A School Redistricting Integer Programming Model for Achieving Connected School Attendance Zones

Joshua C. Detwiler, Layne T. Watson, Naren Ramakrishnan, and David S. Bieri

Virginia Polytechnic Institute and State University, Blacksburg VA 24061, USA,
jdetwiler@vt.edu

Abstract. School redistricting is an optimization problem of geographically partitioning students to public schools. A school board chooses the set of criteria used to compare and evaluate redistricting plans, which often include measures and constraints such as compactness of school attendance zones (SAZs), proximity of students to their schools, connectivity of the SAZs, and so on. Connectivity has been heavily explored in redistricting literature. This work presents an integer nonlinear programming model that expresses SAZ connectivity exactly using the Laplacian matrices of SAZs, as well as another (possibly more computationally tractable) indirect formulation that lazily adds cutting planes.

Keywords: school redistricting, integer programming, graph connectivity, graph Laplacian matrix

1 Introduction

School redistricting is a continuing process of geographically partitioning students to public schools. Often at the city or county level in the United States, a school board is responsible for deciding how to redistrict. The process is invoked for a number of reasons including the building of a new school, the decommissioning of an old school, local population growth or decline, changes in schools' available resources, attempts at improving equality and diversity, etc. [6]. Thus, there are many natural reasons to change the schools that students attend.

Changing students' schools is a highly political process. Statistically, parents are willing to pay more in housing costs in return for better test scores [5]. This reflects general perceptions that some schools are better than others, and it is a major factor in how parents shop for housing. Children too, of course, are stakeholders in the redistricting process. It is often an objective of school planners to minimize student displacement in new school district plans. Administratively, the school board can handle this by only enacting the new plans as students transition to the next school level.

Districting is commonly referred to as a partitioning problem, a clustering problem, and an assignment problem. In school districting, school planning areas

(SPAs) are assigned to school attendance zones (SAZs). SPAs are often neighborhoods or smaller geographic units that a school board has decided on. The SPAs assigned to an SAZ make up the population of students that will attend a school. There are assignments for each school level of elementary, middle, and high, which makes school districting separable respectively to each of the school levels.

Fair redistricting is then a problem of "optimally" choosing such an assignment according to an agreed upon set of criteria. Common criteria in school redistricting include compactness, student proximity, staying within schools' student capacities, extending this capacity constraint to future projected enrollments, and ensuring that SAZs are geographically connected (contiguous). Additionally, school planners increasingly consider ethical criteria, especially as it relates to equitability: racial diversity in schools, socioeconomic diversity, language diversity, student displacement, etc.

Some of these criteria share a foundation in the related problem of political redistricting. Compactness arises from the desire for unbiased districting as well as transportation costs. Proximity measures also reinforce minimizing costs and assigning students to their nearest schools, when possible. Connectivity constraints resemble those seen in addressing political gerrymandering. However, the key difference between the two is that schools act as fixed points in the redistricting problem. Schools anchor their SAZ while political districts are not geographically fixed; hence, school districting may be seen as a special case of political districting.

Redistricting as an assignment problem is one of combinatorial optimization. The formalized problem for a city school district can require thousands of binary variables and the resulting search space has exponentially many feasible solutions. Redistricting and several of its separate objectives were proven to be NP-hard by Altman in 1997 [1]. Puppe and Tasnádi later proved fair redistricting was NP-complete in 2008 [13]. So, not only is the general problem of redistricting NP-complete, but the connectivity requirement in redistricting is at least NP-hard as well [1].

It is this intractability that has led many past researchers to leave connectivity as a "desirable" feature in a solution, but not considered, or one that is only solved approximately [6]. In our work, we introduce an integer nonlinear programming model that can achieve exact connectivity in the problem of school redistricting via a graph's algebraic connectivity, as defined by a graph Laplacian matrix.

