Exploring Tradeoffs in Automated School Redistricting: Computational and Ethical Perspectives

Fanglan Chen, Subhodip Biswas, Zhiqian Chen,
Shuo Lei, Naren Ramakrishnan, Chang-Tien Lu

The Thirteenth Symposium on Educational Advances in Artificial Intelligence (EAAI-23)

Outline

- Introduction
- Problem Formulation
- Proposed Framework
- Experiments
- Conclusion

Introduction

Draw Attendance Boundaries — Group small spatial units, *Student Planning Areas* (**SPA**s), to form school attendance boundaries, *School Attendance Zones* (**SAZ**s), such that the areas inside a region are geographically contiguous.

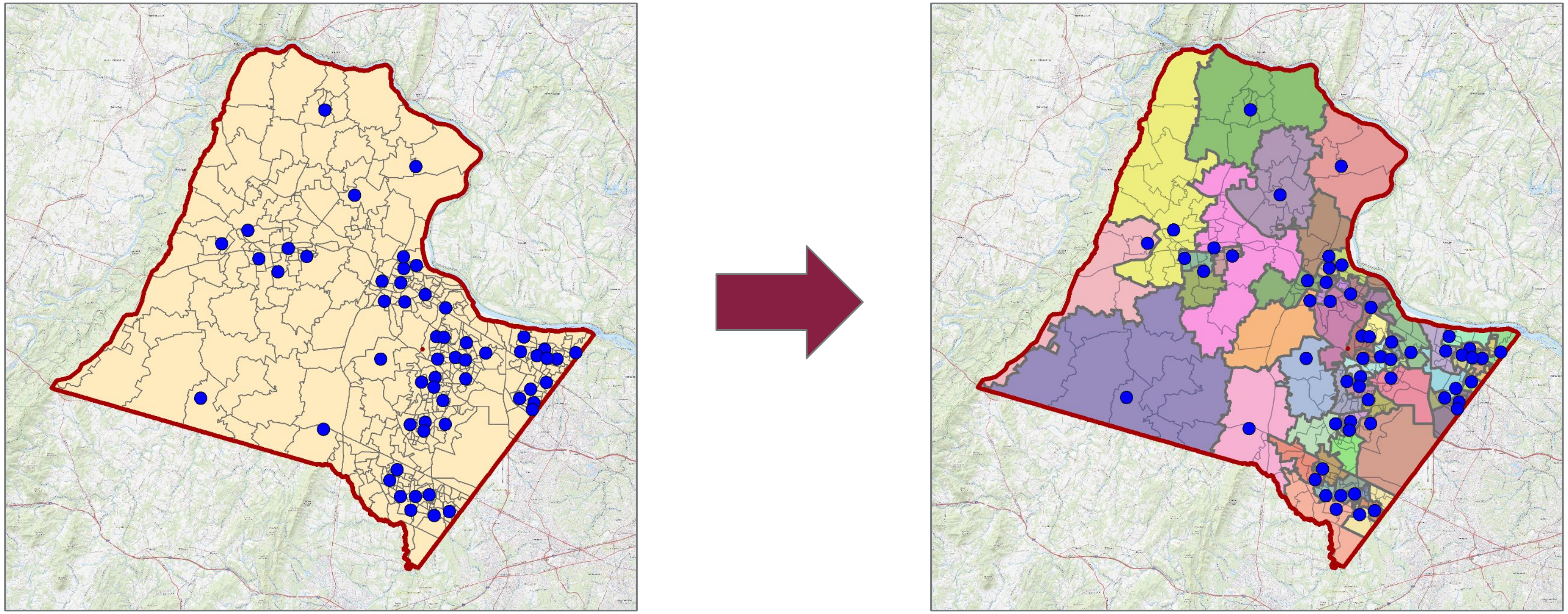


Figure 1. A GIS visualization showing the school district of Loudoun County, VA.

