SIAM J. SCI. COMPUT.
Vol. 42, No. 5, pp. A3340–A3366

GEOMETRY OF GRAPH PARTITIONS VIA OPTIMAL TRANSPORT*

TARA ABRISHAMI†, NESTOR GUILLEN‡, PARKER RULE§, ZACHARY SCHUTZMAN¶, JUSTIN SOLOMON‖, THOMAS WEIGHILL§, AND SI WU#

Abstract. We define a distance metric between partitions of a graph using machinery from optimal transport. Our metric is built from a linear assignment problem that matches partition components, with assignment cost proportional to transport distance over graph edges. We show that our distance can be computed using a single linear program without precomputing pairwise assignment costs and derive several theoretical properties of the metric. Finally, we provide experiments demonstrating these properties empirically, specifically focusing on the metric's value for new problems in ensemble-based analysis of political districting plans.

Key words. partitions, optimal transport, network flows, convex optimization

AMS subject classifications. 65K10, 90B06, 05C21

DOI. 10.1137/19M1295258

1. Introduction.

Several mathematical and computational problems involve collections of graph partitions with fixed numbers of components. Example application scenarios include tracking and clustering of evolving communities in a network, as well as analysis of political redistricting plan ensembles—an application we will study in detail below. Because specifying a single partition requires a label for every vertex, however, it can be difficult to visualize and navigate such a collection. Additionally, because the number of possible partitions typically exponentiates in the size of the underlying graph, collections of partitions usually are extremely large.

Enriching the set of partitions with a geometric structure helps understand the vast space of partitions. A distance allows us to quantify similarity or difference between two partitions and provides insight into the structure of the space of partitions as a whole. For instance, a large family of partitions might be close in a given metric to elements in a smaller family; in this case, we can efficiently infer information about the larger collection from a representative subsample. More broadly, a distance can

*Submitted to the journal's Methods and Algorithms for Scientific Computing section October 24, 2019; accepted for publication (in revised form) July 27, 2020; published electronically October 20, 2020. This work was initiated during the summer 2019 Voting Rights Data Institute in a team project on distances between partitions (faculty lead: Justin Solomon); authors are in alphabetical order by last name. Additional team members are listed in the Acknowledgments.
https://doi.org/10.1137/19M1295258
Funding: The work of the second author was supported by the NSF through grant DMS-1700307. The work of the fifth author was supported by the NSF through grant IIS-1838071, by the Air Force Office of Scientific Research through award FA9550-19-1-0319, by the Army Research Office through grant W911NF-12-R-001, and by the Prof. Amar G. Bose Research Grant. The work of the sixth author was supported by the NSF through grant OIA-1937095. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of these organizations.
†Corresponding author. Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08540 USA (taraa@princeton.edu).
‡Mathematics, Texas State University, San Marcos, TX 78666 USA (nestor@txstate.edu).
§Tufts University, Medford, MA 02155 USA (parker.rule@tufts.edu, thomas.weighill@tufts.edu).
¶University of Pennsylvania, Philadelphia, PA 19104 USA (ianzach@seas.upenn.edu).
‖Massachusetts Institute of Technology, Cambridge, MA 02139 USA (jsolomon@mit.edu).
#Boston University, Boston, MA 02215 USA (siwu1@bu.edu).

evaluate whether a sample spreads over the space of partitions or concentrates in a smaller region. A metric also yields a visualization tool: Given a finite family of partitions, one may compute pairwise distances and use them with an embedding algorithm to create a two-/three-dimensional Euclidean visualization.

Motivated by these challenges, we present a distance on the space of graph partitions motivated by optimal transport. Our model uses transport to measure pairwise relationships between the components of two partitions, a linear assignment problem then extracts the minimum cost (perfect) matching between partition components. Our formulation is a *hierarchical* transport problem that is invariant to the ordering of the individual components and sensitive to geometry, in contrast to simpler overlap-based measures, e.g., those based on Kullback–Leibler divergence or total variation. We derive theoretical properties of our distance and provide an extension to unbalanced problems where the components are weighted unequally.

Our target application is in *political redistricting*, where we can use these tools to compare districting plans for some geographic region. In ensemble-based approaches to districting plan analysis, a collection of feasible districts is generated computationally as a baseline for evaluating a proposed plan; the baseline samples achievable properties for plans given the political geography of a state. An issue in current ensemble-based redistricting pipelines is that the only two options for visualizing and navigating the ensemble are (1) showing a few randomly selected example plans or (2) plotting the empirical distribution of the values of a given measure, such as the number of districts won by a political party or district perimeter, over the ensemble. The first option shows an exceptionally small subset of ensembles that can number in the millions, while the second is an indirect means of understanding the relationships between plans. Here, we show that our transport metric—coupled with embeddings like multidimensional scaling—provides a third alternative, giving a direct and intuitive means of visualizing an entire ensemble. Our experiments confirm the value of this approach on both synthetic and real-world datasets.

Outline. In section 2, we comment on relevant literature, including that from optimal transport and ensemble-based redistricting. Next, section 3 reviews terms and notation. Section 4 introduces optimal transport and the Wasserstein metric, including the Kantorovich and Beckmann problems. In section 5 and section 6, we introduce our distance metric between partitions and prove basic properties. In section 7, we compare our metric to existing alternatives. Finally, section 8 demonstrates the metric on simulated and real geographic data, and section 9 summarizes our work, including open problems and avenues for future research.

2. Related work.

Optimal transport. Optimal transport (OT) is a field of mathematics and computer science dating back at least to the late eighteenth century when Monge posed the problem of determining the most efficient way of transporting a distribution of a given material from one configuration to another [31], a problem whose understanding was limited until over a century later, when Kantorovich reformulated it as a linear program [24]. Beyond this concrete problem lies the broader problem of determining optimal rearrangements (or matchings) between distributions, a problem that arises in many fields. Accordingly, OT-based methods are used today in economics, probability, statistics, machine learning, fluid mechanics, computer graphics, and other fields. The interested reader can find thorough discussion of these domain-applications in [19, 34, 36, 39].

In this paper, we view each component of a graph partition as a kind of distribution by assigning a mass to each vertex in that component. We can then view a graph partition as a collection of these distributions and measure distances between different graph partitions by using OT methods on the constituent distributions. This departs from the context where OT was initially developed (matching probability measures), but more recent work has considered similar instances like partial transport [5] and matching measures with unequal mass [18]. The latter is an example of *unbalanced* OT, a popular topic in modern theoretical/applied transport [9, 10, 26]. This being said, we are unaware of work using OT methods to measure the similarity between two different partitions of a graph.

Spurred by applications in machine learning, computer vision, and other disciplines, several algorithms have been developed for approximate OT and derived quantities; see [34] for a survey. Particularly relevant to our work is transport over graphs with shortest-path distance as the cost, known as *minimum-cost flow without edge capacities* [1] and—in OT—as the 1-Wasserstein distance or Beckmann problem [3, 35]. See [17] for a survey of computational methods for this problem.

Geometry of partitions. Several works imbue the set of partitions with a metric and study the resulting geometry. In geometric analysis, specifically the study of partitions of minimal perimeter, Leonardi and Tamanini introduced a metric on the space of (measurable) partitions of subsets of Euclidean space [25]. Their metric is based on symmetric differences, producing a complete and separable metric space. In the discrete setting, there has been more extensive research on the geometry of set-theoretic partitions. This includes examining the complexity of comparing two partitions [12, 28] and considering the geometry of partitions using machinery from information theory [15].

Ensemble-based redistricting. A growing body of research centers on ensemble-based redistricting, which uses algorithms to generate and analyze thousands or millions of candidate districting plans that meet some criteria [2, 6, 7, 8, 13, 22, 23]. These methods/analyses are increasingly used by legal experts, policymakers, and the public at large to inform debates around redistricting. In the legal context, ensembles of voting maps generated with these methods are being submitted as evidence to state and federal courts in redistricting and voting rights cases, underlining a pressing need to understand quantitative and qualitative properties of ensembles. This is challenging because the space of maps meeting reasonable criteria is large and poorly understood, making it difficult to quantify the "diversity" of a collection of plans. Previous works compute distributions of statistics of interest, such as the number of districts won by a particular political party or a compactness metric, and perform statistical analysis in this lower-dimensional space.

For such analyses to be robust, they should be performed on a sample of plans representative of the universe of valid plans, such as those that meet legal criteria. If an ensemble contains many plans that are all slight variations of one another, a projection to summary statistics may be misleading, since those similar plans likely yield similar statistics, which would in turn be overrepresented in the analysis. Assessing a sample's diversity requires a measure of dissimilarity, and a rigorous development of such a measure is not present in the previous literature. This paper presents a novel direction within the "geometry of redistricting" that is orthogonal to the primary direction of the field, which focuses on analyzing the shapes of the districts themselves; for a survey of classical approaches in shape analysis for political districts, see [40].

3. Preliminaries.

A graph will be denoted by $G = (V, E)$ with vertices V and edges E. If G is weighted with a weight $\omega : E \to \mathbb{R}$, we will write $G = (V, E, \omega)$. The signed incidence matrix associated to the graph G will be denoted by P:

TABLE 1
Notation.

Notation	Definition
$G = (V, E)$	Graph with vertices V and edges E
$G = (V, E, \omega)$	Weighted graph with vertices V, E, and weights $\omega : E \to \mathbb{R}$
$d(v, w)$	Shortest-path distance between $v, w \in V$
$M(V)$	Set of all mass distributions on V
$M(V)^k$	Ordered k-tuples of mass distributions on V
$M(V)^{k*}$	The set $M(V)^k$ modulo index rearrangements
$\text{Prob}(V)$	Set of all probability distributions on V
$\text{Prob}(V)^k$	Ordered k-tuples of probability distributions on V
$\text{Prob}(V)^{k*}$	The set $\text{Prob}(V)^k$ modulo index rearrangements
$W_1(f, g)$	Wasserstein distance between f and g

$$P_{ev} := \begin{cases} -1 & \text{if } e = (v, w) \text{ for some } w \in V, \\ 1 & \text{if } e = (w, v) \text{ for some } w \in V, \\ 0 & \text{otherwise.} \end{cases}$$

We denote by $\text{M}(V)$ the set of all mass distributions over V and by $\text{Prob}(V) \subset \text{M}(V)$ the set of all probability distributions over V. Specifically,

$$\text{M}(V) = \{x \in \mathbb{R}^{|V|} \mid x(v) \geq 0 \; \forall \; v \in V\} \quad \text{and}$$
$$\text{Prob}(V) = \left\{x \in \mathbb{R}^{|V|} \mid x(v) \geq 0 \; \forall \; v \in V \text{ and } \sum_{v \in V} x(v) = 1\right\}.$$

We will also consider the set of mass distributions over the product $V \times V$:

$$\text{M}(V \times V) = \{x \in \mathbb{R}^{|V|^2} \mid x(v, w) \geq 0 \; \forall \; v, w \in V\}.$$

Let $\text{Prob}(V)^k$ be the set of k-tuples of elements of $\text{Prob}(V)$, and let $\text{Prob}(V)^{k*} \cong \text{Prob}(V)^k / S_n$ be the set of k-tuples of elements of $\text{Prob}(V)$ up to reordering. Similarly, let $M(V)^k$ be the set of k-tuples of elements of $M(V)$, and let $M(V)^{k*} \cong M(V)^k / S_n$ be the set of k-tuples of $M(V)$ up to reordering. Table 1 provides relevant notation.

4. Transport distances.

In this section, we review some notions from the theory of OT that will be relevant to our discussion. We limit to a few basic results from transport over graph domains; see [36, 39] for the general case.

Let $x, y \in \text{M}(V)$. A *coupling* or *transport plan* between x and y is a function $\pi : V \times V \to \mathbb{R}_+$ such that

$$\sum_{w \in V} \pi(v, w) = x(v) \;\; \forall \; v \in V \text{ and } \sum_{v \in V} \pi(v, w) = y(w) \;\; \forall \; w \in V.$$

We will use $\Pi(x, y)$ to denote the set of such couplings.