2 Related Works

Various redistricting formulations either ignoring or approximating graph connectivity have been proposed, including the Hess model [11, 3, 6], and using integer linear programming (LP) [11] or heuristics such as hill climbing [7], simulated annealing, and tabu search [4]. Drexl and Haase [8] formulated connectivity using an exponential number of linear constraints, applied to sales territory design

[14, 15], political redistricting [12], and school redistricting [3]. Shirabe [16] modelled contiguity via a polynomial number (still intractable) of constraints based on network flow, and there are other network flow models [17]. Any network flow formulation with polynomial complexity is not solving the same contiguity problem stated here, which is known to be in the class NP.

This work proposes using the algebraic connectivity of a graph defined by the second-smallest eigenvalue of its Laplacian matrix [9], which is computationally tractable but not expressible in closed form, and therefore likely not used in LP or integer programming (IP) based formulations. The proposed model combines Drexl and Haase's linear constraints with those based on graph algebraic connectivity.

3 School Redistricting Problem

The public school district of a city or county in the United States may typically be decomposed into school attendance zones (SAZs), which are clusters of school planning areas (SPAs). The SPAs are often defined at a neighborhood or smaller level by a school board. Although SPAs may sometimes need to administratively change, we assume that these units and their geographic boundaries are fixed. This preexisting framework makes redistricting a problem of assigning SPAs to SAZs at each school level: elementary, middle, and high. We also assume that all public schools in the district have K–5, 6–8, and 9–12 grade levels, respectively. This assumption implies separability for redistricting at each school level. Caro et al. consider school redistricting in a city where not all schools have the same splitting of grades [6].

3.1 Problem Formalization

Let the population data

$$\mathcal{P} = \left\{ \left(x, y, {}_{0}g, {}_{1}g, \ldots, {}_{12}g, {}_{E}g, {}_{M}g, {}_{H}g \right) \right\},$$

be the set of SPAs where (x, y) are geographic coordinates of the approximate geographic center of the SPA, ${}_{0}g$, ${}_{1}g$, …, ${}_{12}g$ are the population sizes of grades 0 (K), 1, …, 12, respectively, and ${}_{E}g$, ${}_{M}g$, ${}_{H}g$ are the 5-year projected enrollments for each of the school levels. The total number of SPAs is $n = |\mathcal{P}|$.

The school data for elementary, middle, and high schools are ${}_{E}\mathcal{S}$, ${}_{M}\mathcal{S}$, ${}_{H}\mathcal{S}$, respectively. Let each $\mathcal{S} = \left\{ (x, y, c) \right\}$, where (x, y) are the geographic coordinates of a school and c is its student capacity. Let X denote any of E, M, or H so that ${}_{X}\mathcal{S}$ represents a dataset of a given school level. Let ${}_{E}n_S = |{}_{E}\mathcal{S}|$, ${}_{M}n_S = |{}_{M}\mathcal{S}|$, ${}_{H}n_S = |{}_{H}\mathcal{S}|$ be the numbers of the different types of schools (or equivalently, ${}_{X}n_S = |{}_{X}\mathcal{S}|$). For some $1 \leq i \leq {}_{X}n_S$, let $({}_{X}x_i, {}_{X}y_i, {}_{X}c_i)$ denote the data of school i.

To manage the requirement of geographically connected SAZs, let $\mathcal{G}$ be the $(n \times n)$ adjacency matrix

$$\mathcal{G}_{ij} = \begin{cases} 1, & \text{SPAs } i \text{ and } j \text{ are deemed adjacent,} \\ 0, & \text{otherwise.} \end{cases}$$

Here, "adjacent" means having a common border of more than just a point, which is also referred to as rook adjacency in geospatial literature [3].

