RESEARCH ARTICLE

Bringing Spatial Interaction Measures into Multi-Criteria Assessment of Redistricting Plans Using Interactive Web Mapping

Jacob Kruse [a], Song Gao [a*], Yuhan Ji [a], Daniel P. Szabo [b], Kenneth R. Mayer [c]

[a]Geospatial Data Science Lab, Department of Geography, University of Wisconsin-Madison [b]Eötvös Lóránd Tudományos Egyetem, Department of Operations Research, Budapest
[c]Department of Political Science, University of Wisconsin-Madison

Corresponding author: Song Gao, song.gao@wisc.edu

ABSTRACT

Redistricting is the process by which electoral district boundaries are drawn, and a common normative assumption in this process is that districts should be drawn so as to capture coherent communities of interest (COIs). While states rely on various proxies for community illustration, such as compactness metrics and municipal split counts, to guide redistricting, recent legal challenges and scholarly works have shown the failings of such proxy measures and the difficulty of balancing multiple criteria in district plan creation. To address these issues, we propose the use of spatial interaction communities to directly quantify the degree to which districts capture the underlying COIs. Using large-scale human mobility flow data, we condense spatial interaction community capture for a set of districts into a single number, the interaction ratio (IR), which can be used for redistricting plan evaluation. To compare the IR to traditional redistricting criteria (compactness and fairness), and to explore the range of IR values found in valid districting plans, we employ a Markov chain-based regionalization algorithm (ReCom) to produce ensembles of valid plans, and calculate the degree to which they capture spatial interaction communities. Furthermore, we propose two methods for biasing the ReCom algorithm towards different IR values. Using the ensembles produced from these three methods, we perform a multi-criteria assessment of the space of valid maps, and present the results in an interactive web map. The experiments on Wisconsin congressional districting plans demonstrate the effectiveness of our methods for biasing sampling towards higher or lower IR values. Furthermore, the analysis of the districts produced with these methods suggests that districts with higher IR and compactness values tend to produce district plans that are more proportional with regards to seats allocated to each of the two major parties.

KEYWORDS
redistricting, regionalization, mobility, interactive map

1. Introduction

In the United States, voters within a given electoral district elect a representative to protect and advance their interests in the government. Political redistricting is the process by which a set of geographic units (e.g., census block groups) are partitioned into electoral districts of equal population, producing a map with district boundaries (Williams Jr, 1995). Though most redistricting rules in the United States were written before the invention of computers, maps are now made and contested with the aid of computers (Altman et al., 2005; Turner Jr and LaMacchia, 1999), and regionalization algorithms (DeFord et al., 2019; Aydin et al., 2018; Guo, 2008). Advances in computationally-driven redistricting have, however, exposed the flaws of various redistricting criteria, as well as the conflicts that arise when trying to accommodate multiple criteria. As a result, there has been a vigorous political, academic, and legal debate about the means and ends of political redistricting (Nagle, 2019; Winburn, 2008; Webster, 2013).

A central premise in the debate on redistricting is that districts should be designed so as to capture coherent underlying communities, commonly referred to as communities of interest (COIs). However, there is no agreed upon method for defining such communities, and, consequently, a variety of proxy measurements are used instead. Borrowing from the spatial analysis and economic literature, we propose that spatial interactions be used to quantify COIs for the purposes of redistricting, as COIs defined in this way would be directly based on the spatial distribution of human activities, rather than proxies for it.

With a method to quantitatively define COIs captured in redistricting plans, it is also possible to evaluate how COI for a given plan varies with other common redistricting criteria, such as compactness and efficiency gap (see section 3.2). To perform a multi-criteria analysis of redistricting plans, we employ and modify a state-of-the-art algorithm, the Recombination algorithm (DeFord et al., 2019) to produce ensembles of valid district plans. Finally, as redistricting is a political process that garners attention from the public, legal scholars, and researchers alike, it is important to engage and educate these various groups regarding any proposed redistricting methodologies. Accordingly, we present a public-facing web map which can be used to understand our multi-criteria assessment on various redistricting plans, in particular regarding the trade-offs in other redistricting metric values when trying to optimize for a particular metric.

Bringing together these different research threads, the major contributions of this work are as follows:

(1) Theoretical: the novel use of spatial interactions as the basis of COIs in redistricting, quantified in a global way with the interaction ratio (IR).
(2) Technical: two strategies for producing district plan ensembles in which higher or lower spatial interaction values are preferentially sampled with the Recombination (ReCom) sampling algorithm.
(3) Educational: an interactive mapping website for exploring trade-offs between various redistricting criteria, containing functionality not present in existing redistricting websites.

The remainder of this paper is organized as follows. In section 2 we present a literature review of measurements for COIs, spatial interactions in network-based community detection, redistricting algorithms and district evaluation methods. In section 3, we describe the human mobility flow dataset, and how it is used to calculate the spatial interaction ratio in this study. Next, we describe how to bias the ReCom algorithm sample distribution in the space of redistricting plans towards districts with higher or lower interaction ratios. In section 4, we present experiments used to generate redistricting plans for Wisconsin congressional district maps, which are then used to verify the effectiveness of the proposed sampling strategies and compare the interaction ratio with other redistricting metrics, as well as a usability study on the developed web map. In section 5 we discuss limitations of this work. Finally, in section 6, we conclude this work and suggest areas for future research.

2. Related Works

2.1. Measurements for Community of Interest Quantification

A common normative assumption in redistricting is that districts should be drawn so as to capture coherent underlying communities, commonly referred to as communities of interest (COIs). The process of redistricting aims to draw electoral boundaries in such a way as to keep COIs intact within electoral districts, rather than splitting them. Though the debate on how to define such communities is ongoing in the scientific literature on redistricting (Webster, 2013; Forest, 2004; Malone, 1997), states already employ a variety of proxy measurements (NCSL, National Conference of State Legislatures, 2021) to describe how well underlying communities are captured by districts. Here, we review the problems associated with some of these common proxy measurements.

Compactness metrics, generally speaking, assume that communities form and spread in a geometrically-compact way over space (Stephanopoulos, 2012a; Duchin and Tenner, 2018). While this is reasonable to assume in a general sense, there is no reason to assume that the exact degree of compactness calculated by a given metric should correlate to an exact degree of community coherence. Said another way, compactness can only be said to be a proxy for underlying communities, since it is not actually measuring community directly. Another commonly-used measurement to evaluate redistricting plans is the counting of municipal splits, which relies on the assumption that existing municipal boundary shapes capture underlying communities in a significant way (Winburn, 2008). However, as with compactness measurements, this method fails to directly measure a particular aspect of the underlying communities, such as shared economic activity or religious affiliation (Gimpel and Harbridge-Yong, 2020).

In response to the issues with the proxy metrics described above, researchers in political science have proposed several metrics that try to directly measure community coherence in a quantitative way. Makse (2012) argued that voters themselves should define COIs through ballot initiatives. Stephanopoulos (2012a) introduced the concept of a territorial community test, which evaluates districts based on their spatial and socioeconomic congruence, as well as subjective affiliation to a given community. In a follow-up work, Stephanopoulos (2012b) described a metric called spatial diversity, which they use to measure spatial and socioeconomic congruence. To calculate spatial diversity, factor analysis is performed on census data to derive composite factors for individual census tracts. Then, the weighted standard deviation of each districts' composite socioeconomic categories is calculated, using the percent of variance described by each of the composite factors as the weight.

2.2. Spatial Interactions in Community Detection

In a separate but related branch of research, geographers have been incorporating new, finely-grained spatial interaction datasets into more traditional methods for finding and quantifying communities. Spatial interactions, broadly speaking, are movements over space that result from a human process, and can range from human mobility flows, captured by surveys (Nelson and Rae, 2016), mobile phone calls or location device tracking data (Ratti et al., 2010; Kang et al., 2020; Xu et al., 2022), to commodity and information flows (Haynes and Fotheringham, 2020; Li et al., 2020; Rao et al., 2022). Spatial interaction communities, then, are communities defined by a particularly dense set of spatial interactions. As an example of this, Gao et al. (2013) used a mobile phone call dataset to measure human mobility flows across a city in China, and then used the resulting flows to identify spatial interaction communities. Dong et al. (2015), used mobile phone movement data for traffic zone division. Building on these works, Liang et al. (2022) used a spatial interaction flow ratio to evaluate communities found through various community detection algorithms, where the flow ratio is the sum of intra-community flows divided by the sum of inter-community flows.

Here, we argue that such a ratio is a theoretically-sound way to quantify community coherence for the purposes of redistricting, since spatial interactions are the embodiment of people's

interests, whether social, economic, political, or otherwise. When people move in the same spaces, they are demonstrating involvement in the same geographic community.

In particular, we present this metric under the name *interaction ratio* (IR), a new metric for redistricting plan evaluation which compares spatial interactions within districts to spatial interactions between districts in a global way. Essentially, this is a global metric to describe community strength, i.e., the degree to which district boundaries are aligned with the underlying spatial interaction communities.

2.3. Redistricting Algorithms

In the contexts described above, spatial interactions are used as inputs for community detection algorithms, which produce district assignments for a set of geographic units by grouping regions with dense spatial interactions together. In the redistricting literature, a broader set of algorithms called *regionalization* algorithms are used for similar purposes, namely: 1) for the production of optimal plans, and 2) for the evaluation of proposed plans. In the former category, different techniques are used to produce a single districting plan that optimizes some set of criteria (e.g., compactness and fairness), while meeting constraints (e.g., equal population and geographic contiguity). Some examples of this approach include REDCAP (Guo, 2008), the Divide and Conquer algorithm (Levin and Friedler, 2019), the Simulated Annealing Genetic algorithm (Bergey et al., 2003), and iRedistrict, a geovisual analytical program which incorporates graphical user input into a Tabu-style algorithm (Guo and Jin, 2011). The main problem with this class of optimization algorithms is that the redistricting problem is NP-hard, making the task of finding the globally-optimum plan computationally intractable for most political redistricting situations (Altman, 1997).

Rather than searching for the global optimum, the other category of redistricting algorithms is focused on producing *ensembles* of diverse plans that can be used to evaluate a given plan, where the ensembles are made up of representative samples from the space of valid plans. *Outlier analysis*, in particular, compares a given plan to the distribution of values from an ensemble in several dimensions, such as fairness and compactness, and evaluates if the plan is an outlier or if it is relatively neutral. This type of analysis has been used successfully in several legal challenges to state-level redistricting plans, and is likely to see continued use in redistricting litigation (Ramachandran and Gold, 2018). Importantly, the generation of plans in this way allows for outlier analysis, but is not used to produce actual redistricting maps. This is for two reasons. 1) District maps produced in this way, while valid according to the constraints imposed on the algorithm, might not reflect practical considerations or all legal criteria, such as the requirement to maintain district cores, and 2) gerrymandering is a political process, and so legislators are incentivized to produce maps that favor themselves or their party, which is generally approached as an optimization task (Stephanopoulos and McGhee, 2015; Brunell, 2008).

