Journal of Artificial Intelligence Research 79 (2024) 1113-1166 Submitted 07/2023; published 04/2024

A Map of Diverse Synthetic Stable Matching Instances

Niclas Boehmer NBOEHMER@G.HARVARD.EDU
Klaus Heeger HEEGER@POST.BGU.AC.IL
Technische Universität Berlin, Berlin, Germany

Stanisław Szufa SZUFA@AGH.EDU.PL
AGH University, Kraków, Poland

Abstract

Focusing on Stable Roommates (SR), we contribute to the toolbox for conducting experiments for stable matching problems. We introduce the polynomial-time computable mutual attraction distance to measure the similarity of SR instances, analyze its properties, and use it to create a map of SR instances. This map visualizes 460 synthetic SR instances (each sampled from one of ten different statistical cultures) as follows: Each instance is a point in the plane, and two points are close on the map if the corresponding SR instances are similar with respect to our mutual attraction distance to each other. Subsequently, we conduct several illustrative experiments and depict their results on the map, illustrating the map's usefulness as a non-aggregate visualization tool, the diversity of our generated dataset, and the need to use instances sampled from different statistical cultures. Lastly, we extend our approach to the bipartite Stable Marriage problem.

1. Introduction

Since their introduction by Gale and Shapley (1962), stable matching problems have been extensively studied, both from a theoretical and a practical viewpoint. Numerous practical applications have been identified, and theoretical research has influenced the design of real-world matching systems (Knuth, 1976; Gusfield & Irving, 1989; Manlove, 2013). In addition to the rich theoretical literature, there are also several works containing empirical investigations of stable matching problems.[1] Although the examples given in Footnote 1 indicate that experimental works regularly occur, many papers on stable matchings do not include an experimental part and instead solely focus on the computational or axiomatic aspects of some mechanism or problem. However, to understand the properties of problems and mechanisms in practice, computational experiments are vital. Note that in this paper, we focus on computational experiments (also known as numerical experiments or simulations). This explicitly excludes other forms of experiments such as lab or live experiments.

1. We give a certainly incomplete list of experimental works here: Teo & Sethuraman, 2000; Mertens, 2015; Genc, Siala, Simonin, & O'Sullivan, 2019; Erdem, Fidan, Manlove, & Prosser, 2020 conducted experiments on general one-to-one matchings known as Stable Roommates, Podhradsky, 2010; Manne, Naim, Lerring, & Halappanavar, 2016; Genc, Siala, O'Sullivan, & Simonin, 2017; Cooper & Manlove, 2019; Delorme, García, Gondzio, Kalcsics, Manlove, & Pettersson, 2019; Cooper & Manlove, 2020; Tziavelis, Giannakopoulos, Johansen, Doka, Koziris, & Karras, 2020; Pettersson, Delorme, García, Gondzio, Kalcsics, & Manlove, 2021; Agarwal & Cole, 2023; Boehmer, Heeger, & Niedermeier, 2022b; Brilliantova & Hosseini, 2022 on bipartite one-to-one matchings known as Stable Marriage, Irving & Manlove, 2009; Kwanashie & Manlove, 2013; Delorme et al., 2019; Pettersson et al., 2021; Manlove, Milne, & Olaosebikan, 2022 on bipartite many-to-one matchings known as Hospital Residents; and Siala & O'Sullivan, 2017 on the bipartite many-to-many problem.

Examples of computational experiments include the evaluation of the running time of an algorithm or the quantitative analysis of the properties of a computed matching.

One reason for the lack of experimental work might be the rarity of real-world data (exceptions can be found in Irving & Manlove, 2009; Delorme et al., 2019; Manlove et al., 2022). As a result of this general lack, researchers typically resort to some distribution, known as a statistical culture or synthetic model, for generating synthetic data. Remarkably, the vast majority of previous works simply use random preferences, where all possible valid preferences are sampled with the same probability (out of the 19 works listed above, 17 use this model, nine of them as a single data source). However, as we will see later, instances where preferences are sampled uniformly at random share many properties. Accordingly, conclusions drawn from experiments using only such instances (or, generally speaking, only instances sampled from one model) should be treated with caution, as it is unclear whether their results generalize.

Similar in spirit to the line of works on "maps of elections" started by Szufa, Faliszewski, Skowron, Slinko, and Talmon (2020) in the context of voting, we want to lay the foundation for more experimental work around stable matchings by contributing to the toolbox for conducting computational experiments in various ways. Among other things, we introduce a polynomial-time computable measure for the similarity of instances and create a diverse synthetic dataset for testing together with a convenient framework to visualize and analyze it as a map (see Figure 2 for an example). We focus on instances of the Stable Roommates (SR) problem, where we have a set of agents, and each agent has strict preferences over all other agents. We selected the SR problem for this first, illustrative study because it is the mathematically most natural stable matching problem (agents' preferences do not contain ties and are complete, and there are no different "types" of agents). Consequently, statistical cultures for SR instances are relatively simple. Nevertheless, our general approach and several of our ideas and techniques can also be used to carry out similar studies for other stable matching problems, as demonstrated in Section 7, where we describe how to adapt our results to Stable Marriage (SM) instances (SM is the bipartite analog of SR).

As part of our agenda to empower experimental work on stable matchings, we carry out the following steps:

Distances Between SR Instances (Section 3). To judge the diversity of a dataset for testing and to compare different statistical cultures to each other, a similarity measure of SR instances is needed. For this, we introduce the notion of isomorphism between SR instances and show how distances between preference orders naturally extend to distances between SR instances. Most importantly, we propose the polynomial-time computable mutual attraction distance[2], which we use in the following.

Understanding the Space of SR Instances (Section 4). To better understand the space of SR instances induced by our mutual attraction distance, we introduce four canonical "extreme" instances, which are far away from each other. Moreover, we prove that two of them form a diameter of our space, i.e., they are at the maximum possible distance.

2. Note that we use the terms "distance (measure)" in an informal sense to refer to some function mapping pairs of instances to a positive real number; in particular, all our distance measures are pseudometrics but not all are metrics (see Proposition 11).

A Map of Stable Roommates Instances (Section 5). We define multiple statistical cultures to generate SR instances, striving to put together a diverse dataset for testing that shows a large range of possible behaviors. From them, we generate a diverse test set for experimental work and picture it as a map of SR instances, by first computing the mutual attraction distance between each pair of instances, and subsequently embedding the instances as points in the Euclidean plane such that the Euclidean distances between points resemble the underlying distances between instances as accurately as possible (see Figure 2 for an example). The use cases of the map are multifaceted. For instance, by observing the position of an instance on the map, we can learn something about its nature. Moreover, by coloring points according to the outcome of some experiment on the respective instance, the map can be used as a convenient framework to visualize non-aggregate experimental results, which may also help with designing more focused, follow-up experiments. In this section, we also give intuitive interpretations of different areas on the map. In addition, we analyze where different statistical cultures land on the map and how they relate to each other.

Using the Map (Section 6). To demonstrate possible use cases for the map, we perform some experimental studies as examples and use the map as a visualization tool. We analyze a quality measure for stable matchings, the number of blocking pairs for random matchings, and the running time to compute an "optimal" stable matching using an Integer Linear Program (ILP). In sum, the instance-based view on experimental results provided by the map allows us to identify several important phenomena, for example, that instances sampled from the same culture all behave very similarly in our experiments; an observation that has been neglected in the experimental design of many previous papers. Moreover, we further observe that instances from the same area of the map exhibit a similar behavior, which justifies the distance we use.

A Map of Stable Marriage Instances (Section 7). To demonstrate the general applicability of our framework to draw maps of stable matching instances, we create a map of Stable Marriage (SM) instances—SM is the bipartite analog of SR. For this, we describe how to transfer the mutual attraction distance, extreme instances, and statistical cultures from the SR to the SM setting. Finally, we illustrate the usefulness of the map of SM instances and verify that instances that are close to each other on the map have similar properties by again conducting some example experiments.

From a methodological perspective, our work follows a series of recent papers on (ordinal) elections[3] (Faliszewski, Skowron, Slinko, Szufa, & Talmon, 2019; Szufa et al., 2020; Boehmer, Bredereck, Faliszewski, Niedermeier, & Szufa, 2021b; Boehmer, Bredereck, Elkind, Faliszewski, & Szufa, 2022; Boehmer, Faliszewski, Niedermeier, Szufa, & Was, 2022a): Faliszewski et al. (2019) introduced the problem of computing the distance between elections, focusing on isomorphic distances. Following up on this, Szufa et al. (2020) created a dataset of synthetic elections sampled from a variety of different cultures and visualized them as a map of elections. Subsequently, Boehmer et al. (2021b) added several canonical elections to the map to give absolute positions a clearer meaning and added

3. Note that the term "election" is used in a formal sense here to describe a mathematical object defined by a set of candidates and a set of voters with strict preferences over candidates.

some real-world elections. Recently, Szufa, Faliszewski, Janeczko, Lackner, Slinko, Sornat, and Talmon (2022) created and analyzed a map of approval elections. The usefulness of the maps has already been demonstrated in different contexts. For example, Szufa et al. (2020) identified that for elections from a certain region of the map, election winners are particularly hard to compute, Boehmer et al. (2021b) and Boehmer and Schaar (2023) analyzed the nature and relationship of real-world elections by placing them on the map, and Boehmer, Bredereck, Faliszewski, and Niedermeier (2021a) evaluated the robustness of election winners using the map. Although our general agenda and approach are similar to the works of Faliszewski et al. (2019), Szufa et al. (2020) and Boehmer et al. (2021b), the intermediate steps, used distance measures, cultures, experiments, and technical details are naturally quite different.

The code for generating the map and conducting our experiments is available at `https://github.com/szufix/mapel`. The generated datasets of SR and SM instances are available at `https://github.com/szufix/mapel_data`.