Denoting student planning area i by $\sigma_i = \big(x_i,\, y_i,\, {}_0g_i,\, {}_1g_i,\, \ldots,\, {}_{12}g_i,\, {}_Eg,\, {}_Mg,\, {}_Hg\big)$, an acceptable school districting is three partitions ${}_E\mathcal{Z}$, ${}_M\mathcal{Z}$, ${}_H\mathcal{Z}$ of $\mathcal{P}$ of respective sizes ${}_En_S$, ${}_Mn_S$, ${}_Hn_S$, where each subset (SAZ) of $\mathcal{P}$

$${}_X\mathcal{Z}_i = \big\{\sigma_{j_1}, \ldots, \sigma_{j_{{}_Xm_i}}\big\}, \quad i = 1, \ldots, {}_Xn_S,$$

and ${}_Xm_i$ is the number of SPAs m assigned to school i of type X. Each SAZ ${}_X\mathcal{Z}_i$ is defined by its school planning area index set ${}_X\mathcal{I}_i = \big\{j_1,\, \ldots,\, j_{{}_Xm_i}\big\}$.

3.2 Core Model

The school attendance zones ${}_X\mathcal{Z}_i$ in a partition must be geographically connected, compact, and satisfy physical and political constraints. We optimize on compactness and approximate student travel distances subject to constraints on connectivity of SAZs, student population with respect to capacity, and 5-year projected populations with respect to capacity. There are also viability constraints protecting against non-assignment and multiple assignment of a SPA as well as assignment of a SPA containing a school to another school.

The attendance at school i of type E, M, H is, respectively,

$${}_E\mathcal{A}_i = \sum_{j \in {}_E\mathcal{I}_i} \sum_{k=0}^{5} {}_kg_j, \quad {}_M\mathcal{A}_i = \sum_{j \in {}_M\mathcal{I}_i} \sum_{k=6}^{8} {}_kg_j, \quad {}_H\mathcal{A}_i = \sum_{j \in {}_H\mathcal{I}_i} \sum_{k=9}^{12} {}_kg_j.$$

The physical capacity constraints of the schools are therefore

$$\begin{aligned}
(1-\tau)_Ec_i \le {}_E\mathcal{A}_i \le (1+\tau)_Ec_i, \qquad & i = 1, \ldots, {}_En_S, \\
(1-\tau)_Mc_i \le {}_M\mathcal{A}_i \le (1+\tau)_Mc_i, \qquad & i = 1, \ldots, {}_Mn_S, \\
(1-\tau)_Hc_i \le {}_H\mathcal{A}_i \le (1+\tau)_Hc_i, \qquad & i = 1, \ldots, {}_Hn_S,
\end{aligned}$$

and the capacity constraints from the 5-year projected enrollment would be

$$\begin{aligned}
(1-\tau)_Ec_i \le \sum_{j \in {}_E\mathcal{I}_i} {}_Eg_j \le (1+\tau)_Ec_i,\ & i = 1, \ldots, {}_En_S, \\
(1-\tau)_Mc_i \le \sum_{j \in {}_M\mathcal{I}_i} {}_Mg_j \le (1+\tau)_Mc_i,\ & i = 1, \ldots, {}_Mn_S, \\
(1-\tau)_Hc_i \le \sum_{j \in {}_H\mathcal{I}_i} {}_Hg_j \le (1+\tau)_Hc_i,\ & i = 1, \ldots, {}_Hn_S,
\end{aligned}$$

where τ is a parameter chosen by the school board as a tolerance for going under or over school capacity (e.g. $\tau = 0.2$).

The partition subsets ${}_X\mathcal{Z}_i$ and index sets ${}_X\mathcal{I}_i$ may be conveniently recognized as binary variables

$$ {}_X W_{ij} = \begin{cases} 1, & \text{if SPA } j \text{ is assigned to SAZ } i, \\ 0, & \text{otherwise,} \end{cases} $$

for $i = 1, \ldots, {}_E n_S$ ($X = E$), $i = 1, \ldots, {}_M n_S$ ($X = M$), $i = 1, \ldots, {}_H n_S$ ($X = H$), and $j = 1, \ldots, n$. For viability, each of the 0–1 matrices ${}_X W$ must have exactly one 1 in each column, meaning each SPA must be assigned to a single SAZ:

$$ \sum_{i=1}^{{}_X n_S} {}_X W_{ij} = 1, \quad j = 1, \ldots, n. $$

The school planning areas that contain a school must also be assigned to their corresponding SAZ. These constraints are

$$ {}_X W_{ij} = 1 \text{ if SPA } j \text{ contains school } i, $$

for $i = 1, \ldots, {}_X n_S$, and $j = 1, \ldots, n$. This anchors SAZs to the schools they represent.