In order to produce sample distributions, researchers from many domains have turned to Markov Chain Monte Carlo (MCMC) methods (Diaconis, 2009). In a redistricting context, MCMC methods sample from some target distribution on the space of partition plans, perhaps defined by certain population and compactness thresholds, by making iterative changes to some initial district plan. The sample distribution approaches the target distribution as the number of iterations grows, with each iteration producing one map for the ensemble. Uses of this methodology for outlier analysis in redistricting be found in Herschlag et al. (2017), Chikina et al. (2020), and Fifield et al. (2020). Calculating the number of steps necessary to get a representative sample from the target distribution is impossible for most redistricting tasks, and so DeFord et al. (2019) suggest using convergence heuristics and sensitivity analysis to demonstrate the robustness of the methods and settings used in a given Markov chain.

2.4. ReCom Algorithm for Plan Comparison

One recent MCMC method that has proven useful for redistricting is the Recombination (ReCom) Markov chains algorithm (DeFord et al., 2019). Rather than using a typical flip-walk, which "flips" the assignment of a border unit from one district to another, the ReCom algorithm first merges two adjacent districts and then repartitions them by cutting a "spanning tree" (an acyclic subgraph in which all of the nodes are connected with the minimum possible number of edges) formed over the nodes (e.g., electoral precincts or wards) of the two merged districts. This method, compared to traditional methods, greatly reduces the time needed for convergence to a steady state (i.e., the mixing time), since the boundaries between adjacent districts are completely destroyed when they are merged, allowing for much more efficient movement throughout the space of valid maps. Examples of ReCom chains with state-level districts in the literature show that chains run for as little as 10,000 steps are able to produce well-mixed ensemble distributions, while traditional methods frequently do not pass these tests even after millions of steps (Becker et al., 2021; DeFord et al., 2019).

With an ensemble of plans in hand, a given map can be evaluated against the ensemble distribution for any number of redistricting metrics, allowing for multi-criteria assessment. Here, the term *metric* simply refers to the particular measurement used to quantify a given criterion (e.g. Polsby-Popper compactness is a metric to quantify the criterion of district compactness; IR is a metric to quantify the communities of interest; efficiency gap is a metric to measure the fairness of a redistricting plan). The combination of ensembles of alternatives and multi-dimensional evaluation aids the process of redistricting in two ways: 1) by producing valid maps on actual geometry, we avoid the well-documented issues with the single-number ideals proposed by different metrics (DeFord et al., 2021; Bernstein and Duchin, 2017; Katz et al., 2020). Rather than trying to come up with universal rules that hold any and everywhere, or under only a very specific set of conditions, comparing maps against ensembles of sampled maps allows for comparisons against actual alternatives for the area in question; 2) by using multiple criteria to evaluate district plans, we reduce the risk of over-relying on a particular criterion, which can easily misrepresent the ends or means they propose to support or measure, and embrace the complex, qualitative task of using various quantitative measures to support the pursuit of human ends, such as fairness, as redistricting objectives (Duchin and Walch, 2021).

Finally, although we use the ReCom algorithm as our base sampling algorithm in the research presented here, we also introduce new modifications (biased spanning tree and min-/max-cut, see section 3.4) that bias the sample distribution of spatial interaction ratio values towards higher or lower values to explore the redistricting plans with spatial interaction flow-connected communities.

2.5. Educational Tools for Redistricting

As can be seen in the previous sections, the process of redistricting plan production and evaluation permits an enormous number of options in every step, from setting compactness bounds to picking a evaluation metric. To help educate and engage the public about this complicated process, a number of open web-maps have been developed and made available online. One prominent example is Dave's Redistricting website, commonly referred to as DRA (`https://davesredistricting.org`). The website enables exploration of Congressional, State Senate, and State House districts for each U.S. state. Demographic and voting data are provided for different census boundaries, and there are plots of statistical analyses for common redistricting metrics. The web-map portion allows for basic map interactions, including pan and zoom, as well as the creation of new redistricting plans. Districtr (`https://districtr.org`) and DistrictBuilder (`https://www.districtbuilder.org`) are similar tools which offer similar capabilities. One feature lacking, however, in all of these tools is the ability to compare redistricting plans side-by-side, which would facilitate learning by the end-user by allowing them to visually find differences between maps (Guo and Jin, 2011). For the purposes of this study, maps with different properties, like a high interaction ratio or compactness value, could

be compared in such a way, allowing the user to understand how metric values impact map design. Additionally, if supported with statistical summaries, a web map of redistricting plans would allow for the interaction ratio to be understood in the context of existing metrics. To support these learning goals, we introduce an interactive web-map that allows for multi-criteria evaluation of different redistricting plans.

3. Methods

3.1. Spatial Interaction Dataset Preparation

In order to calculate the spatial interaction ratio, we employ SafeGraph Neighborhood Patterns (`https://docs.safegraph.com/docs/neighborhood-patterns`) as our underlying dataset of spatial interactions. This dataset records the number of anonymized cell phone "visits" from one location to another, where the place visits are recorded at the census block group (CBG) level for each temporal period (e.g., hourly, daily, weekly, and monthly) of the year. To protect privacy, SafeGraph employs differential privacy techniques on the flow dataset, including the removal of flows where counts are very low (i.e., less than 4 devices). In terms of the representativeness of SafeGraph data, the mobile phone users detected by SafeGraph represent about 10% of the mobile phone users across the United States, with local variation, and therefore some scaling is required to estimate the population-level human mobility flows. Following Kang et al. (2020), we infer the population-level human mobility flows (pop_flows) between a given origin and destination (O-D) CBG pair using the ratio of origin CBG population to the number of SafeGraph devices in the origin. The equation is as follows:

$$pop\_flows(o, d) = device\_flows(o, d) \times \frac{pop(o)}{num\_devices(o)} \qquad (1)$$

Using this equation, we take the Neighborhood Pattern dataset for every month of 2020 to produce an O-D flows matrix of the average monthly population flows between and within Wisconsin CBGs, where the i_{th} row and j_{th} column intersection is the average monthly population-level flow counts going from the origin CBG to the destination CBG. Although there are some human mobility pattern changes associated with the COVID-19 lockdown period (Gao et al., 2020), the monthly average OD flows over the year are relatively stable.

3.2. Interaction Ratio and Comparison Metrics

The interaction ratio (IR) is a global measure of spatial interaction community strength for a given redistricting map, where a spatial interaction community is defined by people moving within the same set of CBGs, and community strength is the degree to which district boundaries place geographic units (e.g., census tracts or CBGs) which have a relatively high number of spatial interactions between them within the same district. Specifically, the degree to which a set of district boundaries encapsulates the underlying communities is calculated as the sum of district intra-flows divided by the sum of district inter-flows. Intra-flows are defined as population flows $s_{o,d}$ where the origin CBG o and the destination CBG d are assigned to the same district (i.e., $D_o = D_d$). Inter-flows are defined as population-flows originating from CBGs in one district and ending in another (i.e., $D_o \neq D_d$). Therefore, a given set of districts generated by the ReCom algorithm (section 3.3) has a higher IR if more flows originate and end in the same district relative to the number of flows that originate in one district but end in another. The following equation expresses this idea concisely, where K is the total number of districts:

$$IR = \frac{Sum\ of\ Intra\ Flows}{Sum\ of\ Inter\ Flows} = \frac{\sum_{D_o = D_d} s_{o,d}}{\sum_{D_o \neq D_d} s_{o,d}}; D_o, D_d \in 1, 2, \ldots, K \tag{2}$$

To help demonstrate the calculation of the IR, we provide Figure 1 using a hypothetical region, which shows the intra and inter-flows for a set of four districts. In this example, the underlying CBGs that comprise each district are not depicted so as to not overly-clutter the diagram. Instead, we show only the intra-flows for CBGs assigned to a given district, depicted with black arrows, and the inter-flows for CBGs assigned to different districts, depicted with white arrows. As seen in Equation 3, the sum of intra-flows is the sum of black arrow flows, and the sum of inter-flows is the sum of white arrows flows, yielding an IR of 1.25 for this particular set of districts. Note that the IR is only one of the simple metrics to quantify the spatial interaction strength within communities. A normalized version of IR (i.e., the ratio of intra flows to total flows) can also be used to get a value between 0 and 1.
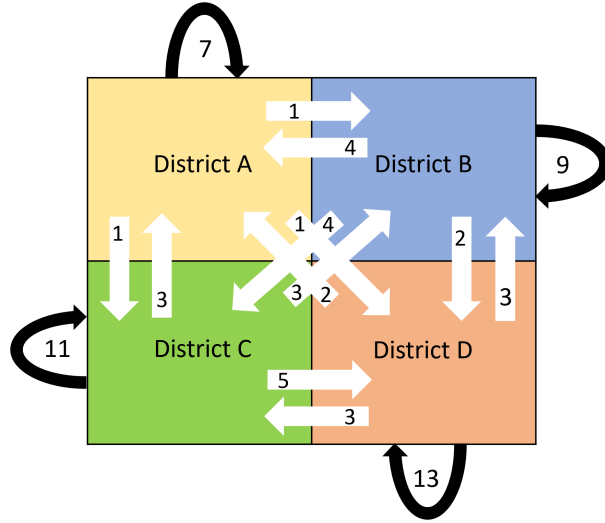

Figure 1.: Intra and inter-flows for an example set of districts. Black arrows depict intra-flows, while white arrows depict inter-flows. [one-column width]

$$IR = \frac{7 + 9 + 11 + 13}{1 + 1 + 1 + 4 + 4 + 2 + 3 + 3 + 5 + 2 + 3 + 3} = \frac{40}{32} = 1.25 \tag{3}$$

To measure how the IR varies with common redistricting metrics, we also calculate compactness and efficiency gap scores for the districts produced by the ReCom algorithm (section 3.3). For compactness, we use the Polsby-Popper measure (Polsby and Popper, 1991) which provides a score in the interval of $[0, 1]$, with 0 being the least compact and 1 being maximally compact. The equation is as follows:

$$Polsby\text{-}Popper\ Compactness = \frac{4\pi Area}{Perimeter^2} \tag{4}$$

In contrast to this area-perimeter ratio definition of compactness, the ReCom algorithm uses a discrete definition of compactness, where the compactness of a given plan is bounded by n times the number of cut edges in the original plan, and n is a multiplier set by the user. However, we report specific values of compactness using the Polsby-Popper metric, as it is a more common measurement of compactness in redistricting evaluation.