Remark 4.1. If for $x, y \in \text{M}(V)$ there is at least one $\pi \in \Pi(x, y)$, then

$$\sum_{v \in V} x(v) = \sum_{v \in V} \sum_{w \in V} \pi(v, w) \text{ and } \sum_{w \in V} y(w) = \sum_{w \in V} \sum_{v \in V} \pi(v, w).$$

The sums on the right-hand sides are finite rearrangements and therefore agree. Conversely, if x and y have the same total mass $m > 0$, then the product distribution $\pi(v, w) = \frac{1}{m} x(v) y(w)$ belongs to $\Pi(x, y)$. This means that there are admissible plans between x and y exactly when these distributions have the same total mass.

The *total transportation cost* of $\pi \in \Pi(x, y)$ is $\sum_{v,w\in V} d(v, w)\pi(v, w)$, where the cost of moving mass between vertices $v, w \in V$ is the shortest-path distance $d(v, w)$. Then, the *transport distance* (or 1-Wasserstein distance) between x and y is the minimum total transportation cost for a plan $\pi \in \Pi(x, y)$, denoted $W_1(x, y)$:

$$W_1(x, y) := \min_{\pi\in\Pi(x,y)} \sum_{v,w\in V} d(v, w)\pi(v, w). \tag{4.1}$$

We take $W_1(x, y) = +\infty$ if x and y do not have the same total mass. A consequence of general OT theory is that W_1 defines a metric on $\mathrm{Prob}(V)$; see [39, Chapter 7] or [36, Chapter 5] for general discussion or [11] for a proof in the discrete case.

In practice, (4.1) can be difficult to solve because it deals with $|V| \times |V|$ pairwise distances. In graph theory, however, this problem is known as *minimum cost flow without edge capacities* and admits an alternative formulation scaling linearly in $|E|$:

$$W_1(x, y) = \begin{cases} \displaystyle\min_{J\in\mathbb{R}^{|E|}} & \sum_{e\in E} \omega(e)|J_e| \\ \text{subject to} & P^\top J = y - x, \end{cases} \tag{4.2}$$

where ω denotes edge weights and P is the incidence matrix. If G is unweighted, we can take $\omega(\cdot) \equiv 1$. The equivalence between (4.1) and (4.2) is discussed in subsection 6.4 in the broader setting of unbalanced transport. See also [17] and references therein for motivation as well as references to relevant algorithms.

5. The distance on partitions: Balanced case.

In this section, we propose a distance between graph partitions that lifts the transport distances described above. This distance is defined in two steps: computing distances between partition components and subsequently finding a minimum-cost matching between the partition components. In particular, we take the distance between components to be the Wasserstein distance and use linear assignment to find the matching. With this definition in place, we prove some basic properties of the lifted distance and give a formulation as a single combined linear program rather than a two-step procedure.

5.1. Distances between components.

Let $G = (V, E)$ be a graph, and let $(V_1, \ldots, V_k)$ be a partition of V. We represent $(V_1, \ldots, V_k)$ by an element of $\mathrm{Prob}(V)^{k*}$ as follows: To every V_i, we associate a vector $x_i \in \mathbb{R}^{|V|}$ such that

$$x_i(v) = \begin{cases} \frac{1}{|V_i|} & \text{if } v \in V_i, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

Then, $X = (x_1, \ldots, x_k) \in \mathrm{Prob}(V)^{k*}$ gives a concrete representation of $(V_1, \ldots, V_k)$. This expression defines a balanced representation of partitions, because $\sum_{v\in V} x_i(v) = 1$ for all $i \in \{1, \ldots, k\}$. The case of unbalanced representations is covered in section 6.

Remark 5.1. Given a strictly positive weight function on the vertices $\omega : V \to \mathbb{R}_+$, we can give an alternative definition of the vector x_i associated to component V_i as

$$x_i(v) = \begin{cases} \frac{\omega(v)}{\sum_{u\in V_i} \omega(u)} & \text{if } v \in V_i, \\ 0 & \text{otherwise.} \end{cases}$$

In our target application of political redistricting, this alternative definition can be useful when incorporating populations associated with census units.

Now that we have a representation of graph partitions in $\mathrm{Prob}(V)^{k*}$, we can define a distance between components of partitions. Let $X = (x_1, \ldots, x_k) \in \mathrm{Prob}(V)^{k*}$ and $Y = (y_1, \ldots, y_k) \in \mathrm{Prob}(V)^{k*}$ be partitions of G. The distance between components of X and Y is given by the (unweighted) Wasserstein distance, defined in section 4. In particular, for any $x_i \in X$ and $y_j \in Y$, we take

$$W_1(x_i, y_j) = \begin{cases} \min\limits_{J \in \mathbb{R}^{|E|}} & \sum\limits_{e \in E} |J_e| \\ \text{subject to} & P^\top J = y_j - x_i. \end{cases}$$

5.2. Distances between partitions. In subsection 5.1, we endowed the space of partition components with the Wasserstein distance. Here, we lift this distance to a distance between partitions using a linear assignment problem.

For ease of notation, we define the relevant constraint set for our problem:

Definition 5.2 (Birkhoff polytope). *The* Birkhoff polytope DS_k *is the set of* $k \times k$ doubly stochastic *matrices, nonnegative matrices whose rows/columns sum to* 1*:*

$$\mathrm{DS}_k = \{S \in \mathbb{R}^{k \times k} \mid S\mathbb{1} = \mathbb{1}, S^\top \mathbb{1} = \mathbb{1}, \text{ and } S \geq 0\}. \tag{5.2}$$

The Birkhoff–von Neumann theorem gives that DS_k is a convex polytope and its vertices are the *permutation matrices*, those elements of DS_k with integer entries.

Definition 5.3 (lifted distance). *Given a distance* $C : \mathit{Prob}(V) \times \mathit{Prob}(V) \to \mathbb{R}$*, the* lifted distance $A : \mathit{Prob}(V)^{k*} \times \mathit{Prob}(V)^{k*} \to \mathbb{R}$ *between partitions* X *and* Y *is*

$$A(X, Y) = \begin{cases} \min\limits_{S \in \mathbb{R}^{k \times k}} & \sum\limits_{ij} S_{ij} C(x_i, y_j) \\ \textit{subject to} & S \in \mathrm{DS}_k, \end{cases} \tag{5.3}$$

where $X = (x_1, \ldots, x_k), Y = (y_1, \ldots, y_k) \in \mathit{Prob}(V)^{k*}$*, and* $C(\cdot, \cdot)$ *is a distance between partition components. Unless otherwise noted, we will take* $C = W_1$ *from* (4.2)*.*

Because the extreme points of DS_k are permutations, the minimizer of this linear program is a matching between the components of X and the components of Y, and the distance is the sum of the pairwise distances between matched components.

Remark 5.4. Many properties of this lifting are independent of the ground metric, which in (5.3) is the transport distance W_1. We can view this construction as an instance of *hierarchical* OT, i.e., a matching problem whose cost comes from another matching problem; see [41] for an example in natural language processing.

Before studying properties of our construction, we verify that (5.3) lifts any distance between components (i.e., a distance on $\mathrm{Prob}(V)$) to a distance on $\mathrm{Prob}(V)^{k*}$.

Proposition 5.5. *Given any metric* $C : \mathit{Prob}(V) \times \mathit{Prob}(V) \to \mathbb{R}$*, the lifted distance* $A : \mathit{Prob}(V)^{k*} \times \mathit{Prob}(V)^{k*} \to \mathbb{R}$ *is a metric on* $\mathit{Prob}(V)^{k*}$*.*

Proof. Since C is a metric and S is nonnegative, it is immediate that A is nonnegative and symmetric.

Let $X = (x_1, x_2, \ldots, x_k), Y = (y_1, y_2, \ldots, y_k) \in \mathrm{Prob}(V)^{k*}$, and suppose $X \equiv Y$ in $\mathrm{Prob}(V)^{k*}$. Then, there exists some permutation P with $P_{ij} = 1$ if and only if $x_i = y_j$. Since $C(x_i, y_j) = 0$ when $x_i = y_j$, we have $\sum_{ij} P_{ij} C(x_i, y_j) = 0$. Hence, when $X \equiv Y$ in $\mathrm{Prob}(V)^{k*}$, $A(X, Y) = 0$. Conversely, suppose $A(X, Y) = 0$, and let S minimize (5.3). If $S_{ij} = 1$, then since the objective is zero we must have $C(x_i, y_j) = 0$. Because C is a metric, $C(x_i, y_j) = 0$ if and only if $x_i = y_j$. Therefore, $X \equiv Y$ in $\mathrm{Prob}(V)^{k*}$.

Now, suppose $X, Y, Z \in \mathrm{Prob}(V)^{k*}$. Let N and W be minimizing permutation matrices with $A(X,Y) = \sum_{ij} N_{ij} C(x_i, y_j)$ and $A(Y,Z) = \sum_{jl} W_{jl} C(y_j, z_l)$. Then,

$$A(X,Y)+A(Y,Z)=\sum_{ij} N_{ij}C(x_i,y_j)+\sum_{jl} W_{jl}C(y_j,z_l)=\sum_{ilj} \left(N_{ij}C(x_i,y_j)+W_{jl}C(y_j,z_l)\right).$$

Because N and W are permutations, for a fixed j, there is a unique i_j such that $N_{i_j j} = 1$ and a unique l_j such that $W_{jl_j} = 1$. Therefore,

$$A(X,Y) + A(Y,Z) = \sum_j \left(C(x_{i_j}, y_j) + C(y_j, z_{l_j})\right) \geq \sum_j C(x_{i_j}, z_{l_j})$$

by the triangle inequality. Let $B = NW$. Then, $\sum_j C(x_{i_j}, z_{l_j}) = \sum_{il} B_{il} C(x_i, z_l)$. Because B is a permutation, $\sum_{il} B_{il} C(x_i, z_l) \geq A(X,Z)$. Therefore, $A(X,Y) + A(Y,Z) \geq A(X,Z)$, verifying a triangle inequality. ∎

Remark 5.6. Since C is a metric, Proposition 5.5 likely follows from general results about discrete transport, e.g., [11, Theorem 1]. We include the direct proof since metric properties follow directly from our definition.

5.3. Basic properties. In this section, we prove several basic properties of the lifted distance. First, we show that if two partitions have a component in common, there exists an optimal matching that fixes the shared component.

PROPOSITION 5.7. *Let $X = (x_1, \ldots, x_k)$ and $Y = (y_1, \ldots, y_k)$ be two partitions of G, and suppose $x_a = y_b$. Then, there exists a matching S such that $S_{ab} = 1$ and S is an optimizer for the lifted distance* (5.3)*.*

Proof. Let P be a permutation matrix that is an optimizer for $A(X,Y)$, so $A(X,Y) = \sum_{ij} C(x_i, y_j) P_{ij}$. Suppose that P maps x_a to some y_d and some x_c to y_b. Let S be the permutation obtained by matching x_a to y_b, x_c to y_d, and every other component in X to its image under P. It is clear that

$$\sum_{ij} S_{ij} C(x_i, y_i) = \sum_{ij} P_{ij} C(x_i, y_j) - C(x_a, y_d) - C(x_c, y_b) + C(x_a, y_b) + C(x_c, y_d).$$

Since C is a metric, the triangle inequality yields $C(x_a, y_d) + C(x_c, y_b) \geq C(x_c, y_d)$. Because $x_a = y_b$ by assumption, $C(x_a, y_b) = 0$, so $-C(x_a, y_d) - C(x_c, y_d) + C(x_a, y_b) + C(x_c, y_d) \leq 0$, and therefore $\sum_{ij} S_{ij} C(x_i, y_i) \leq \sum_{ij} P_{ij} C(x_i, y_j)$. Since P is an optimizer for the lifted distance, we also have that $\sum_{ij} P_{ij} C(x_i, y_j) \leq \sum_{ij} S_{ij} C(x_i, y_j)$. Therefore, S is an optimizer for $A(X,Y)$ with $S_{ab} = 1$. ∎

Our formulation of the lifted distance involves pairwise distances between partition components and a subsequent linear assignment problem to find the minimum cost matching. Below, we formulate the lifted distance using only one linear program.

