Exploring Tradeoffs in Automated School Redistricting: Computational and Ethical Perspectives

Fanglan Chen[1], Subhodip Biswas[1], Zhiqian Chen[2], Shuo Lei[1], Naren Ramakrishnan[1], Chang-Tien Lu[1]

[1]Department of Computer Science, Virginia Tech
[2]Department of Computer Science and Engineering, Mississippi State University
{fanglanc, subhodip, slei, naren, ctlu}@vt.edu, zchen@cse.msstate.edu

Abstract

The US public school system is administered by local school districts. Each district comprises a set of schools mapped to attendance zones which are annually assessed to meet enrollment objectives. To support school officials in redrawing school boundaries, existing approaches have proven promising but still suffer from several challenges, including 1) inability to scale to large school districts, 2) high computational cost of obtaining compact school attendance zones, and 3) lack of ethical considerations underlying the redrawing of school boundaries. Motivated by these challenges, this paper approaches the school redistricting problem from both computational and ethical standpoints. First, we introduce a practical framework based on Markov Chain Monte Carlo methods to solve school redistricting as a graph partitioning problem. Next, the advantages of adopting a modified objective function for optimizing discrete geometry to obtain compact boundaries are examined. Lastly, alternative metrics to address ethical considerations in real-world scenarios are formally defined and thoroughly discussed. Our findings highlight the inclusiveness and efficiency advantages of the designed framework and depict how tradeoffs need to be made to obtain qualitatively different school redistricting plans.

Introduction

In the United States, the public school system is operated locally by school districts. Each school district covers a geographical area usually coterminous with the county or city boundaries administered by publicly elected school boards (Sell 2006). For a given school district, the boundary of a *school attendance zone* designates the geographical area of students attending the same school. Each attendance zone spans multiple smaller-sized geographical units called *student planning areas* (Biswas et al. 2020b). Because school districts witness the inflow and outflow of student populations regularly, school boundaries need to be redrawn on an annual basis (Caro et al. 2004). This school redistricting process is traditionally conducted by urban planners and school officials with inputs from different stakeholders, such as parents and teachers. School attendance boundary design is a non-trivial task, not only due to the need to balance multiple criteria but also due to the potential involvement of multiple, overlapping, redistricting efforts.

Redrawing school boundaries is an intensive process that requires multiple levels of planning, inputs from the community, and tallying of different data sources (Kelly 2019). Even though school planners today are equipped with state-of-the-art geographic information system technologies like ArcGIS (Law and Collins 2015), they could be overburdened when a school district is involved in multiple school redistricting efforts in a single year. One of the key challenges in the process is the construction of a new plan, which remains largely manual despite significant advances in visualization and computational toolkits (Ferland and Guénette 1990; Caro et al. 2004; Bulka et al. 2007). This manual process is hard to scale, especially if the plan-making is to be performed at the level of the entire school district. Besides, consensuses among the stakeholders on what factors are considered important during the process are themselves difficult to reach (Gimpel and Harbridge-Yong 2020).

Existing work on school redistricting from the computational perspective usually approaches the task as a derived transportation optimization or continuous linear programming (LP) problem (Belford and Ratliff 1972; Liggett 1973; Caro et al. 2004; Bouzarth et al. 2018) so as to obtain an optimal or a near-optimal solution. For those seeking to address this issue, the major challenges can be summarized as follows. 1) *Scalability issues in large school districts.* Due to the discrete nature of geographical units, the school redistricting problem suffers from a combinatorial explosion. However, exact algorithms such as integer linear programming can only be used with small input. The ability of these methods to generate desirable school attendance zones, especially in large school districts, is limited. 2) *Inefficient metrics for discrete geometry.* As recognized as a traditional redistricting principle, the compactness of solution plans require precise definitions. For school redistricting, the Polsby-Popper score (Polsby and Popper 1991) is commonly used to measure the compactness of the generated plans. However, as a geographical measurement, the Polsby-Popper score is typically sensitive to the projection by the chosen coordinate system. More importantly, using the Polsby-Popper score to optimize the compactness of a given plan is very expensive because the perimeter and area of each school attendance zone need to be recalculated in each step. 3) *Insufficiency of ethical considerations incorporated in the optimization process.* In practice, the redistricting process is primarily driven by school capacity utilization, but serious considerations need to be given to the number of students displaced, demographic makeup at each school level, and related metrics (Bulka et al. 2007). Existing literature on automated school redistricting primarily focused on meeting school capacities and generating compact solution plans (Biswas et al. 2019, 2020b,a). The potential to incorporate ethical and fairness considerations to generate inclusive redistricting plans in a computational manner has not yet been fully explored.

To simultaneously overcome the aforementioned challenges, we introduce a practical framework to solve the school redistricting problem through a computational and ethical lens. The major contributions of this work can be summarized as follows:

- We design a Markov Chain Monte Carlo (MCMC)-based framework for solving the school redistricting problem. The framework adopts a greedy approach by continually searching for improved solutions one step further from the incumbent solution.
- We use the retained edge ratio as a measure of compactness instead of the classical Polsby-Popper score and the efficiency improvements for multiple algorithms through this modification are analyzed.
- We develop multiple ethical and fairness evaluation metrics and incorporate these considerations to support the decision-making process. Besides the balance and compactness scores in existing literature, we explore other important metrics of solution plans and incorporate these new metrics into the optimization process.
- We conduct extensive experiments on two US school districts and analyze the performance of the proposed framework w.r.t. balance, compactness, and ethical considerations. A case study is conducted to examine the approach's ability to obtain school redistricting plans with desirable properties.