The efficiency gap is a measure of election fairness where wasted votes for the two major parties are counted for each district and then summed across districts (Stephanopoulos and McGhee, 2015). Wasted votes for the winning party in a given district are votes cast in excess of what is needed to win the election (i.e., every vote after 50% of votes plus one), while all votes for the losing party in a given district are considered wasted. Therefore, in this work the efficiency gap is calculated using votes from all districts with the following formula:

$$Efficiency\ Gap = \frac{Wasted\ Votes\ for\ Democrats - Wasted\ Votes\ for\ Republicans}{Total\ Votes} \quad (5)$$

The ReCom algorithm implementation and calculations for Polsby-Popper compactness and the efficiency gap are performed with the GerryChain Python package (2022), while all other calculations are done in Python using standard libraries.

3.3. ReCom Algorithm for Plan Sampling

As shown in Figure 2, the ReCom algorithm draws and cuts spanning trees on merged districts in order to efficiently repartition districts and sample possible arrangements. Since more compact districts allow for the creation of more spanning trees, compact districts are chosen with much greater frequency (Procaccia and Tucker-Foltz, 2022; DeFord et al., 2019). Rather than producing a uniform distribution of district arrangements, then, which would be dominated by non-compact districts, the ReCom algorithm targets the *spanning tree distribution*, which refers to the subset of relatively compact district plans that get sampled when the spanning tree method is used in ReCom. More specifically, the algorithm works in the following way (Figure 2): a) there exists an initial partition in which nodes (e.g., electoral precincts, census block groups or other geographical units) are assigned to different districts; b) when two districts' nodes in the graph are merged, a random spanning tree is built on the induced subgraph's edges; c) cuts are then explored to see which would maintain (near) population equality between the two new partitions; d) a cut is made, creating the new districts. An example of this process with state-level geometry is shown in Figure 3.

a) Initial partition; nodes representing CBGs are assigned to districts

b) Two adjacent districts are merged

c) Random spanning tree formed on nodes of merged districts; balanced cut made

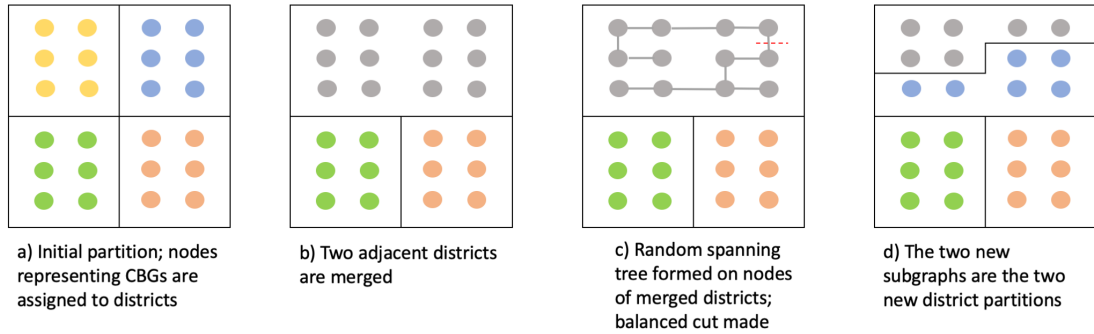d) The two new subgraphs are the two new district partitions

Figure 2.: Demonstration of the ReCom algorithm on an example set of nodes (representing geographic units, e.g. census block groups). District assignments are represented by node color, and district boundaries are represented by black lines. The spanning tree formed on the nodes from the merged districts is represented in gray, and the cut to the spanning tree is represented with the dashed red line. [two-column width]

a) Initial partition of CBGs
into districts

b) Two adjacent districts
are merged (gray)

c) The merged districts are
repartitioned by forming
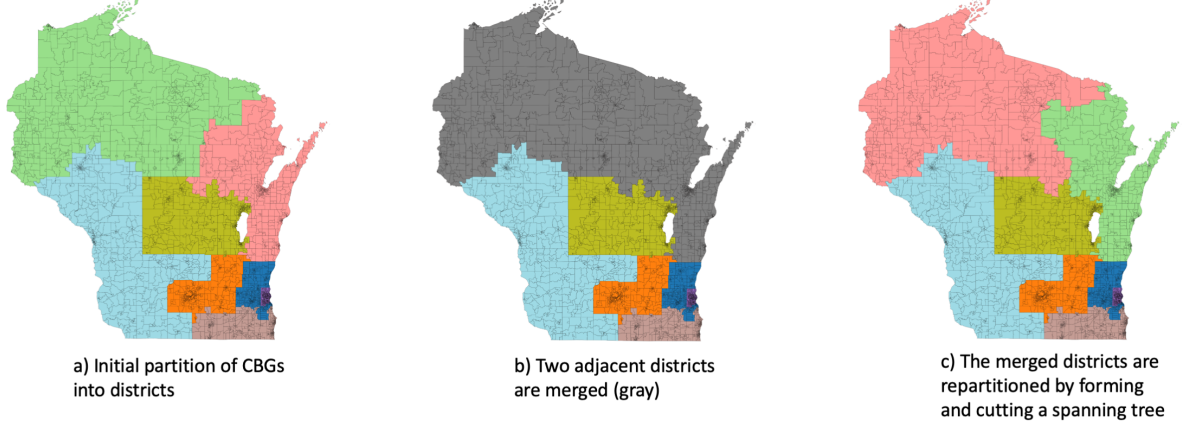and cutting a spanning tree

Figure 3.: An example of Wisconsin congressional districts being merged and repartitioned with the ReCom algorithm, using census block groups as the underlying spatial units. The pink and green districts are merged to form the gray district, and which in turn is repartitioned into two new districts. [two-column width]

3.4. ReCom Algorithm Modifications for Biased Sampling

We note that some weighting scheme for tree or cut-selection could potentially change the sample distribution of districting plans, making this an ideal place for intervention to build upon ReCom's useful properties. In particular, for trying to preferentially sample districts that better capture the underlying spatial communities, as described by the IR, we could choose cuts or trees that favor districts with higher IR values. Modifying the sample distribution of districting plan maps in these ways is important for the same reason that generating an ensemble of districting plan maps is important in the first place: having a distribution of maps allows one to evaluate a given map against a set of alternatives. Being able to modify the target distribution with respect to IR would give redistricting researchers some ability to specify the acceptable target distribution of redistricting plans.

As the focus of our experiments, we produce and analyze the 2020 congressional district maps for the state of Wisconsin. Our goal is to compare trade-offs between compactness, the efficiency gap, and our proposed IR metric for quantifying community strength, and so sampling as widely as possible for each respective metric is desirable. To that end, we allow for up to a 3% population deviation between the largest and smallest districts. For compactness exploration, we consider high and low compactness limits, defined as 1 or 5 times the number of cut edges in the initial plan. With these experimental settings in place, we next describe the ReCom algorithm and our modifications to it.

The steps of the algorithm are as follows (adapted from DeFord et al. (2019)):

Algorithm 1 Recombination (ReCom)

Input: Dual graph $G = (V, E)$, and the current partition $P = W_1, \ldots, W_k$ on G
Output: The next partition Q

    Select two adjacent districts, W_i and W_j, from P
    Form the induced subgraph H of G on the nodes of $W = W_i \cup W_j$
    Form a uniform random spanning tree T on H
    while There is no edge in T that creates contiguous districts with equal population **do**
        Replace T with a new random spanning tree on H
    end while
    Pick a valid cut uniformly at random
    return Q

To both see the range of IR values possible within the spanning tree distribution of congressional district plans and to explore how to most effectively sample from such a range, we employ two methods.

In the first method, we adapt how the random spanning tree (RST) is built on the nodes of the merged districts. While the original method builds a maximum spanning tree on the induced subgraph, whose edge weights have been randomized, our method introduces two additional steps to this process: 1) after every edge is given a random weight from a uniform distribution on the interval $[0, 1]$, we then multiply that edge by the actual population flow weight (i.e., the true edge weight), and 2) we multiply this value again by a bias term. Then, as in the original method, a maximum spanning tree is built with these weights and a random cut is chosen, subject to the population and compactness constraints. Essentially, this method seeks to preferentially build spanning trees that keep neighbors with high flow weights between them within the same district. For exploring higher and lower IR values, we can modify the sign and magnitude of the bias term. We will refer to this method as the *biased RST* method. In Figure 2, this method would impact the third pane, biasing the spanning tree built on the grey nodes such that neighbors with high flows would be more likely to be directly connected.

Our second method tries to bring a fuller picture of the IR into the selection of the partitions made. With this method, we use the standard ReCom method to draw a random spanning tree on the set of nodes from two merged districts, and produce a list of valid cuts. Then, we take this list of cuts and evaluate which one most (or least) improves the IR value of the map, and take that cut. As with the biased RST method, one of our goals with this second method is the exploration of possible IR values, and so we are interested in exploring district plans that have higher IR values as well as district plans that have lower IR values. Accordingly, we refer to the two methods as either the *max-IR cut* or the *min-IR cut*, depending whether the max or min cut was chosen. In Figure 2, these methods would impact where the cut (represented with the red dashed line) is made on the grey random spanning tree.

The ReCom method introduced in DeFord et al. (2019) picks a spanning tree uniformly at random from among all of the spanning trees of the merged districts. However, the GerryChain (2022) implementation of ReCom builds a random maximum spanning tree by default. Internal experiments revealed that the two methods yielded comparable results, but that building random maximum spanning trees was substantially faster, and so the default method is used here. For clarity, the standard ReCom runs without any bias sampling or modifications will be referred to as *RST*, referring to the method that builds random maximum spanning tree.

3.5. Experimental Settings for Baseline and Modified ReCom Algorithms

We take the redistricting plan produced by the People's Map Commission for Wisconsin 2020 Federal House Districts (https://govstatus.egov.com/peoplesmaps/work-records) as our seed map. As noted in DeFord et al. (2018), the ReCom algorithm results do not vary much with the seed map or sample path through the state space. Internal experiments for our research here found similar results, and so the experiments that we present here all use the same seed map. For geometry and demographic data, we employ the CBG-level data provided by the Redistricting Data Hub (https://redistrictingdatahub.org/dataset/wisconsin-block-pl-94171-2020/). To get voting data that is comparable with the human mobility flows, we disaggregate ward-level election results to the CBG-level based on the number of citizens that are at least 18 years old in age, as is a standard practice in the Political Science literature (see Benade et al. (2021) and Becker et al. (2021) for examples). Votes cast for Democrats and Republicans in the 2018 Federal Congressional election are used to calculate the efficiency gap.