2. Preliminaries

Preference Orders. Let A be a finite set of *agents*. We denote by $\mathcal{L}(A)$ the set of all strict and total orders over A which we call *preference orders*. We usually denote elements of $\mathcal{L}(A)$ as $\succ$ and for two agents a and b, we say that *a is preferred to b* if $a \succ b$. We sometimes specify the preferences of some agent a by writing $a : b \succ c \succ d$ to denote that a prefers b to c to d. Moreover, for a preference order $\succ \in \mathcal{L}(A)$ and an agent $a \in A$, let $\mathrm{pos}_{\succ}(a)$ denote the *position of a in $\succ$*, i.e., the number of agents that are preferred to a in $\succ$ plus 1. Furthermore, for $i \in [|A|]$, let $\mathrm{ag}_{\succ}(i)$ be the agent ranked in i-th position in $\succ$, i.e., the agent $b \in A$ such that $i = \mathrm{pos}_{\succ}(b)$.

Distances between Preference Orders. For two preference orders $\succ, \succ' \in \mathcal{L}(A)$, their *swap distance* $\mathrm{swap}(\succ, \succ')$ is the number of agent pairs on whose ordering $\succ$ and $\succ'$ disagree. For two preference orders $\succ, \succ' \in \mathcal{L}(A)$, their *Spearman distance* $\mathrm{spear}(\succ, \succ')$ is $\sum_{a \in A} |\mathrm{pos}_{\succ}(a) - \mathrm{pos}_{\succ'}(a)|$. As proven by Diaconis and Graham (1977), it holds that $\mathrm{swap}(\succ, \succ') \leq \mathrm{spear}(\succ, \succ') \leq 2 \cdot \mathrm{swap}(\succ, \succ')$.

Stable Roommates. A Stable Roommates (SR) instance $\mathcal{I}$ consists of a set A of agents, with each agent $a \in A$ having a preference order $\succ_a \in \mathcal{L}(A \setminus \{a\})$ over all other agents. For simplicity, we will focus on instances with an even number of agents, as otherwise, stable matchings leave one agent unmatched.

(Stable) Matchings. A *matching* of agents A is a subset of all possible agent pairs $\{a, a'\}$ with $a \neq a' \in A$ where each agent appears in at most one pair. We say that an agent is *unmatched* in a matching M if a does not appear in any pair from M; otherwise, we say that a is *matched*. For a matched agent $a \in A$ and a matching M, we write $M(a)$ to denote the *partner of a in M*, i.e., $M(a) = a'$ if $\{a, a'\} \in M$. A pair $\{a, a'\}$ of agents *blocks* a matching M if it simultaneously hold that (i) a is unmatched or prefers a' to $M(a)$ and (ii) a' is unmatched or prefers a to $M(a')$. A matching that is not blocked by any agent pair is called a *stable matching*.

Mapping of Instances. For two sets X and Y with $|X| = |Y|$, we denote by $\Pi(X,Y)$ the set of all bijections $\sigma\colon X \to Y$ between X and Y. Let A and A' be two sets of agents with $|A| = |A'|$ and let $\sigma \in \Pi(A, A')$. Then, for an agent $a \in A$ and a preference order $\succ_a \in \mathcal{L}(A \setminus \{a\})$, we write $\sigma(\succ_a)$ to denote the preference order over $A' \setminus \{\sigma(a)\}$ arising from $\succ_a$ by replacing each agent $b \in A \setminus \{a\}$ by $\sigma(b) \in A' \setminus \{\sigma(a)\}$.

Pearson Correlation Coefficient. In our experiments, to evaluate the correlation between two measures, we use the Pearson Correlation Coefficient (PCC). The PCC is a measure of linear correlation between two quantities x and y. A Pearson correlation coefficient of 1 means that x and y are perfectly positively linearly correlated, i.e., it always holds $y = mx + b$ for some $b \in \mathbb{R}$ and $m > 0$. A Pearson correlation coefficient of 0 indicates no linear correlation, and -1 describes a perfect negative correlation ($m < 0$). As established by Schober, Boer, and Schwarte (2018), an absolute Pearson correlation coefficient between 0.4 and 0.69 indicates a moderate correlation, a value between 0.7 and 0.89 indicates a strong correlation, and a value between 0.9 and 1 indicates a very strong correlation.

3. Distance Measures

This section is devoted to measuring the distance between two SR instances, a key ingredient of our map. Other use cases include the meaningful selection of test instances, the comparison of different statistical cultures, and the analysis of real-world instances. In Section 3.1, we define an isomorphism between two SR instances, show how distance measures over preference orders can be generalized to distance measures over SR instances, and prove that computing the Spearman and swap distance between SR instances is computationally intractable. In Section 3.2, we introduce our mutual attraction distance and make some observations concerning its properties and the associated mutual attraction matrices.

3.1 Isomorphism and Isomorphic Distances

Two SR instances are isomorphic if renaming the agents in one instance can produce the other instance. For this, as each agent is associated with a preference order defined over other agents, a single mapping describing a renaming of agents suffices. Accordingly, we define an isomorphism on SR instances:

Definition 1. *Two SR instances $(A, (\succ_a)_{a\in A})$ and $(A', (\succ_{a'})_{a'\in A'})$ with $|A| = |A'|$ are* isomorphic *if there is a bijection $\sigma \in \Pi(A, A')$ such that $\succ_{\sigma(a)} = \sigma(\succ_a)$ for all $a \in A$.*

We give an example for two isomorphic SR instances:

Example 2. *Let $\mathcal{I}$ with agents $\{a, b, c, d\}$ and $\mathcal{I}'$ with agents $\{x, y, z, w\}$ be two SR instances with the following preferences:*

$$a : b \succ c \succ d, \qquad b : c \succ a \succ d, \qquad c : b \succ d \succ a, \qquad d : a \succ c \succ b,$$
$$x : y \succ w \succ z, \qquad y : z \succ w \succ x, \qquad z : w \succ y \succ x, \qquad w : z \succ x \succ y.$$

$\mathcal{I}$ and $\mathcal{I}'$ are isomorphic as witnessed by the mapping $\sigma(a) = y$, $\sigma(b) = z$, $\sigma(c) = w$, and $\sigma(d) = x$.

One can easily check whether two SR instances $(A, (\succ_a)_{a\in A})$ and $(A', (\succ_{a'})_{a'\in A'})$ are isomorphic: Let us assume that we know that some isomorphism $\sigma^* \in \Pi(A, A')$ maps $a \in A$ to $a' \in A'$. Then, this already completely characterizes σ^*, as for each agent $b \in A \setminus \{a\}$ with $\mathrm{pos}_{\succ_a}(b) = i$, we must have $\sigma^*(b) = \mathrm{ag}_{\succ_{a'}}(i)$. Thus, it suffices to fix an arbitrary agent $a \in A$ and then check for each $a' \in A'$ whether the resulting mapping σ^* is an isomorphism.

Observation 3. *Deciding whether two SR instances with $2n$ agents are isomorphic can be done in $\mathcal{O}(n^3)$ time.*

For any distance measure p between preference orders, our notion of isomorphism can be easily used to extend p to a distance measure over SR instances: The resulting distance between two SR instances is the minimum (over all bijections σ between the agent sets) sum (over all agents) of the distance between the preferences of a and the preferences of $\sigma(a)$ (measured by p):

Definition 4. *Let p be a distance measure between preference orders. Let $\mathcal{I} = (A, (\succ_a)_{a\in A})$ and $\mathcal{I}' = (A', (\succ_{a'})_{a'\in A'})$ be two SR instances with $|A| = |A'|$. Their d_p-distance is: $d_p(\mathcal{I}, \mathcal{I}') := \min_{\sigma\in\Pi(A,A')} \sum_{a\in A} p(\sigma(\succ_a), \succ_{\sigma(a)})$.*

In particular, for all distance measures p between preference orders where, for $\succ, \succ' \in \mathcal{L}(A)$, $p(\succ, \succ') = 0$ if and only if $\succ = \succ'$, for any two SR instances $\mathcal{I}$ and $\mathcal{I}'$ it holds that $d_p(\mathcal{I}, \mathcal{I}') = 0$ if and only if $\mathcal{I}$ and $\mathcal{I}'$ are isomorphic. We call such a distance an *isomorphic distance.*

Example 5. *Applying Definition 4, the Spearman distance* $\mathrm{spear}(\cdot, \cdot)$ *and the swap distance* $\mathrm{swap}(\cdot, \cdot)$ *between preference orders (as defined in Section 2) can be lifted to distance measures d_{spear} and d_{swap} between SR instances. Let $\mathcal{I}$ with agents a, b, c, and d and $\mathcal{I}'$ with agents x, y, z, and w be two SR instances with the following preferences:*

$$a : b \succ c \succ d, \quad b : a \succ c \succ d, \quad c : a \succ b \succ d, \quad d : a \succ b \succ c,$$
$$x : y \succ z \succ w, \; y : x \succ z \succ w, \; z : w \succ y \succ x, \; w : z \succ y \succ x.$$

Then, for the mapping $\sigma(a) = x$, $\sigma(b) = y$, $\sigma(c) = z$, and $\sigma(d) = w$, the Spearman distance of $\mathcal{I}$ and $\mathcal{I}'$ is 8 and the swap distance is 6. While for the Spearman distance this is the optimal mapping (so $d_{\mathrm{spear}}(\mathcal{I}, \mathcal{I}') = 8$) for the swap distance the mapping $\sigma(a) = y$, $\sigma(b) = x$, $\sigma(c) = z$, and $\sigma(d) = w$ results in a smaller distance of 4. Indeed, we have $d_{\mathrm{swap}}(\mathcal{I}, \mathcal{I}') = 4$.

We consider the Spearman distance d_{spear} and the swap distance d_{swap} as "ideal" distances, as they are quite fine-grained and isomorphic. Unfortunately, we will show in the following that both are hard to compute. We first show that computing the Spearman distance between two SR instances is at least as hard as deciding whether two graphs are isomorphic, which is a famous candidate for the complexity class NP-intermediate.

Proposition 6. *There is no polynomial-time algorithm to compute d_{spear}, unless* Graph Isomorphism *is in P.*

Proof. For a graph $G = (V, E)$ and a vertex $v \in V$, let $N_G(v)$ be the set of vertices adjacent to v in G. In the Graph Isomorphism problem, we are given two graphs $G = (V, E)$ and $G' = (V', E')$ with $|V| = |V'|$ and the question is whether there is a bijection $\mu\colon V \to V'$ such that $\{v, u\} \in E$ if and only if $\{\mu(v), \mu(u)\} \in E'$. We will now reduce Graph Isomorphism to the problem of computing d_{spear}.