School Redistricting

Non-trivial Task

- Multi-level plan-making
- balance multiple criteria

Question: can we automatically generate school redistricting plans to facilitate the school redistricting efforts?

Challenges

- Scalability issues in large school districts
- Inefficient metrics for discrete geometry
- Insufficiency of ethical considerations

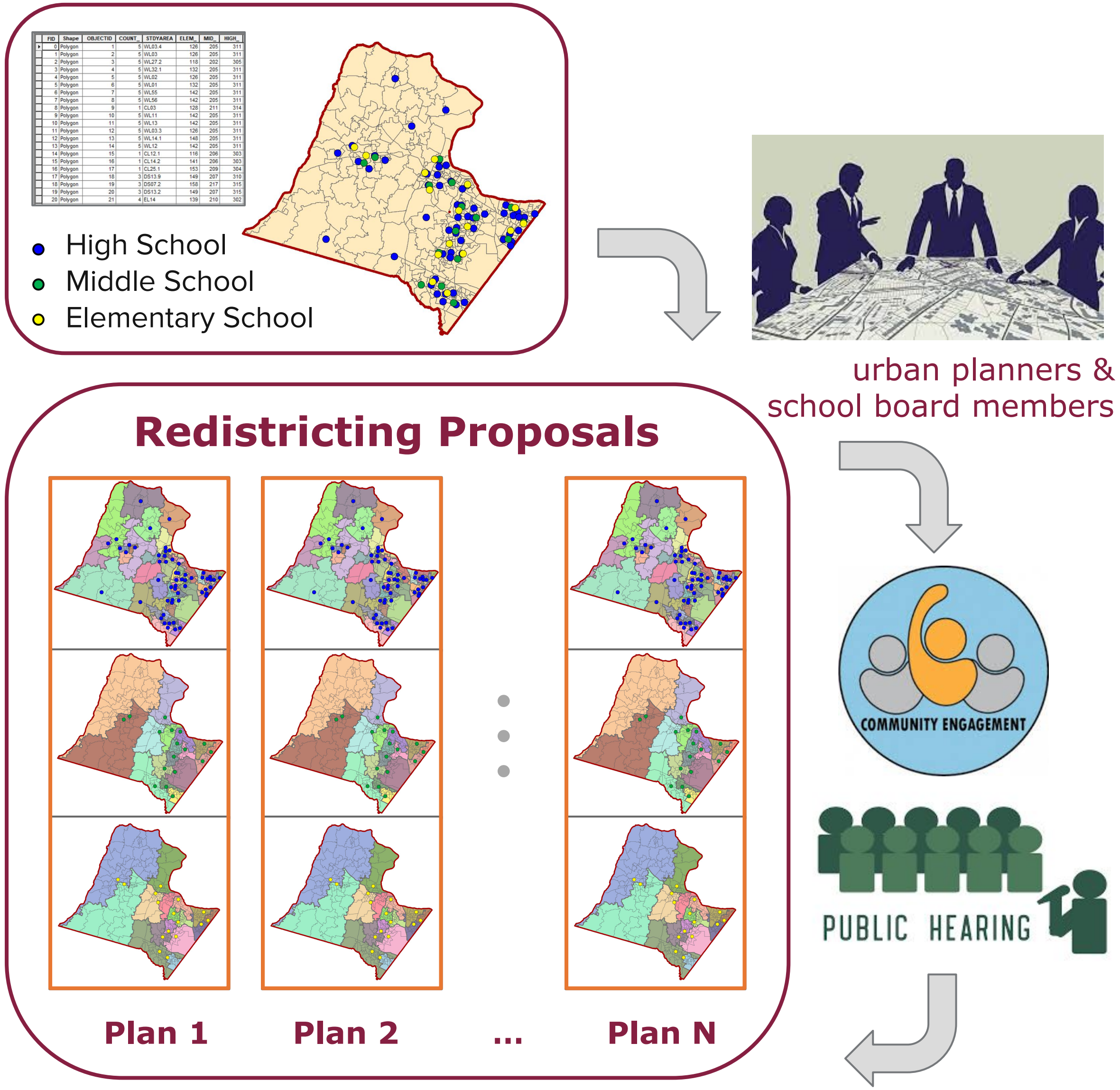


Figure 2. School redistricting process