Using these inputs, we run a total of six chains to 1) test the biased RST and max/min-IR cut methods for sampling higher and lower IR values, and 2) sample a wide range of values for all metrics for use in the multi-criteria assessment of redistricting plans. Prior work by our group had shown that the distribution of IR values is largely dependent on compactness settings (the

same pattern seen in our experiment results in section 4.1), and so we break up the experiments into three chains which evaluate higher IR values, and three chains which evaluate lower IR values, as seen in Table 1.

Table 1.: Experimental settings for the six ReCom chains used to produce redistricting plan ensembles.

Compactness Bound Multiplier	ReCom Method
1×	RST
1×	Biased RST (100 bias factor)
1×	Max-IR cut
5×	RST
5×	Biased RST (-100 bias factor)
5×	Min-IR cut

As mentioned in section 2.4, 10,000 iterations has been shown to be sufficient for convergence with a number of state-level redistricting chain experiments, where each iteration involves two districts being merged and repartitioned. Internal work by our group found that even as few as 3,000 iterations produced relatively stable distributions, and going as high as 12,000 iterations did not continue to change ensemble distributions (stability was evaluated visually by inspecting bar plots of the distributions, similar to what is done in DeFord et al. (2019)). Therefore, each experiment is run for 10,000 iterations.

3.6. Interactive Web Map Development

In addition to describing and analyzing the sampled redistricting ensembles (sections 4.1 and 4.2), we also present some of the results in the interactive web map. In particular, we present the three redistricting plans that maximize the IR, the efficiency gap, and compactness, respectively, such that each of those maps represents the extreme of a given metric for the various district samples drawn. Furthermore, we provide demographic attributes on the district level for each of these maps. To encourage exploration as a learning method for the user, we present two side-by-side interactive maps, as well as a linked, scented widget, which cues users into the attribute values underlying districts. We also present the sampled ensembles of district plans in the form of an interactive 3D cube. Further down the web page, we provide linked text descriptions for each of the calculated metrics. The web map itself (see section 4.3) is developed using a technology stack including JavaScript, HTML, CSS, Leaflet, D3, and Map Sync libraries.

To demonstrate that our web map provides novel functionality in the space of public redistricting resources, we compare our web map to the popular redistricting websites described above, namely the Dave's Redistricting (DRA), Districtr, and DistrictBuilder. The functionalities examined are as follows:

- Compare multiple plan boundaries simultaneously;
- Compare a selected plan to the distribution of plans;
- Compare metrics for multiple maps;
- Plan creation;
- Provide optimal maps for comparison;
- Provide redistricting metrics and demographics for districts;
- Provide reference layers;
- Provide map re-expression.

To assess the usability of the developed web map, we survey 10 users with GIS or web map experience in the University of Wisconsin-Madison to complete a series of six tasks related to map and plot reading. Overall, the tasks in the assessment evaluate if the web map design facilities the use of the main web map features, e.g. are users able to change which map is

12

displayed in each of the map panes. University students with GIS or web map experience, i.e. at least one course in either area, are invited as the participants, as the intended audience for the interactive web map is researchers and members of the public engaged in the redistricting process, and some knowledge of mapping tools is common among these groups. The participants include two undergraduate and eight graduate students, none of whom had experience with political redistricting. For each question, we report the average response accuracy, where responses are either correct or incorrect. The tasks are as follows:

A. Add the CBG overlay to the left map pane;
B. Hover over the district that contains the label ("Milwaukee") in each of the map panes;
C. Change the 2022 Enacted Map panel to show the Max Compactness Map;
D. Hover over the line in the parallel coordinate plot corresponding to the district with the highest $inter\_flows$ value for the Interaction-Ratio map;
E. Read the district-level efficiency gap (i.e., the Dis_EffGap value) for district containing the city of Madison in the 2022 Enacted Map;
F. Hover over the point with the minimum compactness value in the interactive redistricting values cube.

Task A tests the usability to add a map overlay. Task B tests the usability to compare differences in district boundaries between two maps. Task C tests the usability to change which maps are displayed. Task D tests the usability for a user to read the parallel coordinate plot and associate its values with district boundaries. Task E tests the usability to use map boundaries to read values from the associated plots. Task F tests the usability to move the interactive 3D cube in order to read redistricting metric values. To complete the assessment, each participant is read and shown the task description, using one screen, and then they complete the task on another screen. The member of the research team administering the assessment reads the task and responds to any questions while sitting next to the participant. Questions regarding the meaning of the task are answered, while questions regarding the execution of the tasks, i.e. "so I should click right here?" are not answered so as to not provide assistance to the participant beyond the affordances of the web map. Each assessment is performed individually. Task completion accuracy is recorded as 1 for correct or 0 for incorrect, with answers being collected audibly or visually by the researcher administering the test, e.g. "the district-level efficiency gap value is -0.33 for the district containing Madison" and seeing that the participant's mouse is hovering over the correct district.

4. Results

4.1. Baseline and Modified Sampling Distributions

We first present the spanning tree sampling results from the ReCom algorithm and adaptions to it, with respect to IR values. As seen in Figure 4, the most important factor in the IR values sampled is the compactness, with the compactness limit of 1× the number of cut edges in the original map allowing for substantially higher IR values to be sampled when compared with a compactness limit of 5× the number of cut edges. Given that, we compare methods with respect to compactness thresholds.
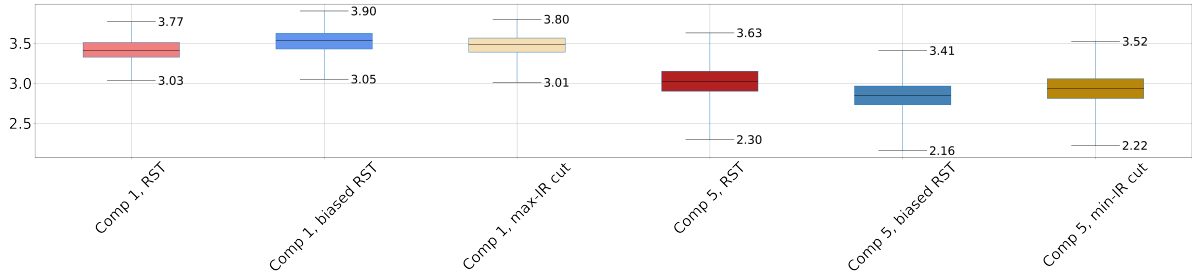
Figure 4.: IR values (y-axis) for ensembles produced by RST (red), biased RST (blue), and max/min-cut RST (brown) algorithms. Each chain was run for 10,000 iterations. Lighter colors on the left are for the tighter compactness threshold of one, while darker colors on the right are for the looser compactness threshold of five. [two-column width]

With a compactness limit of 1×, the RST method sampled IR values ranging from 3.03 to 3.77. Both the 100 biased RST and max-IR cut methods produced distributions that were significantly different from baseline, with KS-tests comparing them to the RST method yielding p-values of 0.00. The biased RST method, in particular, was able to sample IR values as high as 3.90, while the max-IR cut method did not substantially increase the range over baseline.

Similar results were produced with a compactness limit of 5×, with the -100 biased RST method producing a sample with both a lower range and center of IR values, compared to the standard ReCom method, and the min-IR cut producing similar but smaller improvements. The baseline RST method sampled IR values ranging from 2.30 to 3.63, while the biased RST method sampled values ranging from 2.16 to 3.41. KS-test values comparing the biased RST and min-IR cut methods to RST produced p-values of 0.00, suggesting that those samples come from different underlying distributions than that of the standard ReCom sampler.

While the biased RST and max-IR cut methods only shift the range of IR values a modest amount over the spanning-tree distribution, the ability to shift the distribution at all is important for several reasons. For one, legal challenges employing outlier analysis must specify the distribution that is being sampled from (Procaccia and Tucker-Foltz, 2022), and the methods detailed here provide a way to sample districts that better match communities defined by spatial interactions. For another, the difference in performance between the two methods gives us some insight into what mobility flow factors play a larger role in changing IR values. The biased RST method, which was able to shift the distribution within the spanning tree distribution the most, works by randomly weighting all edges, multiplying those edges by the human mobility flows of each edge, respectively, and then building a maximum spanning tree on this new set of weights. This suggests that higher IR districts are, on average, sampled by trying to increase intra-flows between adjacent sub-geographies, rather than trying to decrease inter-flows or both increase intra-flows and decrease inter-flows at the same time, as in the max-IR cut algorithm.

It may have been difficult to sample higher IR values because the ReCom algorithm is already producing samples near the upper range of possible values. Consider, for example, our seed map, the map produced by the People's Map Commission. The map was designed with the intention of keeping COIs intact (`https://evers.wi.gov/Documents/PMCCriteriaMemoFINAL.pdf`), and has an IR value of 3.04. Sampling lower IR values is likely possible, but undesirable, as producing very low-values IR plans would probably involve producing very non-compact districts which would not be used in redistricting.

For one final point, we also argue that even the modest increase in range for the IR values sampled between methods is important. For instance, consider the change in mobility flows from the highest IR value sampled by the ReCom algorithm 3.77, to highest value sampled by biased RST method, 3.90. Going from 3.77 to 3.90 (an increase of about 3%) represents a change of 549,605 human travel trips being moved from inter-flows to intra-flows, which represents about 1% of total flows in our dataset. While this is a modest amount, it is a tangible increase

in community quality. Furthermore, the use of outlier analysis to challenge maps relies more on the center of the IR value distribution, rather than the range, and the sampling methods introduced here significantly shift the center of the sampled distributions.