PROPOSITION 5.8. *The lifted distance $A(X,Y)$ between partitions X, Y satisfies*

$$(5.4) \qquad A(X,Y) = \begin{cases} \min\limits_{Q \in \mathbb{R}^{|E| \times k^2}, S \in \mathbb{R}^{k \times k}} & \sum\limits_{ij} \sum\limits_{e \in E} |Q^e_{ij}| \\ \textit{subject to} & S \in \mathrm{DS}_k, \\ & P^\top Q_{ij} - (x_i - y_j) S_{ij} = 0. \end{cases}$$

Proof. Substituting the transport cost (4.2) into (5.3), we can write

$$A(X,Y) = \begin{cases} \min\limits_{S\in\mathbb{R}^{k\times k}} & \sum\limits_{ij} S_{ij} \sum\limits_{e\in E} |J^{*e}_{ij}| \\ \text{subject to} & S \in \mathrm{DS}_k \\ & J^*_{ij} = \begin{cases} \operatorname*{argmin}\limits_{J_{ij}\in\mathbb{R}^{|E|}} & \sum\limits_{e\in E} |J^e_{ij}|, \\ \text{subject to} & P^\top J_{ij} = (x_i - y_j). \end{cases} \end{cases}$$

Since the inner and outer problems are both minimizations, we can simplify to

$$A(X,Y) = \begin{cases} \min\limits_{S\in\mathbb{R}^{k\times k}, J\in\mathbb{R}^{|E|}} & \sum_{ij} S_{ij} \sum\limits_{e\in E} |J^e_{ij}| \\ \text{subject to} & S \in \mathrm{DS}_k, \\ & P^\top J_{ij} = (x_i - y_j) \end{cases}$$

This is a quadratic program in S and J. Substituting $Q^e_{ij} = S_{ij} J^e_{ij}$ yields (5.4). ∎

The formula in Proposition 5.8 computes the distances between components and the assignment in a single linear program, suggesting an alternative means for computing $A(\cdot,\cdot)$ without precomputing pairwise costs. Even so, S gives the minimum matching of the components of X and Y, and Q represents the per-edge flow.

A standard linear programming duality argument applied to (5.4) shows

$$A(X,Y) = \begin{cases} \max\limits_{\phi,\psi\in\mathbb{R}^k, \gamma\in\mathbb{R}^{|V|\times k^2}} & \mathbb{1}^\top(\phi+\psi) \\ \text{subject to} & \phi_i + \psi_j \le \gamma_{ij}^\top (x_i - y_j), \\ & |\gamma^w_{ij} - \gamma ij^v| \le 1 \ \forall\ (w,v) \in E. \end{cases} \tag{5.5}$$

Similar to the argument in Proposition 5.8, this formula also can be derived directly by substituting the dual of (4.2) into the dual of (5.3).

6. The distance on partitions: Unbalanced case. We now revisit the construction in section 5 to propose a distance between graph partitions when the mass of each component may not be equal. We call these *unbalanced partitions*. Similar to the balanced case, we define the distance in two steps: computing distances between partition components and using a linear assignment to lift the distance between components to a distance between partitions. Our distance between components is a modified transport distance that allows mass to be inserted or removed at vertices with some cost. Lifting the unbalanced distance between components gives a valid metric on the space of unbalanced partitions; we also prove basic properties of the general distance.

6.1. Distances between unbalanced components. Let $G = (V, E, \omega)$ be a weighted graph, and let $(V_1, \dots, V_k)$ be a partition of the vertices of G. We define an unbalanced representation of $(V_1, \dots, V_k)$ in $M(V)^{k*}$ as follows: To every component V_i, we associate a vector $x_i \in \mathbb{R}^{|V|}$ such that

$$x_i(v) = \begin{cases} \omega(v) & \text{if } v \in V_i, \\ 0 & \text{otherwise.} \end{cases}$$

Then, $X = (x_1, \dots, x_k) \in M(V)^{k*}$ gives an unbalanced representation of the partition $(V_1, \dots, V_k)$. For example, in our target application of redistricting, the vertices of the graph correspond to geographic units such as census blocks, and ω might represent populations associated with these units, which typically are balanced between voting districts but not identical from one unit to the next.

We address this first at the level of $\mathrm{M}(V)$ with some inspiration from the formulations of unbalanced OT in [9, 26]. Let $x, y \in \mathrm{M}(v)$. For $p \geq 1$ and $\lambda > 0$, the unbalanced problem minimizes (see subsection 6.4 for further discussion when $p = 1$)

$$\begin{array}{ll} \min\limits_{J \in \mathbb{R}^{|E|}, z \in \mathbb{R}^{|V|}} & \|J\|_1 + \lambda \|z\|_p \\ \text{subject to} & P^\top J = y - x + z. \end{array} \tag{6.1}$$

Then, we introduce a distance function on the space $M(V)$ of unbalanced partition components of a graph $G = (V, E, \omega)$ as follows.

Let $X = (x_1, \ldots, x_k), Y = (y_1, \ldots, y_k) \in M(V)^{k*}$ be unbalanced partitions of G. The distance between components x_i and y_j is defined as

$$C_{\lambda,p}(x_i, y_j) = \begin{cases} \min\limits_{J \in \mathbb{R}^{|E|}, z \in \mathbb{R}^{|V|}} & \|J\|_1 + \lambda \|z\|_p \\ \text{subject to} & P^\top J = y_j - x_i + z, \end{cases} \tag{6.2}$$

where $\lambda \geq 0$ and $p \geq 1$ are parameters of the distance function $C_{\lambda,p}$. The variable z allows slack in the mass transported to or from each vertex under the transport plan, and λ and p determine the weight of z relative to J in the objective. In subsection 6.3, we discuss how λ affects $C_{\lambda,p}$. We first show that $C_{\lambda,p}$ defines a valid metric on $M(V)$.

PROPOSITION 6.1. *$C_{\lambda,p}(x, y)$ is a metric on $M(V)$ when $\lambda \geq 0$ and $p \geq 1$.*

Proof. It is immediate that $C_{\lambda,p}$ is nonnegative and symmetric.

Let $x, y \in M(V)$, and suppose $x = y$. Then, $J = 0$ and $z = 0$ are feasible for $C_{\lambda,p}$, so $C_{\lambda,p} = 0$. Conversely, let $x, y \in M(V)$, and suppose $C_{\lambda,p} = 0$. Then, $J = 0$ and $z = 0$, so the constraint $P^\top J = y - x + z$ implies that $x = y$.

Now, suppose $x, y, w \in M(V)$, and let

$$C_{\lambda,p}(x, y) = \|J_{xy}\|_1 + \lambda \|z_{xy}\|_p,$$

$$C_{\lambda,p}(x, w) = \|J_{xw}\|_1 + \lambda \|z_{xw}\|_p,$$

$$C_{\lambda,p}(w, y) = \|J_{wy}\|_1 + \lambda \|z_{wy}\|_p.$$

Then, $P^\top (J_{xw} + J_{wy}) = y - x + z_{xw} + z_{wy}$, so $J_{xw} + J_{wy}$ and $v_{xw} + v_{wy}$ are feasible for $C_{\lambda,p}(x, y)$. Since J_{xy} and v_{xy} are optimizers for $C_{\lambda,p}(x, y)$, we have

$$\begin{aligned} \|J_{xy}\|_1 + \lambda \|v_{xy}\|_p &\leq \|J_{xz} + J_{zy}\|_1 + \lambda \|v_{xz} + v_{zy}\|_p \\ &\leq \|J_{xz}\|_1 + \|J_{zy}\|_1 + \lambda \|v_{xz}\|_p + \lambda \|v_{zy}\|_p. \end{aligned}$$

Hence, $C_{\lambda,p}(x, y) \leq C_{\lambda,p}(x, z) + C_{\lambda,p}(z, y)$, so $C_{\lambda,p}$ satisfies the triangle inequality. □

6.2. Distances between unbalanced partitions. We now extend the lifted distance from subsection 5.2 to a general lifted distance between unbalanced partitions.

DEFINITION 6.2 (unbalanced lifted distance). *The* unbalanced lifted distance $A_{\lambda,p} : M(V)^{k*} \times M(V)^{k*} \to \mathbb{R}$ *between partitions X and Y is defined as*

$$A_{\lambda,p}(X, Y) = \begin{cases} \min\limits_{S \in \mathbb{R}^{k \times k}} & \sum\limits_{ij} S_{ij} C_{\lambda,p}(x_i, y_j) \\ \textit{subject to} & S \in \mathrm{DS}_k, \end{cases} \tag{6.3}$$

where $X = (x_1, \ldots, x_k), Y = (y_1, \ldots, y_k) \in M(V)^{k*}$ *and* $C_{\lambda,p}$ *is the unbalanced distance between partitions in* (6.2).

Like the balanced case, the unbalanced lifted distance uses a linear assignment problem to find a minimum-cost matching of the partition components. The unbalanced lifted distance induces a valid metric on the space of unbalanced graph partitions.

PROPOSITION 6.3. *The lifted distance* $A_{\lambda,p}$ *in* (6.3) *is a metric on* $M(V)^{k*}$.

The proof of Proposition 6.3 follows from the proof of Proposition 5.5.

6.3. Basic properties. In this section, we prove several properties of the unbalanced lifted distance (6.3). First, we show that $C_{\lambda,p}$ and $A_{\lambda,p}$ are monotonic in λ.

PROPOSITION 6.4. *The unbalanced cost function* $C_{\lambda,p}$ *is monotonic in* λ.

Proof. Suppose $\lambda_1, \lambda_2 \geq 0$ and $\lambda_2 > \lambda_1$. Let J_2 and z_2 be optimizers for $C_{\lambda_2,p}(x_i, y_j)$. Then, J_2 and z_2 are feasible for $C_{\lambda_1,p}(x_i, y_j)$, so

$$C_{\lambda_1,p}(x_i, y_j) \leq \sum_e |J_2^e| + \lambda_1 \|z_2\|_p \leq \sum_e |J_2^e| + \lambda_2 \|z_2\|_p \leq C_{\lambda_2,p}(x_i, y_j).$$

Therefore, if $\lambda_2 > \lambda_1$, we have $C_{\lambda_2,p} \geq C_{\lambda_1,p}$. ∎

COROLLARY 6.5. *The unbalanced lifted distance* $A_{\lambda,p}$ *is monotonic in* λ.

Proof. Suppose $\lambda_1, \lambda_2 \geq 0$ and $\lambda_2 > \lambda_1$. Let S_2 be an optimizer for $A_{\lambda_2,p}$. Since S_2 is feasible for $A_{\lambda_1,p}$,

$$A_{\lambda_1,p} \leq \sum_{ij} S_{2ij} C_{\lambda_1,p}(x_i, y_j) \leq \sum_{ij} S_{2ij} C_{\lambda_2,p}(x_i, y_j) = A_{\lambda_2,p},$$

as desired. ∎

In the following proposition, we show that the norm of the optimizer of the mass difference z for $C_{\lambda,p}$ is monotonic in λ.

PROPOSITION 6.6. *The* p*-norm of the optimizer* z_λ *for the unbalanced cost function* $C_{\lambda,p}$ *is monotonic in* λ.

Proof. Suppose $\lambda_1, \lambda_2 \geq 0$ and $\lambda_2 > \lambda_1$. Let J_1 and z_1 be optimizers for $C_{\lambda_1,p}$, and let J_2 and z_2 be optimizers for $C_{\lambda_2,p}$. Since J_2 and z_2 are feasible for $C_{\lambda_1,p}$, $\|J_1\|_1 + \lambda_1\|z_1\|_p \leq \|J_2\|_1 + \lambda_1\|z_2\|_p$, and thus $\|J_1\|_1 - \|J_2\|_1 \leq \lambda_1(\|z_2\|_p - \|z_1\|_p)$. By an identical argument, $\|J_1\|_1 - \|J_2\|_1 \geq \lambda_2(\|z_2\|_p - \|z_1\|_p)$. Combining these expressions, $\lambda_2(\|z_2\|_p - \|z_1\|_p) \leq \lambda_1(\|z_2\|_p - \|z_1\|_p)$. Since $\lambda_2 > \lambda_1$ by assumption, we must have $\|z_2\|_p - \|z_1\|_p \leq 0$, as needed. ∎

In Proposition 5.8, we formulate the balanced lifted distance as a combined linear program. We can derive a similar program for the unbalanced lifted distance.