Problem Formulation

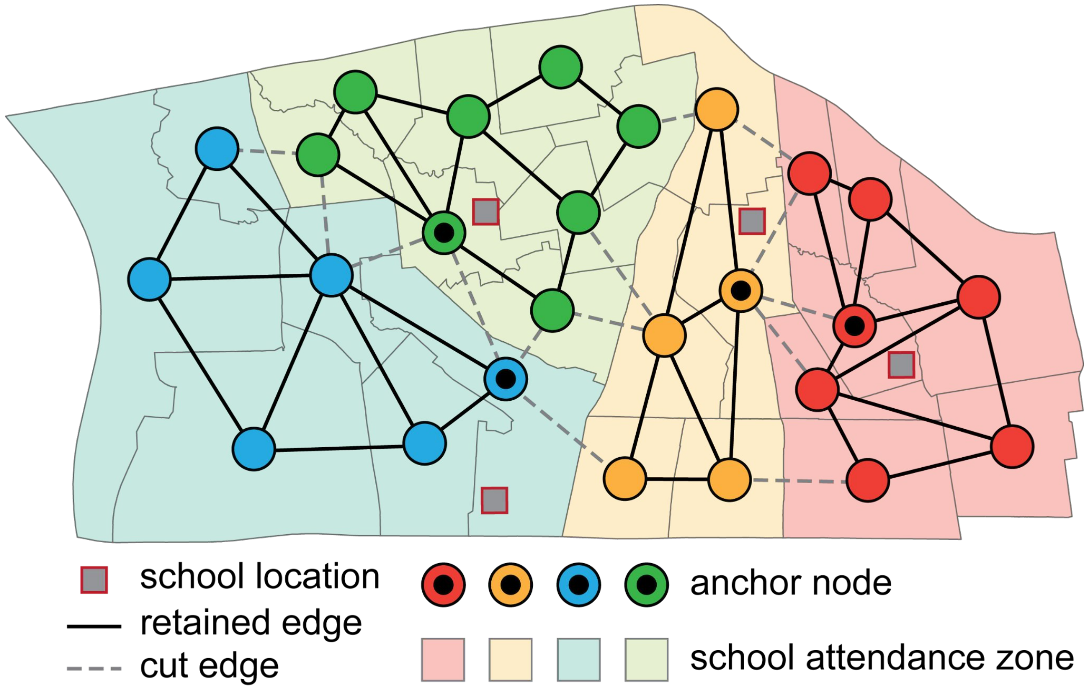


Figure 3. Transform a school district into a planar graph. A small example illustrating the partitioning of N nodes into K connected subgraphs. (N = 26 and K = 4 in the case)

Node	Subgraph	Graph
Student Planning Areas (SPAs)	School Attendance Zones (SAZs)	School boundary configuration
Spatial units	School boundaries	School district plan

