The Fortieth AAAI Conference on Artificial Intelligence (AAAI-26)

Putting Fair Division on the Map

Paula Böhm,[1] Robert Bredereck,[1] Paul Gölz,[2] Andrzej Kaczmarczyk[3], and Stanisław Szufa[4,5]
[1]Institut für Informatik, TU Clausthal, Germany
[2]Cornell University, Ithaca, NY, USA
[3]University of Chicago, Chicago, IL, USA
[4]AGH University, Kraków, Poland
[5]CNRS, LAMSADE, Université Paris Dauphine – PSL, France
{paula.boehm, robert.bredereck}@tu-clausthal.de, mail@paulgoelz.de, akaczmarczyk@uchicago.edu, s.szufa@gmail.com

Abstract

The fair division of indivisible goods is not only a subject of theoretical research, but also an important problem in practice, with solutions being offered on several online platforms. Little is known, however, about the characteristics of real-world allocation instances and how they compare to synthetic instances. Using dimensionality reduction, we compute a *map* of allocation instances: a 2-dimensional embedding such that an instance's location on the map is predictive of the instance's origin and other key instance features. Because the axes of this map closely align with the utility matrix's two largest singular values, we define a second, explicit map, which we theoretically characterize.

1 Introduction

Over the past 20 years, the field of fair division has made great advances in studying allocations of indivisible goods (Amanatidis et al. 2023). To illustrate this progress, consider the axiom of *envy-freeness*, which demands that no agent prefer another agent's bundle of allocated goods to their own. By the end of the 20th century, economists already understood envy-freeness well in settings with *divisible* goods. For example — assuming that preferences are additive, as we do throughout the paper — the allocation maximizing the Nash welfare is envy-free in these settings (Varian 1974; Shafer and Sonnenschein 1982). Little was known, however, about indivisible goods, a domain whose combinatorial structure poses additional challenges to mathematical investigation. Since envy-free allocations need not exist for all indivisible allocation instances[1], could envy at least be limited, or were large amounts of envy unavoidable?

Since then, we have gained a refined understanding of the degree to which envy can (and cannot) be avoided. Notably, the field has coalesced around an attractive relaxation of envy-freeness — *envy-freeness up to one good* (*EF1*) (Budish 2011) — and identified elegant algorithms (Lipton et al. 2004; Caragiannis et al. 2019) that construct EF1 allocations for any instance. Though intriguing open questions remain,[2] these questions only extend a solid understanding of the landscape of allocations. A similar progress has been made for alternative families of fairness axioms, like the maximin share or proportionality (Amanatidis et al. 2023).

In parallel to these theoretical advances, algorithms for allocating indivisible goods have entered practical usage, raising new questions for fair division. The Course Match system, for example, assigns course seats to MBA students at Wharton (Budish et al. 2017), and thousands of users have used the website Spliddit (Goldman and Procaccia 2014) to divide up estates or joint possessions. The deployment of such systems makes it more pressing to study not only worst-case instances but instances typically encountered in practice as well. For example, though envy-free allocations do not exist for all instances, should algorithms not aim for envy-free allocations for the 71% of Spliddit instances (Bai et al. 2022) where envy-freeness is possible? If so, how to choose among envy-free allocations? Or, as a second example, which algorithms can be implemented in practice? After all, the algorithms deployed on Course Match (Budish et al. 2023), on Spliddit (Caragiannis et al. 2019) as well as other proposed methods (Bredereck et al. 2021) run fast on practical inputs, seemingly defying (worst-case) computational hardness results. Answers to these questions cannot be found through worst-case analysis alone.

Whereas most work in fair division follows the worst-case paradigm, a noteworthy exception is some work in the paradigm of *distributional analysis*, which assumes that allocation instances are drawn from a probability distribution (Roughgarden 2020). Typically, these distributional models assume that all agent–good values are drawn independently, either from a single distribution (Amanatidis et al. 2017; Manurangsi and Suksompong 2019, 2021), a distribution that depends on the agent (Kurokawa, Procaccia, and Wang 2016; Farhadi et al. 2019; Bai and Gölz 2022), or a distribution that depends on the alternative (Dickerson et al. 2014; Farhadi et al. 2019).

On the upside, when m and n are large, these distributions generate highly structured instances, for which fair allocations are more prevalent. For example, basic algorithms yield envy-free allocations for these instances, with high probability (Dickerson et al. 2014; Manurangsi and Suksompong 2021). On the flip side, we are not aware of any em-

[1]Under the standard assumption that all goods must be allocated, which we also make throughout the paper.

[2]For example, whether envy-freeness up to *any* good (EFX) can be guaranteed (Chaudhury, Garg, and Mehlhorn 2020; Plaut and Roughgarden 2020; Amanatidis et al. 2021).

pirical work that has tested if the structures of these random instances are present in practical allocation problems.[3] In recent work, Bai et al. (2022) try to overcome some of these concerns through *smoothed analysis*, which means that their probability distributions are defined by adding random noise to a worst-case utility profile. While they extend the possibility results for envy-freeness to more general distributions, they leave "whether our smoothed model accurately describes the properties of real-world utility profiles that possess envy-free allocations" to future empirical analysis.

We bridge this gap between theoretical work and practical instances in fair allocation by providing a principled framework facilitating the analysis of the structure and properties of indivisible resource allocation instances. We draw inspiration from Szufa et al. (2020), who created a *map of elections* — a two-dimensional embedding of election instances. Together with following works (Boehmer et al. 2021; Faliszewski et al. 2023), Szufa et al.'s technique significantly increased the understanding of the structure and features of distributional models of elections as well as their relation to real-world election instances. Besides founding the framework itself, our most important contributions include:

- Proposing two types of *maps of allocation instances* utilizing dimensionality reduction for analyzing structures of allocation instances;
- Giving a conceptual interpretation to the two axes of the map as well as to the boundaries of the map, which can be traced by natural "extreme" instances;
- Showing that an instance's position on the map is highly informative by demonstrating that multiple key features of allocation instances vary continuously over our maps;
- Naturally clustering instances generated by a specific probability distribution or coming from specific data sources, thus allowing for relating various data sources and their features to each other;
- Proposing a new distribution of allocation instances capable of covering vast areas of our maps, which can easily be applied to future experiments to generate diverse data.

