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

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

Virginia Polytechnic Institute and State University

19th International Conference on Scientific Computing
CSC'21
July 28, 2021

School Districting

- **School districting** is traditionally a geographical partitioning problem: draw boundaries, making an attendance region for each public school.
  - Equivalently: an **assignment** problem.
- **School attendance zones (SAZs)** are clusters of **student planning areas (SPAs)**.
- SPAs are often defined at a neighborhood level or smaller by a school board.
- So, the assignment problem is to

assign SPAs to SAZs

Motivations and Impacts of Redistricting

Redistricting is a recurring process.

- New schools are built,
- The local population changes,
- etc.

GIS-based data is still relatively new.

Rucho v. Common Cause (June 2019): "partisan gerrymandering claims present political questions beyond the reach of the federal courts."

Prelude: Objectives and Constraints

"Fair" redistricting typically uses an agreed upon set of criteria to evaluate redistricting maps.

- Balancing attendance sizes with school capacities
- Balancing projected attendance sizes with school capacities
- Proximity of students to their schools
- Geographic compactness of SAZs
- Geographic contiguity/connectivity of SAZs

Related Works

- “Fair” redistricting is NP-complete.
- Connectivity has been
  - ignored and fixed in postprocessing,
  - maintained heuristically from a known good starting point (often the existing school zones),
  - repaired during the local search phase itself,
  - used with conditions that weren’t globally sufficient,
  - and more.
- Drexl and Haase had an exact method using exponentially many linear constraints.
- There are formulations with network flow constraints (polynomial complexity; conflicts with known NP problem).

Our Contribution

We propose an exact method of connectivity that uses the second-smallest eigenvalue of a graph Laplacian matrix.

We also describe an adapted model that uses Drexl and Haase's formulation as cutting planes, similar to TSP subtour elimination constraints.

Definitions: The Datasets

Let $\mathcal{P} = \{(x, y, {}_0g, {}_1g, \ldots, {}_{12}g, {}_Eg, {}_Mg, {}_Hg)\}$ represent the planning areas (SPAs), where

- x and y are the approximate center of the SPA;
- ${}_0g, {}_1g, \ldots, {}_{12}g$ are the SPA's population sizes for grades K-12; and
- ${}_Eg, {}_Mg, {}_Hg$ are 5-year projections for each school type.

Let ${}_X\mathcal{S} = \{(x, y, c)\}$ represent the schools, where

- x and y are the school's coordinates;
- c is the school capacity; and
- X is one of E, M, or H for the different school types.

Let $n = |\mathcal{P}|$ and ${}_Xn_S = |{}_X\mathcal{S}|$.

Definitions: Connectivity

Let $\mathcal{G}$ be the adjacency matrix for the SPAs where

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

Here, "adjacent" means having a common border of more than just a point.

Definitions: Optimization Variables

Let ${}_XW \in \{0,1\}^{x n_S \times n}$ be the 3 binary assignment/partition matrices where

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

- The rows capture which SPAs are assigned to school i.

School Capacity with Current Populations

Define ${}_XP$ to be vectors of length n, where for $j = 1, \ldots, n$,

$$ {}_EP_j = \sum_{k=0}^{5} {}_kg_j, \qquad {}_MP_j = \sum_{k=6}^{8} {}_kg_j, \qquad {}_HP_j = \sum_{k=9}^{12} {}_kg_j. $$

Constraint: School Capacity

τ is a threshold chosen by the School Board.
For all $X \in \{E, M, H\}, i = 1, \ldots, {}_Xn_S$,

$$ (1-\tau)_Xc_i \le ({}_XW_XP)_i \le (1+\tau)_Xc_i. $$

School Capacity with Projected Populations

Define ${}_XQ$ to be vectors of length n, where for $j = 1, \ldots, n$,

$$ {}_EQ_j = {}_Eg_j, \qquad {}_MQ_j = {}_Mg_j, \qquad {}_HQ_j = {}_Hg_j. $$

Constraint: School Capacity

For all $X \in \{E, M, H\}, i = 1, \ldots, {}_Xn_S$,

$$ (1-\tau)_X c_i \leq ({}_XW\,{}_XQ)_i \leq (1+\tau)_X c_i. $$

SANGHANI CENTER FOR ARTIFICIAL INTELLIGENCE & DATA ANALYTICS
VIRGINIA TECH.

Compactness and Proximity

The **population barycenter** is a population-based centroid of an SAZ.

- Weighting comes from each SPA's contribution to the school's total attendance.

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 is

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

Compactness and Proximity

Compactness: Minimize the 2-norms between the SPAs and the population barycenter of their SAZ.

Proximity: Minimize the 1-norms between schools and the population barycenter of their SAZ.