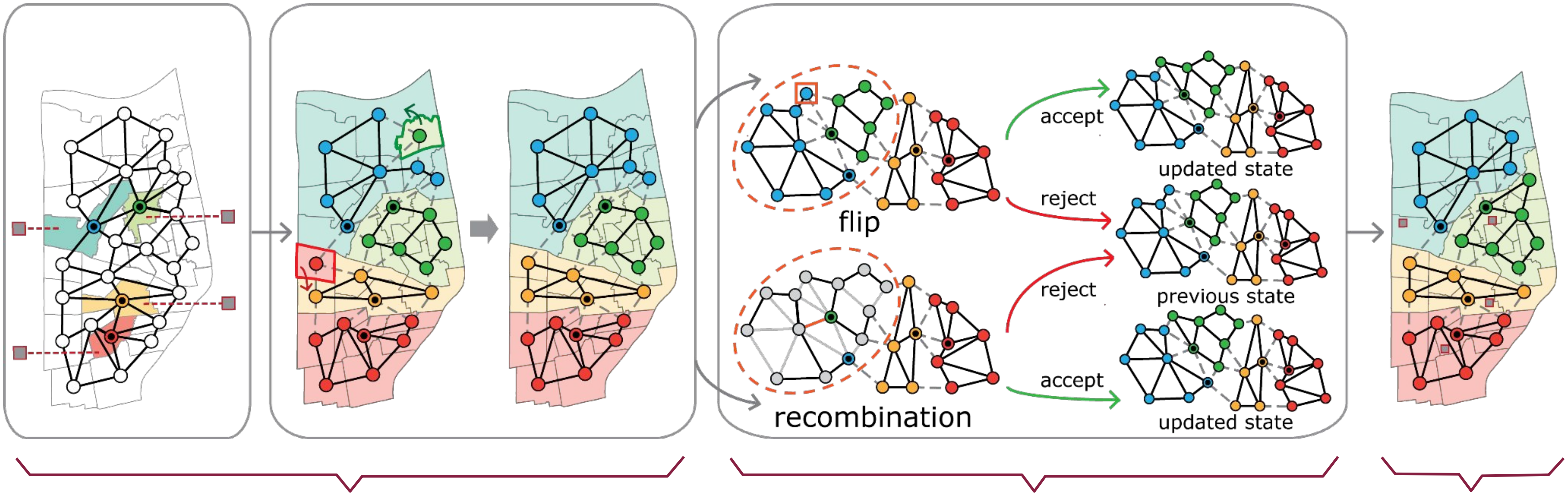
Problem objective: $\underset{\xi}{\text{maxmize}} \; \mathcal{F}(\xi)$

Constraints:

- Each node can be assigned to only one subgraph
- Each subgraph can contain only one anchor node
- Every subgraph must be connected

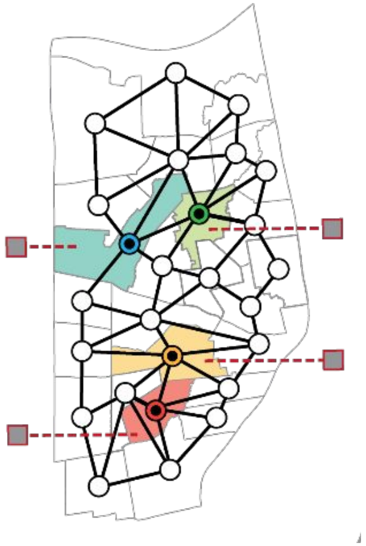
VIRGINIA TECH

Proposed Framework



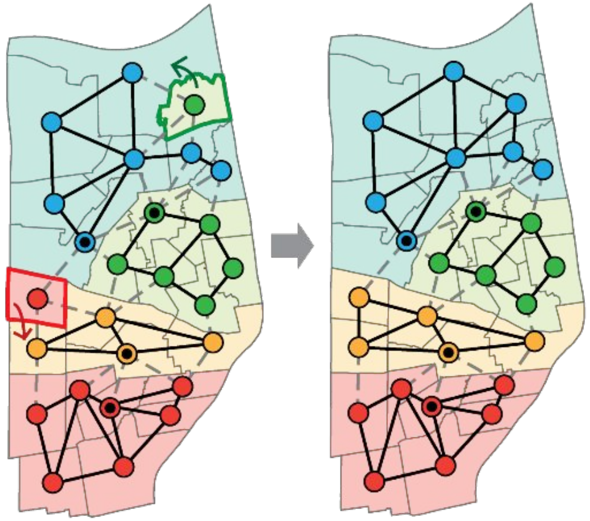
initialization improvement Solution Plan