4.2. Multi-Criteria Assessment of Ensembles of Redistricting Plans

Taking the ensembles of redistricting maps described in the previous section, we show the combined redistricting results in a multi-dimensional fashion, considering IR, compactness, and efficiency gap values. The goals of this analysis are to explore the space of valid redistricting plans and the trade-offs between the various criteria. Combining the sampled ensembles in this way allows us to explore a wider space of valid maps, but also means that the resulting distribution should not be used for outlier analysis. To highlight the relationships between the three metrics used, we produce a 3D-cube that plots sample values in three dimensions. While this cube is interactive in the web-map described below, we present two different views of the cube here to show the sampling results. In Figure 5, the 3D-cube is oriented to show the range of IR and compactness values found by the various ensembles of redistricting plans, with higher IR values colored in yellow, and lower IR values colored in purple. Higher compactness values are shown with larger dot markers. As seen in the figure, IR has a monotonic relationship with compactness, which is measured with the Polsby-Popper formula (eq. 4). Notably, a given compactness value can still allow for a modest range of IR values.
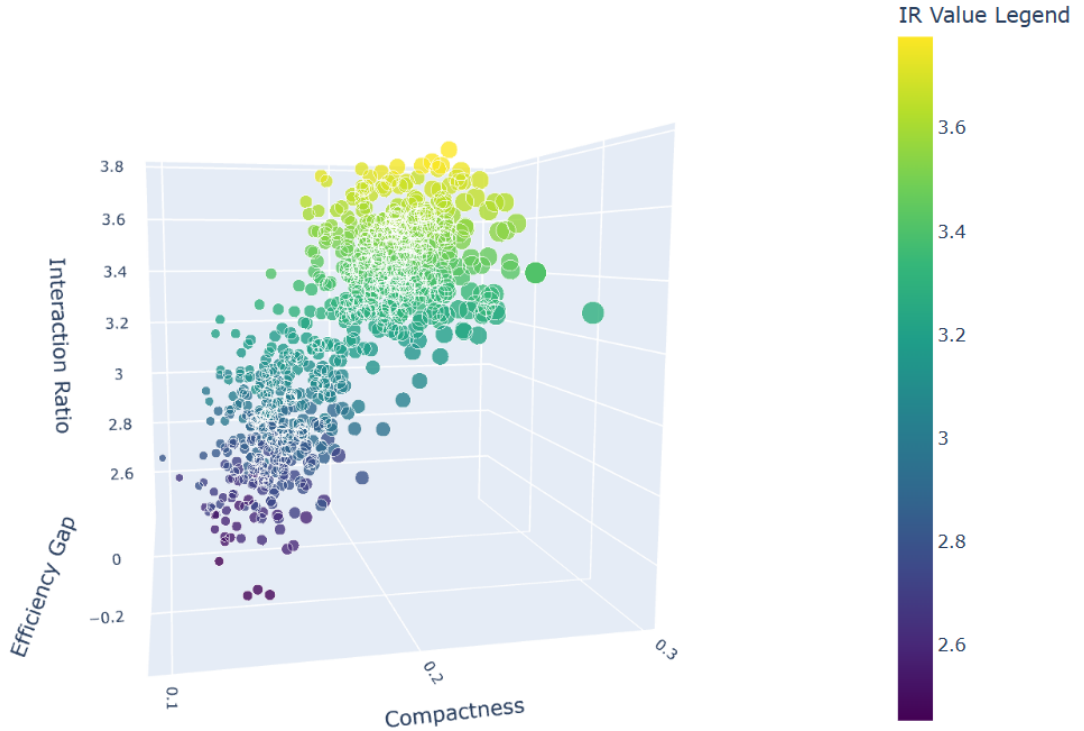

Figure 5.: The 3D-cube visualization focused on the IR and compactness dimensions, where IR value is shown in color, and compactness is shown with marker size. Compactness has a monotonic relationship with the IR. [two-column width]
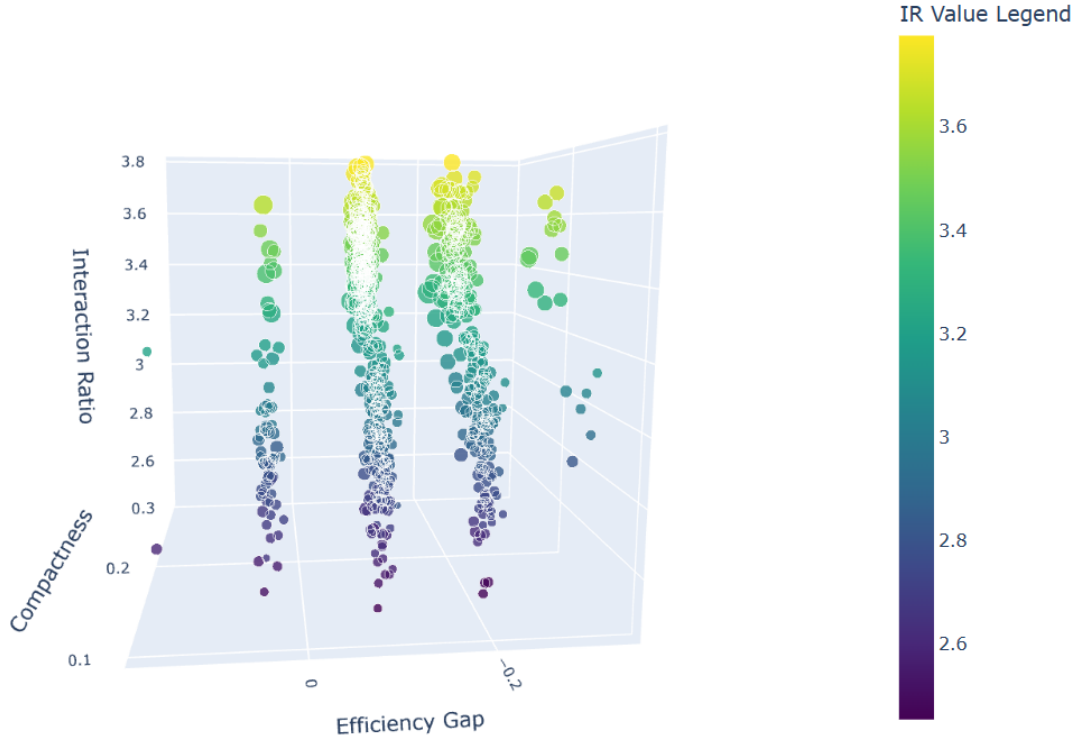
15

Figure 6.: The 3D-cube visualization focused on the efficiency gap-dimension. The visible bands correspond to different seat allocations between the two major parties. Negative efficiency gap values indicate that more Democrat votes are wasted, and that Republicans gained a disproportionate number of seats relative to their votes. [two-column width]

In Figure 6, we show the 3D-cube visualization focused on variability in the efficiency gap. Negative values indicate that more Democrat votes are wasted, and that Republicans gained a disproportionate number of seats relative to their votes, while positive values indicate the other way around. The majority of values are negative, showing that valid district plans generally favor Republicans in Wisconsin. All efficiency gap values, however, cluster into a few visible bands. This reflects that the efficiency gap is largely tied to seat allocation (Duchin, 2018; Bernstein and Duchin, 2017), where each band represents another Congressional seat flipping from one party to another. The noise in each band is due to differences in voter turnout and population variation between districts, the latter of which is allowed only for sampling purposes (legally, districts must be of equal population).

The fact that the majority of plans favor Republicans, even while Democrats received more statewide votes for the election used in this analysis, is generally considered to be a reflection of the partisan gerrymandering in Wisconsin political geography (Chen and Rodden, 2013; Herschlag et al., 2017; Wang, 2016): Democratic voters tend to naturally pack themselves in cities, namely Madison and Milwaukee, while Republican votes tend be more spread out in rural areas. In order to overcome this settling bias and produce maps that are more favorable to Democrats, districts encompassing rural areas nearby cities would essentially need to trade some of their rural area for a portion of city area, which often necessitates the creation of non-compact districts. What exact sorts of spatial clustering in political geography result in different partisan leans, however, is still an open question (Duchin and Walch, 2021; Guo, 2008).

Indeed, appeals to partisan proportionality need to contend with how communities of interest are impacted by such practices. Using the ensembles sampled here, we can see the trade-offs between compactness, IR, and partisan bias by comparing low and high compactness biased RST chains, which together have the lowest and highest range of IR values, respectively, of any of the chains. Since compactness generally has a monotonic relationship with IR value, as seen in the 3D-cube visualization, we can consider these two chains to represent a low compactness and

low IR ensemble, and high compactness and high IR ensemble, respectively. To compare these two ensembles of redistricting plans in terms of efficiency gap and seat allocation, we count the number of times in each ensemble, respectively, that a given seat allocation occurred. Then, we calculate the average efficiency gap for all of the district plans in each seat allocation category, with the resulting averages corresponding to the efficiency gap bands noted in the 3D-cube. For reference, statewide vote percentages for the election in use (US House 2018) were ∼54% and ∼46%, for Democrats and Republicans, respectively, resulting in a slightly negative efficiency gap for district plans where equal seat allocation was achieved.

As seen in Figure 7, the low compactness and low IR ensemble produced seat allocations varying between 2-Democrats, 6-Republicans (in total 8 seats) and the inverse, with most values falling in 3-5, 4-4, and 5-3 seat allocation categories. The *proportionality* (4-4 seat allocation) is achieved in a relatively small majority of redistricting maps (compared to Figure 8), and the same category has a relatively neutral efficiency gap average. In contrast, the high compactness and high IR ensemble produced a much larger fraction of maps that had proportional outcomes overall, as seen in Figure 8, suggesting that pursuing higher IR maps to better encapsulate COIs may also help produce results that have, on average, more proportional outcomes. Importantly, this analysis applies to the spatial scale used (US congressional districts composed of CBG-level geometry) and the particular region used (Wisconsin); the relationships between IR, compactness, and proportionality may differ for analyses at different spatial scales or in other regions.
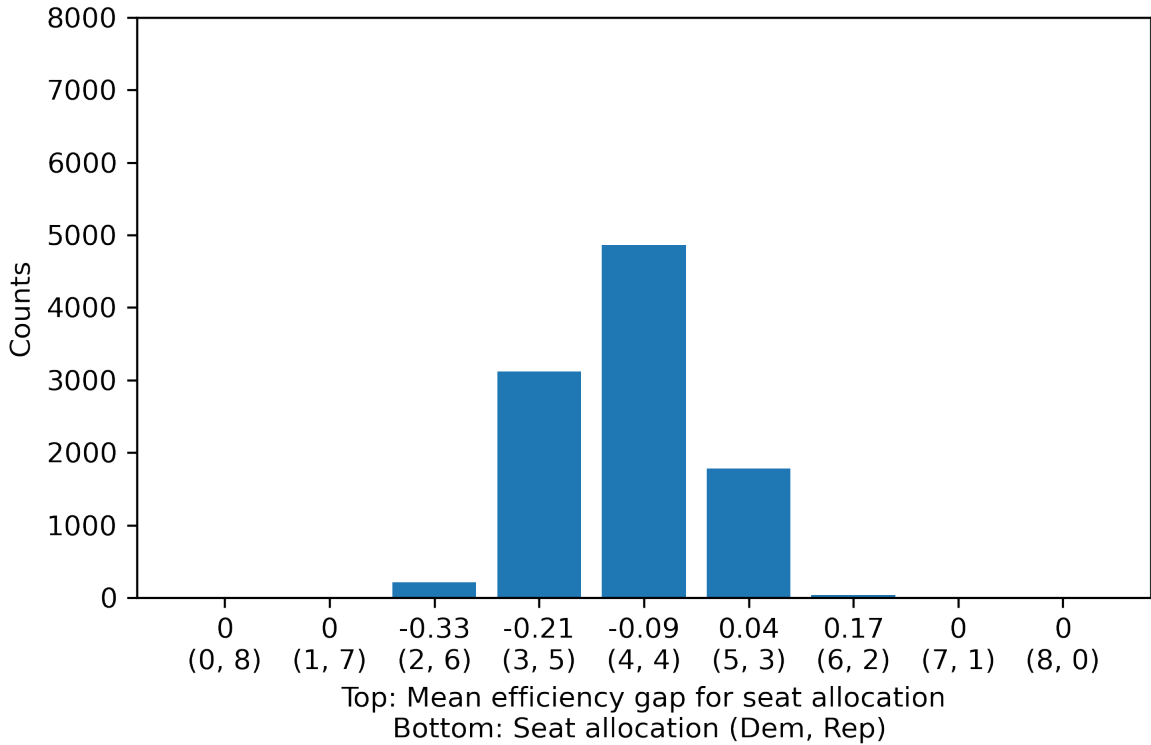