Construction. Given an instance $(G = (V, E), G' = (V', E'))$ of Graph Isomorphism, we construct two SR instances as follows. Without loss of generality, we assume that there are no isolated vertices in G and G' and that $\nu := |V| = |V'| > 2$. From G, we construct an SR instance $\mathcal{I}$ with agent set A as follows: First, we add each vertex $v \in V$ as an agent to A. Moreover, we add a set D of ν^4 dummy agents. We now describe the preferences of the agents. In order to do so, we denote, for a set B of agents by $[B]$ an arbitrary but fixed strict and total order of agents from B. For an agent $b \in B$, we denote by $[B] \setminus \{b\}$ the order arising from $[B]$ by removing b. The preferences of the agents are as follows:

$$\begin{aligned} &v : [N_G(v)] \succ [D] \succ [V \setminus (N_G(v) \cup \{v\})] && \forall v \in V \\ &d : [D] \setminus \{d\} \succ [V] && \forall d \in D \end{aligned}$$

From $G' = (\{v'_1, \dots, v'_\nu\}, E')$, we construct the second SR instance $\mathcal{I}'$ with agent set A'. We add each vertex $v' \in V'$ as an agent to A' and for each $i \in [\nu]$ a set D'_i of ν^3 dummy agents. We set $D' := \bigcup_{i \in [\nu]} D'_i$. The preferences of the agents are as follows:

$$\begin{aligned} &v' : [N_{G'}(v')] \succ [D'] \succ [V' \setminus (N_{G'}(v') \cup \{v'\})] && \forall v' \in V' \\ &d' : [D'] \setminus \{d'\} \succ v'_i \succ v'_{i+1} \succ v'_{i+2} \succ \dots \\ &\qquad \succ v'_\nu \succ v'_1 \succ \dots \succ v'_{i-1} && \forall d' \in D'_i, \forall i \in [\nu] \end{aligned}$$

We now prove that the given Graph Isomorphism instance is a yes-instance if and only if $d_{\text{spear}}(\mathcal{I}, \mathcal{I}') \leq \nu^3 \cdot (\sum_{j \in [\nu]} \sum_{i \in [\nu]} |i - j|) + \nu^3 = \frac{1}{3}\nu^4(\nu^2 - 1) + \nu^3$.

Proof of Correctness. ($\Rightarrow$) Let $\pi\colon D \to D'$ be the mapping that maps for $i \in [\nu^4]$ the dummy agent ranked in position i in $[D]$ to the dummy agent ranked in position i in $[D']$. Assume that G and G' are isomorphic as witnessed by the bijection $\mu\colon V \to V'$. Then, we construct a bijection $\sigma\colon A \to A'$ by mapping v to $\mu(v)$ for all $v \in V$ and d to $\pi(d)$ for all $d \in D$. To estimate $d_{\text{spear}}(\mathcal{I}, \mathcal{I}')$, we start by upper-bounding the Spearman distance between $\sigma(\succ_v)$ and $\succ_{\sigma(v)}$ for $v \in V$. As μ is an isomorphism between G and G', we have that $\{\sigma(w) \mid w \in N_G(v)\} = N_{G'}(\sigma(v))$. Moreover, we have $\{\sigma(w) \mid w \in V \setminus (N_G(v) \cup \{v\})\} = V' \setminus (N'_G(\sigma(v)) \cup \{\sigma(v)\})$. Thus, in $\sigma(\succ_v)$ the same agents from V' appear before the first dummy agent as in $\succ_{\sigma(v)}$ and the same agents from V' appear after the last dummy agent. Moreover, note that all dummy agents are ranked in the same position in the two preference orders. Thus, we can upper bound $\text{spear}(\sigma(\succ_v), \succ_{\sigma(v)}) \leq \nu^2$: For each of the $\nu - 1$ agents from $V' \setminus \{\sigma(v)\}$ their position in the two preference orders can differ by at most ν, since in both preference orders the same at most ν agents appear before the first dummy agent and the same at most ν agents appear after the last dummy agent. Overall, we get that

$$\sum_{v \in V} \text{spear}(\sigma(\succ_v), \succ_{\sigma(v)}) \leq \nu^3 . \tag{1}$$

Turning to the preferences of the dummy agents, note that for each $d \in D$, d ranks all dummy agents in the same position as $\sigma(d)$. Thus, only the different ordering of the agents from V' in $\sigma(\succ_d)$ and $\succ_{\sigma(d)}$ contribute to the Spearman distance between the two. Observe that for each two $b, d \in D$, each agent from V' appears in the same position in $\sigma(\succ_b)$ as in $\sigma(\succ_d)$. Moreover, observe that considering the preference orders of all agents from D', each agent $v' \in V'$ appears exactly ν^3 times in position $(\nu^4 - 1) + i$ for each $i \in [\nu]$.

Let us now focus on agent $v' := \sigma(v) \in V'$ where v is ranked in position $(\nu^4 - 1) + j$ for $j \in [\nu]$ by $\succ_d$ for each $d \in D$. Then v' contributes $|j - i|$ to $\mathrm{spear}(\sigma(\succ_d), \succ_{\sigma(d)})$ for each $d \in D$ where v' is ranked in position $(\nu^4 - 1) + i$ in $\succ_{\sigma(d)}$. Together with our previous observation that each vertex agent appears ν^3 times in position $(\nu^4 - 1) + i$ for each $i \in [\nu]$ in the preferences of agents from D' this implies that the misplacement of agent v' overall contributes $\nu^3 \cdot (\sum_{i\in[\nu]} |i - j|)$ to the Spearman distance between the mapped preference orders of dummy agents. Summing up over all $j \in [\nu]$, we get that the total Spearman distance between the mapped preference orders of dummy agents is $\nu^3 \cdot \sum_{j\in[\nu]} \sum_{i\in[\nu]} |i - j|$. Combining this with Equation (1), we get that $d_{\mathrm{spear}}(\mathcal{I}, \mathcal{I}') \leq \nu^3 \cdot (\sum_{j\in[\nu]} \sum_{i\in[\nu]} |i - j|) + \nu^3$, which proves the forward direction.

($\Leftarrow$) Let $\sigma\colon A \to A'$ witness $d_{\mathrm{spear}}(\mathcal{I}, \mathcal{I}') \leq \frac{1}{3}\nu^4(\nu^2 - 1) + \nu^3$.

We first show that σ does not map any agent from D to an agent from V'. To show this, let $X \subseteq D$ be the subset consisting of all agents from D which are mapped to agents from V' in σ and $Y' := \sigma(X) \subseteq V'$ be the subset of agents from V' to which an agent from $X \subseteq D$ is mapped in σ. Assume for the sake of contradiction that $x := |X| = |Y'| > 0$. Now we compute the summed distance between the preferences of the agents from $D \setminus X$ and the preferences of the agents they are mapped to in σ and show that this distance already exceeds the given budget. In particular, we give a lower bound on

$$\begin{aligned}\sum_{d\in D\setminus X} \mathrm{spear}(\sigma(\succ_d), \succ_{\sigma(d)}) \geq & \sum_{d\in D\setminus X} \sum_{v'\in V'\setminus Y'} |\mathrm{pos}_{\succ_d}(\sigma^{-1}(v')) - \mathrm{pos}_{\sigma(\succ_d)}(v')| \\ & + \sum_{d\in D\setminus X} \sum_{y'\in Y'} |\mathrm{pos}_{\succ_d}(\sigma^{-1}(y')) - \mathrm{pos}_{\sigma(\succ_d)}(y')| \,. \end{aligned} \tag{2}$$

We first give a lower bound on the first summand, i.e., $\sum_{d\in D\setminus X} \sum_{v'\in V'\setminus Y'} |\mathrm{pos}_{\succ_d}(\sigma^{-1}(v')) - \mathrm{pos}_{\sigma(\succ_d)}(v')|$. Note that for each two agents $b, d \in D \setminus X$ and each $v' \in V' \setminus Y'$ it holds that $\sigma(\succ_b)$ and $\sigma(\succ_d)$ rank v' in the same position. In other words, for each agent $v' \in V' \setminus Y'$, there is some $j \in [\nu]$ such that v' is ranked in position $(\nu^4 - 1) + j$ in $\sigma(\succ_d)$ for each $d \in D \setminus X$. We define $\mathrm{id}(v') := j$. Then, for each $d \in D \setminus X$ with v' being ranked in position $(\nu^4 - 1) + i$ in $\succ_{\sigma(d)}$, agent v' contributes $|\mathrm{id}(v') - i|$ to $\mathrm{spear}(\sigma(\succ_d), \succ_{\sigma(d)})$. From this together with the facts that there are ν^3 agents from D' ranking v' in position i for each $i \in [\nu]$, $|X| = x$, and $|\mathrm{id}(v') - i| \leq \nu$, we get that

$$\sum_{d\in D\setminus X} |\mathrm{pos}_{\succ_d}(\sigma^{-1}(v')) - \mathrm{pos}_{\sigma(\succ_d)}(v')| \geq \nu^3 \cdot \big(\sum_{i\in[\nu]} |i - \mathrm{id}(v')|\big) - x\nu \,. \tag{3}$$

We now turn to the second summand of Equation (2), i.e., $\sum_{d\in D\setminus X} \sum_{y'\in Y'} |\mathrm{pos}_{\succ_d}(\sigma^{-1}(y')) - \mathrm{pos}_{\sigma(\succ_d)}(y')|$. Note that for each agent $y' \in Y'$, we have that it is placed in position $\nu^4 - i \leq \nu^4 - 1$ for some $i \in [\nu^4 - 1]$ in $\sigma(\succ_d)$ for all $d \in D \setminus X$, as a dummy agent is mapped to y' and dummy agents appear only in the first $\nu^4 - 1$ positions in $\sigma(\succ_d)$. Thus, as there are ν^3 agents in D' that rank y' in position $(\nu^4 - 1) + j$ for each $j \in [\nu]$ and as $|X| = x$, we get that

$$\begin{aligned}&\sum_{d\in D\setminus X} |\mathrm{pos}_{\succ_d}(\sigma^{-1}(y')) - \mathrm{pos}_{\sigma(\succ_d)}(y')|\\ &\geq -x\nu + \nu^3 \sum_{j=1}^{\nu} j = \nu^3 \cdot \frac{\nu\cdot(\nu+1)}{2} - x\nu = \nu^4 + \nu^3\frac{\nu^2-\nu}{2} - x\nu\,. \end{aligned} \tag{4}$$