Initialization



Identify the anchor nodes corresponding to the spatial units

- Assign capacity as a node feature merely to those anchor nodes
- Use existing plan as the starting plan



Ensure each subgraph is connected

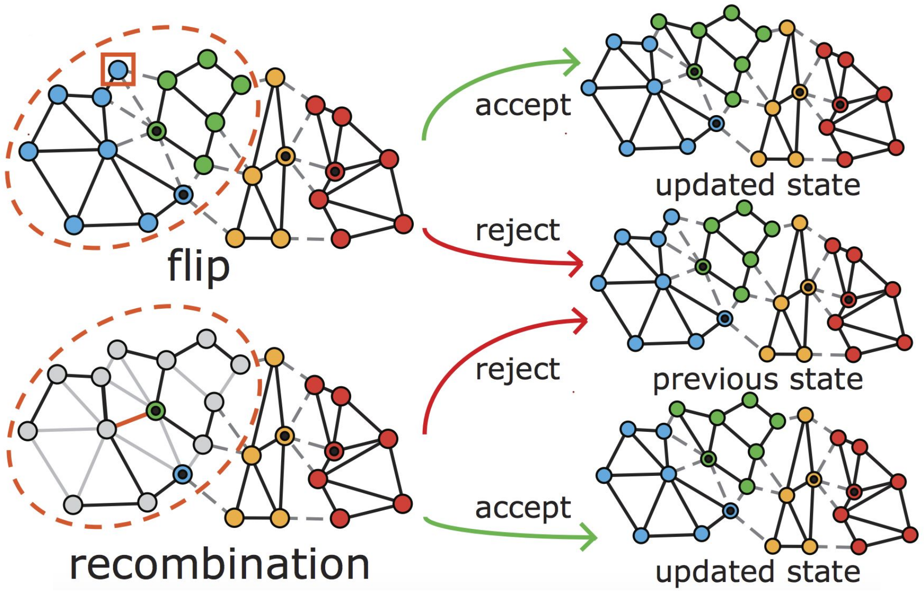
- Fix that plan to obtain a new valid plan only a few hops away from the starting plan
- Adopt a path-linking method

- **Valid plan**

  Satisfy all the constraints

  Lead to exploration of the search space

Improvement



Improve the quality of plan though Markov chain.

Each step in transition function is:

1. From a given state, determine the set S of all pairs (ξ, ξ′) where ξ refers to the current node assignment, and ξ′ refers to the modified assignment obtained by transition proposals.
2. From S, choose one pair (ξ, ξ′) uniformly at random.
3. Transition ξ → ξ′ if ξ′ satisfies the acceptance criteria.

- **Acceptance criteria**

  Pass the contiguity check

  Improve the objective

Objective Function

Linear combination of objectives:

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

$$\text{where} \quad \alpha + \beta + \gamma = 1.0$$

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

$$\mathrm{SD}\left(G_s(\xi)\right) = -\frac{\sum_{j=1}^{c} p_j \cdot \ln(p_j)}{\ln(c)}$$

$$\mathrm{SD}(\xi) = \frac{1}{k} \sum_{s=1}^{s=k} \mathrm{SD}\left(G_s(\xi)\right)$$

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

measures how well-utilized is a school's capacity with respect to its student enrollment.

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

$$\mathrm{RER}(\xi) = \frac{\sum_{s=1}^{s=k} \mathrm{count}\left(E_s(\xi)\right)}{\mathrm{count}(E)}$$

measures how the school attendance zones are tightly packed for a consideration of proximity

measures the percentage of students need to be replaced due to redistricting and student diversity index.