Our Approach and Results

In Section 3, we create a map by following the methodology of Szufa et al. (2020): we define a natural distance between fair division instances (the ℓ^1 distance between utility matrices up to row and column permutations), and use multi-dimensional scaling (Kruskal 1964) (a common technique for dimensionality reduction) to find a 2-dimensional *distance-embedding map* of a given set of allocation instances that approximately preserves the pairwise distances. To scale this approach to allocation instances with many agents and goods, we also propose a computationally tractable proxy distance, which leads to almost identical maps at a much smaller computational cost.

Using data from three real-world data sources and several synthetic distributions over approval instances, we show that the map picks up on common properties of instances from the same data source. We also show that key features of the allocation instance (e.g., the maximum achievable Nash welfare or the existence of envy-free allocations) are distributed in clear patterns across the map. Whereas two natural probability distributions over instances fail to cover the whole range of real-world instances, a new distribution we propose covers the whole map in our experiments and so is a natural choice for future synthetic experiments in fair division.

In Section 4, we go beyond the heuristically generated embedding described above and provide an *explicit map*, i.e., an explicit function from allocation instances into $\mathbb{R}^2$, which reproduces the general layout of the distance-embedding map. Since an instance's position on this explicit map is given by the two largest singular values of its utility matrix, our explicit map is amenable to theoretical analysis. In particular, we tightly characterize the range of the map, and identify (up to rounding terms) the most extreme instances in the map's corners. We also show that the explicit map similarly segments instance sources and features.

Omitted proofs, results, experiments,[4] and details can be found in the full version of the paper (Böhm et al. 2025).

2 Preliminaries

Let $[n]$ be a set of agents and $[m]$ be a set of goods. For ease of exposition, we assume that $m \geq n \geq 2$, which arguably includes all interesting allocation instances. An *allocation instance* (of indivisible goods) is described by a *utility matrix* $U \in \mathbb{R}_{\geq 0}^{n \times m}$ whose entries $u_{i,j}$ describe agent i's utility for good j. We refer to the ith row of this matrix as i's *utility vector* $\vec{u}_i \in \mathbb{R}_{\geq 0}^{m}$. We assume additive preferences, so agent i gets utility $u_i(S) := \sum_{j \in S}$ for a *bundle* $S \subseteq [m]$ of goods is. Since we only consider tasks in which agents' utilities $u_i([m])$ for the whole bundle are normalized to 1, we deal exactly with the set of *row-stochastic* matrices U.

Characteristic Instances

As useful signposts for navigating through the space of allocation instances, we define several *characteristic instances*. Each of these instances represents an intuitively extreme scenario with easily-understood, symmetric structure. For any n and m, our three characteristic instances are the following (for a matrix representation, see full version):

Indifference (IND) models the situation where each good is equally valuable to each agent. Thus, all entries of its utility matrix are $1/m$.

Separability (SEP) captures the scenario in which each agent values exactly one good, distinct from the goods that all other agents value. Thus, its utility matrix is a matrix with ones on the diagonal and zeros everywhere else. In particular, if $m > n$, all but the first n columns have all-zero entries.

Contention (CON) considers one single good valued by all agents, and all other goods yielding no value to any agent. Hence, its utility matrix has a first column of ones, and is zero everywhere else.

In Section 4, the explicit map will lead us to introduce three new characteristic instances: two variants of separability and an entirely new instance called *bicontention*.

[3] Bai et al. (2022) voice doubts, but also do not provide data.

[4] The code is publicly available at https://github.com/Project-PRAGMA/Map-of-Allocations--AAAI26

Real-World Instances

Three of the instance sources we consider generate instances derived from real-world preferences over goods. Two of these sources have not been previously analyzed in the fair-division literature:

Spliddit. The heart of our real-world data is a dataset of all allocation instances submitted to Spliddit as of June 2025.[5] This dataset is particularly valuable because it contains instances from real Spliddit users. Since most of the 5574 Spliddit instances are small, our evaluation will focus on two combinations of n, m that are relatively well represented: First, we will study the setting $n = 3$, $m = 6$, for which the number of instances (namely, 3152) is highest. Since, for larger dimensions, the number of Spliddit instances drops precipitously, only 36 Spliddit instances exist for our second evaluation scenario of $n = m = 5$.

Island. To obtain this dataset, Benadè et al. (2018) elicited additive utilities for private goods (though they were motivated by a public-goods setting), by asking 572 crowd workers to spread 100 points between 10 items in proportion to how much they would value these items (a map, pocket knife, compass, etc.) if they were stranded on a deserted island.[6] By sampling sets of n agents and m goods and rescaling agents' utilities, we simulate (hypothetical) allocation scenarios in which only those m goods stand to be allocated between those n agents.

Candy. Our final dataset (Kaczmarczyk 2023) has a similar shape and consists of the additive preferences over 10 types of snacks indicated by 48 teenagers attending a summer camp. We again obtain instances by subsampling, assuming that only one snack of each type is available.

Distributions over Synthetic Instances

In addition to the above instances derived from practical data, we consider three types of synthetic instances:

I.i.d. As described in the introduction, i.i.d. valuations have been empirically and theoretically studied in the literature. Each agent i's utility vector is independently generated by sampling i's values for the m goods independently from some fixed distribution $\mathcal{D}$, and then rescaling this vector to sum up to 1. In our experiments, we choose $\mathcal{D}$ as the uniform distribution over $[0, 1]$ and as the exponential distribution. (The exponential distribution's rate is inconsequential since utilities are normalized.)