Define the population barycenter of school attendance zone ${}_X\mathcal{Z}_i$ for each school type X as

$$ \begin{aligned} {}_E(\bar{x}_i, \bar{y}_i) &= \sum_{j \in {}_E\mathcal{I}_i} \left(\sum_{k=0}^{5} {}_k g_j \Big/ {}_E\mathcal{A}_i \right) (x_j, y_j), \\ {}_M(\bar{x}_i, \bar{y}_i) &= \sum_{j \in {}_M\mathcal{I}_i} \left(\sum_{k=6}^{9} {}_k g_j \Big/ {}_M\mathcal{A}_i \right) (x_j, y_j), \\ {}_H(\bar{x}_i, \bar{y}_i) &= \sum_{j \in {}_H\mathcal{I}_i} \left(\sum_{k=10}^{12} {}_k g_j \Big/ {}_H\mathcal{A}_i \right) (x_j, y_j). \end{aligned} $$

The barycenter is a population-based centroid of the SAZ that weights each SPA's contribution to a school's total attendance.

Achieving optimal SAZ compactness is then similar to K-means clustering of the SPA centers using the 2-norm. Achieving student proximity to schools uses the 1-norm, which captures travel time better than the Euclidean 2-norm distance. Composing these two objectives, assign SPAs σ_j to ${}_X n_S$ SAZs ${}_X\mathcal{Z}_i$ for each school level X to minimize

$$ \begin{aligned} \Phi({}_E W, {}_M W, {}_H W) = &\sum_{i=1}^{{}_E n_S} \sum_{j \in {}_E\mathcal{I}_i} \left\| (x_j, y_j) - {}_E(\bar{x}_i, \bar{y}_i) \right\|_2^2 \\ &+ \sum_{i=1}^{{}_M n_S} \sum_{j \in {}_M\mathcal{I}_i} \left\| (x_j, y_j) - {}_M(\bar{x}_i, \bar{y}_i) \right\|_2^2 \end{aligned} $$

$$
\begin{aligned}
&+\sum_{i=1}^{_H n_S}\sum_{j\in {_H\mathcal{I}_i}}\left\|(x_j,y_j)-{_H(\bar{x}_i,\bar{y}_i)}\right\|_2^2\\
&+\gamma\Bigg(\sum_{i=1}^{_E n_S}\left\|({_E x_i},{_E y_i})-{_E(\bar{x}_i,\bar{y}_i)}\right\|_1\\
&+\sum_{i=1}^{_M n_S}\left\|({_M x_i},{_M y_i})-{_M(\bar{x}_i,\bar{y}_i)}\right\|_1\\
&+\sum_{i=1}^{_H n_S}\left\|({_H x_i},{_H y_i})-{_H(\bar{x}_i,\bar{y}_i)}\right\|_1\Bigg),
\end{aligned}
$$

subject to the capacity, projected enrollment, connectivity, and viability constraints. The first three terms correspond to compactness and the last three correspond to proximity. The parameter $\gamma \gg 1$ weights the relative importance of proximity to compactness. Note that the objective is separable into three disjoint minimizations for each of the school levels.

There are several ways of formulating the compactness and proximity objectives. For example, proximity may be written as the distances between SPA centers and schools rather than the single SAZ barycenter to school distance as above. One could also formulate a multiobjective optimization problem with a compactness objective and a proximity objective; however, each Pareto optimal solution of this multiobjective formulation corresponds to some choice of γ in the above formulation.