Figure 7.: Negatively biased RST ensemble results, with low compactness and IR values, grouped by seat allocations. Seat allocation is reported on the lower level of labels on the x-axis in the format of "(Democrat, Republican)" seats won under a given plan, e.g. the label "(2,6)" means that Democrats won two seats and Republicans won six. The average efficiency gap is reported for each seat allocation ratio. [two-column width]
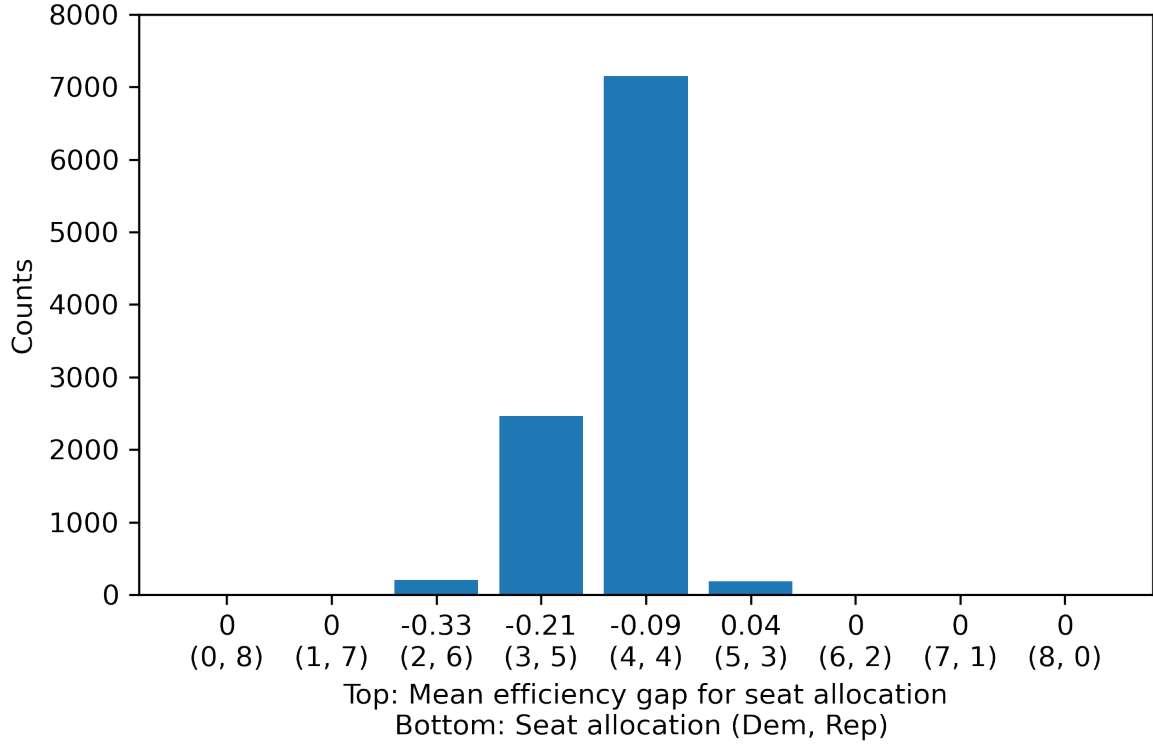
Figure 8.: Positively biased RST ensemble results (with high compactness and IR values) grouped by seat allocations, where seat allocation is reported as Democrat, Republican seats won under a given plan. The average efficiency gap is reported for each seat allocation ratio. Compared to the negatively biased RST ensemble, this ensemble produced many more maps with proportional representation (4:4). [two-column width]

4.3. Interactive Web Map

In our developed web map, as seen in Figure 9, public users are allowed to choose from Wisconsin congressional district plans that are optimized for compactness, the IR, and the efficiency gap, respectively. We also provide the currently enacted district map, and a districting plan map from the People's Map Commission, for reference. Existing works have shown the importance of demographic and socioeconomic variables for public policy making (Kind and Buckingham, 2018; Hou et al., 2021), and so demographic attributes and mobility flow values for each district can be visualized on the choropleth maps by selecting the corresponding header (with bold text) on the parallel coordinate plot. The same values can also be resymbolized as bar plots or proportional symbol maps. Finally, an OpenStreetMap base map provides spatial context by showing city and road names. The two map panels are synchronized, such that the panning or zooming on one map is performed in equal measure on the other map. Map interaction, synchronization and the ability to select which variable to display are key features in making this web map educational. By exploring the redistricting maps and trying out different interactions on the website, users can get a better grasp of how district map attributes relate to each other. Similarly, users can interactively reposition the 3D cube for the associated results of ensemble plans. An additional feature of the interactive form of the 3D cube is the ability to view point attributes, displaying the compactness, efficiency gap, and IR values for any point that is hovered over (Figure 10). We also include text descriptions and diagrams below the interactive maps and charts, providing redistricting context and implementation details for the evaluation metrics used in the web map. The map is available to the public at `https://geography.wisc.edu/geods-file/redistricting/`.
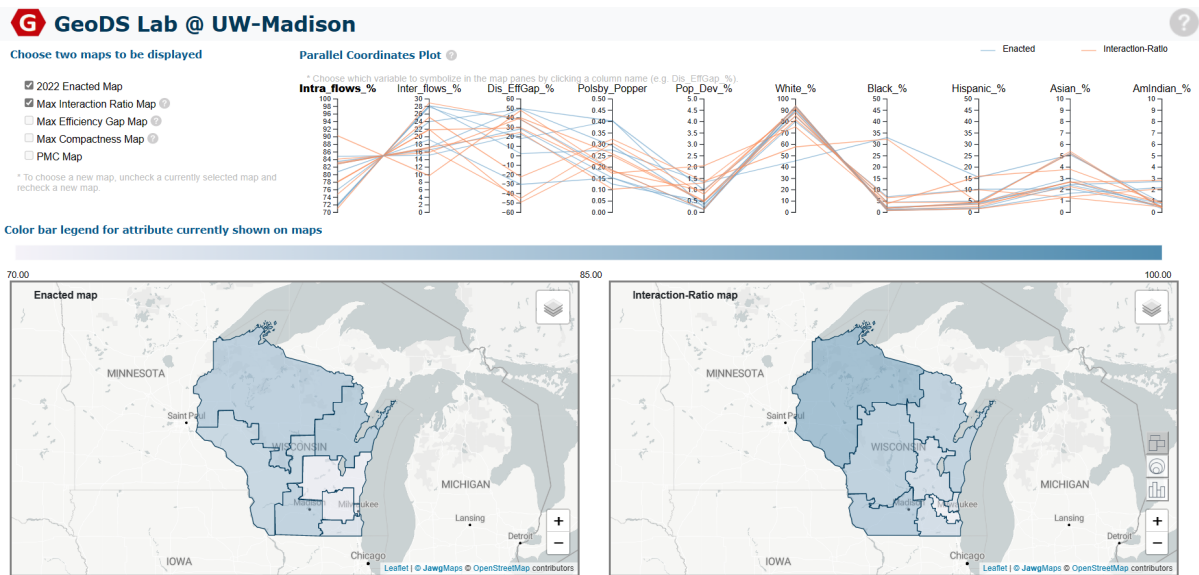
Figure 9.: Our developed interactive web map, with the 2022 enacted district plan displayed in the left map panel, and the maximum Interaction Ratio map displayed on the right. [two-column width]

Interactive redistricting values cube

While the above map panes allow us to explore multiple redistricting, there are in fact trillions of maps that could be made for the eight Congressional House districts. To get a sense of the properties of district plans from this vast space of redistricting possibilities, we sample some of the maps and plot them in three dimensions, allowing for a multi-dimensional analysis. This also also helps us know how well a given map performs relative to reasonable alternatives.
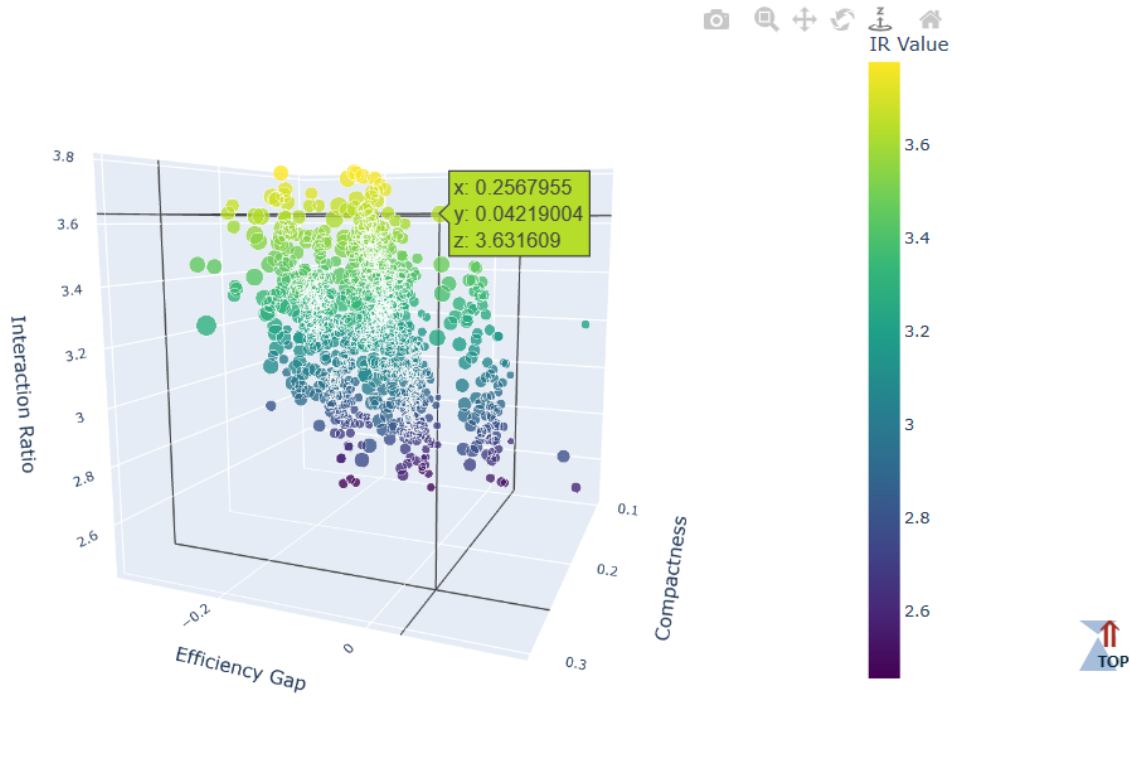