Summing Equations (3) and (4) over all $v' \in V'$ and using (for the second inequality) that $\sum_{i\in[\nu]} |i-j| \leq \frac{\nu^2-\nu}{2}$ for all $j \in [\nu]$, we get

$$\begin{aligned}&\sum_{d\in D\setminus X} \mathrm{spear}(\sigma(\succ_d), \succ_{\sigma(d)}) \geq\\ &\sum_{v'\in V'\setminus Y'} \Big(\nu^3 \cdot \big(\sum_{i\in[\nu]} |i - \mathrm{id}(v')|\big) - x\nu\Big) + \sum_{y'\in Y'} \Big(\nu^4 + \nu^3\frac{\nu^2-\nu}{2} - x\nu\Big)\\ &\geq |Y'|\cdot\nu^4 + \sum_{v'\in V'} \Big(\nu^3\cdot\big(\sum_{i\in[\nu]}|i-\mathrm{id}(v')|\big) - x\nu\Big)\\ &= x\nu^4 + \nu^3\cdot\big(\sum_{j\in[\nu]}\sum_{i\in[\nu]}|i-j|\big) - x\nu^2\\ &> \nu^3\cdot\big(\sum_{j\in[\nu]}\sum_{i\in[\nu]}|i-j|\big) + \nu^3,\end{aligned}$$

where we used our assumption $\nu > 2$ as well as $x > 0$ for the last inequality. Thus, we have reached a contradiction to σ witnessing a solution as the resulting distance exceeds the given upper bound, implying that $|X| = 0$. Consequently, we may assume in the following without loss of generality that σ matches dummy agents from D to dummy agents from D' and vertex agents from V to vertex agents from V' in σ.

Observe that the arguments given in the forward direction of the proof imply that independent of how σ maps vertex agents to vertex agents and dummy agents to dummy agents we have that $\sum_{d\in D} \mathrm{spear}(\sigma(\succ_d), \succ_{\sigma(d)}) \geq \frac{1}{3}\nu^4\cdot(\nu^2-1)$. Thus, it needs to hold that $\sum_{v\in V} \mathrm{spear}(\sigma(\succ_v), \succ_{\sigma(v)}) \leq \nu^3 < \nu^4$. As we have ν^4 dummy agents, this implies that for each $v \in V$, we need to have that $\sigma(\succ_v)$ and $\succ_{\sigma(v)}$ rank the same agents before the first dummy agent: The position difference of an agent that appears in one preference order before the dummy agents and in the other after the dummy agents would be at least ν^4, which is not possible. Thus, we have that $\{\sigma(w) \mid w \in N_G(v)\} = N_{G'}(\sigma(v))$. Thus, restricting the mapping σ to the agents from V leads to a mapping $\mu\colon V \to V'$ that induces an isomorphism from G to G'. $\square$

We leave it as an open question whether computing d_{spear} is indeed NP-hard. For d_{swap}, the NP-hardness follows from the NP-hardness of computing the Kemeny score of an election (Dwork, Kumar, Naor, & Sivakumar, 2001; Biedl, Brandenburg, & Deng, 2005):

Proposition 7. *Given two* SR *instances $\mathcal{I}$ and $\mathcal{I}'$ and an integer ℓ, deciding whether $d_{\mathrm{swap}}(\mathcal{I},\mathcal{I}') \leq \ell$ is NP-complete.*

Proof. In the NP-hard Kemeny Score problem, we are given an election $E = (C, V)$[4] and an integer k, and the question is whether there is a central order $v^* \in \mathcal{L}(C)$ such that $\sum_{v \in V} \mathrm{swap}(v, v^*) \leq k$ (Dwork et al., 2001).

Claim 8. Kemeny Score *is NP-complete even if* $|C| = |V|$ *and each candidate appears as the top choice of exactly one voter.*

Proof of Claim 8. We reduce from the Kemeny Score problem with four voters which was shown to be NP-hard by Dwork et al. (2001). Let $(E = (C, V), k)$ with $V = \{v_1, v_2, v_3, v_4\}$ be an instance of this problem (where we assume that C is even, as otherwise we can add a candidate and add it to all votes in the last position). We construct an instance $(E' = (C', V'), k')$ of Kemeny Score with candidate set is $C' := C \cup \{c_1, c_2, c_3, c_4\}$. We construct the set V' of voters as follows: As the first voter, we add to V' a voter whose preferences start with $c_1 \succ c_2 \succ c_3 \succ c_4$ followed by the candidates from C in the order in which they appear in v_1. We also insert such a voter for each of v_2, v_3, and v_4, whose preferences start with $c_2 \succ c_1 \succ c_3 \succ c_4$, $c_3 \succ c_1 \succ c_2 \succ c_4$, and $c_4 \succ c_1 \succ c_2 \succ c_3$, respectively. By adding $c_1 \succ c_2 \succ c_3 \succ c_4$ to the beginning of some central order over C, the summed distance of this central order to the four voters increases by 6.

Let $P \subseteq C \times C$ be an arbitrary partitioning of the candidates from P into pairs. For each $(c, c') \in P$, let $v \in \mathcal{L}(C')$ be an arbitrary vote starting with c and ending with c'. We add v to V' as well as a vote ranking the candidates in the opposite order as they are ranked in v. Each of these pairs has a summed swap distance of $\binom{|C'|}{2}$ to any central order. In the constructed election $E' = (C', V')$, we have $|C'| = |V'|$ and each voter has a different top-choice. With $k' := k + 6 + \frac{|C|}{2} \cdot \binom{|C'|}{2}$, it is easy to see that the two instances are equivalent. ♦

Given an instance $(E = (C, V), k)$ of the NP-hard restricted version of the Kemeny Score problem from Claim 8, let $\pi \colon C \to V$ be the bijection between candidates and voters where each candidate is mapped to the unique voter where this candidate appears in the first position.

We construct two SR instances $\mathcal{I}$ and $\mathcal{I}'$ with $|C|$ agents as follows. The first SR instance $\mathcal{I}$ has agent set $A = \{a_1, \cdots, a_{|C|}\}$. All agents from A rank the other agents increasingly by their index in their preferences. The second SR instance $\mathcal{I}'$ has agent set $A' = C$. Each agent $c \in A'$ orders the agents except for themselves as they are ordered in $\pi(c)$. We set $\ell = k - \sum_{i=0}^{|C|-1} i$.

($\Rightarrow$) Assume that there is a central order $v^* \in \mathcal{L}(C)$such that $\sum_{v \in V} \mathrm{swap}(v, v^*) \leq k$. Let σ be a bijection between the agents A and A', where agent $a_i \in A$ is mapped to agent $c \in A'$ if c appears in position i in v^*. Then for $i \in [|C|]$, we have that

$$\mathrm{swap}(\sigma(\succ_{a_i}), \succ_{\sigma(a_i)}) = \mathrm{swap}(v^*, \pi(\sigma(a_i))) - (i - 1),$$

as $\sigma(\succ_{a_i})$ is the central order v^* without $\sigma(a_i)$ and the preferences of $\succ_{\sigma(a_i)}$ is the vote $\pi(\sigma(a_i))$ without $\sigma(a_i)$, which appears in position i in v^* and in position one in $\pi(\sigma(a_i))$.

4. An election $E = (C, V)$ is defined by a set C of candidates and a set V of voters. Each voter $v \in V$ is identified with a preference order, also known as their vote, from $\mathcal{L}(C)$.

Consequently, we have:

$$\sum_{i\in[|C|]} \operatorname{swap}(\sigma(\succ_{a_i}), \succ_{\sigma(a_i)}) \leq k - \sum_{i=0}^{|C|-1} i.$$

($\Leftarrow$) Assume that there is a bijection between the agents A and A' with $\sum_{i\in[|C|]} \operatorname{swap}(\sigma(\succ_{a_i}), \succ_{\sigma(a_i)}) \leq k - \sum_{i=0}^{|C|-1} i$. Let $v^* \in \mathcal{L}(C)$ be a vote where candidate $c \in C$ appears in position i if a_i is mapped to c by σ. Then analgous to as argued above, for $i \in [|C|]$, we have that $\operatorname{swap}(v^*, \pi(\sigma(a_i))) = \operatorname{swap}(\sigma(\succ_{a_i}), \succ_{\sigma(a_i)}) + (i-1)$ and consequently that $\sum_{v\in V} \operatorname{swap}(v, v^*) = d_{\text{swap}}(\mathcal{I}, \mathcal{I}') + \sum_{i=0}^{|C|-1} i \leq k$. □

Note that in the context of voting, similar to our approach, Faliszewski et al. (2019) extended the Spearman and swap distances between preference orders to distances between elections. They proved that for both swap and Spearman, computing the distance between two elections is NP-hard and hard to approximate unless Graph Isomorphism is in P. While their reductions are from the same problems as ours, our constructions are quite different and more involved, as we no longer have both candidates and voters but just agents.

3.2 Mutual Attraction Distance

In this section, we introduce and discuss our main distance measure, which we call *mutual attraction distance.*

Intuition. One characteristic of SR instances is that each agent is associated with a preference order and also appears in the preference order of other agents. Thus, when considering, for instance, stable matchings, for an agent a it is not only important which agents a likes, but also whether they like a as well. Accordingly, our mutual attraction distance focuses on how pairs of agents rank each other. In particular, each agent a is characterized by a mutual attraction vector whose i-th entry contains the position in which a appears in the preferences of the agent who is ranked in i-th position by a. In the new mutual attraction distance, we match the agents from two different instances such that the ℓ_1-distance between the mutual attraction vectors of matched agents is minimized.