The binary variables simplify the expressions and calculations. For instance, if a population vector ${_E P}$ of length n were precomputed with jth element being $\sum_{k=0}^{5} {_k g_j}$, then the capacity constraint on school i

$$(1-\tau){_E c_i} \le {_E\mathcal{A}_i} \le (1+\tau){_E c_i}$$

would become simply

$$(1-\tau){_E c_i} \le ({_E W}\,{_E P})_i \le (1+\tau){_E c_i}$$

in terms of ${_E W}$. Similarly, take ${_E Q}$ to be another vector of length n with jth element being ${_E g_j}$. Then, the 5-year projected enrollment constraint on school i becomes

$$(1-\tau){_E c_i} \le ({_E W}\,{_E Q})_i \le (1+\tau){_E c_i}.$$

Let ${_X B_x} \in \mathbb{R}^n$ be precomputed with jth element being ${_X P_j} x_j$, and ${_X B_y} \in \mathbb{R}^n$ with jth element being ${_X P_j} y_j$. Then the population barycenter simplifies to

$${_X(\bar{x}_i,\bar{y}_i)} = \left(\frac{({_X W}\,{_X B_x})_i}{({_X W}\,{_X P})_i}, \frac{({_X W}\,{_X B_y})_i}{({_X W}\,{_X P})_i}\right).$$

SAZ connectivity can be expressed using the Laplacian matrix of the graph formed by the SPAs. Let $e = (1, \ldots, 1) \in \mathbb{R}^n$ and $\mathrm{diag}(\mathcal{G}e)$ be the $n \times n$ diagonal degree matrix whose diagonal elements are the degrees of the nodes in the graph G with adjacency matrix $\mathcal{G}$. The Laplacian matrix of the graph G is defined as

$$\Lambda(G) = \operatorname{diag}(\mathcal{G}e) - \mathcal{G},$$

and G is connected if and only if $\mathcal{G}$ is irreducible. Also, $\mathcal{G}$ is irreducible if and only if the second-smallest eigenvalue of $\Lambda(G)$ is not zero. The second-smallest eigenvalue of the graph Laplacian matrix is known as the graph's algebraic connectivity, or its Fiedler value [9]. With this, the SAZ ${}_X\mathcal{Z}_i$ is connected if and only if the second-smallest eigenvalue of the graph Laplacian

$$\operatorname{diag}\big(\mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\, e\big) - \mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}$$

of the subgraph of G corresponding to ${}_X\mathcal{Z}_i$ is not zero; $e \in \mathbb{R}^{{}_X m_i}$ here. The graph Laplacian matrix is symmetric and positive semidefinite. If we order its real-valued non-negative eigenvalues as λ_1, λ_2, ..., $\lambda_{{}_X m_i}$ (smallest to largest), the connectivity constraints for the SAZs can be taken as

$$0.001 - \lambda_2\big(\operatorname{diag}\big(\mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\, e\big) - \mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\big) \le 0$$

for $i = 1, \ldots, {}_X n_S$.

In practice, the following model can be solved separately in the binary variables ${}_E W$, ${}_M W$, and ${}_H W$ since the objective function Φ is separable, and the school districting constraints for elementary, middle, and high schools are completely independent of each other. Having said this, the full problem is now stated.

For convenience, define population vectors ${}_E P$, ${}_M P$, ${}_H P$ of length n with jth element being $\sum_{k=0}^{5} {}_k g_j$, $\sum_{k=6}^{8} {}_k g_j$, $\sum_{k=9}^{12} {}_k g_j$, respectively. Define 5-year projected population vectors ${}_E Q$, ${}_M Q$, ${}_H Q$ of length n with jth element being ${}_E g_j$, ${}_M g_j$, ${}_H g_j$, respectively. The variables are ${}_E W \in \{0,1\}^{{}_E n_S \times n}$, ${}_M W \in \{0,1\}^{{}_M n_S \times n}$, and ${}_H W \in \{0,1\}^{{}_H n_S \times n}$. The optimization problem is