Attributes. It is a natural explanatory model of how agents' utilities arise (also used by Boehmer, Heeger, and Szufa (2023)). Let $d \in \mathbb{N}$ be a fixed number of *attributes*. For each good j, we sample a vector $\vec{g}_j$ from $[0, 1]^d$ uniformly at random (higher coordinates indicate that the good is more desirable along an attribute). For each agent i, we sample a *priority vector* $\vec{a}_i$ over the attributes also from $[0, 1]^d$ (higher coordinates indicate that the agent cares more about an attribute). Then, agent i's utility for good j is proportional to the dot product of $\vec{a}_i$ and $\vec{g}_j$.

Resampling-Dirichlet. The former two distributions will end up only covering small areas of the map. Hence, we introduce a third distribution, inspired by that over approval elections (Szufa et al. 2022), which will cover the range of real-world allocation instances. Given three parameters $p \in [0, 1]$, $\phi \in [0, 1]$, and $t \geq 0$, we proceed in three phases. First, we choose the instance's *central approval set* V^* by uniformly drawing $\lfloor p \cdot m \rfloor$ goods. Second, for each agent i, we determine independently across i and j whether i approves j as follows: with probability $1 - \phi$, i approves j iff $j \in V^*$; else, i approves j with probability p. For the third step, let $A = \{a_1, a_2, \dots\} \subseteq [m]$ be the goods approved by some agent i, labeled such that $a_1 < a_2 < \dots$. Then, for a fixed parameter t, we draw agent i's utilities from a Dirichlet distribution parameterized by an m-dimensional vector $\vec{v}$, such that $\vec{v}[j] = 10^{-10}$ for $j \in [m] \setminus A$ and $\vec{v}[a_j] = \frac{2}{(j/|A|+0.01)^t}$ for $j \in [|A|]$. Should this process leave i without any approved goods, they approve one good uniformly at random. For a more detailed description of this distribution, see full version.

3 Distance-Embedding Map

We create our first map by performing the following two steps, introduced by Szufa et al. (2020). First, we define a notion of distance between pairs of allocation instances, and compute all pairwise distances for a collection of instances from the sources described above.[7] Second, we use *multi-dimensional scaling* (Kruskal 1964) to embed this collection of instances into the plane in a way that approximately preserves their pairwise distances, which will allow us to see patterns in the similarity of instances.

Distances between Allocation Instances

Naïvely, we would like to measure the distance between two instances (with equal n and m) through the entry-wise ℓ^1 norm. That is, if the instances' utility matrices are U^1 and U^2, we would calculate their distance $||U^1 - U^2||_{1,1}$ by summing up, over all $n \cdot m$ coordinates, the absolute difference between U^1's and U^2's entry at this coordinate. This distance is, however, not desirable, since the ordering of rows and columns in a utility matrix is arbitrary but would greatly impact this distance. Instead, an appropriate distance between allocation instances ought to remain invariant when reordering the utility matrices' rows (i.e., agents) or columns (i.e., goods).

Our *valuation distance* achieves these goals of anonymity and neutrality by explicitly minimizing over all row and column permutations. To express this without matrix notation, suppose that we have bijections $\pi_{agents}\colon [1, n] \to [1, n]$ and $\pi_{goods}\colon [1, m] \to [1, m]$ between, respectively, the agents and the goods of the first and second allocation instance. By summing up, for each agent i and each good j, the absolute difference $|U^1_{i,j} - U^2_{\pi_{agents}(i), \pi_{goods}(j)}|$ between i's utility for j and the utility of i's matched agent $\pi_{agents}(i)$ for j's matched good $\pi_{goods}(j)$, we calculate the entry-wise ℓ^1 distance; the valuation distance is defined as the minimum of this distance, taken over all matchings π_{agents} and π_{goods}. A

[5] Dataset available upon request from Nisarg Shah.

[6] Dataset available upon request from Gerdus Benadè.

[7] The number and parameters of the instances we study from each source are described in the full version.

nice property of the valuation distance is that two instances U^1 and U^2 have distance zero exactly if they are identical up to relabeling agents and goods, i.e., they are *isomorphic*.

The valuation distance confirms our intuition that the characteristic instances introduced in Section 2 are indeed "extreme points" in the space of allocation instances in the sense that, if $n = m$, the characteristic instances are mutually equidistant at distance $2\,(m-1)$; and this distance is maximal (see full version).

Studying the Distance-Embedding Map

To plot distance-embedding maps, we first generate a collection of instances from all of our sources — as mentioned in Section 2, we consider two combinations of n, m: for $n = 3, m = 6$ ("3×6" from now on) and for $n = m = 5$ ("5×5") (see full version for details). Then, we compute the valuation distance between all pairs of instances in each collection, and embed the distances in 2D Euclidean space using multi-dimensional scaling (implemented in scikit-learn, using default parameters).[8] Fig. 1 displays the resulting embeddings, where each point represents an instance, and an instance's location stays fixed within all maps of the same dimension. Note that the characteristic instances lie in distinct corners of the map, reflecting our observation at the end of the previous subsection. The triangle they form spans most of the map, so they serve as useful reference points.

We can also immediately make out that the instance sources are spread in different patterns across the map. Among the real-world sources (Figs. 1a and 1e), the Spliddit instances are concentrated near indifference for 3×6 instances but spread evenly across the map among 5×5 instances. The island instances tend to vary between indifference and contention whereas the candy instances tend to lie closer to separability, which suggests that the agents' preferences over survival items are more aligned than the children's preferences over snacks. Among the synthetic distributions (Figs. 1b and 1f), the i.i.d. distributions and attributes model generate concentrated clusters close to indifference. The observation that these synthetic distributions do not cover the range of the real-world data (and even of just the Spliddit data) raises concerns about the degree to which distributional-analysis results for i.i.d. instances (or attributes instances) can be applied to real-world fair division problems. Since the resampling-Dirichlet instances cover the map to a much higher degree, the resampling-Dirichlet distribution appears to be a more fruitful proxy for real-world instances for future studies.[9]