Related Work

Research in computational redistricting can be traced back to the 1960s (Vickrey 1961; Hess et al. 1965). Since then, the redistricting task has been approached as an optimization problem (Duchin and Walch 2022), and a line of algorithms (Duchin and Walch 2022) have been proposed, including hill climbing and simulated-annealing (Altman and McDonald 2011), local search (King, Jacobson, and Sewell 2015), spatial evolutionary algorithms (Liu, Cho, and Wang 2016), and Voronoi diagrams (Levin and Friedler 2019; Gawrychowski et al. 2021). The second line of research has explored generating a large ensemble of redistricting plans with predefined desirable properties. These algorithmic approaches include Flood Fill (Cirincione, Darling, and O'Rourke 2000; Magleby and Mosesson 2018), Column Generation (Gurnee and Shmoys 2021), and Markov Chain Monte Carlo simulation (Tam Cho and Liu 2016; Fifield et al. 2020; DeFord, Duchin, and Solomon 2019; Procaccia and Tucker-Foltz 2022). As the aforementioned methods are primarily developed for political redistricting, they cannot be directly applied to school redistricting because the soft and hard constraints that capture the varied school capacity, fixed facility locations, and other problem-specific considerations are not easily satisfied in the computation.

As early attempts to solve the school redistricting problem, several studies formulated it as a continuous linear programming or derived transportation problem (Sutcliffe, Board, and Cheshire 1984; Belford and Ratliff 1972; Liggett 1973). Among the few works that have been published in this direction, Schoepfle and Church coined the term *Generic School Redistricting Problem* (Schoepfle and Church 1989) to generalize a series of school boundary problems in which students are assigned to schools while minimizing a proximity or cost function along with a set of balancing constraints. Caro et al. formulated school redistricting as a problem in integer programming that minimizes the overall student travel distance (Caro et al. 2004). This method was inspired by the sales territory alignment model (Zoltners and Sinha 1983) and the spatial connectivity is considered. However, these methods suffer from scalability issues and cannot perform on the whole school district.

To overcome the salable bottleneck of exact methods, Bulka et al. presented a heuristic search approach and information visualization techniques to generate and assess redistricting plans (Bulka et al. 2007). Along the same line of heuristics methods, the REGAL framework (Biswas et al. 2019) was proposed to tackle the school redistricting problem. Following an initial assignment of clusters, REGAL attempted to swap the boundary spatial units between clusters for an improved plan. GeoKM (Biswas et al. 2020b) is a spatial clustering method that handles the problem in a constrained setting through a hybrid strategy to assign polygons to clusters. As the assignment starts from scratch and geodesic distance is adopted as its selection criterion, the balance and compactness of generated plans by the algorithm cannot be guaranteed. The current state-of-the-art approach to the school redistricting problem is SPATIAL (Biswas et al. 2020a), a population-based metaheuristics method for solving spatial optimization problems. By leveraging domain knowledge in spatially-aware search, this algorithm facilitated the look for an optimal solution in discrete search space while satisfying the spatial constraints. However, the ability of these algorithms to achieve compact school attendance zones, especially in large school districts is limited due to the large computational cost of updating the compactness measure at each iteration. As balance and compactness are the primary focus of the methods, the incorporation of other important criteria is not fully discussed.

In this work, we expand the standard view of school redistricting as partitioning a planar graph on spatial units into contiguous, compact, balanced, and inclusive school attendance zones. Our goal is to introduce a flexible framework that could incorporate different objectives including ethical metrics and specific constraints to facilitate efficient and inclusive automatic school redistricting. Specifically, we explore a scalable MCMC-based approach to improve the quality and efficiency of generating school redistricting plans and assess the reduced computational cost of adopting a refined objective function for measuring the discrete geometry of the obtained school district.

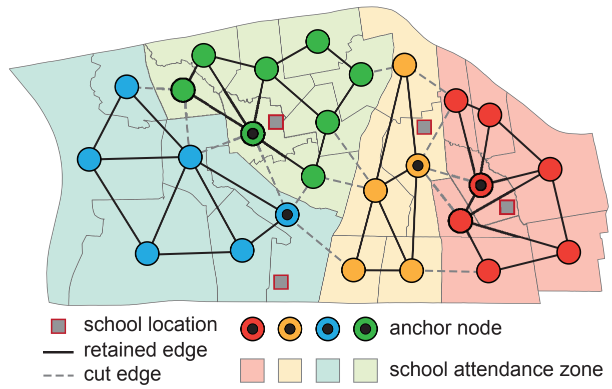

Figure 1: Transform a school district into a planar graph.

The School Redistricting Problem

Designing school boundaries can be understood as dividing a connected geographical area into a given number of contiguous school attendance zones. Figure 1 illustrates this basic idea through a small school district example. In such a planar graph, a node is a student planning area (SPA) and edges between two nodes imply their spatial contiguity. Thus, partitioning a school district into k school attendance zones (SAZs) is equivalent to removing sets of edges from a connected graph and forming k subgraphs.

Let $G = (V, E)$ denotes an undirected graph, and $V = \{1, 2, \dots, v\}$ is the set of nodes. Anchor nodes refer to a subset of spatial units, each of which contains one and only one school. The nodes in graph G have features including student population, area/perimeter of the spatial unit, and so forth. E is the set of undirected edges. $E_{i,j} \in E$ denotes the edge between the corresponding nodes of two contiguous spatial units i and j. In context of graph partitioning problem (Bichot and Siarry 2013), a school district G is partitioned into k school attendance zones $\xi = \{V_1, V_2, \dots, V_k\}$, and each zone V_s is a nonempty subset of V. The partition ξ induces a subgraph $G_s(\xi) = (V_s, E_s(\xi)) \subseteq G$ where an edge $E_{i,j} \in E(\xi)$ if the connected nodes i and j are assigned to the same partition V_s:

$$E(\xi) = \{E_{i,j} \in E : \exists V_s \in \xi \text{ s.t. } i, j \in V_s\}. \quad (1)$$

In a school district graph G, SAZs are obtained by removing edges from the set of edges E. In this work, we formulate the school redistricting problem as a modified graph partitioning function $\mathcal{F}(\xi) : E(\xi) \to E'(\xi)$ to generate partitions(in each solution plan) of desired quality.