$$\min_{{}_E W, {}_M W, {}_H W} \Phi\big({}_E W, {}_M W, {}_H W\big)$$

subject to

$$\begin{aligned}
&(1-\tau)\,{}_X c_i \le ({}_X W\, {}_X P)_i \le (1+\tau)\,{}_X c_i, \\
&\qquad\qquad \forall X \in \{E, M, H\}, \quad \forall i = 1, \ldots, {}_X n_S, && (1)\\
&(1-\tau)\,{}_X c_i \le ({}_X W\, {}_X Q)_i \le (1+\tau)\,{}_X c_i, \\
&\qquad\qquad \forall X \in \{E, M, H\}, \quad \forall i = 1, \ldots, {}_X n_S, && (2)\\
&0.001 - \lambda_2\big(\operatorname{diag}\big(\mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\, e\big) - \mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\big) \le 0, \\
&\qquad\qquad \forall X \in \{E, M, H\}, \quad \forall i = 1, \ldots, {}_X n_S, && (3)\\
&\sum_{i=1}^{{}_X n_S} {}_X W_{ij} = 1, \\
&\qquad\qquad \forall X \in \{E, M, H\}, \quad \forall j = 1, \ldots, n, && (4)\\
&{}_X W_{ij} = 1 \text{ if SPA } j \text{ contains school } i, \\
&\qquad\qquad \forall X \in \{E, M, H\}, \forall i = 1, \ldots, {}_X n_S, \forall j = 1, \ldots, n. && (5)
\end{aligned}$$

Equation (1) is the capacity constraint that bounds student attendance at each of the schools. Equation (2) is the 5-year enrollment projection constraint that similarly bounds schools' expected student population in the future. Equation (3) is the SAZ connectivity constraint that represents geographic connectivity exactly for the SPAs in each SAZ. Equation (4) is a viability constraint requiring that each SPA belong to exactly one SAZ for each school level, preventing non-assignment and multiple assignment. Lastly, equation (5) is the other viability constraint that requires that schools are assigned their own SPAs.

3.3 Reformulation for Computability

The SAZ connectivity constraint above (equation (3)) presents computational challenges using current integer programming solvers. There are no existing non-commercial software that the authors know of which explicitly support solving a model with this nonlinear, black-box constraint.

However, some integer programming solvers support user callback functions. These have been used by solutions to the traveling salesman problem (TSP) to eliminate subtours [2]. They evaluate a feasibility constraint outside the solver at believed solution nodes in the branch and bound tree. The user may then append violated cutting planes lazily, updating the model. With inspiration from the TSP, we can introduce another connectivity formulation as the cutting planes and use the eigenvalue constraint in the feasibility check.

The formulation for graph connectivity used by Drexl and Haase in sales territory design has an exponential number of linear constraints [8]. For brevity, denote the nodes of the graph G by the student planning area indices, and for $\emptyset \neq S \subset \{1, \ldots, n\}$, define the neighborhood of S by

$$N(S) = S \cup \big\{m \mid m \text{ is adjacent to some } j \in S\big\},$$

i.e., $N(S)$ represents all student planning areas either represented by S or adjacent to some student planning area represented by S. Then, the SAZs ${}_X\mathcal{Z}_i$ are connected if $\forall i = 1, \ldots, {}_Xn_S$, $\forall \ell = 1, \ldots, n$, $\forall S \subset \{1, \ldots, n\} \backslash N(\{\ell\}) \neq \emptyset$,

$$\sum_{j \in N(S) \backslash S} {}_XW_{ij} - \sum_{j \in S \cup \{\ell\}} {}_XW_{ij} \geq 1 - (|S| + 1),$$

which is $\mathcal{O}({}_Xn_S n 2^n)$ constraints.

Note that these subset-based constraints are violated exactly when each SPA represented in S is assigned to SAZ i, as well as another SPA l, but no neighbors of S are also assigned to that SAZ. In other words, a constraint is violated when there is no path from S to l among the SPAs assigned to i. The formulation is the same as written by Drexl and Haase except for adding a term for l to both sides of the inequalities, which is more readily useful to redistricting as opposed to routing. For example, a disconnected SPA, l, assigned to a larger connected component of the SAZ, S, can be explicitly formed into a violated constraint.