Next, we discuss how several natural features of allocation instances vary across the map, which we compute using constraint programming. In Figs. 1c and 1g, we study to which degree the instances allow for (almost) envy-free allocations. Specifically, denote an allocation of all goods over the agents by $[m] = S_1 \,\dot\cup\, S_2 \,\dot\cup \cdots \dot\cup\, S_n$, where S_i denotes agent i's bundle. The *minimax envy* is the minimum, over all allocations, of the largest amount by which some agent envies another, i.e., $\max_{i \neq i'} u_i(S_{i'}) - u_i(S_i)$. An instance has envy-free allocations if and only if the minimax envy is at most 0 (we highlight these instances with cross markers). But the minimax envy gives a gradual measure of how far envy-freeness is from being achievable (or how much it can be overattained). As we can see, an instance's position on the map is highly informative for the minimax envy and the existence of envy-free allocations. For 3×6 instances, envy-freeness seems to be hopeless near contention (and for contention itself) and easy near separability. For the rest of the map, minimax envy is close to zero, which means that almost envy-free allocations exist widely, and exactly envy-free allocations generally exist below the upper outline of the map. 5×5 instances are less hospitable to envy-freeness: envy-free allocations exist only near the lower border of the map, and the minimax envy becomes higher (i.e., worse), the further up on the map an instance lies.

Finally, Figs. 1d and 1h show that the maximum Nash welfare achievable by any allocation also varies smoothly over the map, increasing the closer an instance lies to separability. It is noteworthy that this map differs only slightly from the maximum utilitarian welfare that can be achieved (see Fig. 2b below). We show the distribution of various additional features in the full version.

A Faster Distance for Large Instances

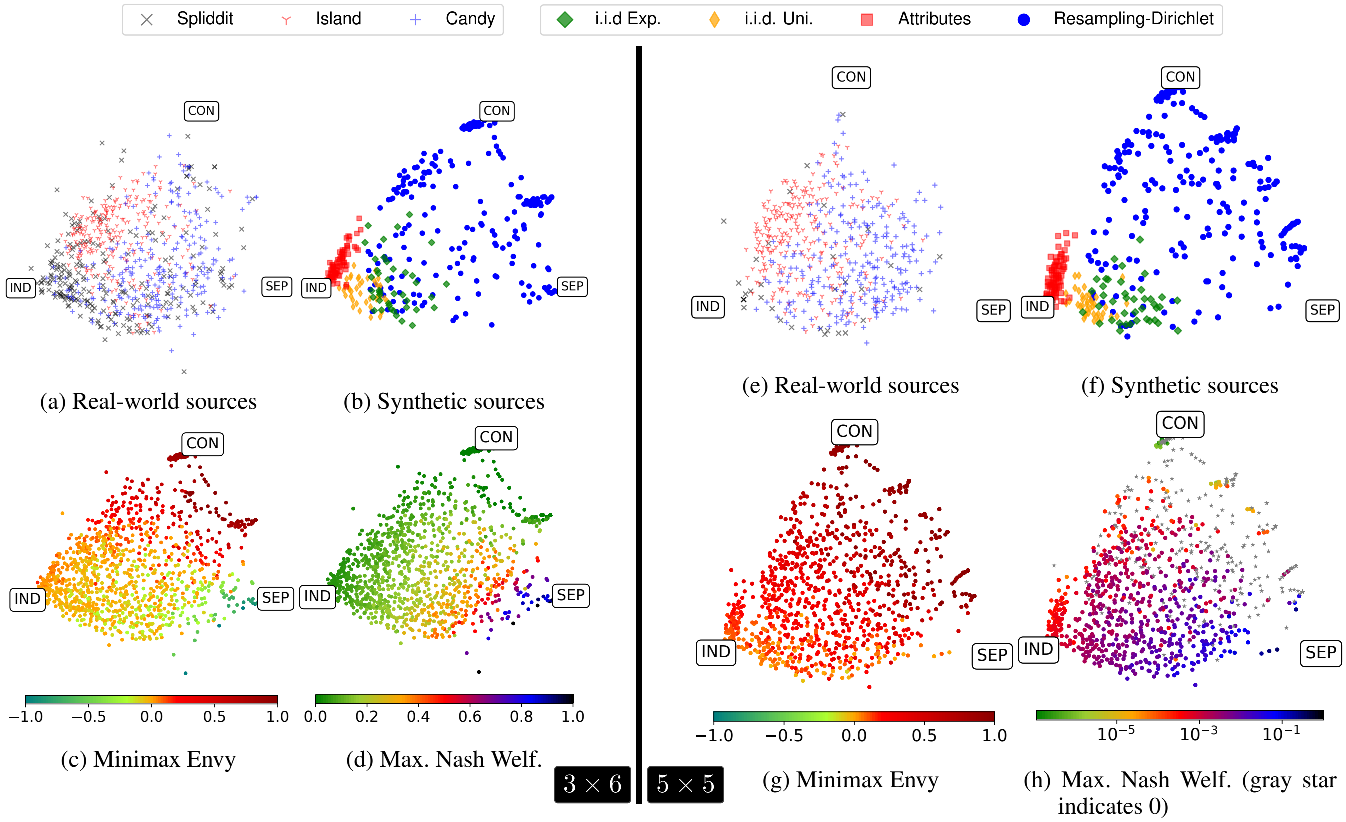
Creating these distance-embedding maps required computing the valuation distance for many pairs of instances. We were able to do this because the instance dimensions we have focused on, i.e., the dimensions that regularly appear on Spliddit, are rather small. In general, however, computing the valuation distance is NP-hard (see full version), and would be prohibitively slow to compute for, say, instances with dimensions $n = m = 10$ (even with an ILP solver).