PROPOSITION 6.7. *The unbalanced lifted distance* $A_{\lambda,p}(X, Y)$ *satisfies*

$$(6.4) \quad A_{\lambda,p}(X,Y) = \begin{cases} \min\limits_{Q \in \mathbb{R}^{|E| \times k^2}, S \in \mathbb{R}^{k \times k}, u \in \mathbb{R}^{|V| \times k^2}} & \sum\limits_{ij} \left(\sum\limits_{e \in E} |Q_{ij}^e| + \lambda \|u_{ij}\|_p \right) \\ \textit{subject to} & S \in \mathrm{DS}_k, \\ & P^\top Q_{ij} - (x_i - y_j) S_{ij} = u_{ij}. \end{cases}$$

Proof. This proof proceeds identically to the proof of Proposition 5.8. Plugging the unbalanced transport cost (6.2) into the unbalanced lifted distance (6.3), we get

$$A_{\lambda,p}(X,Y) = \begin{cases} \min\limits_{S\in\mathbb{R}^{k\times k}} & \sum\limits_{ij} S_{ij}\left(\sum\limits_{e\in E}|J^{*e}_{ij}| + \lambda\|z^*_{ij}\|_p\right) \\ \text{subject to} & S\in \mathrm{DS}_k \\ & J^*_{ij}, z^*_{ij} = \begin{cases} \operatorname*{argmin}\limits_{J_{ij},z_{ij}} & \sum\limits_{e\in E}|J^e_{ij}| + \lambda\|z_{ij}\|_p \\ \text{subject to} & P^\top J_{ij} = x_i - y_j + z_{ij}. \end{cases} \end{cases}$$

Since the inner and outer problems are both minimizations, this simplifies to

$$A_{\lambda,p}(X,Y) = \begin{cases} \min\limits_{S\in\mathbb{R}^{k\times k}, J\in\mathbb{R}^{|E|\times k^2}, z\in\mathbb{R}^{|V|\times k}} & \sum\limits_{ij} S_{ij}\left(\sum\limits_{e\in E}|J^e_{ij}| + \lambda\|z_{ij}\|_p\right) \\ \text{subject to} & S\in \mathrm{DS}_k, \\ & P^\top J_{ij} = x_i - y_j + z_{ij} \end{cases}$$

The variable substitutions $Q^e_{ij} = S_{ij}J^e_{ij}$ and $u_{ij} = S_{ij}z_{ij}$ give the desired result. ∎

The dual of (6.4) is given by

$$(6.5)\qquad A_{\lambda,p}(X,Y) = \begin{cases} \max\limits_{\phi,\psi\in\mathbb{R}^k, \gamma\in\mathbb{R}^{|V|\times k^2}} & \mathbb{1}^\top(\phi+\psi) \\ \text{subject to} & \phi_i + \psi_j \le \gamma_{ij}^\top(x_i - y_j), \\ & |\gamma^w_{ij} - \gamma^v_{ij}| \le 1\ \forall\ (w,v)\in E, \\ & \|\gamma_{ij}\|_q \le \lambda, \end{cases}$$

where the q-norm, satisfying

$$q = \begin{cases} \frac{p}{p-1} & p > 1, \\ \infty & p = 1, \end{cases}$$

is the dual of the p-norm in the primal objective. Slater's condition [38] gives that strong duality holds for (6.4), since we can write the inner problem (6.2) as

$$(6.6)\qquad C_{\lambda,p}(x_i,y_j) = \begin{cases} \min\limits_{J\in\mathbb{R}^{|E|\times k^2}, z\in\mathbb{R}^{|V|}, m\in\mathbb{R}} & \|J\|_1 + \lambda m \\ \text{subject to} & P^\top J = y_j - x_i + z, \\ & m \ge ||z||_p. \end{cases}$$

The final constraint is the only nonlinear one, and for any solution satisfying the linear constraints, we can choose m to be large enough that this solution satisfies the nonlinear constraint with strict inequality. The linear program that computes the optimal matching does not introduce any additional nonlinear constraints, so Slater's condition is satisfied for (6.4), and strong duality holds.

Next, we show that for balanced partitions, the general distance is an extension of the balanced distance. Specifically, if we take $p = 1$, there exists λ sufficiently large such that the unbalanced distance between two balanced partition components is equal to the balanced distance. The following proposition formalizes this notion.

PROPOSITION 6.8. *If X and Y are balanced partitions of a graph G and $\lambda \ge$ diam$(G)/2$, then $C_{\lambda,1}(x_i, y_j) = W_1(x_i, y_j)$ for $x_i \in X$, $y_j \in Y$.*

Proof. Suppose $\bar{J}$ and $\bar{z}$ are optimizers for $C_{\lambda,1}(x_i, y_j)$. We know $\sum_{v\in V} x_i(v) + \sum_{v\in V} \bar{z}(v) = \sum_{v\in V} y_j(v)$, so when X and Y are balanced, $\sum_{v\in V} \bar{z}(v) = 0$. Let $\bar{z}^+$ and $\bar{z}^-$ be defined such that

$$\bar{z}^+(v) = \begin{cases} z(v) & z(v) > 0, \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad \bar{z}^-(v) = \begin{cases} |z(v)| & z(v) < 0, \\ 0 & \text{otherwise.} \end{cases}$$

Let J^* represent the OT plan between $\bar{z}^-$ and $\bar{z}^+$. Then,

$$\|J^*\|_1 \leq \text{diam}(G) \sum_{v\in V} |\bar{z}^-(v)| = \frac{\text{diam}(G)}{2} \sum_{v\in V} |\bar{z}(v)|.$$

Therefore,

$$\|\bar{J}\|_1 + \|J^*\|_1 \leq \|\bar{J}\|_1 + \frac{\text{diam}(G)}{2}\|\bar{z}\|_1 \leq C_{\lambda,1}(x_i, y_j). \tag{6.7}$$

Since $P^T(\bar{J} + J^*) = (x_i - y_j + \bar{z}) + (\bar{z}^- - \bar{z}^+) = x_i - y_j$, we know $\bar{J} + J^*$ is feasible for $W_1(x_i, y_j)$, and

$$W_1(x_i, y_j) \leq \|\bar{J} + J^*\|_1 \leq \|\bar{J}\|_1 + \|J^*\|_1. \tag{6.8}$$

Combining (6.7) and (6.8), we get $W_1(x_i, y_j) \leq C_{\lambda,1}(x_i, y_j)$. In the balanced case, $C_{\lambda,1}(x_i, y_j) \leq W_1(x_i, y_j)$, so $W_1(x_i, y_j) = C_{\lambda,1}(x_i, y_j)$, as desired. □

We can extend the previous result to show that for $\lambda \geq \text{diam}(G)/2$, the unbalanced lifted distance is equal to the balanced lifted distance for balanced partitions.

COROLLARY 6.9. *If X and Y are balanced partitions of a graph G and $\lambda \geq$ diam$(G)/2$, then $A_{\lambda,1}(X, Y) = A(X, Y)$.*

Proof. When $W_1(x_i, y_j) = C_{\lambda,1}(x_i, y_j)$ for all $x_i \in X$, $y_j \in Y$, the definition of lifted distance in (5.3) is the same as the unbalanced lifted distance in (6.3). □

6.4. The unbalanced transport problem when $p = 1$. When $p = 1$, the unbalanced problem (6.1) is equivalent to a discrete version of the transportation problem with boundary studied by Figalli and Gigli [18], which is to (6.1) what the Kantorovich problem in (4.1) is to (4.2). The idea is to modify the Kantorovich problem by adding vertices that serve as infinite-capacity sinks/reservoirs that can receive or provide mass to compensate for unequal total masses between $x, y \in \text{M}(V)$.

In our case, we expand the graph $G = (V, E, \omega)$ by adding one extra auxiliary vertex v_s. We also add edges between every vertex and v_s, all with the the same weight $\lambda > 0$. Concretely, we define $G_* = (V_*, E_*, \omega_*)$ as follows:

$$V_* := V \cup \{v_s\}, E_* := E \cup V \times \{v_s\} \cup \{v_s\} \times V, \text{ and}$$
$$\omega_*(e) := \begin{cases} \omega(e) & \text{if } e \in E, \\ \lambda & \text{if } e \in E_* \setminus E. \end{cases}$$

We denote by $d_\lambda(v, w)$ the resulting graph distance in V^*; this simply extends the distance in G via $d(v, v_s) = \lambda$ for every $v \in V$.

Given $x, y \in \text{M}(V)$, we will say that $\pi \in \text{M}(V \times V)$ is an admissible transport plan for x and y with sink at v_s if

$$\sum_{w\in V_*} \pi(v, w) = x(v) \ \ \forall\, v \in V \ \text{ and } \sum_{v\in V_*} \pi(v, w) = y(w) \ \ \forall\, w \in V.$$

This condition is similar to the usual Kantorovich problem from (4.1), except in our larger space we do not impose the marginal constraint at v_s; this means we are free to move any amount of mass to, from, or through v_s. The set of such admissible plans will be denoted $\Pi_*(x, y)$. Then, the analogue of the Kantorovich problem is

$$\min_{\pi \in \Pi_*(x,y)} \sum_{v,w \in V_*} d_\lambda(v, w)\pi(v, w). \tag{6.9}$$

We will show that problem (6.9) is equivalent to problem (6.1) when $p = 1$.

LEMMA 6.10. *The minimum for* (6.9) *is the same as the minimum for the problem*

$$\begin{array}{rl} \min_{J \in \mathbb{R}^{|E|}} & \|J\|_1 + \lambda \|z\|_1 \\ \textit{subject to} & P^\top J = y - x + z. \end{array} \tag{6.10}$$

Moreover, from any π *that minimizes* (6.9) *we can construct* (J, z) *that minimizes* (6.10).

To prove Lemma 6.10, let us make some preliminary observations. The essence of the proof lies in the following construction, which is commonly used to prove the equivalence between (4.1) and (4.2) (see [36, section 4.2]). For every $v, w \in V_*$, choose a minimal path from v to w, and denote by $E(v, w) \subset E_*$ the set of edges in this path. That is, if the minimal path between v and w is $v = v_0, \ldots, v_N = w$, then $E_*(v, w) = \{(v_0, v_1), (v_1, v_2), \ldots, (v_{N-1}, v_N)\}$. Then, given any $\pi \in M(V_* \times V_*)$ we define $J_\pi : E_* \to \mathbb{R}$ and $z_\pi : V_* \to \mathbb{R}$ as follows:

$$J_{\pi,e} := \sum_{v \in V_*} \sum_{w \in V_*} \mathbb{1}_{E(v,w)}(e)\pi(v, w), \quad z_\pi(v) := \pi(v, v_s) - \pi(v_s, v). \tag{6.11}$$

The proof of Lemma 6.10 boils down to showing that if π minimizes (6.9), then (J_π, z_π) given by (6.11) minimizes (6.10). We start by showing (J_π, z_π) is admissible.

PROPOSITION 6.11. *Let* $x, y \in M(V)$. *If* $\pi \in \Pi_*(x, y)$, *then* $(P^\top J_\pi)(v) = y(v) - x(v) + z_\pi(v)$ *for* $v \in V$. *Moreover, we have, with* P_* *denoting the incidence matrix for the graph* G_*, $(P_*^\top J_\pi)(v) = y(v) - x(v)$ *for* $v \in V$.