These linear subset-based constraints express connectivity exactly if all can be incorporated into the model directly. Taking any subset of the constraints constitutes an approximation to connectivity, and this is indeed how some have approached graph connectivity in past integer programs [8]. Combining the eigenvalue constraints with these subset constraints, we obtain a modified optimization model with a way of achieving connected SAZs exactly that's more suitable to existing software.

4 Discussion

Previous districting methods of addressing compactness include those based on sums of distances and those based on perimeter. The compactness part of Φ is similar to the former class, but is based on distances between SPAs and their "barycenter," a centroid of the SAZ. This measure offers a variable center point rather than taking the distances to their assigned school, allowing for globular SAZs whose centers are not fixed. This may also be preferable to methods based on the perimeter of an SAZ, as those are based on the shape of SPAs. Perimeter-based compactness is known to unfairly penalize highly irregular boundaries when this is generally out of school planners' immediate control [10]. This more expressive form of compactness comes at the cost of nonlinearity, which presents technical challenges with currently available integer programming software.

A similar argument may be made regarding the chosen proximity metric. Taking distances between SPAs and schools is an equally viable option; however, the barycenter captures part of the surrounding population density. An unweighted metric based in SPA–school distances favors SPA location and quantity, whereas barycenter–school distances factor in the students.

Incorporating the eigenvalue formulation of graph connectivity in districting is one of the main contributions of this work. These nonlinear constraints, linear in the number of schools, are relatively quick to compute, but not directly viable with available integer programming solvers. Hence, we propose a cutting planes technique that is capable of closing off disconnected districting plans as necessary during optimization.

This formulation of linear constraints for connectivity leads to a natural algorithm of identifying violated SAZ connectivity. As the objective will tend to produce compact SAZs, any disconnected SAZ will tend to have a connected component in the corresponding subgraph that contains a majority of its assigned SPAs. Using the variables ${}_XW$, choose this largest connected component to be S. Breadth-first search can be used to construct S in $\mathcal{O}({}_Xm_i)$, linear in the number of SPAs assigned to the school. Since the SAZ is disconnected, any other assigned SPA not part of S may be chosen as l. The resulting constraint is verifiably violated and updates the model.

There are also considerably many more ways of selecting from the exponentially many linear constraints. If the chosen constraint above is violated, then it is also violated for each subset of S too. One could just as easily take multiple constraints for each choice of l.

The callback must also handle an edge case of having too few eigenvalues. Equation (5) ensures that each SAZ has at least one assigned SPA, but a solver might intermediately have only that single SPA assigned. The corresponding Laplacian matrix of this SAZ only has one eigenvalue. The trivially connected SAZ has no "second-smallest" eigenvalue.

There are still computational challenges even with this formulation that enables exact connectivity. While some mixed-integer nonlinear programming solvers can handle specific classes of nonlinear objectives, such as Gurobi and Couenne, Φ is not known to have one of these supported forms, largely due to the quotient formed by the barycenter formulation. Options for solving this optimization problem include seeking out a solver that meets the required conditions, developing new algorithms to meet these conditions, or settling for heuristically solving the problem to a local optimum instead.

The school data itself and parameters chosen can be limitations to feasibility too. For instance, if there are many schools, then it may be difficult or impossible to find a feasible solution that satisfies the capacity constraints. This arises from requiring connectivity of SPAs, but the schools are too densely colocated. An integer programming solver will continue to extensively explore solutions in the case where capacity cannot be satisfied; relaxing τ to a greater tolerance may be necessary to find feasible solutions.

5 Conclusion

We propose a model for school redistricting that optimizes compactness of school attendance zones (SAZs) and proximity of students to assigned schools while satisfying physical and political constraints. Inspired by the traveling salesman problem, the constraint of SAZ connectivity can be solved exactly. The use of cutting planes to close off disconnected solutions makes the problem more computationally viable and is a technique that may be extended to other geographic partitioning problems requiring geographic connectivity.

Acknowledgment

This work is supported in part by the National Science Foundation via grant #DGE-1545362, UrbComp (Urban Computing): Data Science for Modeling, Understanding, and Advancing Urban Populations.