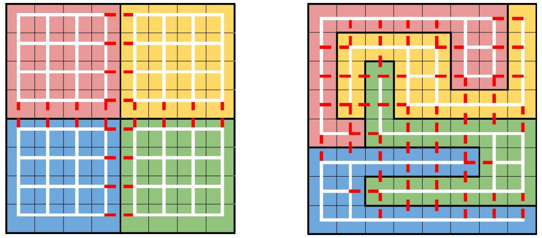


Figure 4. Two ways to partition a graph into 4 subgraphs.

Image courtesy: https://fisherzachary.github.io/public/r-output.html

VIRGINIA TECH

Experiments

- ## Datasets

Table 1. Summary statistics of the two school districts for the year 2020-21.

School District	# SPA	# Schools		
		Elementary	Middle	High
A	453	57	17	16
B	1,313	138	26	24

- ## Baselines

Heuristic methods: Stochastic Hill Climbing (SHC), Simulated Annealing (SA), Tabu Search (TS).

Meta-heuristic method: Swarm-based sPAtial memeTIc ALgorithm (SPATIAL).

- ## Settings

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.

All the algorithms were made 25 runs with the same random seeds.

Note: Our proposed framework is built on the GerryChain library(https://github.com/mggg/GerryChain) by the Metric Geometry and Gerrymandering Group.

Tradeoff of balance and compactness

Weight study

$$\mathcal{F}(\xi) = \alpha \cdot \mathrm{BAL}\,(\xi) \;\; + (1-\alpha) \cdot \mathrm{RER}\,(\xi) + 0 \cdot \mathrm{ETH}\,(\xi)$$