To verify that the patterns we described above extend to large instances, and to ready our mapping approach for a future in which larger fair-division problems might be routinely solved, we define the *demand distance*, a heuristic approximation to the valuation distance, in the full version. Crucially, the demand distance between two instances can be computed in polynomial time by finding a maximum-weight bipartite matching, which is also fast in practice. Though the demand distance may, in principle, deviate substantially from the valuation distance, we find that both correlate very well, with a Pearson correlation coefficient of at least 97% across our dimensions. Maps resulting from both distances are virtually indistinguishable, which is demonstrated by numerous juxtaposition figures in the full version.

Using this demand distance, we compute a distance-embedding map together with multiple allocation features for $n = 10$ and $m = 20$ (see full version). Since none of the real-world allocation instances are this large, for demonstration purposes we decided to adapt another four large datasets related to scenarios other than standard indivisible goods al-

[8] We made the implementation of generation of our distance-embedding maps publicly available as a python package *mapof-allocations* (https://pypi.org/project/mapof-allocations) extending *mapof* (https://pypi.org/project/mapof), formerly *mapel*, a framework for computing maps of elections (Szufa et al. 2020).

[9] In the full version, we show that resampling instances continue to cover the embedding map for substantially larger instance dimensions ($n = 10, m = 20$; $n = 25, m \in \{25, 75, 100\}$), using the methodology of Section 3.



(a) Real-world sources (b) Synthetic sources (c) Minimax Envy (d) Max. Nash Welf. (e) Real-world sources (f) Synthetic sources (g) Minimax Envy (h) Max. Nash Welf. (gray star indicates 0)

Figure 1: Distribution of instance sources and two features on the distance-embedding maps.

location (details in full version). We also dropped i.i.d. and attributes instances as they were even more clustered around indifference. The broad patterns and data-source separation we observed in 3×6 and 5×5 instances continue to hold, which we have also confirmed for even larger instances. Creating maps using the demand distance scales readily to larger sizes — even, say, to $n = m = 100$.

4 Explicit Maps

Generating maps through a distance embedding entails several inherent disadvantages:

Instability. The distance-embedding map may potentially change non-continuously as the result of slight changes to the random seed or the set of mapped instances (though we did not observe this, see full version).

Data dependence. Suppose that you want to place an allocation instance on the map to predict its properties. This would require data for all other instances and computing pairwise distances, which would be more difficult than directly computing your instance's properties.

Theoretical intractability. Which instances are "most extreme"? Where do instances from a probability distribution lie on the map? One can answer such questions empirically, but not theoretically.

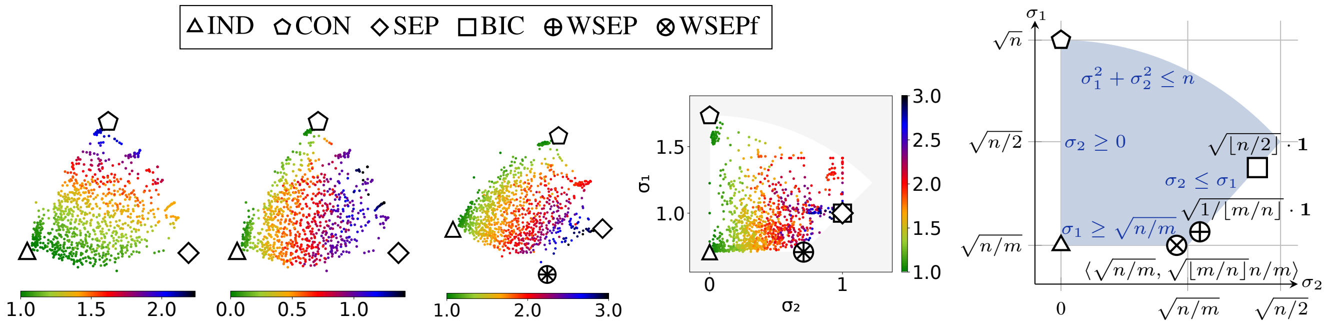
To overcome these challenges, we propose an *explicit map* of fair division instances: a function μ from allocation instances to $\mathbb{R}^2$, which replicates the general layout of the distance-embedding map. Specifically, this function maps $n \times m$ utility matrices as follows:

$$\mu : \mathbb{R}^{n \times m} \to \mathbb{R}^2 \qquad U \mapsto \big(\sigma_1(U), \sigma_2(U)\big),$$

where $\sigma_1(U)$ and $\sigma_2(U)$ are the largest and second-largest singular values of the matrix U, respectively. As Fig. 2a shows (on the same 5×5 map as in Fig. 1), these two values closely capture the vertical and horizontal ordering of instances in our distance-embedding map, ensuring that the two maps are closely aligned (for the corresponding 3×6 map, see the full version). In this section, we show that the explicit map is similarly informative as the distance-embedding map, while being stable, data independent, and theoretically tractable by design.

Demystifying the Singular Value Map

We begin by recalling facts about singular values that make them suitable components for our explicit map function. First, the singular values are invariant under permutations of rows or columns in the utility matrix, so that relabeling agents or goods will not change the map embedding. Second, σ_1 and σ_2 are 1-Lipschitz continuous in the entries of the matrix, which together with the previous point implies that two instances with small valuation distance must be placed near each other on the explicit map. Third, adding a column of zeros, i.e., a good that no agent values, does not change the singular values, which means that instances



(a) Distribution of σ_1 (left) and σ_2 (right) on our distance-embedding map for 5×5.

(b) Distribution of max. utilitarian welfare on our distance-embedding (left) and explicit (right) map for 3×6.

(c) Bounding inequalities of the map and locations $\langle \sigma_1, \sigma_2 \rangle$ of characteristic instances, with $\mathbf{1} = \langle 1, 1 \rangle$.

Figure 2: Distributions of the σ_1, σ_2 values (a), comparison of the two maps (b), and the general shape of the explicit map (c).

can be naturally compared across different m. Finally, implementations of efficient algorithms for computing singular values are readily available (e.g., in numpy), which makes it easy to compute a given instance's position on the map.