References

1. Altman, M.: The computational complexity of automated redistricting: Is automation the answer? Rutgers Comput. Technol. L. J. **23**(1), 81–142 (1997)
2. Applegate, D.L., Bixby, R.E., Chvátal, V., Cook, W.J.: The Traveling Salesman Problem: A Computational Study. Princeton University Press, Princeton NJ USA (2007). DOI 10.1515/9781400841103

3. Biswas, S., Chen, F., Chen, Z., Lu, C.T., Ramakrishnan, N.: Incorporating domain knowledge into Memetic Algorithms for solving Spatial Optimization problems. In: Proc. 28th Int. Conf. Adv. Geogr. Inf. Syst., pp. 25–35. ACM, Seattle WA USA (2020). DOI 10.1145/3397536.3422265
4. Biswas, S., Chen, F., Chen, Z., Sistrunk, A., Self, N., Lu, C.T., Ramakrishnan, N.: REGAL: A regionalization framework for school boundaries. In: Proc. 27th ACM SIGSPATIAL Int. Conf. Adv. Geogr. Inf. Syst., pp. 544–547. ACM, Chicago IL USA (2019). DOI 10.1145/3347146.3359377
5. Black, S.E.: Do better schools matter? parental valuation of elementary education. Q. J. Econ. **114**(2), 577–599 (1999). DOI 10.1162/003355399556070
6. Caro, F., Shirabe, T., Guignard, M., Weintraub, A.: School redistricting: embedding GIS tools with integer programming. J. Oper. Res. Soc. **55**(8), 836–849 (2004). DOI 10.1057/palgrave.jors.2601729
7. desJardins, M., Bulka, B., Carr, R., Jordan, E., Rheingans, P.: Heuristic search and information visualization methods for school redistricting. AI Mag. **28**(3), 59–72 (2007). DOI 10.1609/aimag.v28i3.2055
8. Drexl, A., Haase, K.: Fast approximation methods for sales force deployment. Manag. Sci. **45**(10), 1307–1323 (1999). DOI 10.1287/mnsc.45.10.1307
9. Fiedler, M.: Algebraic connectivity of graphs. Czech. Math. J. **23**(2), 298–305 (1973). DOI 10.21136/CMJ.1973.101168
10. Haunert, J.H., Wolff, A.: Area aggregation in map generalisation by mixed-integer programming. Int. J. Geogr. Inf. Sci. **24**(12), 1871–1897 (2010)
11. Hess, S.W., Weaver, J.B., Siegfeldt, H.J., Whelan, J.N., Zitlau, P.A.: Nonpartisan political redistricting by computer. Oper. Res. **13**(6), 998–1006 (1965)
12. King, D.M., Jacobson, S.H., Sewell, E.C., Cho, W.K.T.: Geo-Graphs: An efficient model for enforcing contiguity and hole constraints in planar graph partitioning. Oper. Res. **60**(5), 1213–1228 (2012). DOI 10.1287/opre.1120.1083
13. Puppe, C., Tasnádi, A.: A computational approach to unbiased districting. Math. Comput. Model. **48**(9-10), 1455–1460 (2008). DOI 10.1016/j.mcm.2008.05.024
14. Ríos-Mercado, R.Z., Fernández, E.: A reactive GRASP for a commercial territory design problem with multiple balancing requirements. Comput. Oper. Res. **36**(3), 755–776 (2009). DOI 10.1016/j.cor.2007.10.024
15. Salazar-Aguilar, M.A., Ríos-Mercado, R.Z., González-Velarde, J.L.: GRASP strategies for a bi-objective commercial territory design problem. J. Heuristics **19**(2), 179–200 (2013). DOI 10.1007/s10732-011-9160-8
16. Shirabe, T.: Districting modeling with exact contiguity constraints. Environ. Plann. B: Plann. Des. **36**(6), 1053–1066 (2009). DOI 10.1068/b34104
17. Validi, H., Buchanan, A., Lykhovyd, E.: Imposing contiguity constraints in political districting models. Oper. Res. (in press)