Framework

In this section, we introduce an automated school redistricting framework as illustrated in Figure 2. The subsections detail our design of the modules, including seeding and repairing phases to construct valid initial plans, MCMC phase comprised of state transition and acceptance criteria, and the objective function for efficient and inclusive concerns.

Seeding

Different from other redistricting problems (Ricca and Simeone 2008; Ricca, Scozzari, and Simeone 2013), school redistricting is required to assign an anchor node and only to one partition. To satisfy this hard constraint, the seeding phase aims to identify the anchor nodes corresponding to the spatial units that contain schools and assign capacity as a node feature merely to those anchor nodes. A seed's corresponding school-level capacity C_s should be positive, and a partition should contain only one anchor node (Biswas et al. 2019). In addition, other hard constraints that each node is exclusively assigned to a partition should be satisfied.

Repairing

Spatial contiguity is another important requirement in designing school boundaries, so we need to ensure each subgraph is connected in initialization. Since we are focusing on improving the current school attendance zones, ideally, starting from the existing plan should satisfy all the constraints. However, we noticed that, in some school districts, the existing plan failed to satisfy the spatial contiguity constraint due to special real-world cases or geospatial data quality issues. Hence, we have to fix that plan to obtain a new valid plan only a few hops away from the starting plan. To achieve the goal, we adopt a path-linking method that helps to repair the solutions if they fall into the infeasible search space (Glover, Laguna, and Marti 2003; Biswas et al. 2020a). Upon repairment is finished, the initial plan is valid in regards to the aforementioned constraints and thus leads to an efficient exploration of the search space through the MCMC process.

State Transition

The state transition in a Markov chain starts from a given state(a full set of subgraphs), then determine the set S of all pairs $(E(\xi), E'(\xi))$ where $E(\xi)$ refers to the current graph partitions, and $E'(\xi)$ refers to the modified partitions obtained by choosing one pair $(E(\xi), E'(\xi))$ uniformly at random from the set S within the predefined transition space. The transition $E(\xi) \to E'(\xi)$ is conducted if $E'(\xi)$ satisfies the acceptance criteria which is explained in a later subsection. The transition space is completely determined by the design of transition proposals. To examine the potential of the proposed framework for school redistricting, two transition proposals, flip and recombination, are explored.

Flip. For the design of the flip transition proposal, we follow the simplest version, in which the assignment of a single boundary node is modified at each step in the chain under the contiguity constraint. To ensure spatial contiguity, it is intuitive to start by choosing an arbitrary boundary node, but this may potentially introduce non-uniformity to the process due to the varying degrees. Inspired by (Chikina, Frieze, and Pegden 2017), we sample uniformly from the set of (node, subgraph) pairs (i, V_s) where the node i resides on the boundary of subgraph V_s and there exists a cut edge $E_{i,v} \in \xi$ with $\xi(v) = V_s$ to enable a reversible chain. This can be understood as a simple random walk on the plan ξ, in which two states are connected if they are varied at one single boundary node.

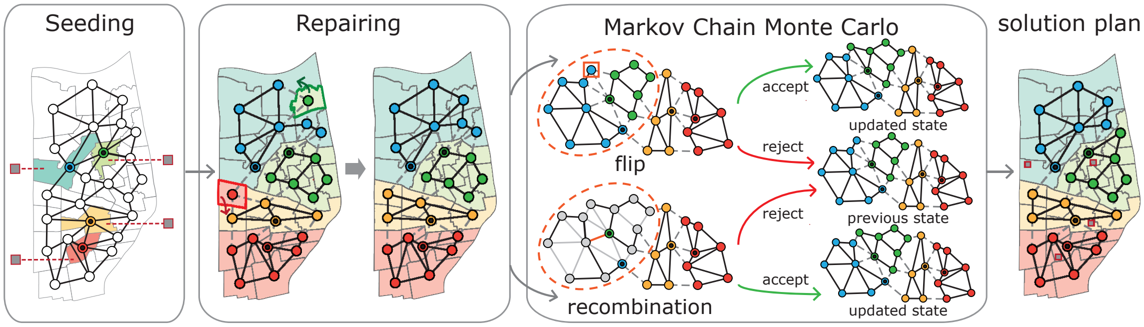

Figure 2: Proposed automated school redistricting framework.

Recombination. The second variant transition proposal is recombination which adopts a spanning tree approach to bipartition a merged graph (DeFord, Duchin, and Solomon 2019). The first step is to construct a spanning tree uniformly on the area merged by two randomly picked adjacent partitions of a graph. This form of recombination merely combines two adjacent partitions. Our experiments with ReCom-Chain employ the loop-erased random walk approach, and the details can be referred to Wilson's algorithm (Wilson 1996). The next step is to cut the merged graph to form two new subgraphs. The target is to find an edge cut in the constructed spanning tree to ensure the complementary partition is within the predefined tolerance of the population beyond or below the capacity of the affected school attendance zone within a certain number of node repeats. More than one edge that could be cut to generate new partitions with the predefined tolerance exists, so we perform uniform sampling among those potential cut edges.

Acceptance Criteria

In view of the school redistricting problem, spatial contiguity and each school attendance zone containing one and only one school are both hard constraints. The contiguity-based rejection sampling is instantiated through a function that checks the contiguity of the generated plan at each step. It returns *True* if the newly generated plan is contiguous, otherwise, it returns *False*. To incorporate the one anchor node for a SAZ constraint, we only assign the anchor nodes with positive capacity and 0 for the other nodes. This criterion can be satisfied by checking the accumulated capacity of each subgraph, and rejecting the transition if the capacity of any subgraph drops to 0 in the state transition phase.