Objective Function for High Schools

$$\Phi({}_H W) = \sum_{i=1}^{n} \sum_{j \in {}_H\mathcal{I}_i} \|(x_j, y_j) - {}_X(\bar{x}_i, \bar{y}_i)\|_2^2 + \gamma \left(\sum_{i=1}^{n} \|({}_H x_i, {}_H y_i) - {}_X(\bar{x}_i, \bar{y}_i)\|_1 \right)$$

SAZ Connectivity

The **Laplacian matrix** of a graph G with adjacency matrix $\mathcal{G}$ is defined as

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

where e is the ones vector.

Connectivity comes from the following sketch of an argument:

- When G is a disconnected graph, $\mathcal{G}$ is permutation similar to a block diagonal matrix with one block for each connected component.
- Each row/column of $\Lambda(G)$ sums to zero; and, e is an eigenvector with eigenvalue 0.
- $\Lambda(G)$ will have a 0 eigenvalue for each connected component.

SAZ Connectivity

An SAZ is connected iff the second-smallest eigenvalue of its Laplacian matrix is not zero.

Notation takes the subgraph for an SAZ from the adjacency matrix $\mathcal{G}$.

Constraint: Connectivity

For all $X \in \{E, M, H\}, i = 1, \ldots, {}_X n_S$,

$$0.001 - \lambda_2 \left(\text{diag}\left(\mathcal{G}_{{}_X\mathcal{I}_i, {}_X\mathcal{I}_i} e\right) - \mathcal{G}_{{}_X\mathcal{I}_i, {}_X\mathcal{I}_i}\right) \leq 0.$$

All Together

Optimization Problem

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

subject to

$$(1-\tau)_X c_i \leq ({}_X W_X P)_i \leq (1+\tau)_X c_i, \quad \forall X \in \{E, M, H\}, \quad \forall i,$$
$$(1-\tau)_X c_i \leq ({}_X W_X Q)_i \leq (1+\tau)_X c_i, \quad \forall X \in \{E, M, H\}, \quad \forall i,$$
$$0.001 - \lambda_2 \left(\text{diag}\left(\mathcal{G}_{{}_X\mathcal{I}_i, {}_X\mathcal{I}_i} e\right) - \mathcal{G}_{{}_X\mathcal{I}_i, {}_X\mathcal{I}_i}\right) \leq 0 \qquad \forall X, \quad \forall i,$$
$$\sum_{i=1}^{{}_X n_S} {}_X W_{ij} = 1 \qquad \forall X, \quad \forall j,$$
$${}_X W_{ij} = 1 \text{ if SPA } j \text{ contains school } i \qquad \forall X, \quad \forall i, \quad \forall j.$$

Challenges

- The eigenvalue constraint for connectivity is nonlinear, as well as black-box.
- There are no open source integer programming solvers that support black-box constraints, like those in real optimization do.

The Travelling Salesman Problem: Subtour Elimination Constraints

- TSP: Find the shortest tour of a graph. NP-hard.
- Can be modelled as a minimization problem, and everything is linear in the model until...
- It returns tours with cycles. Easy to identify by algorithm, but...
- Fixing the subtours requires exponentially many additional linear constraints.
- Solution: Fix violations incrementally and evaluate feasibility outside the solver.

Learning From the TSP

Use Drexl and Haase's exponentially many linear constraints as cutting planes!

For $\emptyset = S = \{1, \ldots, n\}$, define

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

Constraint: Connectivity

Each SAZ is connected if
$\forall i = 1, \ldots, {}_X n_S, \forall l = 1, \ldots, n, \forall S \subset \{1, \ldots, n\} \setminus N(\{l\}) \neq \emptyset,$

$$\sum_{j \in N(S) \setminus S} {}_X W_{ij} - \sum_{j \in S \cup \{l\}} {}_X W_{ij} \geq -|S|\,.$$

SANGHANI CENTER FOR ARTIFICIAL INTELLIGENCE & DATA ANALYTICS
VIRGINIA TECH.

More Challenges

- The objective is still nonlinear, and discrete + nonlinear programs still have little support from solvers. Those that exist require specific forms such as quadratic, etc.
- Needs a solver with this requirement plus our others such as binary variables and user callbacks for the connectivity cutting planes.
- There isn't an easy way of breaking down the problem into something smaller; each component has its own complexity.
- Feasible solutions may not exist due to capacity constraints at odds with requiring connectivity.

Future Work

- We found a constraint programming solver that supports the problem, so we can now try solving the model on our data.
- Equitability and diversity: I'm currently working with others at VT to model these into the problem.
- The political districting problem.

Questions?

Joshua C. Detwiler

jdetwiler@vt.edu

https://jdetwiler.tech