	Voters V_i	Rep. R_i	Demo. D_i	Other U_i	Latitude h_i	Longitude v_i
1	Abbeville	26,167	8374	4450	3766	158	34.1891	82.4753
2	Aiken	142,552	50,782	33,203	16,409	1170	33.6006	81.6035
3	Allendale	11,211	3340	967	2338	35	33.006	81.3839
4	Anderson	165,740	56,681	35,827	19,606	1248	34.3052	82.3856
5	Bamberg	16,658	5551	2047	3451	53	33.2074	81.0755
6	Barnwell	23,478	8276	4521	3661	94	33.2799	81.4718
7	Beaufort	120,397	44,148	25,561	17,487	1100	32.35	80.69
8	Berkeley	142,651	43,316	24,796	17,707	813	33.1261	80.0088
9	Calhoun	15,185	6373	3216	3063	94	33.6739	80.7658
10	Charleston1	103,323	37,159	19,410	16,507	1242	32.7365	80.1875
11	Charleston2	103,323	37,159	19,410	16,507	1242	32.7765	79.9311
12	Charleston3	103,323	37,159	19,410	16,507	1242	32.8165	79.6951
13	Cherokee	52,537	16,323	9900	6138	285	35.012	81.6035
14	Chester	34,068	10,432	4986	5242	204	34.6708	81.1196
15	Chesterfield	42,768	12,526	6266	6111	149	34.6831	80.1429
16	Clarendon	32,502	11,290	5186	5999	105	33.6734	80.1875
17	Colleton	38,264	13,390	6767	6449	174	32.8242	80.7214
18	Darlington	67,394	21,863	11,290	10,253	320	34.365	80.0088
19	Dillon	30,772	8990	3975	4930	85	34.3596	79.4254
20	Dorchester	96,413	33,641	20,734	12,168	739	33.168	80.5438
21	Edgefield	24,595	8834	4760	3950	124	33.8016	81.9535
22	Fairfield	23,454	8398	3011	5263	124	34.4478	81.0755
23	Florence	125,761	41,437	23,678	17,157	602	33.9829	79.6502
24	Georgetown	55,797	20,351	10,535	9445	371	33.4213	79.2902
25	Greenville1	126,539	46,764	30,905	14,603	1256	35.0444	82.38556
26	Greenville2	126,539	46,764	30,905	14,603	1256	34.8444	82.38556
27	Greenville3	126,539	46,764	30,905	14,603	1256	34.6444	82.38556
28	Greenwood	66,271	20,860	12,193	8139	528	34.1673	82.1278
29	Hampton	21,386	7759	2798	4896	65	32.8085	81.1196
30	Horry	196,629	71,265	40,300	29,113	1852	33.9197	78.9288
31	Jasper	20,678	6469	2414	3646	409	32.4914	81.0755
32	Kershaw	52,647	19,677	11,911	7428	338	34.3672	80.5883
33	Lancaster	61,351	20,705	11,676	8782	247	34.7253	80.6771
34	Laurens	69,567	20,410	12,102	7920	388	34.5203	82.1278
35	Lee	20,119	6642	2675	3899	68	34.163	80.2321
36	Lexington1	108,007	41,541	29,048	11,415	1078	33.991	81.3019
37	Lexington2	108,007	41,541	29,048	11,415	1078	33.871	81.2019
38	Marion	35,466	12,149	4687	7358	104	34.088	79.3353
39	Marlboro	28,818	7883	2699	5060	124	34.5582	79.6951
40	Newberry	36,108	12,372	7492	4428	452	34.3091	81.6035
41	Oconee	66,215	23,575	15,364	7571	640	34.7492	82.9932
42	Orangeburg	91,582	32,734	12,657	19,802	275	33.5635	81.0755
43	Richland1	160,339	58,241	25,082	31,590	1569	34.0612	81.0029
44	Richland2	160,339	58,241	25,082	31,590	1569	34.0212	80.8829
45	Saluda	19,181	6891	4098	2682	111	34.0087	81.7349
46	Spartanburg1	126,896	41,777	26,057	14,780	940	34.8806	81.9535
47	Spartanburg2	126,896	41,777	26,057	14,780	940	34.8406	81.9535
48	Sumter	104,646	30,672	15,915	14,365	392	33.9123	80.4104
49	Union	29,881	10,589	5768	4662	159	34.6613	81.6035
50	Williamsburg	37,217	11,331	4524	6723	84	33.6294	79.6502
51	York	164,614	54,351	33,776	19,251	1324	34.9871	81.2519
Total		3,890,807	1,345,534	760,041	555,216	30,277		
Percentage				0.56486198	0.41263617	0.022502		

References

Bacao, F., Lobo, V., & Painho, M. (2005). Applying genetic algorithms to zone design. *Soft Computing, 9*(5), 341–348.