In addition, we propose an objective function that returns a higher score to those redistricting plans that better adhere to the design criteria. The function is used to define the search space of generating redistricting plans during the MCMC process.

Balance. As one of the most important objectives in school redistricting, a school district's balance is defined as:

$$\text{BAL}\,(\xi) = \frac{1}{k}\sum_{s=1}^{s=k}\left|\frac{\sum_{v_i \in V_s} \text{pop}(v_i)}{\sum_{v_i \in V_s} \text{cap}(v_i)}\right|, \tag{2}$$

where $\text{pop}(v_i)$ and $\text{cap}(v_i)$ represent the population and capacity of node v_i, respectively.

Compactness. How tightly the area of a polygon shape is packed into its boundary is another criterion in assessing redistricting plans. Here, we discuss a widely used metric and a modified alternative metric based on graph theory.

Polsby-Popper. As the most popularly used compactness metric (Polsby and Popper 1991), the PP score of a school district is defined as:

$$\text{PP}\,(\xi) = \frac{1}{k}\sum_{s=1}^{s=k}\left|4\pi \cdot \frac{\text{Area}\,(G_s(\xi))}{\text{Peri}\,(G_s(\xi))^2}\right|, \tag{3}$$

where $\text{Area}\,(G_s(\xi))$ and $\text{Peri}\,(G_s(\xi))$ correspond to the area and perimeter of partition s. This score asymptotically approaches 1 as the partition shape approaches that of a circle.

Retained edge ratio. From a graph-theoretic standpoint, a less number of edges removed, the more will be the overall compactness of the plan (West et al. 2001; Becker and Solomon 2020). To incorporate this knowledge, we refined a retained edges ratio metric to compute the ratio of edges maintained to obtain a solution plan ξ as:

$$\text{RER}\,(\xi) = \frac{\sum_{s=1}^{s=k} \text{count}\,(E_s\,(\xi))}{\text{count}\,(E)}, \tag{4}$$

where plans with higher values of $\text{RER}\,(\xi)$ can be interpreted as plans with more compact partitions.

Ethical consideration. Besides balance and compactness, other metrics are also prevalent in the assessment of school redistricting plans. Through our discussion with urban planners and existing literature, we propose additional metrics which evaluate a plan from an ethical perspective.

Student diversity. To capture the racial/ethnic diversity of the student population, we consider employing the Shannon diversity index (White 1986). Another widely adopted diversity metric, Simpson's diversity index (Tóthmérész 1995), accounts primarily for the balanced proportion of the particular population in a sample. The Shannon diversity index is based on randomness present in an area and considers both population richness and balance in the distribution of a sample (Kim et al. 2017). In our case, the student population of racial/ethnic groups is inherently imbalanced. Hence, the inclusiveness of different racial/ethnic groups is our main focus. The normalized version of the Shannon diversity index at the partition level is defined as:

$$\text{SD}\,(G_s(\xi)) = -\frac{\sum_{j=1}^{c} p_j \cdot \ln\,(p_j)}{\ln(c)}, \tag{5}$$

	#SPA	#ES	#MS	#HS
School District A	453	57	17	16
School District B	1,313	138	26	24

Table 1: Summary statistics of the two school districts.

where c is the number of racial/ethnic student groups, which is 5 corresponding to ASIAN, BLACK, HISPANIC, WHITE, and OTHERS in our case. p_j denotes a particular racial/ethnic group's proportion of the student population. The student diversity at the plan level, addressing the diversity among the student population can be calculated as:

$$\text{SD}\left(\xi\right) = \frac{1}{k}\sum_{s=1}^{s=k} \text{SD}\left(G_s(\xi)\right), \tag{6}$$

where a large value of SD (ξ) indicates the presence of high student diversity across the school district.

Retained student ratio. In school redistricting, a number of students may get reassigned to new schools. In such a situation, students may lose social connections with their academic cohort. The impact of this reassignment can be captured by a quantitative comparison of the modified plan ξ with the existing plan ζ as:

$$\text{RSR}\left(\xi\right) = \frac{\sum_{s=1}^{s=k}\sum_{v \in V_s} I\left(\zeta(v) == \xi(v)\right) \cdot \text{pop}(v)}{\sum_{s=1}^{s=k}\sum_{v \in V_s} \text{pop}(v)}, \tag{7}$$

where $I(x)$ is an indicator function which evaluates to 1 if x is true, else it is 0.

Objective function. In this work, we examine the tradeoffs in improving the aforementioned considerations of the existing school district plan, and the objective function is defined as a weighted sum of three considerations:

$$\mathcal{F}(\xi) = \underbrace{\alpha \cdot \text{BAL}\left(\xi\right)}_{\text{population balance}} + \underbrace{\beta \cdot \text{COM}\left(\xi\right)}_{\text{boundary compactness}} + \underbrace{\gamma \cdot \text{ETH}\left(\xi\right)}_{\text{ethical consideration}}, \tag{8}$$

where α, β, and γ are positive weights with their sum as 1. COM(ξ) is calculated by either PP(ξ) or RER(ξ). ETH(ξ) can be composed of SD(ξ) or RSR(ξ) or a weighted sum of them. Potentially, other objectives such as student commute distance, test score distributions, and future enrollment projections at each school level can be easily incorporated into the proposed framework with data available and metric scales defined in the range from 0 to 1.

Experimentation

In this section, we evaluated the performance of our framework on two school district datasets and tradeoffs in incorporating alternative metrics in the objective function.

Experimental Setup

Datasets. The experiments were conducted on two US school districts, and the enrollment data for the school year 2020-21 was used for this study. The statistics of the two datasets are summarized in Table 1.