Proof. Let $E_{\text{in}}(v)$ and $E_{\text{out}(v)}$ denote the sets of the incoming and outgoing edges of vertex v, respectively; then

$$(P^\top J_\pi)(v) = \sum_{e \in E_{\text{in}(v)}} J_{\pi,e} - \sum_{e \in E_{\text{out}(v)}} J_{\pi,e}.$$

Fix $v_0, w_0 \in V$, and let π_0 be the function

$$\pi_0(v, w) = \begin{cases} 1 & \text{if } (v, w) = (v_0, w_0), \\ 0 & \text{otherwise.} \end{cases}$$

Then, we have

$$\sum_{e \in E_{\text{out}(v_0)}} J_{\pi,e} - \sum_{e \in E_{\text{in}(v_0)}} J_{\pi,e} = 1,$$

$$\sum_{e \in E_{\text{out}(w_0)}} J_{\pi,e} - \sum_{e \in E_{\text{in}(w_0)}} J_{\pi,e} = -1, \text{ and}$$

$$\sum_{e \in E_{\text{out}(v)}} J_{\pi,e} - \sum_{e \in E_{\text{in}(v)}} J_{\pi,e} = 0 \text{ if } v \neq v_0, w_0.$$

From a linear combination of these identities for each pair $(v_0, w_0) \in V \times V$ we obtain the following formula for any $\pi \in M(V \times V)$:

$$\sum_{e \in E_{\text{out}(v)}} J_{\pi,e} - \sum_{e \in E_{\text{in}(v)}} J_{\pi,e} = \sum_{w \in V_*} \pi(w,v) - \sum_{v \in V_*} \pi(v,w).$$

Now, if π is an admissible plan, we have

$$\begin{aligned}\sum_{e \in E_{\text{out}(v)}} J_{\pi,e} - \sum_{e \in E_{\text{in}(v)}} J_{\pi,e} &= \sum_{w \in V} \pi(w,v) - \sum_{w \in V} \pi(v,w) + \pi(v_s, v) - \pi(v, v_s) \\ &= x(v) - y(v) - z_\pi(v).\end{aligned}$$

It follows that $(P^\top J_\pi)_v = y(v) - x(v) + z_\pi(v)$, proving the first identity. For the second, observe

$$\begin{aligned}(P_*^\top J_\pi)_v = \sum_{e \in E_*} P_{ev} J_{\pi,v} &= \sum_{e \in E} P_{ev} J_{\pi,v} + \sum_{w \in V} P_{(w,v_s)v} J_{\pi,w} + \sum_{w \in V} P_{(v_s,w)v} J_{\pi,w} \\ &= (P^\top J_\pi)(v) - J_{\pi,(v,v_s)} + J_{\pi,(v_s,v)}.\end{aligned}$$

Using that $J_{\pi,(v,v_s)} = \pi(v, v_s)$ and $J_{\pi,(v_s,v)} = \pi(v_s, v)$ together with the formula for $(P^\top J_\pi)_v$, we obtain

$$(P_*^\top J_\pi)_v = y(v) - x(v) + z_\pi(v) - \pi(v, v_s) + \pi(v_s, v) = y(v) - x(v),$$

and the second formula is proved. ◻

With this, we are ready to prove the equivalence between the two problems.

Proof of Lemma 6.10. Let π be a minimizer for problem (6.9). According to Proposition 6.11, (J_π, z_π) is an admissible pair for problem (6.10). Therefore,

$$\|J_\pi\|_1 + \lambda\|z_\pi\|_1 \geq \begin{cases} \min_{J \in \mathbb{R}^{|E|}} & \|J\|_1 + \lambda\|z\|_1 \\ \text{subject to} & P^\top J = y - x + z. \end{cases}$$

We have $J_{\pi,e} \geq 0$ for every e, and hence

$$\begin{aligned}\sum_{e \in E_*} |J_{\pi,e}|\omega(e) = \sum_{e \in E_*} J_{\pi,e}\omega(e) &= \sum_{e \in E_*} \sum_{v \in V_*} \sum_{w \in V_*} \mathbb{1}_{E(v,w)}(e)\omega(e)\pi(v,w) \\ &= \sum_{v \in V_*} \sum_{w \in V_*} \left(\sum_{e \in E} \mathbb{1}_{E(v,w)}(e)\omega(e) \right) \pi(v,w).\end{aligned}$$

From the definition of the sets $E(v, w)$, for any $v, w \in V_\infty$ we have

$$\sum_{e \in E_*} \mathbb{1}_{E(v,w)}(e)\omega(e) = d_\lambda(v,w).$$

Therefore, $\sum_{e \in E_*} |J_{\pi,e}|\omega(e) = \sum_{v \in V_*} \sum_{w \in V_*} d_\lambda(v,w)\pi(v,w)$. On the other hand, the sum on the left can be decomposed as

$$\sum_{e \in E_*} J_{\pi,e}\omega(e) = \sum_{e \in E} J_{\pi,e}\omega(e) + \sum_{v \in V} J_{\pi,(v,v_s)}\omega(v,v_s) + \sum_{v \in V} J_{\pi,(v_s,v)}\omega(v_s,v).$$

Since $\omega(v_s, v) = \omega(v, v_s) = \lambda$ for every $v \in V$,

$$\sum_{e\in E_*} J_{\pi,e}\omega(e) = \sum_{e\in E} J_{\pi,e}\omega(e) + \lambda \sum_{v\in V} J_{\pi,(v,v_s)} + J_{\pi,(v_s,v)}.$$

From the definition of $J_{\pi,e}$, $J_{\pi,(v_s,v)} = \pi(v_s, v)$ and $J_{\pi,(v,v_s)} = \pi(v, v_s)$ for every $v \in V$. The minimizer π can always be modified so that for every v at most one of $\pi(v, v_s)$ and $\pi(v_s, v)$ is nonzero. In this case $|z_\pi(v)| = \pi(v, v_s) + \pi(v_s, v)$ for every $v \in V$, and

$$\sum_{e\in E_*} J_{\pi,e}\omega(e) = \sum_{e\in E} J_{\pi,e}\omega(e) + \lambda \sum_{v\in V} |z_\pi(v)|.$$

This shows that

$$\sum_{v\in V_*}\sum_{w\in V_*} d_\lambda(v, w)\pi(v, w) = \|J_\pi\|_1 + \lambda\|z_\pi\|_1,$$

which shows the minimum for (6.9) is no smaller than the minimum for (6.10).

For the reverse inequality we will implicitly use the dual problem to (6.9). Consider pairs of functions $\phi, \psi : V_* \to \mathbb{R}$ such that $\phi(v_s) = \psi(v_s) = 0$ and for every $v, w \in V_*$

$$\phi(v) + \psi(w) \le d_\lambda(v, w). \tag{6.12}$$

Following [20, Appendix A], the dual problem to (6.9) is maximizing the functional

$$\sum_{v\in V} \phi(v)x(v) + \sum_{w\in V} \psi(w)y(w)$$

over all pairs ϕ, ψ described above. Let us show that the minimum of (6.10) is larger than this for any ϕ, ψ. Without loss of generality, we may assume that ϕ is such that

$$\phi(v) = \min_{w\in V^*} d_\lambda(v, w) - \psi(w).$$

In this case it is easy to see that $|\phi(v) - \phi(v)| \le d_\lambda(v, w)$ for every v and w. In particular, if $e = (v, w)$ is an edge, we have $|\phi(w) - \phi(v)| \le d_\lambda(v, w) = \omega(e)$. Since $(P_*\phi)_e = \phi(w) - \phi(v)$, this shows $|(P_*\phi)_e| \le \omega(e)$ for every $e \in E_*$. Combining these inequalities for each $e \in E_*$ and using the dual of P_*, we have

$$\sum_{e\in E_*} \omega(e)|J_e| \ge -\sum_{e\in E_*} J_e(P_*\phi)_e = \sum_{v\in V_*} (P_*^\top J)_v\phi(v).$$

Since $\phi(v_s) = 0$ and $(P_*^\top J)_v = y(v) - x(v)$ when $v \ne v_s$, it follows that

$$\sum_{e\in E_*} \omega(e)|J_e| \ge \sum_{v\in V} \phi(v)(x(v) - y(v)) = \sum_{v\in V} \phi(v)x(v) - \sum_{v\in V} \phi(v)y(v).$$

On the other hand, applying (6.12) with $v = w$ yields the inequality $\psi(v) \le -\phi(v)$ for every v, from where it follows that

$$\sum_{v\in V} \phi(v)x(v) - \sum_{v\in V} \phi(v)y(v) \ge \sum_{v\in V} \phi(v)x(v) + \sum_{v\in V} \psi(v)y(v).$$

In conclusion, for every admissible pair ϕ and ψ we have the inequality

$$\sum_{e\in E_*} |J_e|\omega(e) \ge \sum_{v\in V} \phi(v)x(v) + \sum_{v\in V} \psi(v)y(v).$$

Taking the supremum over all admissible ϕ and ψ we have, by duality,

$$\sum_{e\in E_*} |J_e|\omega(e) \ge \inf_{\pi\in\Pi_*} \sum_{v\in V_*}\sum_{w\in V_*} d_\lambda(v, w)\pi(v, w),$$

and this finishes the proof. ∎

6.5. Overview of the distance. Informally, the lifted distance between partitions $X = (x_1, \ldots, x_k)$ and $Y = (y_1, \ldots, y_k)$ quantifies the notion of how difficult it is to change partition X into partition Y, working componentwise. Suppose, for example, that $x_1 = y_1$. Intuitively, then, when changing X to Y, the component x_1 should remain the same; this idea is formalized in Proposition 5.7. Similarly, when changing x_2 into a component of Y, we would prefer to turn x_2 into a component y_i that is almost equal to x_2 instead of a component y_j that is very different from x_2. The linear assignment performs this calculation over all the components at once, matching each x_i to a component y_j such that the total difference across all matched pairs is as small as possible.

In the unbalanced case, not all of the components have the same weight. Therefore, it may not be possible to change component x_i into any component y_j exactly. For this reason, the unbalanced component distance between x_i and y_j allows x_i to fail to be matched exactly to y_j, and the cost incorporates the extent of this failure; this is represented by the variable z in the unbalanced component distance (6.2). When the penalty for a component x_i failing to exactly become y_j is high, we prefer to match x_i to a component y_j with similar total weight, to minimize the failure. If the penalty for x_i failing to exactly become y_j is sufficiently high, the unbalanced distance will ensure that x_i perfectly becomes y_j when X and Y are balanced; this is proven in Proposition 6.8.

7. Bounds. We can relate our distance on partitions to other constructions in the literature: the *Hamming distance* and the *total variation distance*.

The Hamming distance between two binary strings of equal length is the number of positions in which they differ. Inspired by this definition, [22] computes a notion of Hamming distance between two graph partitions X and Y. Using $v \in x_i$ to indicate that vertex v belongs to component i of partition X, the Hamming distance is

$$(7.1) \qquad \mathrm{dist}_{\mathrm{HAM}}(X, Y) = \begin{cases} \min\limits_{S \in \mathbb{R}^{k \times k}} & \sum\limits_{v \in V} \sum\limits_{ij} S_{ij} \mathbb{1}\left[v \in x_i \wedge v \notin y_j\right] \\ \text{subject to} & S \in \mathrm{DS}_k. \end{cases}$$

That is, for a given matching of the components, we count the number of vertices whose label differs and take the minimum over all matchings. Generalizing to nonbinary functions for the weights on vertices, we can formulate a distance as the sum of the vertexwise differences in weights over the matched components. First, we can write the L_1 or *total variation* distance between two components x_i and y_j as

$$(7.2) \qquad \ell_1(x_i, y_j) = \frac{1}{2} \sum_{v \in V} |x_i(v) - y_j(v)|,$$

and we can lift this to a distance between partitions by solving the assignment problem using ℓ_1 as the cost function. We write

$$(7.3) \qquad L_1(X, Y) = \begin{cases} \min\limits_{S \in \mathbb{R}^{k \times k}} & \sum\limits_{v \in V} \sum\limits_{ij} S_{ij} \ell_1(x_i, y_j) \\ \text{subject to} & S \in \mathrm{DS}_k. \end{cases}$$

These distances are of the same form as (5.3) but with different cost functions C.

We can show that, for any choice of weight on the vertices, the L_1 distance lower-bounds the transport distance between partitions.