Figure 10.: Variables x, y, and z on the label produced by mouse-over correspond to compactness, efficiency gap, and IR values, respectively. [two-column width]

4.4. Interactive Web Map Usability Assessment

The results of the functionality comparison between our web map and other public redistricting websites can be seen in Figure 11, where dark blue circles represent full functionality, half-filled circles represent partial functionality, and light blue circles represent no functionality. This type of functionality comparison method has been widely adopted in the cartography and visual analytics communities (Roth et al., 2015; Tu et al., 2023). As seen in the first column, our website and DRA provide more functionality than the other two websites. However, ours is the only web map that allows for two maps to be compared side by side, which is important for understanding the impact of plans on specific cities and regions. While our website does provide an interactive 3D cube, which has ensemble distribution values displayed in three dimensions, it does not highlight the provided maps (optimal or otherwise) in that cube. In DRA, the selected maps for presentation are shown against box plots for ensemble distributions. For displaying district-level attributes, our website is the only website that provides metrics for two maps in a parallel coordinate plot, while DRA did allow for comparison between the metrics of various maps, but without being able to view the maps themselves. DRA provided the ability to decide which attribute is expressed in the choropleth map, but ours is the only one that provides

re-expression in the form of bar plots or proportional symbol maps. Notable, all of the other maps supported plan creation, while ours does not, though that was not a goal of our website as it mainly focuses on multi-criteria redistricting plan assessment.

	Our web map	DRA	Districtr	DistrictBuilder
Compare multiple maps	●	○	○	○
Compare plan to distribution	◐	●	○	○
Compare metrics for multiple plans	●	◐	○	○
Provide map re-expression	●	◐	○	○
Plan creation	○	●	●	●
Provide optimal maps	●	●	○	○
Provide reference layers	●	●	◐	◐
Provide redistricting metrics	●	●	●	●

Figure 11.: Comparison of our interactive web map with other redistricting websites. Dark blue circles represent full functionality, half filled circles represent partial functionality, and light blue circles represent no functionality. [two-column width]

Figure 12 shows the average accuracy results (in percentage) of the usability assessment on our developed web map using the six tasked introduced in section 3.6. Response accuracy was generally very high for all map and plot reading tasks, indicating effective web design to facilitate knowledge extraction. The most difficult task for participants involved interacting with the 3D cube, suggesting that some sort of visual affordance is needed to make users aware that the 3D cube can be repositioned with mouse control, which poses a spatial cognition challenge to certain degree. Users who were able to use the cube interactively had used similar plots before in other contexts, and so assumed that the cube in the assessment was also interactive, while users who were not able to interact with the cube had no such experience. However, users were generally still able to interpret the meaning of points in the cube, as related to the three metrics used in the cube axes.
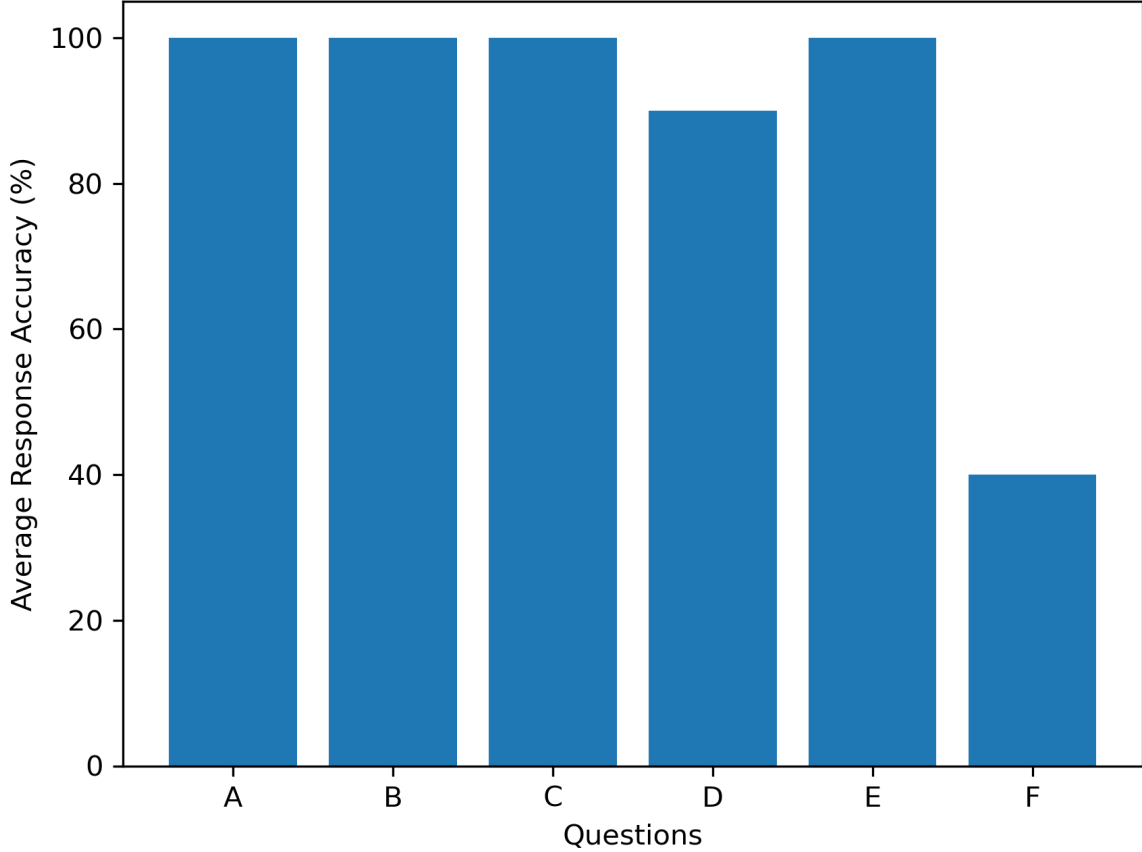
Figure 12.: Average response accuracy for the usability survey. The questions are as follows: A) Add the CBG overlay to the left map pane, B) Hover over the district that contains the label "Milwaukee" in each of the map panes, C) Change the 2022 Enacted Map panel to show the Max Compactness Map, D) Hover over the line in the parallel coordinate plot corresponding to the district with the highest inter_flows value for the Interaction-Ratio map, E) Read the Dis_effGap value for district containing the city of Madison in the 2022 Enacted Map, F) Hover over the point with the minimum compactness value in the interactive redistricting values 3D cube. [two-column width]

5. Discussion

5.1. ReCom Sampling Modifications

As shown in section 4.1, the biased RST and max-IR cut methods are able to shift the sampled district plan distributions towards higher or lower IR values, as desired. We believe the reason that the algorithm was unable to sample more widely has to do with the spanning tree distribution. While this distribution yields districts that are visually compact, it is also likely restricting the space explored by our adapted algorithms. For instance, it is easy to imagine that some very non-compact districts could produce the highest IR values, if the districts were well lined-up with the underlying communities. In fact, we did set the compactness threshold even lower, to $7\times$ the number of cut edges in the original plan, but it did not change the IR distribution over $5\times$ the number of cut edges. This is likely because non-compact districts are rarely produced by random spanning trees, even if the compactness threshold allows for them. More generally, further research is needed to understand the spanning tree distribution used in ReCom (Procaccia and Tucker-Foltz, 2022), and such research may shed light on how to bias the distribution of IR values sampled.

5.2. Relationship Between Interaction Ratio and Current Redistricting Criteria

While a fairly monotonic relationship exists between compactness scores and IR values in this study, that may not hold for other regions. We have compared IR values to other traditional redistricting metrics (e.g., Polsby-Popper compactness and efficiency gap) so that IR can be understood in context, but, in actual use for redistricting, we would hope that the presence of actual IR-defined communities would be used as the basis for districts, rather than as a means to justify proxies like compactness. Still, the adoption of the IR for redistricting would require further study into how it relates to existing redistricting goals and criteria.

Another common redistricting constraint, which we have not addressed explicitly in this paper, is geographic contiguity (although the spanning tree method already considers the spatial connectedness of adjacent nodes/districts). Enshrined in U.S. redistricting custom is a territorial basis for districts (Makse, 2012). However, there are types of communities that, while sharing common political interests, may not be geographically adjacent. Examples of this may be minority communities or virtual communities. The IR could incorporate other types of flows between districts, such as phone calls (Ratti et al., 2010; Gao et al., 2013) or friendship link strength (Bailey et al., 2018), to help quantify non-contiguous districts. More broadly, we note that spatial interaction communities may only encompass some portion of what different parties consider to be COIs, since COIs have been considered to be racial, geographic, economic, and political (Williams and Grofman, 2019).

Finally, we note, as other have, that better capturing communities of interest may not always result in more equal seat allocation between the two major parties or more competitive districts. Indeed, the particular spatial arrangement (Duchin and Walch, 2021; Guo, 2008) and partisan nature (Gimpel and Harbridge-Yong, 2020) of certain types of communities may inherently make districts less competitive.

5.3. Interactive Web Map Assessment

While our developed website provides functionality not present in other redistricting websites and scored highly in the usability assessment, there are limitations in the design of the assessment. Most notable, users familiar with GIS and web maps were surveyed, rather than redistricting experts with experience or interest in redistricting. Therefore, the tasks were mainly focused on usability, rather than on the sorts of tasks that people familiar with redistricting would be trying to accomplish, such as comparing the district shapes or identifying particular communities. However, our current assessment is still useful in demonstrating effective web design for the basic map-based functionalities that would facilitate insight by people exploring the map specifically for redistricting. Our future work will seek community engagement with the state-level redistricting committee to further assess our web map product.