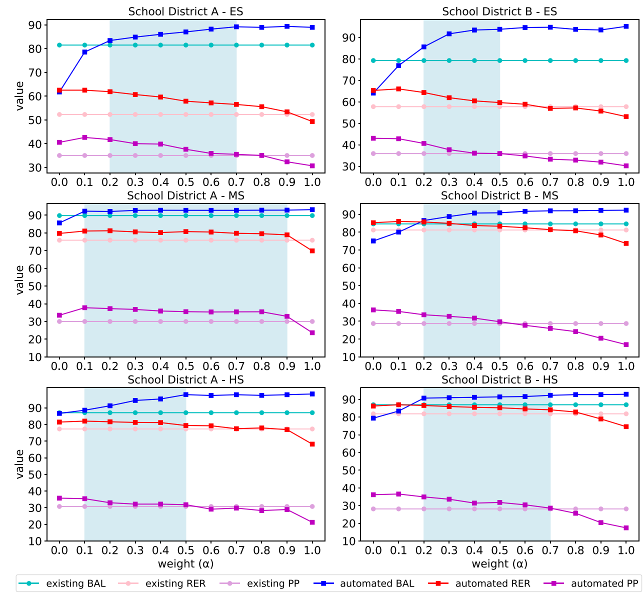

Figure 3: Study of different weight parameters.

Comparison models. We compared our framework with the following baselines that consider balance and compactness in their optimization objective, including: (1) *Stochastic Hill Climbing (SHC)* (Juels and Wattenberg 1995), (2) *Simulated Annealing (SA)* (Kirkpatrick, Gelatt, and Vecchi 1983), (3) *Tabu Search (TS)* (Glover and Laguna 1998), and (4) *SPATIAL* (Biswas et al. 2020a). Our framework was built on top of the GerryChain library[1]. For FlipChain, we set the number of steps to $10,000,000$. For RecomChain, we set the number of steps to $1,000,000$ with 10 node repeats in one step. To ensure a fair comparison, all the algorithms were made 25 runs with the same random seeds.

Tradeoff of Balance and Compactness

School redistricting performance. In this experiment, we primarily focus on the population balance and boundary compactness measured by retained edge ratio as the weight γ for ethical consideration in Equation 8 is set to 0. To appropriately choose the weight parameters α and β, we conducted a weight study as shown in Figure 3. Based on the plot, a higher share of balance in the objective results in a lower retained edge ratio. Also, the changing trend of the Polsby-Popper score mostly aligns with the retained edge ratio, which justifies our rationale for using retained edge ratio as an efficient compactness metric. We can also observe that beyond a specific setting of weight parameters, the gain of increase in balance is negligible, but the drop in retained edge ratio is significant. Hence, an appropriate setting of weight α is important. The light blue shaded area illustrates the desired range of weight setting in which the balance, retained edge ratio, and PP score are all above those of the existing plan. We choose 0.5 as the weights α and β in the objective function that combines balance and retained edge

[1]Available at https://github.com/mggg/GerryChain

School / Models	Elementary			Middle			High		
School District A									
	BAL	RER	PP	BAL	RER	PP	BAL	RER	PP
existing	83.50 (0.00)	50.80 (0.00)	32.53 (0.00)	89.74 (0.00)	75.44 (0.00)	26.77 (0.00)	87.07 (0.00)	76.88 (0.00)	27.35 (0.00)
SHC	86.00 (0.60)	56.52 (0.46)	34.82 (0.75)	92.61 (0.07)	79.82 (0.54)	32.65 (1.92)	96.25 (0.74)	79.45 (0.76)	29.93 (1.53)
SA	86.23 (0.87)	56.83 (1.16)	35.21 (1.45)	92.65 (0.15)	80.55 (1.06)	33.48 (2.57)	96.29 (1.18)	78.82 (0.89)	29.02 (1.53)
TS	84.50 (0.43)	56.57 (0.56)	34.99 (0.40)	92.60 (0.03)	79.99 (0.20)	32.33 (0.28)	96.79 (0.00)	78.74 (0.03)	29.07 (0.32)
SPATIAL	86.56 (0.58)	56.53 (0.62)	35.05 (0.75)	92.71 (0.05)	81.17 (0.29)	35.76 (1.00)	**97.42 (0.79)**	80.43 (0.72)	30.68 (1.25)
FlipChain	87.60 (0.50)	58.80 (0.54)	37.68 (1.43)	**92.72 (0.05)**	80.19 (0.46)	36.20 (1.74)	96.61 (0.77)	79.52 (0.85)	31.64 (1.16)
RecomChain	**88.05 (0.47)**	**60.43 (0.62)**	**40.22 (1.29)**	92.67 (0.86)	**82.08 (0.50)**	**39.38 (1.94)**	97.36 (0.56)	**81.16 (0.45)**	**35.16 (1.43)**
School District B									
	BAL	RER	PP	BAL	RER	PP	BAL	RER	PP
existing	82.11 (0.00)	56.35 (0.00)	35.92 (0.00)	84.23 (0.00)	81.27 (0.00)	27.71 (0.00)	86.95 (0.00)	81.80 (0.00)	26.80 (0.00)
SHC	86.22 (0.86)	59.97 (0.26)	37.63 (0.43)	88.23 (1.34)	84.63 (0.35)	29.44 (0.83)	86.87 (2.19)	85.57 (0.32)	29.76 (0.81)
SA	87.14 (1.35)	59.87 (0.31)	37.04 (0.53)	90.93 (0.70)	84.98 (0.41)	30.04 (1.03)	90.97 (1.88)	86.13 (0.41)	29.73 (1.22)
TS	84.61 (0.54)	59.99 (0.16)	38.07 (0.32)	87.50 (1.11)	84.64 (0.19)	30.03 (0.83)	85.94 (1.83)	85.83 (0.14)	30.16 (0.32)
SPATIAL	90.85 (0.61)	59.24 (0.39)	36.09 (0.54)	91.68 (0.11)	84.56 (0.22)	29.94 (1.12)	91.47 (0.10)	86.47 (0.17)	30.51 (0.49)
FlipChain	93.67 (0.52)	59.50 (0.28)	35.52 (0.58)	91.56 (0.28)	83.71 (0.51)	30.36 (1.49)	91.40 (0.14)	85.37 (0.28)	31.69 (1.53)
RecomChain	**94.25 (0.54)**	**61.18 (0.30)**	**38.27 (0.68)**	**92.20 (0.12)**	**85.25 (0.47)**	**32.54 (0.92)**	**92.47 (0.10)**	**86.53 (0.33)**	**34.55 (1.04)**

Table 2: A comparison of the automated plan generated and the existing plan (all metrics reported in %).