We now aim to give the reader an intuition for what information σ_1 and σ_2 express about an allocation instance and why. We begin with σ_1, which can be expressed as

$$\sigma_1 = \max_{\vec{v}_1 \in \mathbb{R}^m, \|\vec{v}_1\|=1} \|U\,\vec{v}_1\|, \quad (1)$$

where $\|\cdot\|$ is the Euclidean (ℓ^2) norm. Since we rarely think about utility matrices as linear functions over unitary vectors, it is instructive to pretend that the norms in Eq. (1) were ℓ^1-norms. In this case (choosing $\vec{v}_1$ nonnegative w.l.o.g.), the $U\,\vec{v}_1$ being optimized over are the convex combination of U's columns, for the coefficients given by $\vec{v}_1$. If we were indeed maximizing the ℓ^1-norm of $U\,\vec{v}_1$, σ_1 would be the largest column sum, or *maximum demand*. Though the ℓ^2 norm slightly complicates the picture,[10] σ_1 and the maximum demand are very highly correlated: across our 3×6 instances, for example, the correlation coefficient is 97%. Thus, σ_1 can be understood as good approximation of the maximum demand, up to shifting and rescaling.

To interpret the second-largest singular value σ_2, we recall how the singular value decomposition of an $\mathbb{R}^{n \times m}$ matrix U can be used to find a low-dimensional embedding of the row vectors (in our case, the agents' utility vectors).[11] For example, the line through the origin $\mathit{span}(\{\vec{v}_1\})$, spanned by the argmax of Eq. (1), is the best 1-dimensional space to embed the rows in, in the following sense: if we sum up, for each row $\vec{u}_i \in \mathbb{R}^m$, the squared length of its projection onto this space, $\mathit{span}(\{\vec{v}_1\})$ maximizes this sum across all 1-dimensional subspaces. In fact, this sum is σ_1^2, which means that σ_1 measures "how much" of the row vectors can be captured by a 1-dimensional embedding. Similarly, σ_2, which can be calculated as

$$\max_{\vec{v}_2 \in \mathbb{R}^m, \|\vec{v}_2\|=1, \vec{v}_2 \perp \vec{v}_1} \|U\,\vec{v}_2\|,$$

measures how much the row embedding improves when going from the optimal 1-dimensional space $\mathit{span}(\{\vec{v}_1\})$ to the optimal 2-dimensional space $\mathit{span}(\{\vec{v}_1, \vec{v}_2\})$.

Thus, as a first approximation, σ_2 measures how diverse the agents' utilities are. It is zero if all agents have the same utility vector, and large when there are blocks of agents that completely disagree on which goods have nonzero value. To again find a more elementary correlate, we define an instance's *preference diversity* as the mean ℓ^2 distance between utility vectors, averaged over all pairs of agents in the instance. Again, we find a very high correlation (96% correlation coefficient for 3×6).

Theoretical Properties of the Map

We theoretically characterize the image of our map function μ for given dimensions n, m. Our task — characterizing the combinations of singular values in stochastic rectangular matrices — is of interest independently to our fair-division setting, but, to our knowledge, has not previously been undertaken. This process will give us a more precise understanding of what makes instances extreme along either dimension of our map. Figure 2c summarizes both the outlines of the map and the positions of characteristic instances, which can guide the reader through this section. We orient on the page such that σ_1 grows in the "North" and σ_2 in the "East" direction, which by Fig. 2a generally aligns with how we have presented the distance-embedding map.

Whereas CON and IND still mark the left corners of our map, the other two corners lead us to new characteristic instances. For the lower-right corner, we refine our definition of separability since SEP (with $\sigma_1 = \sigma_2 = 1$) only lies on

[10] It gives an advantage to combinations of columns in which several columns have positive coefficients, and it encourages making a few coordinates of $U\,\vec{v}_1$ large rather than all.

[11] See Chapter 3 by Blum, Hopcroft, and Kannan (2020) for a detailed explanation. Though singular values are closely connected to dimensionality reduction, our use is non-standard. Applying value decomposition directly to find a 2D embedding of utility matrices would result in embeddings highly sensitive to row and column permutations and would thus not be fruitful. One way to understand the discussion above is that we map each utility matrix to the square roots of the top-two eigenvalues in its principal component analysis; except that we do not shift column sums to zero, since this would, e.g., make IND and CON indistinguishable.

the lower boundary if $n = m$. If m is a proper multiple of n, the lower-left corner is instead inhabited by *wide separability*, in which every agent values m/n disjoint goods, giving equal value n/m to each of them. In the full version, we extend wide separability to $n \nmid m$ in two slightly different ways: one, WSEP, always lies on the right border while the other, WSEPf always lies on the lower border.

The final characteristic instance is *bicontention* (BIC); here, half of the agents place all utility on one common good and half of the agents on a second common good (for odd n, one agent places all value on a third good). Since this instance combines highly demanded goods with sharply distinct utility vectors, it always lies on the right border and, for even n, is exactly located in the upper-right corner.

The main results of this section address the boundaries of the map. For each side, we bound the map by an inequality and show that the inequality is sharp using our characteristic instances. For three of the sides, we give simple, necessary-and-sufficient conditions for an instance lying on the boundary. If n is even and divides m, our characteristic instances lie exactly in the corner points of the map, and we can exactly trace three of the four sides by interpolating between corner instances. If the divisibility conditions do not hold, as shown in Fig. 2c, the characteristic instances lie in the corner up to rounding terms. Proofs of our characterizations tend to be short and cute, but are deferred to the full version.