6. Conclusion

In this study, we propose that spatial interaction communities be used as the basis of COIs in redistricting. To quantify spatial interaction communities, we present a new metric termed the interaction ratio (IR), which compares the global sum of intra-district population flows to the global sum of inter-district population flows for redistricting. To explore the range of IR values associated with valid district plans, we employ the ReCom algorithm on the space of graph partitions to produce an ensemble of district plans, and then calculate the IR values for all of the plans in the ensemble. As such ensembles are used frequently in legal challenges to district plans, we also propose two methods, the biased RST and min/max-IR cut methods, which allow the user to better specify the distribution of IR values that should be sampled. Using the results from these methods, we turn to a multi-criteria assessment of redistricting plans, where we show that an ensemble with higher IR values produced more proportional seat allocation outcomes for the 2018 Federal Congressional election in Wisconsin, when compared to an ensemble with lower

23

average IR-values. We also developed a web map which allows for exploration of the redistricting results from the multi-criteria assessment produced in this research. Future research could focus on more comparisons of the IR with existing redistricting criteria, and as well as methods for changing the distribution of sampled IR values. Additionally, the relationships between IR, compactness, and proportionality in voting allocation could be explored at different spatial scales and in other regions.

This research contributes to the political science literature on redistricting by introducing spatial interaction communities as the basis of COIs. It also contributes to computational redistricting research and practice by proposing two methods for biasing the sample distribution towards higher or lower IR values. Through the multi-criteria assessment described here and made available in the interactive web map, we provide important context for understanding how IR values relate to the values of existing redistricting metrics, which is necessary for the adoption of this new metric in redistricting and regionalization practices.

Data availability statement

The mobility flow dataset used in this research is publicly available on GitHub: `https://github.com/GeoDS/COVID19USFlows` and from SafeGraph. The other aggregated data that support the findings of this study are available from the U.S. census bureau. Due to the privacy protection policies of the data providers, the voting data used here are not publicly available.

Acknowledgments

References

Altman, M. (1997). The computational complexity of automated redistricting: Is automation the answer. *Rutgers Computer & Tech. LJ*, 23:81.

Altman, M., MacDonald, K., and McDonald, M. (2005). From crayons to computers: The evolution of computer use in redistricting. *Social Science Computer Review*, 23(3):334–346.

Aydin, O., Janikas, M. V., Assunçao, R., and Lee, T.-H. (2018). Skater-con: Unsupervised regionalization via stochastic tree partitioning within a consensus framework using random spanning trees. In *Proceedings of the 2nd ACM SIGSPATIAL international workshop on AI for geographic knowledge discovery*, pages 33–42.

Bailey, M., Cao, R., Kuchler, T., Stroebel, J., and Wong, A. (2018). Social connectedness: Measurement, determinants, and effects. *Journal of Economic Perspectives*, 32(3):259–80.

Becker, A., Duchin, M., Gold, D., and Hirsch, S. (2021). Computational redistricting and the voting rights act. *Election Law Journal: Rules, Politics, and Policy*, 20(4):407–441.

Benade, G., Buck, R., Duchin, M., Gold, D., and Weighill, T. (2021). Ranked choice voting and minority representation. *Available at SSRN 3778021*.

Bergey, P. K., Ragsdale, C. T., and Hoskote, M. (2003). A simulated annealing genetic algorithm for the electrical power districting problem. *Annals of Operations Research*, 121(1):33–55.

Bernstein, M. and Duchin, M. (2017). A formula goes to court: Partisan gerrymandering and the efficiency gap. *Notices of the AMS*, 64(9):1020–1024.

Brunell, T. L. (2008). *Redistricting and Representation: Why Competitive Elections are Bad for America*. Routledge.

Chen, J. and Rodden, J. (2013). Unintentional gerrymandering: Political geography and electoral bias in legislatures. *Quarterly Journal of Political Science*, 8(3):239–269.

Chikina, M., Frieze, A., Mattingly, J. C., and Pegden, W. (2020). Separating effect from significance in markov chain tests. *Statistics and Public Policy*, 7(1):101–114.

DeFord, D., Dhamankar, N., Duchin, M., Gupta, V., McPike, M., Schoenbach, G., and Sim, K. W. (2021). Implementing partisan symmetry: Problems and paradoxes. *Political Analysis*, pages 1–20.

DeFord, D., Duchin, M., and Solomon, J. (2018). Comparison of districting plans for the virginia house of delegates. *MGGG Technical Report*.

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

Diaconis, P. (2009). The markov chain monte carlo revolution. *Bulletin of the American Mathematical Society*, 46(2):179–205.

Dong, H., Wu, M., Ding, X., Chu, L., Jia, L., Qin, Y., and Zhou, X. (2015). Traffic zone division based on big data from mobile phone base stations. *Transportation Research Part C: Emerging Technologies*, 58:278–291. Big Data in Transportation and Traffic Engineering.

Duchin, M. (2018). Outlier analysis for pennsylvania congressional redistricting. *LWV vs. Commonwealth of Pennsylvania Docket No. 159 MM 2017*.

Duchin, M. and Tenner, B. E. (2018). Discrete geometry for electoral geography. *arXiv preprint arXiv:1808.05860*.

Duchin, M. and Walch, O. (2021). *Political Geometry*. Springer.

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.

Forest, B. (2004). Information sovereignty and gis: The evolution of "communities of interest" in political redistricting. *Political Geography*, 23(4):425–451.

Gao, S., Liu, Y., Wang, Y., and Ma, X. (2013). Discovering spatial interaction communities from mobile phone data. *Transactions in GIS*, 17(3):463–481.

Gao, S., Rao, J., Kang, Y., Liang, Y., and Kruse, J. (2020). Mapping county-level mobility pattern changes in the united states in response to covid-19. *SIGSpatial Special*, 12(1):16–26.

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.

Guo, D. (2008). Regionalization with dynamically constrained agglomerative clustering and partitioning (redcap). *International Journal of Geographical Information Science*, 22(7):801–823.

Guo, D. and Jin, H. (2011). iredistrict: Geovisual analytics for redistricting optimization. *Journal of Visual Languages & Computing*, 22(4):279–289.

Haynes, K. E. and Fotheringham, A. S. (2020). *Gravity and spatial interaction models*. Regional Research Institute, West Virginia University.

Herschlag, G., Ravier, R., and Mattingly, J. C. (2017). Evaluating partisan gerrymandering in wisconsin. *arXiv preprint arXiv:1709.01596*.

Hou, X., Gao, S., Li, Q., Kang, Y., Chen, N., Chen, K., Rao, J., Ellenberg, J. S., and Patz, J. A. (2021). Intracounty modeling of covid-19 infection with human mobility: Assessing spatial heterogeneity with business traffic, age, and race. *Proceedings of the National Academy of Sciences*, 118(24):e2020524118.

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., and Kruse, J. (2020). Multiscale dynamic human mobility flow dataset in the us during the covid-19 epidemic. *Scientific data*, 7(1):1–13.

Katz, J. N., King, G., and Rosenblatt, E. (2020). Theoretical foundations and empirical evaluations of partisan fairness in district-based democracies. *American Political Science Review*, 114(1):164–178.

Kind, A. J. and Buckingham, W. R. (2018). Making neighborhood-disadvantage metrics accessible—the neighborhood atlas. *The New England journal of medicine*, 378(26):2456.

Lab, M. R. (2022). Gerrychain. *https://github.com/mggg/GerryChain*.

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

Li, B., Gao, S., Liang, Y., Kang, Y., Prestby, T., Gao, Y., and Xiao, R. (2020). Estimation of regional economic development indicator from transportation network analytics. *Scientific reports*, 10(1):1–15.

Liang, Y., Zhu, J., Ye, W., and Gao, S. (2022). Region2vec: Community detection on spatial networks using graph embedding with node attributes and spatial interactions. In *Proceedings of the 30th International Conference on Advances in Geographic Information Systems*, pages 1–4.

Makse, T. (2012). Defining communities of interest in redistricting through initiative voting. *Election Law Journal*, 11(4):503–517.

Malone, S. J. (1997). Recognizing communities of interest in a legislative apportionment plan. *Va. L. Rev.*, 83:461.

Nagle, J. F. (2019). What criteria should be used for redistricting reform? *Election Law Journal: Rules, Politics, and Policy*, 18(1):63–77.

NCSL, National Conference of State Legislatures (2021). Redistricting criteria.

Nelson, G. D. and Rae, A. (2016). An economic geography of the united states: From commutes to megaregions. *PloS one*, 11(11):e0166083.

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)*, pages 3754–3771. SIAM.

Ramachandran, G. and Gold, D. (2018). Using outlier analysis to detect partisan gerrymanders: A survey of current approaches and future directions. *Election Law Journal: Rules, Politics, and Policy*, 17(4):286–301.

Rao, J., Gao, S., Miller, M., and Morales, A. (2022). Measuring network resilience via geospatial knowledge graph: a case study of the us multi-commodity flow network. In *Proceedings of the 1st ACM SIGSPATIAL International Workshop on Geospatial Knowledge Graphs*, pages 17–25.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., Claxton, R., and Strogatz, S. H. (2010). Redrawing the map of great britain from a network of human interactions. *PloS one*, 5(12):e14248.

Roth, R. E., Ross, K. S., and MacEachren, A. M. (2015). User-centered design for interactive maps: A case study in crime analysis. *ISPRS International Journal of Geo-Information*, 4(1):262–301.

Stephanopoulos, N. O. (2012a). Redistricting and the territorial community. *University of Pennsylvania Law Review*, pages 1379–1477.

Stephanopoulos, N. O. (2012b). Spatial diversity. *Harvard Law Review*, pages 1903–2010.

Stephanopoulos, N. O. and McGhee, E. M. (2015). Partisan gerrymandering and the construction of american democracy. *University of Chicago Law Review*, 82(1):1–50.

Tu, Y., Wang, X., Qiu, R., Shen, H.-W., Miller, M., Rao, J., Gao, S., Huber, P. R., Hollander, A. D., Lange, M., et al. (2023). An interactive knowledge and learning environment in smart foodsheds. *IEEE Computer Graphics and Applications*.

Turner Jr, M. L. and LaMacchia, R. A. (1999). The us census, redistricting, and technology: A 30-year perspective. *Social science computer review*, 17(1):16–26.

Wang, S. S.-H. (2016). Three practical tests for gerrymandering: Application to maryland and wisconsin. *Election Law Journal*, 15(4):367–384.

Webster, G. R. (2013). Reflections on current criteria to evaluate redistricting plans. *Political Geography*, 32:3–14.

Williams, B. and Grofman, B. (2019). Community of interest criteria in redistricting: Legal and institutional frameworks. *Election Law Journal*, 18(4):477–499.

Williams Jr, J. C. (1995). Political redistricting: a review. *Papers in Regional Science*, 74(1):13–40.

Winburn, J. (2008). *The realities of redistricting: Following the rules and limiting gerrymandering in state legislative redistricting*. Lexington Books.

Xu, Y., Zou, D., Park, S., Li, Q., Zhou, S., and Li, X. (2022). Understanding the movement predictability of international travelers using a nationwide mobile phone dataset collected in south korea. *Computers, Environment and Urban Systems*, 92:101753.