Model / Objective	Elementary			Middle			High		
	SPATIAL	FlipChain	RecomChain	SPATIAL	FlipChain	RecomChain	SPATIAL	FlipChain	RecomChain
Retained Edge Ratio	175.36	50.07	70.45	102.66	60.35	1770.82	65.04	63.65	1891.97
Polsby-Popper score	443.80	66.19	87.82	228.67	67.40	2194.94	119.31	70.51	2266.06

Table 3: A comparison of the computational time (minute/run) of methods with varied compactness measures.

ratio across three school levels. The PP score is reported for analytical purposes. Under this setting of weights, the performance metrics are tabulated in Table 2.

We observe that both FlipChain and RecomChain achieve better performance compared to the baseline methods. In the comparison with the SPATIAL algorithm, while the improvement is marginal in terms of balance, there is a noticeable improvement in the retained edge ratio and PP score by the proposed framework. The recombination proposal adopted in RecomChain results in more compact partitions as compared to FlipChain. The performance superiority of the proposed framework is more obvious in elementary schools, in which the number of school attendance zones is much more than that of the other two levels of schools.

Runtime efficiency. To examine the efficiency of the alternative compactness metric, the computational time for the top three methods and their variants of adopting retained edge ratio instead of the Polsby-Popper score on School District B was studied. Based on Table 3, FlipChain is the most efficient approach in regards to the time cost per one run and minimal difference on different levels of schools with both of the compactness measures in the combined optimization objective. Although RecomChain obtains the best performance reported in Table 2, its computational time is comparatively long, especially in the high and middle schools. The main bottleneck is that the construction of a spanning tree grows exponentially with an increased number of spatial units per partition. RecomChain is comparatively efficient with FlipChain and SPATIAL in redrawing elementary school boundaries because the number of spatial units per partition is much smaller as the number of elementary schools is much larger than the other two school levels. In addition, the average running time of all the methods was greatly reduced by adopting retained edge ratio as an alternative metric to optimize compactness. The efficiency of the SPATIAL approach also received a huge improvement with the proposed metric compared to the traditional Polsby-Popper score with an average 119.75% improvement across three school level. It also brings an average 20.51% efficiency enhancement for FlipChain and RecomChain.

Ethical considerations. The promise of automated plan generating entails ethical considerations of the process itself. A plan that displaces a higher proportion of students is likely to cause dissent amongst parents and students alike. Similarly, a plan that decreases student diversity is not preferred. To this effect, we analyze the automated plans in terms of two metrics that have strong implications for the stakeholders. We compare the ethical metrics of baselines with FlipChain (FC) and RecomChain (RC) in Figure 4.

On closer inspection, we notice that the plans produced by FlipChain displace a lower number of students, especially for high and middle schools. It is thus desirable when planners aim to minimize student displacements. It is worth mentioning that even though RecomChain achieves the best balance, retained edge ratio, and PP score, as reported by Table 2, its retained student ratio is the worst due to the recombination proposal that has the potential to make lots of assignment modifications at each step. Also, all the algorithms achieve close student diversity scores, which indicates that good diversity might be difficult to obtain.

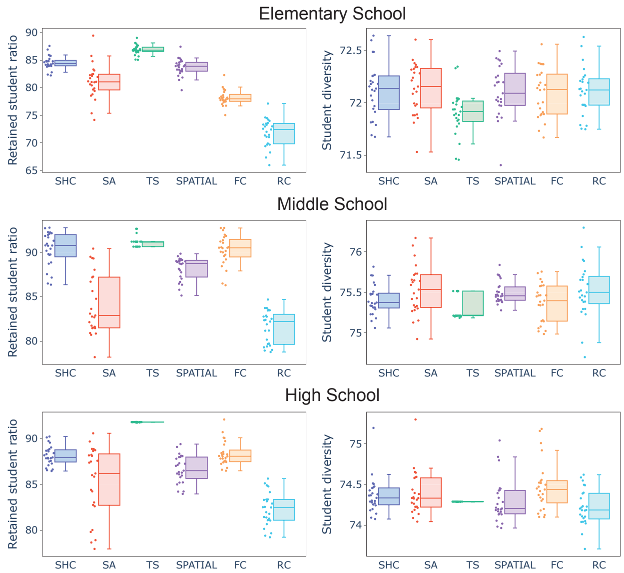

Figure 4: A comparison of automated plans in terms of ethical metrics—retained student ratio and student diversity.

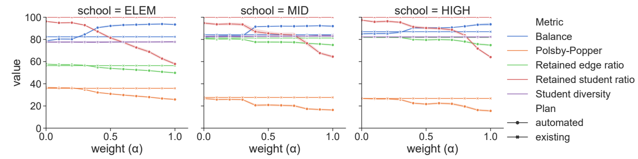

Figure 5: Tradeoff plot of the different performance metrics corresponding to the plans with varied weights.