Notation. For $p, q \in \mathbb{N}$, some $i \in [p]$, and a matrix $M \in \mathbb{N}^{p\times q}$, let M_i denote the i-th row of M. For an SR instance $\mathcal{I} = (A = \{a_1, \dots a_{2n}\}, (\succ_a)_{a\in A})$, an agent $a \in A$, and some $i \in [2n-1]$, let $\mathcal{MA}^{\mathcal{I}}(a, i)$ be the position of a in the preference order of the agent a' that a ranks in position i, i.e., $\mathcal{MA}^{\mathcal{I}}(a, i) := \operatorname{pos}_{\succ_{a'}}(a)$ where $a' := \operatorname{ag}_{\succ_a}(i)$. The *mutual attraction vector* of agent a is $\mathcal{MA}^{\mathcal{I}}(a) = \big(\mathcal{MA}^{\mathcal{I}}(a, 1), \dots, \mathcal{MA}^{\mathcal{I}}(a, 2n-1)\big)$. Lastly, the *mutual attraction matrix* $\mathcal{MA}^{\mathcal{I}}$ of $\mathcal{I}$ is the matrix whose i-th row is the vector $\mathcal{MA}^{\mathcal{I}}(a_i)$.

Definition 9. *For two mutual attraction matrices $\mathcal{MA}^{\mathcal{I}}$ and $\mathcal{MA}^{\mathcal{I}'}$ of SR instances $\mathcal{I}$ and $\mathcal{I}'$ on $2n$ agents, we define their* mutual attraction distance *as*

$$\mathrm{d}_{\mathrm{MAD}}(\mathcal{MA}^{\mathcal{I}}, \mathcal{MA}^{\mathcal{I}'}) := \min_{\sigma\in\Pi([2n],[2n])} \sum_{i\in[2n]} \ell_1\big(\mathcal{MA}^{\mathcal{I}}_i, \mathcal{MA}^{\mathcal{I}'}_{\sigma(i)}\big).$$

The mutual attraction distance $\mathrm{d_{MAD}}(\mathcal{I},\mathcal{I}')$ *between two SR instances* $\mathcal{I}$ *with agents* A *and* $\mathcal{I}'$ *with agents* A' *with* $|A| = |A'|$ *is the mutual attraction distance of their mutual attraction matrices.*

Example 10. *Consider the two SR instances* $\mathcal{I}$ *and* $\mathcal{I}'$ *defined in Example 5. Their mutual attraction matrices are:*

$$\mathcal{MA}^{\mathcal{I}} = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 2 & 3 \\ 3 & 3 & 3 \end{bmatrix} \end{array}, \quad \mathcal{MA}^{\mathcal{I}'} = \begin{array}{c} \\ x \\ y \\ z \\ w \end{array}\!\!\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 3 & 3 \\ 1 & 2 & 2 \\ 1 & 2 & 2 \\ 1 & 3 & 3 \end{bmatrix} \end{array}$$

Their mutual attraction distance is $2+0+2+2 = 6$ *as witnessed by the mapping* $\sigma(a) = z$, $\sigma(b) = y$, $\sigma(c) = x$, *and* $\sigma(d) = w$.

Pseudometric. We now show that the mutual attraction distance is a pseudometric (note that we will prove in Observation 14 that two non-isomorphic instances can be at mutual attraction distance zero, which implies that it is not a metric).

Proposition 11. *The mutual attraction distance* $\mathrm{d_{MAD}}$ *is a pseudometric.*

Proof. For two SR instances $\mathcal{I}$ and $\mathcal{I}'$, it trivially holds that $\mathrm{d_{MAD}}(\mathcal{I},\mathcal{I}) = 0$ and $\mathrm{d_{MAD}}(\mathcal{I},\mathcal{I}') = \mathrm{d_{MAD}}(\mathcal{I}',\mathcal{I})$, so it remains to verify the triangle inequality. Consider three SR instances $\mathcal{I}_1$, $\mathcal{I}_2$, and $\mathcal{I}_3$ with $2n$ agents. Let δ and σ be the matchings that minimize the mutual attraction distances between $\mathcal{I}_1$ and $\mathcal{I}_2$ and between $\mathcal{I}_2$ and $\mathcal{I}_3$, respectively. We have that:

$$\begin{aligned} \mathrm{d_{MAD}}(\mathcal{I}_1,\mathcal{I}_3) &\leq \textstyle\sum_{i\in[2n]} \ell_1\big(\mathcal{MA}^{\mathcal{I}_1}_i, \mathcal{MA}^{\mathcal{I}_3}_{\sigma(\delta(i))}\big) \\ &\leq \textstyle\sum_{i\in[2n]} \ell_1\big(\mathcal{MA}^{\mathcal{I}_1}_i, \mathcal{MA}^{\mathcal{I}_2}_{\delta(i)}\big) + \sum_{i\in[2n]} \ell_1\big(\mathcal{MA}^{\mathcal{I}_2}_{\delta(i)}, \mathcal{MA}^{\mathcal{I}_3}_{\sigma(\delta(i))}\big) \\ &= \mathrm{d_{MAD}}(\mathcal{I}_1,\mathcal{I}_2) + \mathrm{d_{MAD}}(\mathcal{I}_2,\mathcal{I}_3), \end{aligned}$$

where the second inequality follows from the fact that the ℓ_1-distance fulfills the triangle inequality. $\square$

Computation. Given two SR instances $\mathcal{I}$ over agents A and $\mathcal{I}'$ over agents A' with $|A| = |A'|$, computing their mutual attraction distance reduces to finding a minimum-weight perfect matching in a complete bipartite graph $G = (A \cup A', E)$ where edge $\{a, a'\} \in E$ has weight $\ell_1(\mathcal{MA}^{\mathcal{I}}(a), \mathcal{MA}^{\mathcal{I}'}(a'))$.

Observation 12. *Given two SR instances* $\mathcal{I}$ *and* $\mathcal{I}'$ *with* $2n$ *agents each,* $\mathrm{d_{MAD}}(\mathcal{I},\mathcal{I}')$ *can be computed in* $\mathcal{O}(n^3)$ *time.*

Unsuitability of Positionwise Distance. The papers of Szufa et al. (2020) and Boehmer et al. (2021b) on the map of elections used a different distance measure defined over the so-called position matrices. In a position matrix of an election, we have one row for each candidate and one column for each position, and an entry contains the fraction of voters that rank the respective candidate in the respective position. This distance naturally extends to SR instances by introducing a row for each agent capturing in which positions the agent is ranked by the other agents. Intuitively, this representation might appear appealing, as it captures the general popularity/quality of agents in the instance. However, the position matrix ignores that each agent is not only ranked by other agents, but also associated with a preference order itself. Consequently, the positionwise distance disregards mutual opinions, i.e., what agents think of each other, which are essential for stability-related considerations. The unsuitably of the positionwise distance for SR instances is also illustrated in a Pearson correlation coefficient (PCC) of only 0.457 with the Spearman distance (see also Figure 1).[5] Lastly, note that the mutual attraction matrix of an SR instance captures all information contained in its position matrix, as $\mathcal{MA}^{\mathcal{I}}(a)$ for some agent a contains the positions in which a is ranked by the other agents in $\mathcal{I}$.

3.2.1 Realizable Mutual Attraction Matrices

Not every $(2n) \times (2n-1)$-matrix is the mutual attraction matrix of some SR instance. Accordingly, we call a matrix M *realizable* if there is an SR instance $\mathcal{I}$ with $\mathcal{MA}^{\mathcal{I}} = M$. Realizable matrices exhibit certain characteristics. For example, since each agent ranks exactly one agent in position j for every $j \in [2n-1]$, every realizable matrix $M \in \mathbb{N}^{(2n)\times(2n-1)}$ contains each number from $[2n-1]$ exactly $2n$ times. Unfortunately, checking whether a matrix is realizable is NP-hard, which means that we (presumably) cannot hope for a polynomial-time checkable "well-behaved" characterization (Woeginger, 2003).

Theorem 13. *Given a $(2n) \times (2n-1)$ matrix M, deciding if there is an SR instance $\mathcal{I}$ with $\mathcal{MA}^{\mathcal{I}} = M$ is NP-complete.*

Proof. We reduce from the NP-complete problem of deciding whether the edge set of a 3-regular graph can be partitioned into three edge-disjoint perfect matchings (Holyer, 1981).

Construction. Given a 3-regular graph $G = (V = \{v_1, \dots, v_\nu\}, E)$, let $\{v_{p_1}, v_{q_1}\}, \{v_{p_2}, v_{q_2}\}, \dots, \{v_{p_z}, v_{q_z}\}$ be a list of all vertex pairs that are not adjacent in G (this means that $z = \binom{\nu}{2} - \frac{3\nu}{2}$).

To construct matrix M, we first construct a dummy SR instance $\mathcal{J}$ consisting of dummy and vertex agents: We introduce one *vertex agent* a_v for each vertex $v \in V$. Moreover, we introduce one *dummy agent* $d_{i,j}$ for $i \in [z]$ and $j \in [\nu] \setminus \{p_i, q_i\}$. Concerning the agent's preferences, we start by constructing the preferences of some vertex agent a_{v_ℓ} for some $\ell \in [\nu]$. Vertex agent a_{v_ℓ} ranks the three agents corresponding to the three vertices adjacent to v_ℓ in G in the first three positions in arbitrary order. For the subsequent positions for $i \in [z]$, if $\ell = p_i$ or $\ell = q_i$, then a_{v_ℓ} ranks a_{q_i}, respectively, a_{p_i} in position $i+3$; otherwise a_{v_ℓ} ranks $d_{i,\ell}$ in position $i+3$. All remaining agents are appended to the preferences in some fixed, arbitrary order. Concerning the dummy agents, agent $d_{i,j}$ for $i \in [z]$ and $j \in [\nu] \setminus \{p_i, q_i\}$ ranks agent a_{v_j} in the first position. Moreover, the dummy agents rank all other dummy agents in an arbitrary order in the subsequent positions such that no two dummy agents rank each other in the same position (this can be achieved by performing cyclic shifts). Subsequently, they rank all vertex agents in some arbitrary ordering.