PROPOSITION 7.1. *For balanced X and Y, we have $L_1(X, Y) \le A(X, Y)$.*

Proof. Consider a pair of components x_i and y_j. From (4.1), we know that

$$W_1(x_i, y_j) = \min_{\pi \in \Pi(x_i, y_j)} \sum_{v,w \in V} d(v, w)\pi(v, w).$$

Because $d(v, w)$ is the shortest path distance between v and w, $d(v, w) = 0$ if and only if $v = w$, and $d(v, w) \geq 1$ otherwise. Therefore, for any $\pi \in \Pi(x_i, y_j)$,

$$\sum_{v,w \in V} d(v, w)\pi(v, w) = \sum_{v \neq w} d(v, w)\pi(v, w) \geq \sum_{v \neq w} \pi(v, w).$$

Let π^* be the element of $\Pi(x_i, y_j)$ that minimizes $W_1(x_i, y_j)$. Then,

$$\sum_{v \neq w} \pi^*(v, w) = \sum_{v,w \in V} \pi^*(v, w) - \sum_{v \in V} \pi^*(v, v) = \sum_{v \in V} (x_i(v) - \pi^*(v, v)).$$

Because π^* moves as little mass as possible, $\pi^*(v, v) = \min(x_i(v), y_j(v))$. Then,

$$x_i(v) - \pi^*(v, v) = \begin{cases} 0 & \text{if } x_i(v) \leq y_j(v), \\ x_i(v) - y_j(v) & \text{otherwise,} \end{cases}$$

and hence $\sum_{v \neq w} \pi^*(v, w) = \sum_{v \in V, x_i(v) \geq y_j(v)} x_i(v) - y_j(v) = \frac{1}{2} \sum_{v \in V} |x_i(v) - y_j(v)|$. Therefore, we have shown that

$$\min_{\pi \in \Pi(x_i, y_j)} \sum_{v,w \in V} d(v, w)\pi(v, w) \geq \frac{1}{2} \sum_{v \in V} |x_i(v) - y_j(v)|,$$

so $W_1(x_i, y_j) \geq \ell_1(x_i, y_j)$.

Now, let S denote the optimal matching for $A(X, Y)$ given in (5.3). Then, $A(X, Y) = \sum_{ij} S_{ij} W_1(x_i, y_j) \geq \sum_{ij} S_{ij} \ell_1(x_i, y_j) \geq L_1(X, Y)$, as desired. ◻

Qualitatively, we expect some differences between the L_1 and transport distances. L_1 does not see the graph structure, since it is computed from only overlap. For this reason, qualitatively similar components with little overlap are as far apart in L_1 as components on opposite sides of the graph, whereas transport recognizes that the former are closer than the latter. This can happen if we take one partition with thin components (i.e., nearly every vertex is on the boundary of a component) and construct a new partition by slightly perturbing the first. We expect little overlap between these partitions, but this small perturbation can be corrected by moving mass a short distance, which gives rise to high L_1 distance but low transport distance. We give a concrete example in subsection 8.3, where two partitions of the grid with "snakey" components are far apart in L_1 but not in transport distance because the first partition is a small perturbation of the second.

8. Experiments and empirical evaluation. In this section, we provide a suite of empirical applications of our metric using both synthetic data from grid partitions and real geographic data including election results.

8.1. Implementation and experimental setup. We run our experiments using Python on consumer-grade hardware; our code is available on GitHub.[1] We use CVXPY [16] and `scikit-learn` [33] for optimization and embedding. The grid

[1] https://github.com/vrdi/geometry-of-graph-partitions.

partitions in subsection 8.2 and Markov chain sampling in subsections 8.4 to 8.6 rely on the `enumerator` [37] and `GerryChain` [29] packages available on GitHub. Spatial and electoral data comes from the `mggg-states` repository [30], the National Historical Geographic Information System database [27], and the data accompanying [22] from the Quantifying Gerrymandering group [21].

Many of our examples embed multiple partitions of a fixed graph onto the plane to visualize an ensemble, using the method outlined below. We store pairwise distances between the n partitions in a matrix D, where D_{ij} is the distance between partitions i and j. We then would like to find a set of points $P_1, \ldots, P_n \in \mathbb{R}^2$ such that $\|P_i - P_j\|_2 \approx D_{ij}$. Doing so with zero distortion may be impossible or may require using a high-dimensional ambient space. To resolve this issue, we use *multidimensional scaling* (MDS), which computes $P_1, \ldots, P_n$ in a way that (approximately) minimizes $(D_{ij} - \text{dist}(P_i, P_j))^2$. For a modern treatment of MDS, see [4].

8.2. Grid partitions. We begin by examining distances between partitions of a *grid graph*. While the number of feasible districting plans for a US state is unfathomably large, for small grids the number of partitions is more manageable; for example, there are only 117 ways to partition a 4×4 grid into four connected components with four vertices in each component. We can therefore compute the transport distance between all or a large portion of the possible partitions and embed them in the plane.

In Figure 1, we show the MDS embedding of the pairwise transport distances between the ten partitions of a 3×3 grid into three connected components of size three (i.e., the *triomino tilings*). This visualization reveals several features of the metric space. The two partitions into horizontal and vertical "stripes" are the furthest apart, and the remaining eight partitions cluster in pairs based on which straight triomino is included in the partition. For example, the two partitions near the middle-top both have a straight triomino along the top row and two "L"-triominos covering the lower two rows. Furthermore, the partitions that include a horizontal straight triomino fall in the top-left half, and the ones with a vertical straight triomino fall in the lower-right half. The two top-right partitions share an "L"-triomino, as does the lower-left pair.

Similar phenomena can be observed in the embedding of the 117 partitions of the 4×4 grid into four connected components of size four (i.e., the *tetromino tilings*) in Figure 2. Again, the two 'striped' partitions are the furthest apart, and there are visible clusters of partitions with similar compositions. Additionally, there are two

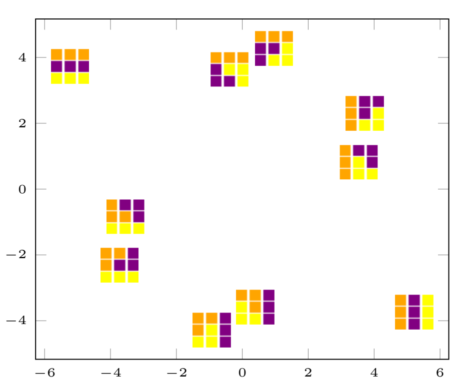


FIG. 1. *The ten partitions of the* 3×3 *grid into three equal-sized components.*

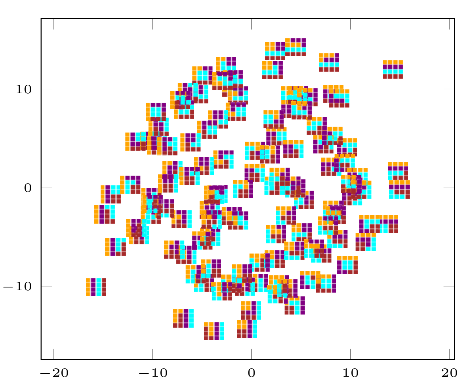


FIG. 2. *The* 117 *partitions of the* 4×4 *grid into four equal-sized components.*

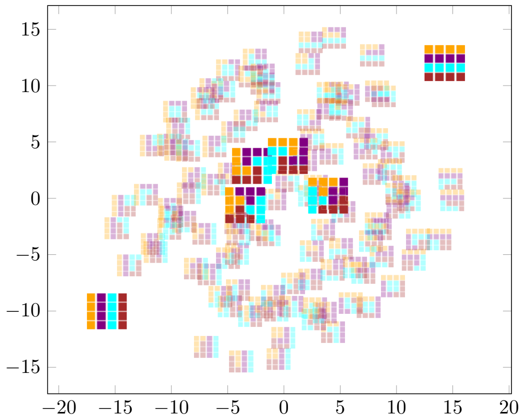


FIG. 3. *The four central and two peripheral partitions of the* 4×4 *grid.*

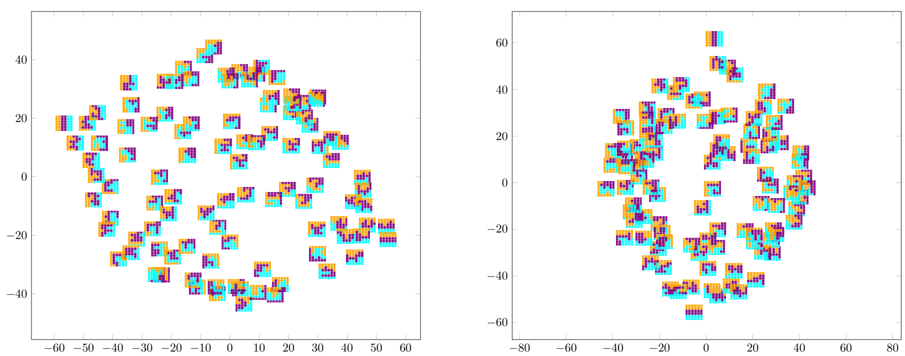


FIG. 4. *Two samples of* 102 *partitions of the* 6×6 *grid into three components.*

partitions each composed of four "T"-tetrominos and two partitions composed of four "L"-tetrominos which are at the center of this metric space, and these appear in the middle of the image. Figure 3 highlights this structure.

To illustrate the balanced case, in Figure 4 we sample from partitions of the 6×6 grid into three equal-sized components. There are 264,500 such partitions, and hence computing pairwise distances is not practical or informative. Rather, we examine a random sample of 100 partitions, plus the two "striped" partitions. The familiar structure emerges: the "striped" partitions are the furthest apart, and visually similar partitions appear near one another in the embedding.

To examine the *unbalanced* problem, we look at partitions of the 3×3 grid into three components of size between one and five, inclusive. There are 170 such partitions, including those in Figure 1. We use (6.4) with $p = 1$. We vary λ to illustrate properties of the unbalanced cost function for different parameter regimes:

- First, we take $\lambda = 0.5$, so the cost of sending flow through an edge is the same as leaving it as unbalanced mass in z. This formulation is similar to the Hamming and total variation distances (section 7), where the cost to match component x_i in the first partition to component y_j in the second is the number of vertices in the set difference $x_i \setminus y_j$. Figure 5 confirms this observation: Nearby partitions tend to have a component in common and the other two components are similar to each other, often differing in the assignment of only one vertex.

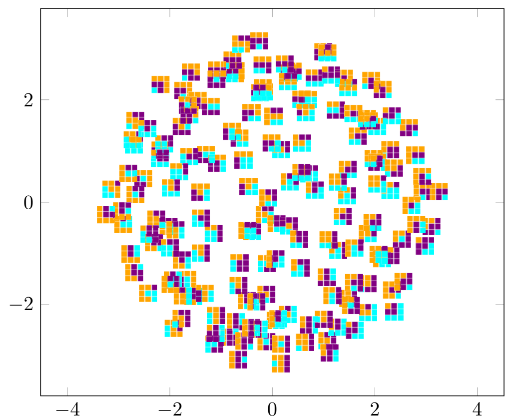


FIG. 5. *The partitions of the* 3×3 *grid into three components of size* 3 ± 2 *with a penalty of* $\lambda = .5$.

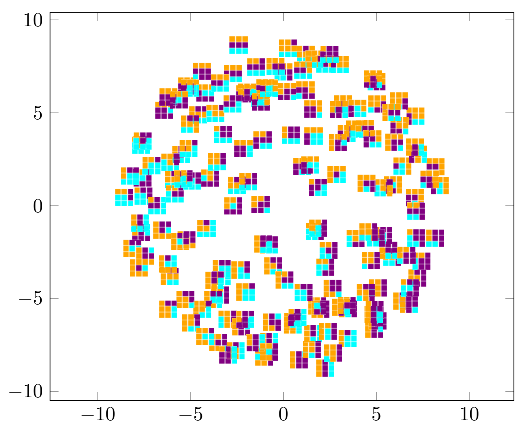


FIG. 6. *The partitions of the* 3×3 *grid into three components of size* 3 ± 2 *with a penalty of* $\lambda = 2$.

- Next, we consider $\lambda = 2$, shown in Figure 6, where the cost of leaving one unit in z equals the graph diameter. As in Proposition 6.8, for any pairing of the partition components, it will be suboptimal to move mass through z unless absolutely necessary, in the case that two components of differing mass are matched.
- The $\lambda = 2$ case contrasts significantly with $\lambda = 5$, shown in Figure 7. Here, the cost of having mass in z is so high that the distance between any pair of unbalanced partitions should be larger than the distance between any pair of balanced partitions. In the embedding, we see the partitions for which there exists a matching of components of equal mass cluster together. Furthermore, we see the familiar structure emerge in each cluster separately. For example, the ten partitions consisting of three components of size three appear together at the top of the Figure 7 in an arrangement similar to the one in Figure 1.