Theorem 1 ("West"). *σ_2 is at least 0. An instance lies on this boundary iff all agents have the same utility vector. In particular,* IND, CON, *and their convex combinations lie on this boundary.*

Theorem 2 ("South"). *σ_1 is at least $\sqrt{n/m}$. An instance lies on this boundary iff all columns of its utility matrix have an equal sum (namely, n/m). In particular,* IND, WSEPf, *and their convex combinations lie on this boundary.*

Theorem 3 ("North"). *σ_1 is at most $\sqrt{n - \sigma_2^2} \leq \sqrt{n}$. An instance lies on this boundary iff each agent values a single good, and if at most two goods are valued by any agent. In particular,* CON *and, if n is even,* BIC *lie on this boundary.*

Theorem 4 ("East"). *σ_2 is at most σ_1. If U, after row and column permutation, has the block matrix structure $\begin{pmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & B \end{pmatrix}$ for rectangular matrices A, B and $\sigma_1(A) \geq \sigma_1(B)$, this is sufficient for lying on the boundary. (If B has height 0, we set $\sigma_1(B) = 0$.) In particular,* WSEP, BIC, *and a suitable interpolation lie on this boundary.*

We conclude by pointing out that existing and future results in the theory of nonnegative random matrices have implications for our explicit map. For example, consider a random process in which a single utility vector is drawn from a flat Dirichlet distribution and duplicated for all agents (thus, $\sigma_2 = 0$). For this distribution over instances, Crumpton, Fyodorov, and Vivo (2022) recently derived that $\mathbb{E}[\sigma_1^2] = 2\,n/(n+1)$ as well as formulas for σ_1^2's higher moments. Brito, Dumitriu, and Harris (2022) study a random process, in which, for fixed integers $d_2 \geq d_1 \geq 3$, an instance is uniformly chosen in which each agent values d_1 goods at value $1/d_1$, and each good is valued by d_2 agents. In this case, σ_1 is always $\sqrt{n/m}$, and the authors show that, as $m, n \to \infty$, σ_2 converges to $(\sqrt{d_1 - 1} + \sqrt{d_2 - 1})/d_1$ in probability.

Comparison of the Maps

Comparing the explicit map to our distance-embedding map (see, e.g., Fig. 2b for the largest achievable utilitarian welfare), we see that the two maps have a similar layout and communicate similar information overall. In the full version, we provide extensive diagrams showing that this similarity extends to other features, the identifiability of instance sources, and to the 5×5 instances as well.

One major difference is in how the density of instances varies across both maps. Whereas the distance-embedding map fills the map at a rather uniform density, which helps legibility, the explicit map clusters some instances very densely (e.g., near the South boundary and the $\sigma_2 = 1$ line in Fig. 2b). But these dense areas of the map seem to highlight meaningful clusters of similar instances, given that instance features tend to be homogeneous within these dense areas. The shape of instances in the explicit map can similarly highlight noteworthy patterns. For example, the straight lines at $\sigma_1 = \sqrt{2}$ and $\sigma_2 = 1$ in Fig. 2b are formed by instances in which several agents only value one good (see full version). The distance-embedding map makes such phenomena much harder to spot.

Because of the advantages of stability, data independence, and theoretical tractability, we see the explicit map as preferable over the distance-embedding map for detailed analysis. Nevertheless, the distance-embedding map plays a crucial role by justifying the explicit map: the relevance of the explicit map rests in large part on the fact that the former map surfaced the two largest singular values as the most salient dimensions of difference between instances.

5 Conclusion

We hope that our exploration of allocation instances and the provided framework initiate discussions about which assumptions on such instances are supported by practice, and how fair-division theory can leverage these assumptions to provide algorithms with stronger fairness properties for the bulk of practical allocation instances. We adopted the standard utility-based notion of fairness — envy-freeness with respect to declared preferences — which does not aim to capture Aristotelian or other ethical conceptions of equity; relating these perspectives, especially under imperfect information or bounded rationality, is left for future work.

The main limitation of our study is that — despite tapping into unconventional data sources — we were unable to test our approach on large real-world instances of typical resource allocation settings (we did so only on datasets adapted from similar yet different domains — rankings and course allocation). This seems rooted in a broader limitation of the practice of fair division: large allocation problems are hardly ever solved, or the preference data are not made available. We believe that our community should strive to collect and share such datasets, as has been recently done for election data (Mattei and Walsh 2013).

Acknowledgments

This project has received funding from the French government under the management of Agence Nationale de la Recherche as part of the France 2030 program, reference ANR-23-IACL-0008, and from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101002854). The work was partly supported by the Office of Naval Research (ONR) under Grant Number N00014-23-1-2802 (the views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Office of Naval Research or the U.S. Government). Paul Gölz was supported by the National Science Foundation under Grant No. DMS-1928930 and by the Alfred P. Sloan Foundation under grant G-2021-16778 while in residence at the Simons Laufer Mathematical Sciences Institute (formerly MSRI) in Berkeley, CA. Paula Böhm and Robert Bredereck acknowledge support by the Deutsche Forschungsgesellschaft (German Research Foundation, DFG), project COMSOC-MPMS (grant agreement No. 465371386).

References

Amanatidis, G.; Aziz, H.; Birmpas, G.; Filos-Ratsikas, A.; Li, B.; Moulin, H.; Voudouris, A. A.; and Wu, X. 2023. Fair Division of Indivisible Goods: Recent Progress and Open Questions. *Artificial Intelligence*, 322: 103965.

Amanatidis, G.; Birmpas, G.; Filos-Ratsikas, A.; Hollender, A.; and Voudouris, A. A. 2021. Maximum Nash Welfare and Other Stories about EFX. *Theoretical Computer Science*, 863: 69–85.

Amanatidis, G.; Markakis, E.; Nikzad, A.; and Saberi, A. 2017. Approximation Algorithms for Computing Maximin Share Allocations. *ACM Transactions on Algorithms (TALG)*, 13(4): 1–28.

Bai, Y.; Feige, U.; Gölz, P.; and Procaccia, A. D. 2022. Fair Allocations for Smoothed Utilities. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC'22)*.

Bai, Y.; and Gölz, P. 2022. Envy-Free and Pareto-Optimal Allocations for Agents with Asymmetric Random Valuations. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI '22)*, 53–59.

Benadè, G.; Itzhak, N.; Shah, N.; Procaccia, A. D.; and Gal, Y. 2018. Efficiency and Usability of Participatory Budgeting Methods. Unpublished.