5. To compute this value we used the test dataset of 460 instances that we will describe in Section 5.1 for twelve agents. We computed the Spearman distance by iterating over all possible agent mappings σ and twelve was the largest number of agents we could handle within weeks.

Let $M' := \mathcal{R}^{\mathcal{J}}$. To obtain M we modify M': For each vertex agent, we set the first entry of its vector to one, the second entry to two, and the third entry to three.

Proof of Correctness. ($\Rightarrow$) Given a partitioning of E into three perfect matching M_1, M_2, and M_3, for each $\ell \in [\nu]$ and $i \in [3]$, let $v_{\ell,i}$ be the vertex adjacent to v_ℓ in M_i. To construct an SR instance $\mathcal{I}$ realizing M, we start with the above-constructed instance $\mathcal{J}$ and modify the first three positions in the preference order of each vertex agent as follows: For $\ell \in [\nu]$ and $i \in [3]$, agent a_{v_ℓ} ranks $a_{v_{\ell,i}}$ in position i. Note that as M_1, M_2, and M_3 are perfect disjoint matchings, in the resulting instance each agent ranks all other agents in its preferences. We now claim that $\mathcal{R}^{\mathcal{I}} = M$. Note that $\mathcal{R}^{\mathcal{I}}$ and $\mathcal{R}^{\mathcal{J}}$ are identical up to the first three entries in rows corresponding to vertex agents. As for each edge $\{v, w\} \in M_i$ for $i \in [3]$ agent a_v ranks a_w in position i and a_w ranks a_v in position i, in $\mathcal{R}^{\mathcal{I}}$ the mutual attraction vector of each vertex agent starts with $1, 2, 3$. Thus, $\mathcal{R}^{\mathcal{I}} = M$.

($\Leftarrow$) Assume that there is an SR instance $\mathcal{I}$ with $\mathcal{R}^{\mathcal{I}} = M$. For convenience, we assume that the names of agents in $\mathcal{I}$ are the same as in our construction. First observe that in M for each $i \in [4, z+3]$ there are exactly two rows which contain an i at position i, that are, the two rows corresponding to vertex agents $a_{v_{p_{i-3}}}$ and $a_{v_{q_{i-3}}}$: All other vertex agents rank a dummy agent in this position, which in turn ranks the vertex agent first. Moreover, by construction, we have that dummy agents rank only other dummy agents in position 4 to $z + 3$ and that no two dummy agents rank each other in the same position. Thus, it follows that in $\mathcal{I}$ $a_{v_{p_{i-3}}}$ and $a_{v_{q_{i-3}}}$ rank each other in position i. This implies that a vertex agent a_v ranks all agents corresponding to vertices that are not adjacent to v in G between positions 4 to $z + 3$. We claim that this further implies that a_v ranks the agents corresponding to vertices adjacent to v in G in the first three positions in $\mathcal{I}$. By construction, for no dummy agent does its mutual attraction vector contain an i at position i for $i \in [3]$. Thus, a_v needs to rank vertex agents in the first three positions, and the vertex agents for adjacent vertices are the only remaining ones. Furthermore, observe that if a_v ranks a_w in position i for $i \in [3]$, then by the construction of M, agent a_w ranks a_v in position i. For $i \in [3]$, let $M_i := \{\{v, w\} \mid a_v \text{ and } a_w \text{ rank each other in position } i \text{ in } \mathcal{I}\}$. Note that M_i is clearly a matching as each agent can only rank one other agent in each position. Moreover, by our above observations, M_i is perfect. Furthermore, M_1, M_2, and M_3 need to be disjoint again because each agent can rank only one agent on each position. Thus, we found a partition of the given graph into three perfect matchings M_1, M_2, and M_3. □

Another implication of this result is that we cannot easily move between SR instances and their mutual attraction matrices. This is in contrast to the "map of elections" context where elections' aggregate representations admit a nice characterization and the realizability problem is polynomial-time solvable (Boehmer et al., 2021b).

3.2.2 Further Properties of the Mutual Attraction Distance

Unfortunately, in contrast to the swap and Spearman distance, the mutual attraction distance is not isomorphic, i.e., there exist multiple non-isomorphic SR instances having the same mutual attraction matrix:[6]

6. Note that the positionwise distance between elections used in the map of elections (Szufa et al., 2020; Boehmer et al., 2021b) is also not isomorphic.

Observation 14. *The mutual attraction distance is not an isomorphic distance.*

Proof. Let $\mathcal{I}$ with agents a, b, c, and d and $\mathcal{I}'$ with agents a', b', c', and d' be two SR instances with the following preferences:

$$\begin{array}{llll} a : b \succ c \succ d, & b : a \succ c \succ d, & c : a \succ d \succ b, & d : a \succ b \succ c, \\ a' : b' \succ c' \succ d', & b' : a' \succ d' \succ c', & c' : a' \succ b' \succ d', & d' : a' \succ c' \succ b'. \end{array}$$

The mutual attraction matrices of the two instances are:

$$\mathcal{MA}_{\mathcal{I}} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} a \\ b \\ c \\ d \end{array} & \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 2 & 3 & 2 \\ 3 & 3 & 2 \end{bmatrix} \end{array}, \mathcal{MA}_{\mathcal{I}'} = \begin{array}{c c} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} a' \\ b' \\ c' \\ d' \end{array} & \begin{bmatrix} 1 & 1 & 1 \\ 1 & 3 & 2 \\ 2 & 3 & 2 \\ 3 & 3 & 2 \end{bmatrix} \end{array}.$$

So we have $\mathrm{d}_{\mathrm{MAD}}(\mathcal{I}, \mathcal{I}') = 0$, yet $\mathcal{I}$ and $\mathcal{I}'$ are not isomorphic: Every isomorphism would need to map a to a', implying that the isomorphism also maps b to b', c to c', and d to d'. However, the second choice of b is c, but c is matched to c' while the second choice of b' is d'. Thus, $\mathcal{I}$ and $\mathcal{I}'$ are not isomorphic. □

Thus, we say that a matrix has a *unique realization* if each pair of SR instances realizing the matrix are isomorphic. Unfortunately, there even exist mutual attraction matrices realized by two non-isomorphic SR instances $\mathcal{I}_1$ and $\mathcal{I}_2$ where $\mathcal{I}_1$ admits a stable matching but $\mathcal{I}_2$ does not. This indicates that the mutual attraction distance between two instances has only a limited predictive value for their relationship in terms of their (distance to) stability. However, this is in turn not too surprising given that stability is dependent on local configurations.

Observation 15. *There are two non-isomorphic instances $\mathcal{I}_1$ and $\mathcal{I}_2$ such that $\mathrm{d}_{\mathrm{MAD}}(\mathcal{I}_1, \mathcal{I}_2) = 0$, $\mathcal{I}_1$ admits a stable matching, and $\mathcal{I}_2$ does not admit a stable matching.*

Proof. Consider the following two instances:

$$\begin{array}{ll} a_1 : a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 & b_1 : b_2 \succ b_6 \succ b_4 \succ b_5 \succ b_3 \\ a_2 : a_3 \succ a_1 \succ a_6 \succ a_4 \succ a_5 & b_2 : b_3 \succ b_1 \succ b_6 \succ b_4 \succ b_5 \\ a_3 : a_1 \succ a_2 \succ a_5 \succ a_6 \succ a_4 & b_3 : b_4 \succ b_2 \succ b_5 \succ b_6 \succ b_1 \\ a_4 : a_5 \succ a_6 \succ a_2 \succ a_1 \succ a_3 & b_4 : b_5 \succ b_3 \succ b_2 \succ b_1 \succ b_6 \\ a_5 : a_6 \succ a_4 \succ a_1 \succ a_3 \succ a_2 & b_5 : b_6 \succ b_4 \succ b_1 \succ b_3 \succ b_2 \\ a_6 : a_4 \succ a_5 \succ a_3 \succ a_2 \succ a_1 & b_6 : b_1 \succ b_5 \succ b_3 \succ b_2 \succ b_4 \end{array}$$

For both instances, the mutual attraction matrix is the following:

$$\begin{bmatrix} 2 & 1 & 4 & 3 & 5 \\ 2 & 1 & 4 & 3 & 5 \\ 2 & 1 & 4 & 3 & 5 \\ 2 & 1 & 4 & 3 & 5 \\ 2 & 1 & 4 & 3 & 5 \\ 2 & 1 & 4 & 3 & 5 \end{bmatrix}$$

The left instance does not admit a stable matching, while the right instance admits the two stable matchings $M_1 = \{\{b_1, b_2\}, \{b_3, b_4\}, \{b_5, b_6\}\}$ and $M_2 = \{\{b_2, b_3\}, \{b_4, b_5\}, \{b_6, b_1\}\}$. $\square$

This is in partial contrast to the Spearman distance, where instances at distance zero are isomorphic, and thus either both or neither of them admits a stable matching. Regarding pairs of instances that are at a non-zero distance, for Spearman, there also exist SR instances at distance 2 (which is the smallest achievable non-zero distance) where one admits a stable matching and the other one does not. However, for the Spearman distance it holds that if a matching M is stable in instance $\mathcal{I}$, then M admits at most $d_{\text{spear}}(\mathcal{I}, \mathcal{I}')$ blocking pairs in instance $\mathcal{I}'$ (Ostrovsky & Rosenbaum, 2015). Thus, under Spearman, if two instances are close to each other and one of them admits a stable matching, then the other instance is also guaranteed to contain a matching that is almost stable. An analogous statement holds for the swap distance: If matching M is stable in instance $\mathcal{I}$, then M admits at most $\mathrm{d}_{\text{swap}}(\mathcal{I}, \mathcal{I}')$ blocking pairs in instance $\mathcal{I}$ (as any swap can create at most one blocking pair). However, this is not the case for the mutual attraction distance: By Observation 15, there are two instances $\mathcal{I}$ and $\mathcal{I}'$ such that $\mathcal{I}$ admits a stable matching while $\mathcal{I}'$ does not. We can merge k copies of $\mathcal{I}$ respectively $\mathcal{I}'$ into an instance $\mathcal{I}_k$ respectively $\mathcal{I}'_k$ (where the preferences are extended to the agents from other copies in the same way for $\mathcal{I}_k$ and $\mathcal{I}'_k$) such that $\mathrm{d}_{\text{MAD}}(\mathcal{I}_k, \mathcal{I}'_k) = 0$. Then, $\mathcal{I}_k$ admits a stable matching, but any stable matching for $\mathcal{I}'_k$ has at least k blocking pairs. In Section 6.1.1, we will demonstrate that all our synthetically generated instances are anyway close to admitting a stable matching in the sense that in all instances there is a matching blocked by only a few pairs.