8.3. Hamming distance vs. transport distance on a grid. Because the Hamming distance depends only on the amount of overlap between two districts in a partition, it is insensitive to the distances between vertices in the graph. To illustrate

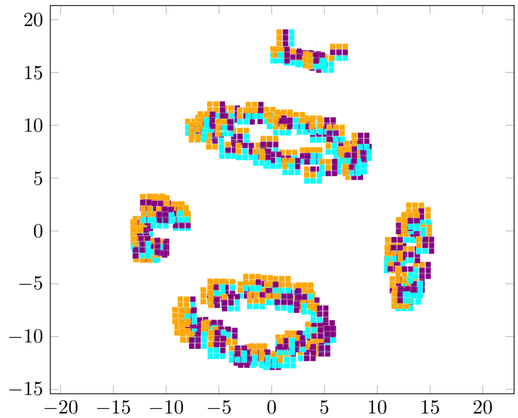
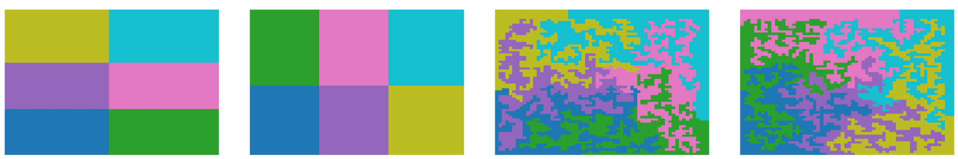


FIG. 7. *The partitions of the* 3×3 *grid into three components of size* 3 ± 2 *with a penalty of* $\lambda = 5$.



(a) Chain 1 start (b) Chain 2 start (c) Chain 1 end (d) Chain 2 end

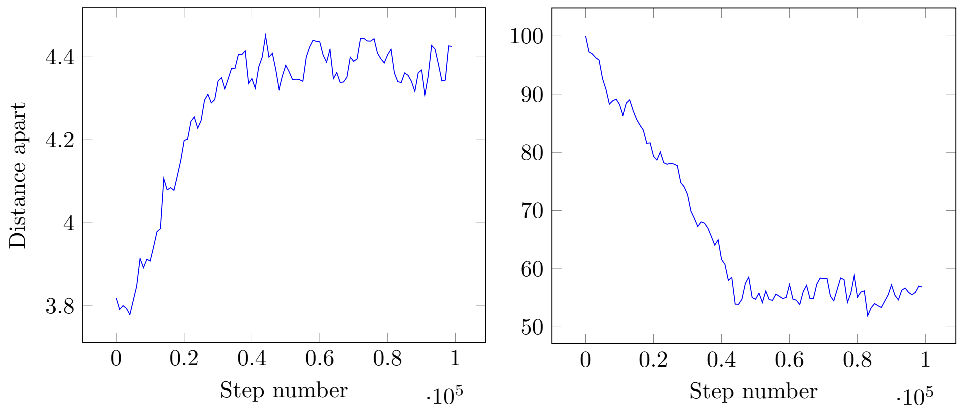
FIG. 8. *Two Markov chains on a* 60*-by-*60 *grid.*

the difference between the Hamming and transport distances, we run two Markov chains of partitions of a 60×60 grid into six components.

These Markov chains, at each time step, relabel a random vertex in the graph. If the resultant partition consists of connected components that all have nearly the same number of vertices, this partition is accepted and we propose another vertex to relabel. Otherwise, the step is rejected and we retry. As shown in [32], this procedure results in partitions with geometrically irregular and highly noncompact components; as an example, the start and end positions of each chain are shown in Figure 8.

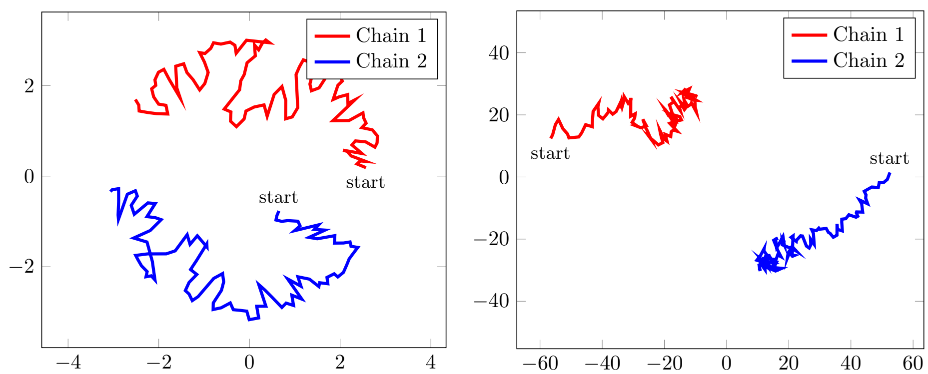
We let both chains run and compute the distance between them at every 1,000th step, using Hamming and transport distance; Figure 9 shows the result. Over time, Hamming distance increases while transport distance decreases. The plots in Figure 10 show that in this situation, transport and Hamming distinguish qualitatively different properties. The Hamming distance highlights that the initial partitions many overlapping vertices while the final partitions do not. Contrastingly, the transport distance detects that transforming the initial partition into the other requires moving mass a large distance through the graph, while less work is needed to match the final plans since their boundaries are interleaved. Put differently, partitions whose boundaries are long and intertwined are similar under the transport metric because mass does not have to be displaced a long distance to convert one into the other, while Hamming distances simply count overlapping vertices.

8.4. Simulated annealing on Arkansas. A more sophisticated technique for generating partitions meeting a specific criterion uses simulated annealing, where a weighting function is tuned over time to first allow rapid exploration of the space of possibilities and later to settle into a local optimum.



(a) Hamming distance (b) Transport distance

FIG. 9. *Distance between Chain* 1 *and Chain* 2.

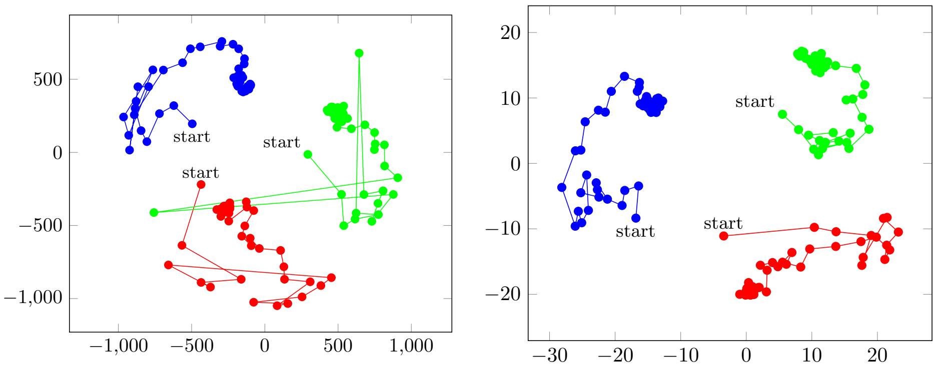


(a) Hamming distance (b) Transport distance

FIG. 10. *MDS embeddings of Chain* 1 *and Chain* 2.

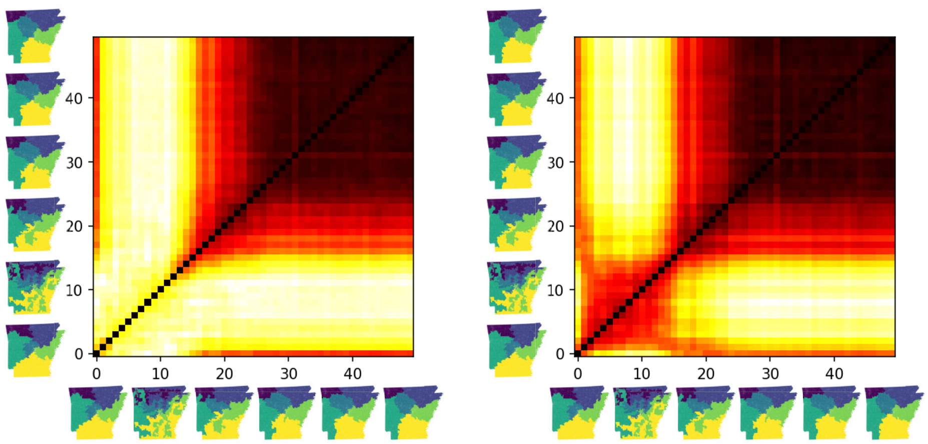
To show how our distance can be used to analyze such a process, we run a Markov chain to partition Arkansas into four congressional districts, focusing on producing compact plans. We take 500,000 steps along the Markov chain using the random relabelling proposal as in subsection 8.3. To perform annealing, the first 100,000 are taken without weighting, and for steps 100,000 to 500,000 we accept a proposed step with probability proportional to $\exp(\beta|\partial P|)$, where ∂P denotes the number of edges in the graph which join two vertices in different components, a discretization of *boundary length*. The parameter β ranges linearly from $\beta = 0$ (no weighting) to $\beta = 3$ from step 100,000 to step 400,000 and is fixed at 3 for the final 100,000 steps.

We show the Hamming distance and transport distance embeddings of every 10,000th plan in three different annealing chains in Figure 11. We also show the pairwise distances for the red chain in Figure 11 as a heatmap in Figure 12. Note the difference between the Hamming and transport distance matrices in the early stages of the chain, which results in a qualitative difference in the embeddings. In particular, the transport distance shows a clear clustering of plans before and after plan 15 (i.e., the 150,000th step in the chain, early on in the cooling phase). This transition



(a) Hamming distance (b) Transport distance

FIG. 11. *Three Markov chain-generated walks in the space of partitions with simulated annealing.*



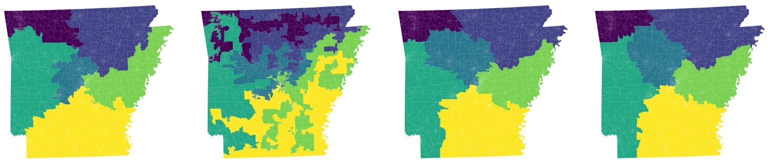
(a) Hamming distance (b) Transport distance

FIG. 12. *Pairwise distance matrices for every* 10,000*th step in one simulated annealing walk (the red chain in Figure* 11*).*

could be related to the phase transition in sampling from a distribution proportional to an exponent of total perimeter, observed in [32]. Both embeddings in Figure 11 show that when the chain is unrestrained by the Metropolis weighting, it moves more quickly through the state space. Also, the chain initially moves away from the starting plan before being brought closer to the initial plan once Metropolis weighting is introduced. This observation can be confirmed in the snapshots in Figure 13.

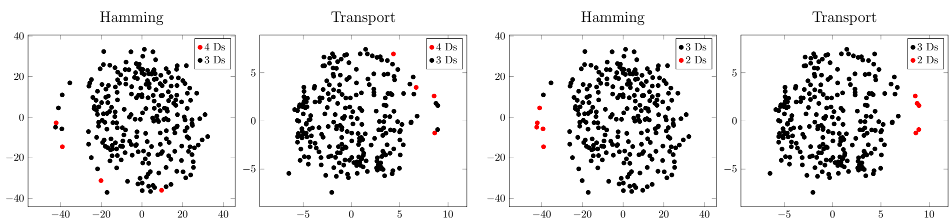
In terms of computational cost, the 11,175 pairwise distances needed for the embedding in Figure 11 took 7 seconds to compute for Hamming distance, and about 15 hours to compute with transport distance. This illustrates the substantial computational cost of the transport distance, although this particular application (along with the others in this section) lends itself naturally to parallelization since the pairwise distances can all be computed independently.