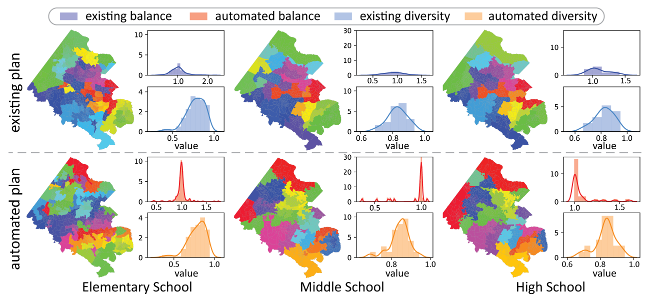

Figure 6: A comparison of the existing plans and automated plans generated by FlipChain.

Tradeoff of Balance and Retained Student Ratio

We also explored the feasibility to find a balance between the population balance and retained students ratio. To do this, we varied the value of weight α in intervals of 0.1, and simulated 25 runs for each value for school district B. Besides our defined metrics, we also kept showing the existing plan to serve as a reference. The results are shown in Figure 5. It is observed that when the weight of the balance objective is set close to 0, the plans are generated by maximizing the retained student ratio. As the weight increases, we hardly observe a change in any of the metrics until it reaches the value of 0.3. Along with a gradual increase of weight α, we can observe a noticeable deterioration in metrics including the retained student ratio, retained edge ratio, and PP score, while balance receives a slight improvement. This changes when the weight α goes beyond 0.7, where the retained student ratio witnesses a sharp decrease, which denotes diminishing returns on prioritizing balance over student displacement above a certain threshold. Similarly, decreasing compactness scores are also observed. Comparatively, no obvious variation is observed in the student diversity across the varied weight parameters. Additional considerations such as socioeconomic factors can also be included depending on the design criteria of school attendance boundaries.

Case Study

In a real-world scenario, both a balanced capacity and a diverse student population are desirable in a school district. As a case study, we explored the feasibility of using the proposed FlipChain to improve balance and student diversity (with both α and γ set to 0.5 and β set to 0 in Equation 8) with the lower bound of retained student ratio set to 85% in School District B. We used a color ramp to show the different assignments of SPAs in the existing and automated plans across three school levels in Figure 6. In each distribution plot, the x-axis is the value of balance/diversity scores, and the y-axis shows the count of school attendance zones with certain values. Since we define both of the scores in the range from 0 to 1 and formulate the school redistricting as a maximization problem, a plan is more desirable as more zones at each school level obtain balance/diversity scores close to 1. The distribution plots reflect improved balance and student diversity in the automated plans. However, the automated approach results in more arbitrarily shaped school boundaries as compactness is not optimized in this case. This case study demonstrates the potential of using the proposed framework to generate multiple qualitatively different school redistricting plans. Due to the multi-faceted nature of the process, school redistricting can also be thought of as a multi-objective optimization problem where multiple varying (often conflicting) criteria need to be balanced. In such instances, the set of solutions produced by the framework can be used to approximate the Pareto-optimal front.

Conclusion

In this paper, we have discussed a framework for automated school redistricting and analyzed multiple considerations that go into this process. As such, we have introduced an MCMC-based redistricting framework and investigated how it can improve the automated generation of school attendance zones. Further, we have demonstrated the advantages of replacing a widely used compactness measurement with an alternative metric. In addition, we have designed additional metrics motivated by ethical considerations. We hope this research can aid urban planners and school officials in finding multiple qualitatively different redistricting plans that represent different tradeoffs in decision-making and facilitating better utilization of resources in schools by promoting better student-teacher ratio and classroom climate.

Acknowledgements

This research was supported in parts by National Science Foundation (NSF) grants DGE-1545362, IIS-1633363, IIS-2153369, and CNS-2141095. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of any school board, NSF, or the U.S. Government.

References

Altman, M.; and McDonald, M. P. 2011. BARD: Better automated redistricting. *Journal of Statistical Software*, 42: 1–28.

Becker, A.; and Solomon, J. 2020. Redistricting Algorithms. *CoRR*, abs/2011.09504.

Belford, P. C.; and Ratliff, H. D. 1972. A network-flow model for racially balancing schools. *Operations Research*, 20(3): 619–628.

Bichot, C.-E.; and Siarry, P. 2013. *Graph partitioning*. John Wiley & Sons.

Biswas, S.; Chen, F.; Chen, Z.; Lu, C.-T.; and Ramakrishnan, N. 2020a. Incorporating domain knowledge into Memetic Algorithms for solving Spatial Optimization problems. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems*, 25–35.

Biswas, S.; Chen, F.; Chen, Z.; Sistrunk, A.; Self, N.; Lu, C.-T.; and Ramakrishnan, N. 2019. REGAL: A Regionalization framework for school boundaries. In *Proceedings of the 27th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, 544–547.

Biswas, S.; Chen, F.; Sistrunk, A.; Muthiah, S.; Chen, Z.; Self, N.; Lu, C.-T.; and Ramakrishnan, N. 2020b. Geospatial clustering for balanced and proximal schools. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 13358–13365.

Bouzarth, E. L.; Forrester, R.; Hutson, K. R.; and Reddoch, L. 2018. Assigning students to schools to minimize both transportation costs and socioeconomic variation between schools. *Socio-Economic Planning Sciences*, 64: 1–8.

Bulka, B.; Carr, R.; Jordan, E.; Rheingans, P.; et al. 2007. Heuristic search and information visualization methods for school redistricting. *AI Magazine*, 28(3): 59–59.

Caro, F.; Shirabe, T.; Guignard, M.; and Weintraub, A. 2004. School redistricting: Embedding GIS tools with integer programming. *Journal of the Operational Research Society*, 55(8): 836–849.

Chikina, M.; Frieze, A.; and Pegden, W. 2017. Assessing significance in a Markov chain without mixing. *Proceedings of the National Academy of Sciences*, 114(11): 2860–2864.