To better understand the general properties of the mutual attraction distance, we continue by proving upper and lower bounds on the distance of two SR instances.

Proposition 16. *For any two SR instances $\mathcal{I}_1$ and $\mathcal{I}_2$ with $2n$ agents each and $\mathrm{d}_{\text{MAD}}(\mathcal{I}_1, \mathcal{I}_2) > 0$, we have $2 \le \mathrm{d}_{\text{MAD}}(\mathcal{I}_1, \mathcal{I}_2) \le 4 \cdot (n-1) \cdot n^2$.*

Proof. Let $M_1 := \mathcal{MA}_{\mathcal{I}_1}$ and $M_2 := \mathcal{MA}_{\mathcal{I}_2}$. As every realizable matrix $M \in \mathbb{N}^{(2n)\times(2n-1)}$ contains each number from $[2n-1]$ exactly $2n$ times, both M_1 and M_2 contain each number from $[2n-1]$ exactly $2n$ time.

For the upper bound, note that from this it follows that each number from $[2n-1]$ appears exactly $4n$ times in M_1 and M_2 together. Since $|x-y| = \max\{x, y\} - \min\{x, y\}$ holds for all $x, y \in \mathbb{R}$, we can upper bound $\mathrm{d}_{\text{MAD}}(M_1, M_2)$ by summing up the $2n \cdot (2n-1)$ largest numbers appearing in M_1 and M_2 and subtracting the $2n \cdot (2n-1)$ smallest numbers appearing in M_1 and M_2. Consequently, we have

$$\begin{aligned}\mathrm{d}_{\text{MAD}}(M_1, M_2) &\le 4n \cdot (\sum_{j=n+1}^{2n-1} j - \sum_{j=1}^{n-1} j) = 4n \cdot (\sum_{j=1}^{n-1} (n+j) - \sum_{j=1}^{n-1} j) \\ &= 4n \cdot (n \cdot (n-1) + \sum_{j=1}^{n-1} j - \sum_{j=1}^{n-1} j) = 4n^2 \cdot (n-1)\end{aligned}$$

For the lower bound, note that if $\mathrm{d}_{\text{MAD}}(M_1, M_2) \neq 0$, then for each $\sigma \in \Pi([2n], [2n])$ there is some $i \in [2n]$ and $j \in [2n-1]$ with $M_{1\,i,j} \neq M_{2\,\sigma(i),j}$. As each number appears in

M_1 and M_2 the same number of times from this it follows that there also needs to be at least one other pair $i' \in [2n]$ and $j' \in [2n-1]$ with $M_{1i',j'} \neq M_{2\sigma(i'),j'}$ and $(i,j) \neq (i',j')$. From this it follows that $\mathrm{d_{MAD}}(M_1, M_2) \geq 2$. □

In fact, it is easy to see that the lower bound is tight. Later in Proposition 23, we will also establish the tightness of the upper bound.

Observation 17. *There are two SR instances $\mathcal{I}$ and $\mathcal{I}'$ with $\mathrm{d_{MAD}}(\mathcal{I}, \mathcal{I}') = 2$.*

Proof. Let $\mathcal{I}$ with agents a, b, c, and d and $\mathcal{I}'$ with agents a', b', c', and d' be two SR instances with the following preferences:

$$\begin{array}{llll} a : b \succ c \succ d, & b : a \succ c \succ d, & c : a \succ b \succ d, & d : a \succ b \succ c, \\ a' : c' \succ b' \succ d', & b' : a' \succ c' \succ d', & c' : a' \succ b' \succ d', & d' : a' \succ b' \succ c'. \end{array}$$

The mutual attraction matrices of the two instances are:

$$\mathcal{MA}_{\mathcal{I}} = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 2 & 2 & 3 \\ 3 & 3 & 3 \end{bmatrix} \end{array}, \mathcal{MA}_{\mathcal{I}'} = \begin{array}{c} \\ a' \\ b' \\ c' \\ d' \end{array} \begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 2 & 3 \\ 3 & 3 & 3 \end{bmatrix} \end{array}.$$

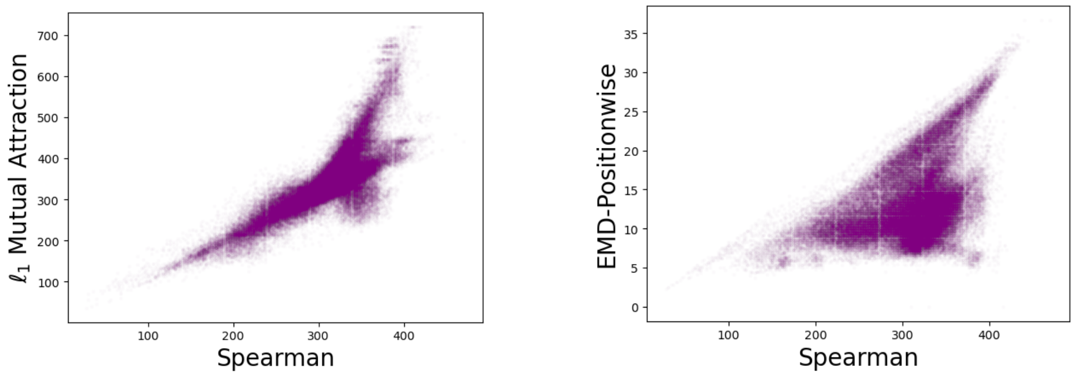
It clearly holds that $\mathrm{d_{MAD}}(\mathcal{MA}_{\mathcal{I}}, \mathcal{MA}_{\mathcal{I}'}) = 2$. □

Correlation of Mutual Attraction and Spearman Distance. As the Spearman distance d_{spear} is a very natural and intuitively appealing distance measure, we checked the correlation between the mutual attraction and Spearman distance. For this, we used the test dataset of 460 instances that we will describe in Section 5.1 for twelve agents.[7] The Pearson Correlation Coefficient (PCC) between the mutual attraction and Spearman distance on our test dataset is 0.801, which is typically regarded as a strong correlation (Schober et al., 2018). In particular, for 95% of instance pairs $(\mathcal{I}, \mathcal{I}')$ we have that $0.82 \cdot \mathrm{d_{MAD}}(\mathcal{I}, \mathcal{I}') \leq d_{\text{spear}}(\mathcal{I}, \mathcal{I}') \leq 1.48 \cdot \mathrm{d_{MAD}}(\mathcal{I}, \mathcal{I}')$. Figure 1 (a) depicts this correlation on the instance level.

Unfortunately, despite the observed correlation in practice, the ratio between the mutual attraction and Spearman distance is unbounded: First, as the Spearman distance is isomorphic but the mutual attraction distance is not, there are instances at mutual attraction distance zero but positive Spearman distance. Second, we show that there are instances with mutual attraction distance zero but unbounded Spearman distance:

Observation 18. *For any $n \geq 2$, there are SR instances $\mathcal{I}_1$ and $\mathcal{I}_2$ on n agents with $d_{\text{spear}}(\mathcal{I}_1, \mathcal{I}_2) = 2$ but $\mathrm{d_{MAD}}(\mathcal{I}_1, \mathcal{I}_2) \geq n - 2$.*

7. We computed the Spearman distance by iterating over all possible agent mappings σ and twelve was the largest number of agents we could handle within weeks. We decided to use the Spearman distance instead of the swap distance here because the latter required more time to compute in practice.



(a) Mutual attraction distance (PCC = 0.801)

(b) Positionwise distance (PCC = 0.457)

Figure 1: Correlation between the Spearman distance and our mutual attraction distance or the positionwise distance of Szufa et al. (2020) on the dataset described in Section 5.1 for twelve agents. Each pair of instances is represented by a point with its x-axis representing their distance according to one of the measures and its y-axis representing their distances according to the other.

Proof. Consider the following SR instance $\mathcal{I}_1$ with n agents.

$$a_1 : a_2 \succ a_n \succ a_{n-1} \succ \cdots \succ a_3$$
$$a_i : a_2 \succ a_3 \succ \cdots \succ a_{i-1} \succ a_1 \succ a_{i+1} \succ a_{i+2} \succ \cdots \succ a_n, \qquad \forall i \in [2, n]$$

Let $\mathcal{I}_2$ be the instance arising from this SR instance by swapping a_2 and a_n in the preferences of a_1. We will denote the agents of the instance belonging to $\mathcal{I}_2$ by a'_1, a'_2, a'_3, …, a'_n. Then $d_{\text{spear}}(\mathcal{I}_1, \mathcal{I}_2) = 2$.

It remains to show that $\mathrm{d}_{\mathrm{MAD}}(\mathcal{I}_1, \mathcal{I}_2) \geq n-2$. Let σ be a bijection from $\{a_1, a_2, \ldots, a_n\}$ to $\{a'_1, a'_2, \ldots, a'_n\}$ corresponding to the minimum distance $\mathrm{d}_{\mathrm{MAD}}(\mathcal{I}_1, \mathcal{I}_2)$. Note that for each $i \in [2, n]$, we have $\mathcal{MA}_{\mathcal{I}_1}(a_i, j) = i - 1 = \mathcal{MA}_{\mathcal{I}_2}(a'_i, j)$ for all $j \neq i$. Thus, if $\sigma(a_i) \neq a'_i$ for some $i \in [2, n]$, then $\mathrm{d}_{\mathrm{MAD}}(\mathcal{I}_1, \mathcal{I}_2) \geq n-2$. Otherwise, we have $\sigma(a_i) = a'_i$ for all $i \in [n]$. However, we have $\mathcal{MA}_{\mathcal{I}_1}(a_1) = (1, n-1, n-2, n-3, \ldots, 2)$ and $\mathcal{MA}_{\mathcal{I}_2}(a'_1) = (n-1, 1, n-2, n-3, \ldots, 2)$. Thus, in this case we would have $\mathrm{d}_{\mathrm{MAD}}(\mathcal{I}_1, \mathcal{I}_2) \geq 2 \cdot (n-2)$. $\square$

Note that the bound from Observation 18 is certainly not tight.

