Public school districts (typically at the city or county level) are decomposed into *school attendance zones* (SAZs), which are clusters of *student planning areas* (SPAs), typically at the neighborhood or smaller level. Let the population data

$$\mathcal{P} = \big\{ \big(x, y, {}_0g, {}_1g, \ldots, {}_{12}g, {}_Eg, {}_Mg, {}_Hg, {}_Ef, {}_Mf, {}_Hf\big) \big\},$$

where (x, y) are geographic coordinates of the approximate geographic center of the student planning area (SPA), ${}_0g$, ${}_1g$, …, ${}_{12}g$ are the population sizes of grades 0(K), 1, …, 12, respectively, ${}_Eg$, ${}_Mg$, ${}_Hg$ are the 5-year projected enrollments for elementary (grades K–5), middle (grades 6–9), and high (grades 10–12) school students, respectively, and ${}_Ef$ ${}_Mf$, ${}_Hf$ are the numbers of free and reduced meal students for elementary, middle, and high school, respectively, for that SPA. The total number of SPAs is $n = |\mathcal{P}|$. The school data for the elementary, middle, and high schools is $\mathcal{S}_E$, $\mathcal{S}_M$, $\mathcal{S}_H$, respectively, where each $\mathcal{S}$ has the form

$$\big\{ (x, y, c) \;\big|\; (x, y) \text{ are the geographic coordinates of the school, } c \text{ is the capacity} \big\}.$$

Let ${}_En_S = |\mathcal{S}_E|$, ${}_Mn_S = |\mathcal{S}_M|$, ${}_Hn_S = |\mathcal{S}_H|$ be the numbers of the different types of schools, and the data for school i of type X will be denoted by $({}_Xx_i,\ {}_Xy_i,\ {}_Xc_i)$.

Assume that the student planning areas, which typically are geographically small and compact with boundaries determined by the practical considerations of bus routes, major streets, etc., are fixed. To manage the requirement of geographically connected school attendance zones (SAZs), let $\mathcal{G}$ be the $(n \times n)$ adjacency matrix

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

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

Denoting student planning area i by $\sigma_i = \big(x_i,\ y_i,\ {}_0g_i,\ {}_1g_i,\ \ldots,\ {}_Hf_i\big)$, an acceptable *school districting* is three partitions ${}_E\mathcal{Z}$, ${}_M\mathcal{Z}$, ${}_H\mathcal{Z}$ of $\mathcal{P}$ of respective sizes ${}_En_S$, ${}_Mn_S$, ${}_Hn_S$ (the numbers of elementary, middle, and high schools, respectively), where each subset (SAZ) of $\mathcal{P}$

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

(X above can be any of E, M, or H) in a partition must be geographically connected, compact, and satisfy other physical, political, and social constraints described later. School attendance zone ${}_X\mathcal{Z}_i$ is defined by its student planning area index set ${}_X\mathcal{I}_i = \big\{ j_1,\ \ldots,\ j_{{}_Xm_i} \big\}$.

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

$${}_E\mathcal{A}_i = \sum_{j \in {}_E\mathcal{I}_i} \sum_{k=0}^{5} {}_kg_j, \quad {}_M\mathcal{A}_i = \sum_{j \in {}_M\mathcal{I}_i} \sum_{k=6}^{9} {}_kg_j, \quad {}_H\mathcal{A}_i = \sum_{j \in {}_H\mathcal{I}_i} \sum_{k=10}^{12} {}_kg_j.$$

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

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

where $\tau = 0.2$ has been chosen by the School Board.

For nonempty $A, B \subset \{1,\dots,n\}$, let $\mathcal{G}_{A,B}$ be the submatrix of $\mathcal{G}$ with rows (columns) indexed by A (B). Thinking of the graph defined by the student planning areas as nodes and arcs defined by student planning area adjacency, each school attendance zone corresponds to a subgraph. The school attendance zone is geographically connected if its corresponding subgraph is connected. In terms of the adjacency matrix $\mathcal{G}$, the constraint that ${}_X\mathcal{Z}_i$ be geographically connected is that the adjacency matrix $\mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}$ be irreducible. How to confirm this matrix property will be discussed in detail later.

It is convenient to represent these partition subsets ${}_X\mathcal{Z}_i$ and index sets ${}_X\mathcal{I}_i$ by binary variables

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

for $i = 1, \dots, {}_En_S$ ($X = E$), $i = 1, \dots, {}_Mn_S$ ($X = M$), $i = 1, \dots, {}_Hn_S$ ($X = H$), and $j = 1, \dots, n$. Each of the 0-1 matrices ${}_XW$ has exactly one 1 in each column. Define the population barycenter of school attendance zone ${}_X\mathcal{Z}_i$ for each school type X as