Bodin, L. D. (1973). A districting experiment with a clustering algorithm. *Annals of the New York Academy of Sciences, 219*(1), 209–214.

Bozkaya, B., Erkut, E., & Laporte, G. (2003). A tabu search heuristic and adaptive memory procedure for political districting. *European Journal of Operational Research, 144*(1), 12–26.

Forest, B. (2005). The changing demographic, legal, and technological contexts of political representation. *Proceedings of the National Academy of Sciences, 102*(43), 15331–15336.

Forman, S. L., & Yue, Y. (2003). Congressional districting using a TSP-based genetic algorithm. Paper presented at the Genetic and Evolutionary Computation Conference, 2072–2083.

Garey, M. R., & Johnson, D. S. (1979). *Computers and intractability: A guide to the theory of NP-completeness.* San Francisco: W.H Freeman and Company.

Garfinkel, R. S., & Nemhauser, G. L. (1970). Optimal political districting by implicit enumeration techniques. *Management Science, 16*(8), 508.

George, J. A., Lamar, B. W., & Wallace, C. A. (1997). Political district determination using large-scale network optimization. *Socio-Economic Planning Sciences, 31*(1), 11–28.

Hess, S. W., Weaver, J. B., Siegfeldt, H. J., Whelan, J. N., & Zitlau, P. A. (1965). Nonpartisan political redistricting by computer. *Operations Research, 13*(6), 998–1006.

Hojati, M. (1996). Optimal political districting. *Computers & Operations Research, 23*(12), 1147–1161.

Kaiser, H. F. (1966). An objective method for establishing legislative districts. *Midwest Journal of Political Science, 10*(2), 200–213.

Kalcsics, J., Nickel, S., & Schröder, M. (2005). Towards a unified territorial design approach—Applications, algorithms and GIS integration. *Top, 13*(1), 1–56.

Karp, R. (1972). Reducibility among combinatorial problems. In R. E. Miller, J. W. Thatcher, & J. D. Bohlinger (Eds.). *Complexity of Computer Computations* (pp. 85–103). New York: Plenum.

King, D. M., Jacobson, S. H., Sewell, E. C., & Cho, W. K. T. (2012). Geo-graphs: An efficient model for enforcing contiguity and hole constraints in planar graph partitioning. *Operations Research, 60*(5), 1213–1228.

Kueng, R., Mixon, D. G., & Villar, S. (2019b). Fair redistricting is hard. *Theoretical Computer Science, 791*, 28–35.

Mehrotra, A., Johnson, E. L., & Nemhauser, G. L. (1998). An optimization based heuristic for political districting. *Management Science, 44*(8), 1100–1114.

Morrill, R. L. (1973). Ideal and reality in reapportionment. *Annals of the Association of American Geographers, 63*(4), 463–477.

Morrill, R. L. (1976). Redistricting revisited. *Annals of the Association of American Geographers, 66*(4), 548–553.

Nagel, S. S. (1964). Simplified bipartisan computer redistricting. *Stanford Law Review, 17*, 863.

Nagel, S. S. (1972). Computers & the law & politics of redistricting. *Polity, 5*(1), 77–93.

Nygreen, B. (1988). European assembly constituencies for wales-comparing of methods for solving a political districting problem. *Mathematical Programming, 42*(1–3), 159–169.

Puppe, C., & Tasnádi, A. (2008). A computational approach to unbiased districting. *Mathematical and Computer Modelling, 48*, 1455–1460.

Ricca, F., Scozzari, A., & Simeone, B. (2013). Political districting: From classical models to recent approaches. *Annals of Operations Research, 204*(1), 271–299.

Ricca, F., & Simeone, B. (2008). Local search algorithms for political districting. *European Journal of Operational Research, 189*(3), 1409–1426.

Thoreson, J. D., & Liittschwager, J. M. (1967). Computers in behavioral science. Legislative districting by computer simulation. *Behavioral Science, 12*(3), 237–247.

US Department of Commerce (2019). US census 2000 resident population, https://www.census.gov/population/www/cen2000/maps/respop.html, accessed on July 12, 2019.

Vickrey, W. (1961). On the prevention of gerrymandering. *Political Science Quarterly, 76*(1), 105–110.

Wikipedia (2019). 2000 United States Presidential Election in South Carolina, https://en.wikipedia.org/wiki/2000_United_States_presidential_election_in_South_Carolina, accessed on July 12, 2019.