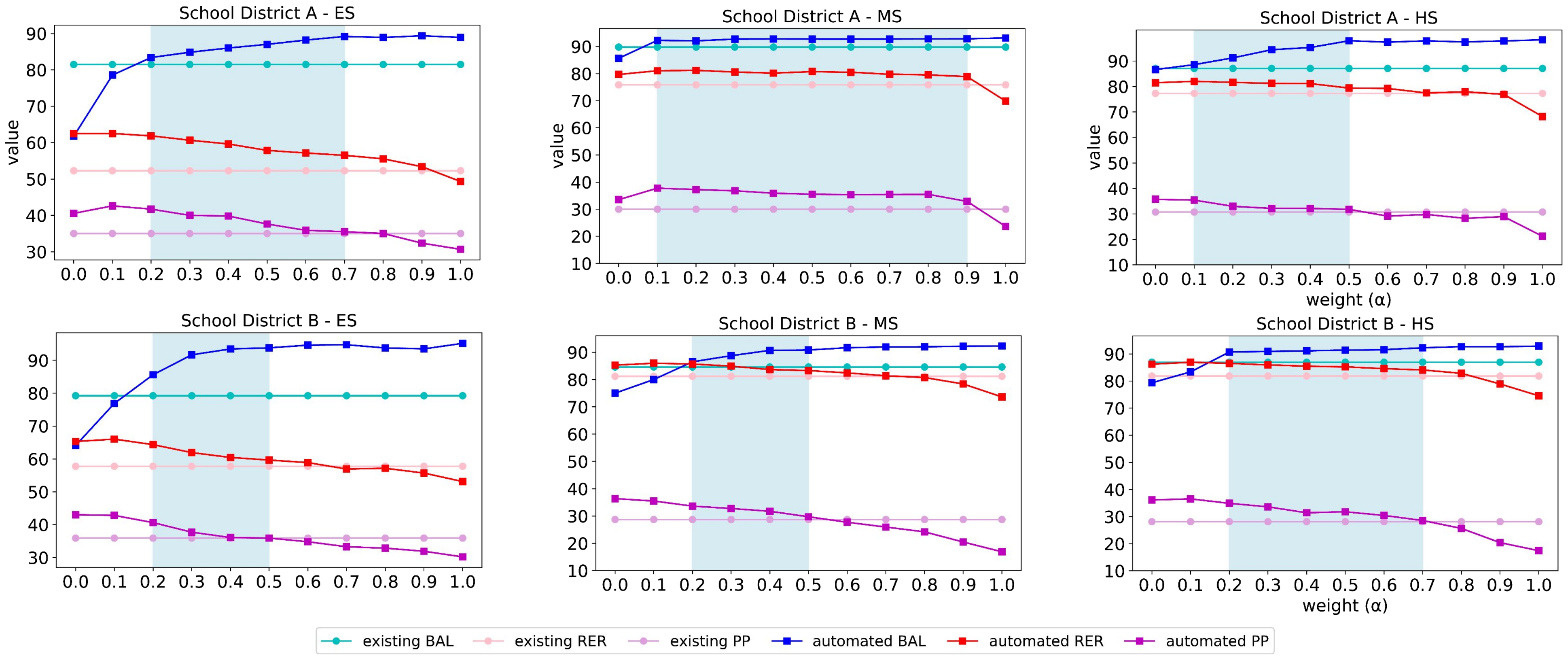


Figure 5. Study of different weight parameters.

VIRGINIA TECH

Tradeoff of balance and compactness

School redistricting performance

$$\mathcal{F}(\xi) = 0.5 \cdot \text{BAL}(\xi) + 0.5 \cdot \text{RER}(\xi)$$

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

School District A

Models \ School	Elementary BAL	Elementary RER	Elementary PP	Middle BAL	Middle RER	Middle PP	High BAL	High RER	High 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

Models \ School	Elementary BAL	Elementary RER	Elementary PP	Middle BAL	Middle RER	Middle PP	High BAL	High RER	High 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)**

Tradeoff of balance and compactness

Runtime study

$$\mathcal{F}(\xi) = 0.5 \cdot \mathrm{BAL}(\xi) + 0.5 \cdot \mathrm{RER}(\xi)$$

$$\mathcal{F}(\xi) = 0.5 \cdot \mathrm{BAL}(\xi) + 0.5 \cdot \mathrm{PP}(\xi)$$

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

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
	(-60.49%	-24.35%	-19.78%	-55.11%	-10.46%	-19.32%	-45.49%	-9.73%	-16.51%)

Tradeoff of balance and retained student ratio

$$\mathcal{F}(\xi) = \alpha \cdot \text{BAL}(\xi) \ + 0 \cdot \text{RER}(\xi) + (1-\alpha) \cdot \text{RSR}(\xi)$$