Blum, A.; Hopcroft, J. E.; and Kannan, R. 2020. *Foundations of Data Science*. Cambridge University Press.

Boehmer, N.; Bredereck, R.; Faliszewski, P.; Niedermeier, R.; and Szufa, S. 2021. Putting a Compass on the Map of Elections. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI '21)*, 59–65.

Boehmer, N.; Heeger, K.; and Szufa, S. 2023. A Map of Diverse Synthetic Stable Roommates Instances. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*, 1003–1011.

Bredereck, R.; Figiel, A.; Kaczmarczyk, A.; Knop, D.; and Niedermeier, R. 2021. High-Multiplicity Fair Allocation Made More Practical. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '21)*, 260–268.

Brito, G.; Dumitriu, I.; and Harris, K. D. 2022. Spectral Gap in Random Bipartite Biregular Graphs and Applications. *Combinatorics, Probability and Computing*, 31(2): 229–267.

Budish, E. 2011. The Combinatorial Assignment Problem: Approximate Competitive Equilibrium from Equal Incomes. *Journal of Political Economy*, 119(6): 1061–1103.

Budish, E.; Cachon, G. P.; Kessler, J.; and Othman, A. 2017. Course Match: A Large-Scale Implementation of Approximate Competitive Equilibrium from Equal Incomes for Combinatorial Allocation. *Operations Research*, 65(2): 314–336.

Budish, E.; Gao, R.; Othman, A.; Rubinstein, A.; and Zhang, Q. 2023. Practical Algorithms and Experimentally Validated Incentives for Equilibrium-Based Fair Division (A-CEEI). arXiv:2305.11406 [cs:GT].

Böhm, P.; Bredereck, R.; Gölz, P.; Kaczmarczyk, A.; and Szufa, S. 2025. Putting Fair Division on the Map. arXiv:2504.20161 [cs.GT].

Caragiannis, I.; Kurokawa, D.; Moulin, H.; Procaccia, A. D.; Shah, N.; and Wang, J. 2019. The Unreasonable Fairness of Maximum Nash Welfare. *ACM Transactions on Economics and Computation*, 7(3): 1–32.

Chaudhury, B. R.; Garg, J.; and Mehlhorn, K. 2020. EFX Exists for Three Agents. In *Proceedings of the 21st ACM Conference on Economics and Computation (EC'20)*, 1–19.

Crumpton, M. J.; Fyodorov, Y. V.; and Vivo, P. 2022. Statistics of the Largest Eigenvalues and Singular Values of Low-Rank Random Matrices with Non-Negative Entries.

Dickerson, J. P.; Goldman, J.; Karp, J.; Procaccia, A. D.; and Sandholm, T. 2014. The Computational Rise and Fall of Fairness. In *Proceedings of the 28th Conference on Artificial Intelligence (AAAI '14)*.

Faliszewski, P.; Kaczmarczyk, A.; Sornat, K.; Szufa, S.; and Wąs, T. 2023. Diversity, Agreement, and Polarization in Elections. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI '23)*, 2684–2692.

Farhadi, A.; Ghodsi, M.; Hajiaghayi, M. T.; Lahaie, S.; Pennock, D.; Seddighin, M.; Seddighin, S.; and Yami, H. 2019. Fair Allocation of Indivisible Goods to Asymmetric Agents. *Journal of Artificial Intelligence Research*, 64: 1–20.

Goldman, J.; and Procaccia, A. D. 2014. Spliddit: Unleashing Fair Division Algorithms. *ACM SIGecom Exchanges*, 13(2): 41–46.

Kaczmarczyk, A. 2023. Camp Candy Allocation Dataset. https://github.com/droodev/Camp-Candy-Allocation-Dataset.

Kruskal, J. 1964. Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis. *Psychometrika*, 29(1): 1–27.

Kurokawa, D.; Procaccia, A. D.; and Wang, J. 2016. When Can the Maximin Share Guarantee Be Guaranteed? In *Proceedings of the 30th Conference on Artificial Intelligence (AAAI '16)*, 523–529.

Lipton, R. J.; Markakis, E.; Mossel, E.; and Saberi, A. 2004. On Approximately Fair Allocations of Indivisible Goods. In *Proceedings of the 5th ACM Conference on Economics and Computation (EC'04)*, 125–131.

Manurangsi, P.; and Suksompong, W. 2019. When Do Envy-Free Allocations Exist? In *Proceedings of the 33rd Conference on Artificial Intelligence (AAAI '19)*, 2109–2116.

Manurangsi, P.; and Suksompong, W. 2021. Closing Gaps in Asymptotic Fair Division. *SIAM Journal on Discrete Mathematics*, 35(2): 668–706.

Mattei, N.; and Walsh, T. 2013. PrefLib: A Library for Preferences. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT '13)*, 259–270.

Plaut, B.; and Roughgarden, T. 2020. Almost Envy-Freeness with General Valuations. *SIAM Journal on Discrete Mathematics*, 34(2): 1039–1068.

Roughgarden, T. 2020. Distributional Analysis. In Roughgarden, T., ed., *Beyond the Worst-Case Analysis of Algorithms*, 167–188. Cambridge University Press.

Shafer, W.; and Sonnenschein, H. 1982. Market Demand and Excess Demand Functions. In Arrow, K. J.; and Intriligator, M. D., eds., *Handbook of Mathematical Economics*, volume 2, 671–693. Elsevier.

Szufa, S.; Faliszewski, P.; Janeczko, L.; Lackner, M.; Slinko, A.; Sornat, K.; and Talmon, N. 2022. How to Sample Approval Elections? In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI '22)*, 496–502.

Szufa, S.; Faliszewski, P.; Skowron, P.; Slinko, A.; and Talmon, N. 2020. Drawing a Map of Elections in the Space of Statistical Cultures. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '20)*, 1341–1349.

Varian, H. R. 1974. Equity, Envy, and Efficiency. *Journal of Economic Theory*, 9(1): 63–91.