Cirincione, C.; Darling, T. A.; and O'Rourke, T. G. 2000. Assessing South Carolina's 1990s congressional districting. *Political Geography*, 19(2): 189–211.

DeFord, D.; Duchin, M.; and Solomon, J. 2019. Recombination: A family of Markov chains for redistricting. *arXiv preprint arXiv:1911.05725*.

Duchin, M.; and Walch, O. 2022. Correction to: Political Geometry–Rethinking Redistricting in the US with Math, Law, and Everything In Between. In *Political Geometry*, C1–C1. Springer.

Ferland, J. A.; and Guénette, G. 1990. Decision support system for the school districting problem. *Operations Research*, 38(1): 15–21.

Fifield, B.; Higgins, M.; Imai, K.; and Tarr, A. 2020. Automated redistricting simulation using Markov chain Monte Carlo. *Journal of Computational and Graphical Statistics*, 29(4): 715–728.

Gawrychowski, P.; Kaplan, H.; Mozes, S.; Sharir, M.; and Weimann, O. 2021. Voronoi Diagrams on Planar Graphs, and Computing the Diameter in Deterministic O(n^5/3) Time. *SIAM Journal on Computing*, 50(2): 509–554.

Gimpel, J. G.; and Harbridge-Yong, L. 2020. Conflicting Goals of Redistricting: Do Districts That Maximize Competition Reckon with Communities of Interest? *Election Law Journal: Rules, Politics, and Policy*, 19(4): 451–471.

Glover, F.; and Laguna, M. 1998. Tabu search. In *Handbook of combinatorial optimization*, 2093–2229. Springer.

Glover, F.; Laguna, M.; and Marti, R. 2003. Scatter search and path relinking: Advances and applications. *Handbook of metaheuristics*, 1–35.

Gurnee, W.; and Shmoys, D. B. 2021. Fairmandering: A column generation heuristic for fairness-optimized political districting. In *SIAM Conference on Applied and Computational Discrete Algorithms (ACDA21)*, 88–99. SIAM.

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

Juels, A.; and Wattenberg, M. 1995. Stochastic Hillclimbing as a Baseline Method for Evaluating Genetic Algorithms. In *Proceedings of the 8th International Conference on Neural Information Processing Systems*, NIPS'95, 430–436. Cambridge, MA, USA: MIT Press.

Kelly, M. G. 2019. A Map Is More Than Just a Graph: Geospatial Educational Research and the Importance of Historical Context. *AERA Open*, 5(1): 2332858419833346.

Kim, B.-R.; Shin, J.; Guevarra, R. B.; Lee, J. H.; Kim, D. W.; Seol, K.-H.; Lee, J.-H.; Kim, H. B.; and Isaacson, R. E. 2017. Deciphering diversity indices for a better understanding of microbial communities. *Journal of Microbiology and Biotechnology*, 27(12): 2089–2093.

King, D. M.; Jacobson, S. H.; and Sewell, E. C. 2015. Efficient geo-graph contiguity and hole algorithms for geographic zoning and dynamic plane graph partitioning. *Mathematical Programming*, 149(1): 425–457.

Kirkpatrick, S.; Gelatt, C. D.; and Vecchi, M. P. 1983. Optimization by simulated annealing. *science*, 220(4598): 671–680.

Law, M.; and Collins, A. 2015. *Getting to know ArcGIS*. ESRI press Redlands.

Levin, H. A.; and Friedler, S. A. 2019. Automated congressional redistricting. *Journal of Experimental Algorithmics (JEA)*, 24: 1–24.

Liggett, R. S. 1973. The application of an implicit enumeration algorithm to the school desegregation problem. *Management Science*, 20(2): 159–168.

Liu, Y. Y.; Cho, W. K. T.; and Wang, S. 2016. PEAR: a massively parallel evolutionary computation approach for political redistricting optimization and analysis. *Swarm and Evolutionary Computation*, 30: 78–92.

Magleby, D. B.; and Mosesson, D. B. 2018. A new approach for developing neutral redistricting plans. *Political Analysis*, 26(2): 147–167.

Polsby, D. D.; and Popper, R. D. 1991. The third criterion: Compactness as a procedural safeguard against partisan gerrymandering. *Yale L. & Pol'y Rev.*, 9: 301.

Procaccia, A. D.; and Tucker-Foltz, J. 2022. Compact Redistricting Plans Have Many Spanning Trees. In *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 3754–3771. SIAM.

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

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

Schoepfle, O. B.; and Church, R. L. 1989. A fast, network-based, hybrid heuristic for the assignment of students to schools. *Journal of the Operational Research Society*, 40(11): 1029–1040.

Sell, S. 2006. Running an effective school district: School boards in the 21st century. *Journal of Education*, 186(3): 71–97.

Sutcliffe, C.; Board, J.; and Cheshire, P. 1984. Goal programming and allocating children to secondary schools in Reading. *Journal of the Operational Research Society*, 35(8): 719–730.

Tam Cho, W. K.; and Liu, Y. Y. 2016. Toward a talismanic redistricting tool: A computational method for identifying extreme redistricting plans. *Election Law Journal*, 15(4): 351–366.

Tóthmérész, B. 1995. Comparison of different methods for diversity ordering. *Journal of vegetation Science*, 6(2): 283–290.

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

West, D. B.; et al. 2001. *Introduction to graph theory*, volume 2. Prentice hall Upper Saddle River.

White, M. J. 1986. Segregation and diversity measures in population distribution. *Population index*, 198–221.

Wilson, D. B. 1996. Generating random spanning trees more quickly than the cover time. In *Proceedings of the twenty-eighth annual ACM symposium on Theory of computing*, 296–303.

Zoltners, A. A.; and Sinha, P. 1983. Sales territory alignment: A review and model. *Management Science*, 29(11): 1237–1256.