8.5. Partisan clustering. Given a large ensemble of districting plans, we can investigate the geographic features of plans with extreme partisan statistics. To



Initial 100,000 steps 400,000 steps Final

Fig. 13. *Snapshots of a simulated annealing Markov chain on Arkansas districts (the red chain in Figure* 11*).*



(a) 2008 Presidential vote data (b) 2012 Presidential vote data

Fig. 14. *Iowa ensemble colored by Democratic seats won.*

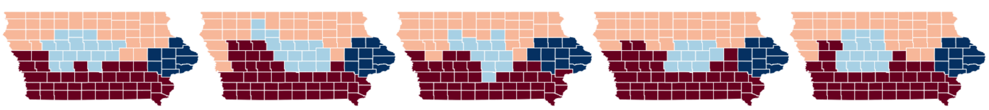
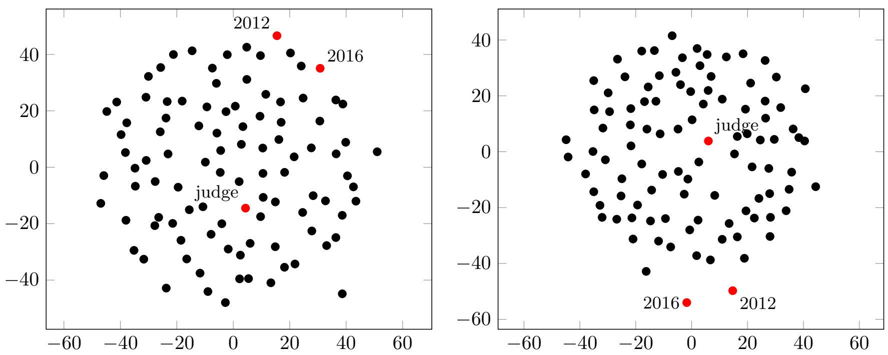


Fig. 15. *Partitions of Iowa with two Democratic seats under* 2012 *Presidential election results. The two blue districts have a Democratic majority, and the two red districts have a Republican majority.*

illustrate how our distance can be used for such an analysis, we consider an ensemble of congressional districting plans for Iowa, generated by a "recombination" Markov chain [14]; rather than choosing a single random vertex to relabel, we instead randomly choose two components of the partition to merge and resplit into two new components.

Figure 14 shows embeddings (using both tranport and Hamming distance) of these plans, colored by the number of seats won under two historical elections: the 2008 Presidential election, in which the Democratic candidate Barack Obama won approximately 55 percent of the two-way vote share against Republican John McCain, and the 2012 Presidential election, in which Barack Obama won approximately 53 percent of the two-way vote share against Republican Mitt Romney. The Hamming and transport embeddings tell the same story, namely, that for each election, most plans produce three Democratic seats but a small number do not, and these cluster near one another, with a stronger pattern for the 2012 voting data. Figure 15 shows the maps of each of the plans which, under the 2012 data, had two majority Democratic districts, which shows the geographic similarity which resulted in the clustering in the embedding.

8.6. Outlier analysis. In [22], the authors demonstrate that the congressional districting plans enacted in 2012 and in 2016 were atypical outliers in terms of their partisan statistics (computed using fixed vote data from multiple elections), relative to a computer-generated ensemble. A third plan by a bipartisan panel of judges was



(a) Ensemble from [22]

(b) ReCom ensemble

FIG. 16. *Transport embeddings of human/computer-drawn plans for North Carolina.*

found to be far more representative. Figure 16(a) shows an MDS embedding of one hundred plans drawn from the the authors' ensemble as well as the three human-drawn plans mentioned above, using our distance. In Figure 16(b), we show the same plot but with an ensemble generated by a "recombination" Markov chain.

The judges' plan lies near the middle of the ensemble in both cases, whereas the 2012 and 2016 plans lie on the edge. This indicates that the judges' plan is far more representative of the ensemble than the 2012 and 2016 plan in terms of its geography. The authors in [22] show that the judges' plan is *politically* representative of the ensemble, and here we see that it is *geographically* representative as well.

9. Conclusion.

Through a straightforward construction, we demonstrate how OT—already a lifting of a geometric structure to the set of probability measures—can be further lifted to a geometry on partitions. Our definition as a transport problem whose cost function is itself the result of solving a transport problem is an intriguing example of *hierarchical* transport in its own right. We demonstrate that several intuitve and classical results about transport and network flow apply in this lifted setting. Moreover, by restricting this distance to the space of partitions rather than general (unordered) collections of measures, we are able to derive some specialized results.

Few works have put a geometry on the space of partitions, and our progress on this problem suggests several avenues for future research:

- A well-known challenge in redistricting is the sheer number of ways to partition a graph, which obstructs global analysis of a districting plan relative to all possible alternatives. By putting a geometry on the space of partitions, we can ask whether the combinatorial count of partitions is truly insurmountable, or if there is a "small neighborhood" phenomenon whereby most partitions are close to a relatively small spanning subset of representatives.
- While our construction puts an intuitive and interpretable distance on the space of partitions, it can be expensive to compute relative to the more naive Hamming distances and total variation distances (see section 7), which have far weaker geometric behavior but are often very easy to compute. It may be the case that alternatives exist with a better compromise between computational efficiency and expressiveness.

- Motivated by our intended applications in redistricting, the constructions in this paper are discrete and restricted to the 1-Wasserstein distance. Our definition readily generalizes to partitions of compact regions in $\mathbb{R}^n$, although the underlying computational problem becomes much more challenging.

Ultimately, as demonstrated in section 8, our distance is not only a valuable mathematical construction but also—perhaps more importantly—a practical tool needed in emerging applications of data analysis to political science. Equipped with our distance and embedding algorithms, we can navigate and judge the extent of a collection of partitions, addressing a significant gap in current methodologies for ensemble-based redistricting.

Acknowledgments. The authors acknowledge the generous support and collaboration of several colleagues. In addition to the authors, the original team at the Voting Rights Data Institute (VRDI) working on distances between districting plans included Kristen Akey, Sonali Durham, Maira Khan, Jasmine Noory, Gabe Schoenbach, and Mélisande Teng. We additionally thank Ruth Buck, Sebastian Claici, Daryl DeFord, Moon Duchin, Lorenzo Najt, David Palmer, and Paul Zhang for valuable discussions throughout the research process.

REFERENCES

[1] R. K. Ahuja, T. L. Magnanti, and J. B. Orlin, *Network Flows*, Prentice-Hall, Englewood Cliffs, NJ, 1988.

[2] S. Bangia, C. V. Graves, G. Herschlag, H. S. Kang, J. Luo, J. C. Mattingly, and R. Ravier, *Redistricting: Drawing the Line*, preprint, arXiv:1704.03360, 2017.

[3] M. Beckmann, *A continuous model of transportation*, Econometrica, 20 (1952), pp. 643–660.

[4] I. Borg, P. J. Groenen, and P. Mair, *Applied Multidimensional Scaling*, Springer Science & Business Media, New York, 2012.

[5] L. A. Caffarelli and R. J. McCann, *Free boundaries in optimal transport and Monge-Ampère obstacle problems*, Ann. of Math. (2), 171 (2010), pp. 673–730.

[6] J. Chen and J. Rodden, *Cutting through the thicket: Redistricting simulations and the detection of partisan gerrymanders*, Election Law J., 14 (2015), pp. 331–345.

[7] J. Chen and J. Rodden, *Unintentional gerrymandering: Political geography and electoral bias in legislatures*, Quart. J. Political Sci., 8 (2013), pp. 239–269.

[8] M. Chikina, A. Frieze, and W. Pegden, *Assessing significance in a Markov chain without mixing*, Proc. Natl. Acad. Sci. USA, 114 (2017), pp. 2860–2864.

[9] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Scaling algorithms for unbalanced optimal transport problems*, Math. Comp., 87 (2018), pp. 2563–2609.

[10] L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard, *Unbalanced optimal transport: Dynamic and Kantorovich formulations*, J. Funct. Anal., 274 (2018), pp. 3090–3123.

[11] M. Cuturi and D. Avis, *Ground metric learning*, J. Mach. Learn. Res., 15 (2014), pp. 533–564.

[12] W. H. Day, *The complexity of computing metric distances between partitions*, Math. Social Sci., 1 (1981), pp. 269–287.

[13] D. DeFord and M. Duchin, *Redistricting reform in Virginia: Districting criteria in context*, Virginia Policy Rev., 12 (2019), pp. 120–146.

[14] D. DeFord, M. Duchin, and J. Solomon, *Recombination: A Family of Markov Chains for Redistricting*, submitted, arXiv:1911.05725, 2019.

[15] L. Denœud and A. Guénoche, *Comparison of distance indices between partitions*, in Data Science and Classification, Springer-Verlag, Berlin, 2006, pp. 21–28.

[16] S. Diamond and S. Boyd, *CVXPY: A Python-embedded modeling language for convex optimization*, J. Mach. Learn. Res., 17 (2016), pp. 1–5.

[17] M. Essid and J. Solomon, *Quadratically regularized optimal transport on graphs*, SIAM J. Sci. Comput., 40 (2018), pp. A1961–A1986.

[18] A. Figalli and N. Gigli, *A new transportation distance between non-negative measures, with applications to gradients flows with Dirichlet boundary conditions*, J. Math. Pures Appl. (9), 94 (2010), pp. 107–130.

[19] A. Galichon, *Optimal Transport Methods in Economics*, Princeton University Press, Princeton, NJ, 2018.

[20] N. Guillen, C. Mou, and A. Święch, *Coupling Lévy measures and comparison principles for viscosity solutions*, Trans. Amer. Math. Soc., 372 (2019), pp. 7327–7370.
[21] G. Herschlag, *gjh/districtingdatarepository*, September 2019, http://git.math.duke.edu/gitlab/gjh/districtingDataRepository.
[22] G. Herschlag, H. S. Kang, J. Luo, C. V. Graves, S. Bangia, R. Ravier, and J. C. Mattingly, *Quantifying Gerrymandering in North Carolina*, Statist. Public Ploicy, 7 (2020), pp. 30–38.
[23] G. Herschlag, R. Ravier, and J. C. Mattingly, *Evaluating Partisan Gerrymandering in Wisconsin*, preprint, arXiv:1709.01596, 2017.
[24] L. V. Kantorovich, *On the translocation of masses*, in Dokl. Akad. Nauk., 37 (1942), pp. 199–201.
[25] G. P. Leonardi and I. Tamanini, *Metric spaces of partitions, and Caccioppoli partitions*, Adv. Math. Sci. Appl., 12 (2002), pp. 725–753.
[26] D. Lombardi and E. Maitre, *Eulerian models and algorithms for unbalanced optimal transport*, ESAIM Math. Model. Numer. Anal., 49 (2015), pp. 1717–1744.
[27] S. Manson, J. Schroeder, D. Van Riper, T. Kugler, and S. Ruggles, *IPUMS National Historical Geographic Information System: Version* 12.0 [database], University of Minnesota, Minneapolis, MN, 2017.
[28] M. Meilă, *Comparing clusterings—an information based distance*, J. Multivariate Anal., 98 (2007), pp. 873–895.
[29] Metric Geometry and Gerrymandering Group, *mggg/gerrychain: v*0.2.12, July 2019, https://github.com/mggg/gerrychain.
[30] Metric Geometry and Gerrymandering Group and R. Buck, *mggg-states*, September 2019, https://github.com/mggg-states.
[31] G. Monge, *Mémoire sur la théorie des déblais et des remblais*, Histoire de l'Académie Royale des Sciences de Paris, 1781.
[32] L. Najt, D. DeFord, and J. Solomon, *Complexity and Geometry of Sampling Connected Graph Partitions*, preprint, arXiv:1908.08881, 2019.
[33] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine learning in Python*, J. Mach. Learn., 12 (2011), pp. 2825–2830.
[34] G. Peyré and M. Cuturi, *Computational optimal transport*, Found. Trends Mach. Learn., 11 (2019), pp. 355–607.
[35] F. Santambrogio, *Prescribed-divergence problems in optimal transportation*, lecture notes presented at the Mathemetical Sciences Research Institute, 2013
[36] F. Santambrogio, *Optimal Transport for Applied Mathematicians*, Springer, Cham, 2015.
[37] Z. Schutzman, *zschutzman/enumerator: v*0.1.5, 10 2019, https://github.com/zschutzman/enumerator.
[38] L. Slater Morton, *Lagrange Multipliers Revisited*, Cowles Commission Discussion Paper: Mathematics 403, Yale University, New Haven, CT, 1950.
[39] C. Villani, *Topics in Optimal Transportation*, Grad. Stud. Math. 58, American Mathematical Society, Providence, RI, 2003.
[40] H. P. Young, *Measuring the compactness of legislative districts*, Legislative Stud. Quart., 13 (1988), pp. 105–115.
[41] M. Yurochkin, S. Claici, E. Chien, F. Mirzazadeh, and J. Solomon, *Hierarchical optimal transport for document representation*, in Proceedings of the Neural Information Processing Systems Conference, 2019.