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

Achieving optimal school attendance zone compactness is then similar to K-means clustering of the SPA centers using the 2-norm, and achieving student proximity to schools uses the 1-norm (which captures travel time better than the Euclidean 2-norm distance): assign student planning areas σ_j to ${}_En_S$, ${}_Mn_S$, ${}_Hn_S$ school attendance zones ${}_E\mathcal{Z}_i$, ${}_M\mathcal{Z}_i$, ${}_H\mathcal{Z}_i$, respectively, to minimize

$$\begin{aligned}\Phi({}_EW,{}_MW,{}_HW) = {} & \sum_{i=1}^{{}_En_S}\sum_{j\in {}_E\mathcal{I}_i}\left\|(x_j,y_j) - {}_E(\bar{x}_i,\bar{y}_i)\right\|_2^2 + \sum_{i=1}^{{}_Mn_S}\sum_{j\in {}_M\mathcal{I}_i}\left\|(x_j,y_j) - {}_M(\bar{x}_i,\bar{y}_i)\right\|_2^2 + \\ & \sum_{i=1}^{{}_Hn_S}\sum_{j\in {}_H\mathcal{I}_i}\left\|(x_j,y_j) - {}_H(\bar{x}_i,\bar{y}_i)\right\|_2^2 + \gamma\left(\sum_{i=1}^{{}_En_S}\left\|({}_Ex_i,{}_Ey_i) - {}_E(\bar{x}_i,\bar{y}_i)\right\|_1 + \right. \\ & \left.\sum_{i=1}^{{}_Mn_S}\left\|({}_Mx_i,{}_My_i) - {}_M(\bar{x}_i,\bar{y}_i)\right\|_1 + \sum_{i=1}^{{}_Hn_S}\left\|({}_Hx_i,{}_Hy_i) - {}_H(\bar{x}_i,\bar{y}_i)\right\|_1\right),\end{aligned}$$

subject to the school capacity and school attendance zone connectivity constraints. $\gamma \gg 1$ weights the relative importance of proximity to compactness. There are several alternative ways to deal with the compactness and proximity objectives. For proximity, one could use all the SPA center to school distances (within a SAZ) rather than the single SAZ barycenter to school distance as above. One could also formulate a multiobjective optimization problem with a compactness objective (first three terms of Φ) and a proximity objective (last three terms of Φ), however, each Pareto optimal solution of this multiobjective formulation corresponds to some choice of γ in the above formulation.

The binary variables ${}_XW_{ij}$ simplify the expressions and calculations. For instance, if a population vector ${}_EP$ of length n were formed with jth element being $\sum_{k=0}^{5} {}_kg_j$, then the capacity constraint

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

would become simply

$$(1-\tau)_Ec_i \le ({}_EW\,{}_EP)_i \le (1+\tau)_Ec_i$$

in terms of ${}_EW$.

Let $e = (1, \ldots, 1) \in \mathbb{R}^n$, and $\operatorname{diag}(\mathcal{G}e)$ be the $n \times n$ diagonal matrix whose diagonal elements are the degrees of the nodes in the graph G with adjacency matrix $\mathcal{G}$. The *graph Laplacian* 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 if and only if the second smallest eigenvalue of $\Lambda(G)$ is not zero. Hence 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; $e \in \mathbb{R}^{{}_Xm_i}$ here. Since the graph Laplacian matrix is symmetric and positive semidefinite, the connectedness constraints for the SAZs can be taken as

$$0.001 - \Big(\text{second smallest eigenvalue of } \operatorname{diag}\big(\mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\, e\big) - \mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\Big) \le 0$$

for $i = 1, \ldots, {}_Xn_S$. Here, as always, X represents any of E, M, or H. An alternative to these eigenvalue constraints, which cannot be expressed in closed form but are computationally cheap, is to express graph connectivity by (exponentially many) linear constraints. For brevity, denote the nodes of the graph G by the student planning area indices, and for $\emptyset \ne S \subset \{1, \ldots, n\}$, 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 school attendance zones ${}_X\mathcal{Z}_i$ are connected if $\forall i = 1, \ldots, {}_Xn_S$, $\forall \ell = 1, \ldots, n$, $\forall S \subset \{1, \ldots, n\} \backslash N(\{\ell\}) \ne \emptyset$,

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

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

The problem is clearly separable in the variables ${}_EW$, ${}_MW$, ${}_HW$, since the objective function Φ is separable, and the school districting constraints for elementary, middle, and high schools are completely independent of each other. While formulated here as a single optimization problem, in practice the problems for the three school types are solved separately. Having said this, the full problem is now stated.

For convenience, define population vectors ${}_EP$, ${}_MP$, ${}_HP$ of length n with jth element being $\sum_{k=0}^{5} {}_kg_j$, $\sum_{k=6}^{9} {}_kg_j$, $\sum_{k=10}^{12} {}_kg_j$, respectively. The variables are ${}_EW \in \{0,1\}^{{}_En_S \times n}$, ${}_MW \in \{0,1\}^{{}_Mn_S \times n}$, and ${}_HW \in \{0,1\}^{{}_Hn_S \times n}$. The optimization problem is

$$\min_{{}_EW,{}_MW,{}_HW} \Phi\big({}_EW, {}_MW, {}_HW\big)$$

subject to

$$(1-\tau)\,{}_Xc_i \le ({}_XW\,{}_XP)_i \le (1+\tau)\,{}_Xc_i, \qquad \forall X \in \{E,M,H\} \quad \forall i,$$

$$(1-\tau)\,{}_Xc_i \le \sum_{j=1}^{n} {}_XW_{ij}\;{}_Xg_j \le (1+\tau)\,{}_Xc_i, \qquad \forall X \in \{E,M,H\} \quad \forall i,$$

$$0.001 - \big(\text{second smallest eigenvalue of } \operatorname{diag}\big(\mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\, e\big) - \mathcal{G}_{{}_X\mathcal{I}_i,{}_X\mathcal{I}_i}\big) \le 0, \quad \forall X \in \{E,M,H\}\ \forall i.$$