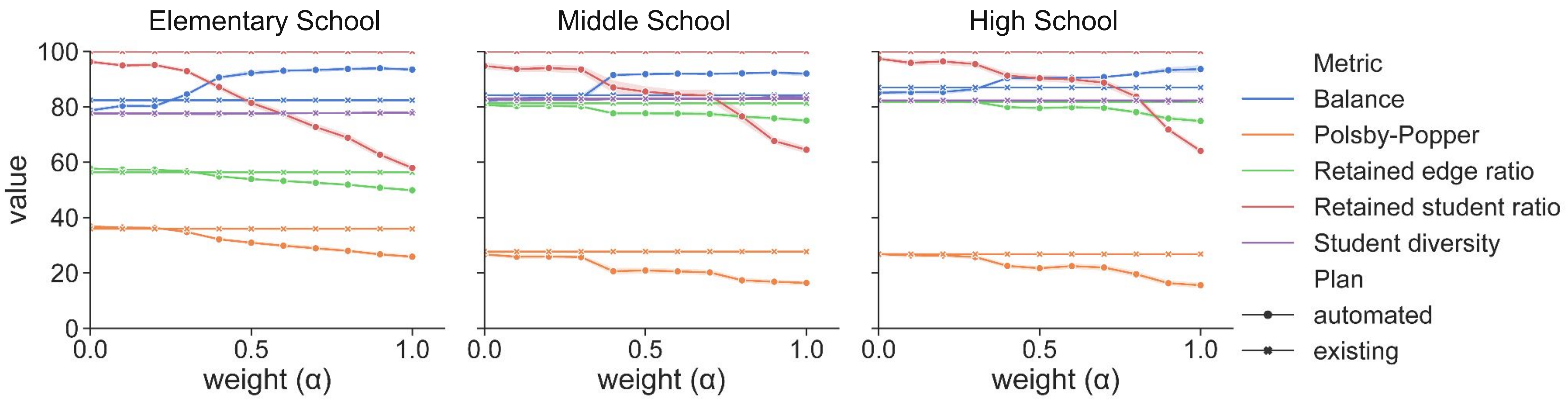


Figure 6. Tradeoff plot of the different performance metrics corresponding to the plans with varied weights.

Case Study

$$\mathcal{F}(\xi) = 0.5 \cdot \mathrm{BAL}(\xi) \; + 0.5 \cdot \mathrm{SD}(\xi)$$

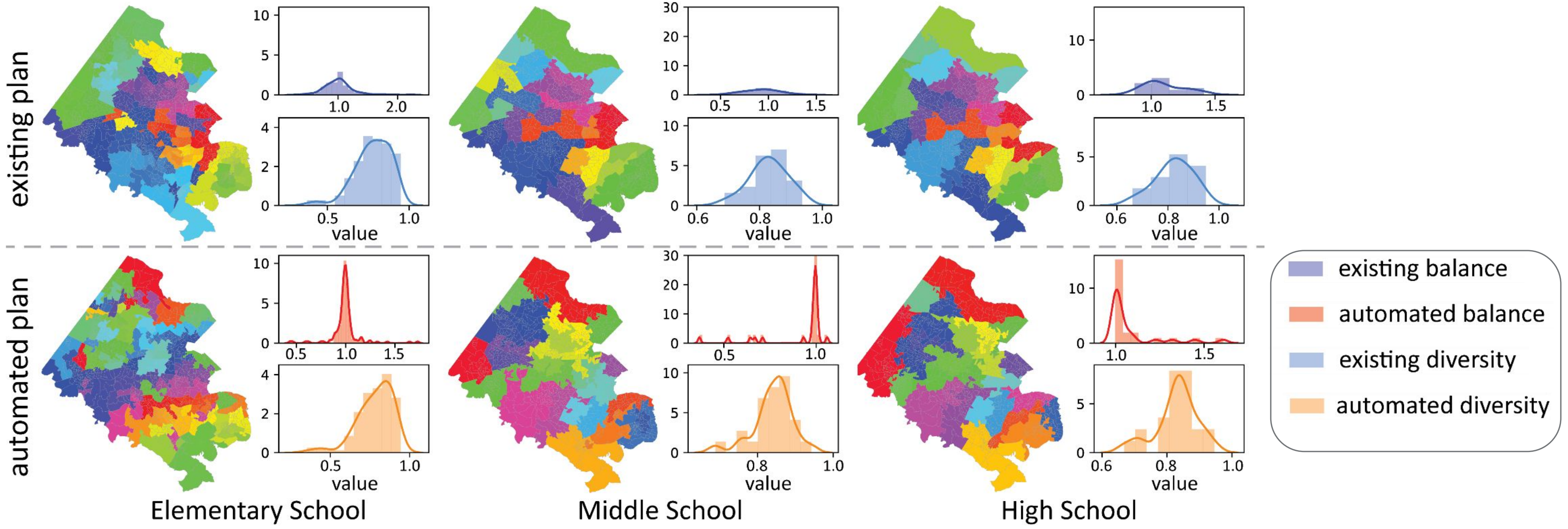


Figure 7. A comparison of the existing plans and automated plans generated by FlipChain.

Conclusion

- Design a flexible framework for solving the school redistricting problem.
- 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.
- Develop multiple ethical evaluation metrics and incorporate these considerations to support the decision-making process.
- Conduct extensive experiments on two US school districts and a case study to examine the approach's ability to obtain school redistricting plans with desirable properties.

THANK YOU!