4. Navigating the Space of SR Instances

In this section we will identify four somewhat "canonical" extreme mutual attraction matrices, which are far away from each other and thus fall into four very different parts of the map. Naturally, the corresponding SR instances also form extreme points of the space of SR instances induced by our distance measure. Such extreme instances have the potential to give meaning to different regions on the map and typically make the embedding more structured (Boehmer, 2023, Section 2.3.1). Related to this, when visualizing the outcome of some experiment by coloring points on the map accordingly, it might be easy to explain why canonical instances behave in a certain way (Boehmer, 2023, Section 8.4.2). Thereby, such canonical instances can give some intuition for why some region on the map behave in a certain way. The proofs of all statements from this section can be found in the appendix.

Identity. Our first extreme case is that all agents have the same preferences, i.e., there exists a central order called master list of the agents A and the preferences of an agent $a \in A$ are derived from the master list by deleting a. Stable matching instances with master lists have already attracted significant attention in the past (Irving, Manlove, & Scott, 2008; Cui & Jia, 2013; Kamiyama, 2019; Bredereck, Heeger, Knop, & Niedermeier, 2020). For $n \in \mathbb{N}$, the identity matrix is defined by

$$\mathrm{ID}^{2n}[i,j] := \begin{cases} i & j \geq i \\ i-1 & j < i \end{cases}$$

for each $i \in [2n]$ and $j \in [2n-1]$. We prove that, in fact, only SR instances where all preferences are derived from a master list realize the identity matrix:

Proposition 19. *For every $n \in \mathbb{N}$, an SR instance $\mathcal{I}$ is a realization of ID^{2n} if and only if the preferences in $\mathcal{I}$ are derived from a master list. In particular, the realization of ID^{2n} is unique.*

Mutual Agreement. Our second extreme case is mutual agreement: For each pair a and a' of agents, a and a' evaluate each other identically, i.e., a ranks a' on the i-th position if and only if a' ranks a on the i-th position.[8] For $n \in \mathbb{N}$, this is captured in the mutual agreement matrix MA^{2n} where we have $\mathrm{MA}^{2n}[i,j] = j$ for each $i \in [2n]$ and $j \in [2n-1]$. At first glance, it is unclear whether the mutual agreement matrix is realizable. It turns out that the realizations of MA^{2n} correspond to Round-Robin tournaments: In a Round-Robin tournament of $2n$ agents, there are $2n-1$ days with each agent competing exactly once each day and exactly once against each other agent (Harary & Moser, 1966). The intuition here is that an agent in the SR instance corresponding to a Round-Robin tournament ranks in the i-th position the agent against whom it competes on the i-th day. Formally, we have:

Proposition 20. *For every $n \in \mathbb{N}$, there is a bijection between realizations of MA^{2n} and the set of Round-Robin tournaments. In particular, there are several non-isomorphic realizations of MA^{2n} for $n = 4$.*

Mutual Disagreement. Our third extreme case is mutual disagreement. For each pair a and a' of agents, their evaluations for each other are diametrical, i.e., a ranks a' in the i-th position if and only if a' ranks a in the $(2n-i)$-th position. For $n \in \mathbb{N}$, this is captured in the mutual disagreement matrix MD^{2n} where we have $\mathrm{MD}^{2n}[i,j] = 2n - j$ for each $i \in [2n]$ and $j \in [2n-1]$. There exists a straightforward realization of MD^{2n} with $2n$ agents $a_1, \ldots, a_{2n}$ where the preferences of agent a_i are derived from the preferences of agent a_{i-1}

8. Notably, in the Stable Marriage with Symmetric Preferences (O'Malley, 2007; Abraham, Levavi, Manlove, & O'Malley, 2008) problem we are given a set of men and women with preferences over each other, where a woman w ranks a man m in position i if and only if m ranks w in position i, an idea very similar to mutual agreement (see O'Malley, 2007, Chapter 6 for additional motivation). O'Malley (2007) and Abraham et al. (2008) study the computational complexity of various traditional questions on such instances (in the presence of ties).

by performing a cyclic shift, i.e., $a_i : a_{i+1} \succ a_{i+2} \succ \cdots \succ a_n \succ a_1 \succ a_2 \succ \cdots \succ a_{i-1}$. However, this realization is not unique:

Proposition 21. *For every $n \in \mathbb{N}$, matrix MD^{2n} is realizable. For $n = 3$, matrix MD^{2n} has multiple non-isomorphic realizations.*

Chaos. Our fourth extreme mutual attraction matrix is the chaos matrix CH^{2n}, which is defined for each $i \in [2n]$ and $j \in [2n-1]$ as

$$\mathrm{CH}^{2n}[i,j] = \begin{cases} j, & \text{for } i = 1 \\ i + nj - n - 1 \mod 2n-1, & \text{otherwise.} \end{cases}$$

Unlike the other three matrices, we have no natural interpretation of the chaos matrix. We added this matrix to the other three because it is far away from each of them and thus falls into an otherwise vacant part of the map. Its name "chaos" stems from the fact that this matrix is typically placed close to instances with uniformly at random sampled preferences on the map. We prove that for infinitely many $n \in \mathbb{N}$, CH^{2n} is realizable (note that in case $2n-1$ is divisible by 3, the matrix is never realizable):

Proposition 22. *For every $n \in \mathbb{N}$ such that $2n-1$ is not divisible by 3, matrix CH^{2n} is realizable, and the realization is unique.*

Distances Between Matrices. The mutual attraction distances between our extreme matrices are as follows:

Proposition 23. *For each $n \in \mathbb{N}$, we have*

1. $\mathrm{d}_{\mathrm{MAD}}(MA^{2n}, MD^{2n}) = 4 \cdot (n-1) \cdot n^2$,
2. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, MA^{2n}) = \mathrm{d}_{\mathrm{MAD}}(MA^{2n}, CH^{2n}) = \frac{8}{3}n^3 - 4n^2 + \frac{4}{3}n$,
3. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, MD^{2n}) = \mathrm{d}_{\mathrm{MAD}}(MD^{2n}, CH^{2n}) = \frac{8}{3}n^3 - 2n^2 - \frac{2}{3}n$,
4. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, CH^{2n}) = \frac{8}{3}n^3 \pm O(n^2)$.

As proven in Proposition 16, $D(2n) := 4 \cdot (n-1) \cdot n^2$ is the maximum possible distance between two mutual attraction matrices of SR instances with $2n$ agents. Thus, the mutual agreement matrix and the mutual disagreement matrix are at the maximum possible distance and therefore form a diameter of our space. For each two matrices X and Y among ID, MA, CH, and MD, we define their asymptotic normalized distance as: $\mathrm{nd}_{\mathrm{MAD}}(X,Y) := \lim_{n\to\infty} \mathrm{d}_{\mathrm{MAD}}(X^{2n},Y^{2n})/D(2n)$. It turns out that for all pairs of matrices $X, Y \in \{\mathrm{ID}, \mathrm{MA}, \mathrm{MD}, \mathrm{CH}\}$ with $\{X,Y\} \neq \{\mathrm{MA}, \mathrm{MD}\}$ we have $\mathrm{nd}_{\mathrm{MAD}}(X,Y) = \frac{2}{3}$, while $\mathrm{nd}_{\mathrm{MAD}}(\mathrm{MA}, \mathrm{MD}) = 1$. This implies that our extreme matrices are indeed far from each other. In the following, we will often consider normalized mutual attraction distances where we divide the computed distance by $D(2n)$.

5. A Map of Stable Roommates Instances

In this section, we present a map of synthetic SR instances. In Section 5.1, we describe how we create the map and how we generate the instances. In Section 5.2, we explain the map by giving the horizontal and vertical axes a natural interpretation and by analyzing where different statistical cultures land.

5.1 Creating the Map

We first describe our dataset of 460 SR instances generated from the following statistical cultures (see Table 1 for an overview). To the best of our knowledge, only the Impartial Culture, Attributes, Mallows, and Euclidean models have been previously considered.

Impartial Culture (IC). Agent $a \in A$ draws its preferences uniformly at random from the set of all possible preference orders over A excluding a, i.e., $\mathcal{L}(A \setminus \{a\})$.

2-IC. Given some $p \in [0, 0.5]$, we partition A into two sets $A_1 \uplus A_2$ with $|A_1| = \lfloor p \cdot |A| \rfloor$. Each agent $a \in A$ samples a preference order $\succ$ from $\mathcal{L}(A_1 \setminus \{a\})$ and preference order $\succ'$ from $\mathcal{L}(A_2 \setminus \{a\})$. Next, if $a \in A_1$, then we let a's preferences start with all agents from $A_1 \setminus \{a\}$ ordered according to $\succ$ and then all agents from A_2 ordered according to $\succ'$. If $a \in A_2$, then it is the other way around, i.e., the preferences start with $\succ'$ and end with $\succ$. The intuition is that there are two groups of different sizes (e.g., representing demographic groups), and each agent prefers all agents from its group to agents from the other group; preferences within each group are random.

Mallows Model. In the original Mallows model (Mallows, 1957), for a parameter $\phi \in [0, 1]$ and a preference order $\succ^* \in \mathcal{L}(A)$, the Mallows distribution $\mathcal{D}^{\succ^*,\phi}_{\text{Mallows}}$ assigns preference order $\succ \in \mathcal{L}(A)$ a probability proportional to $\phi^{\text{swap}(\succ^*,\succ)}$. The intuition is that there is a central order and the probability of sampling a preference order is proportional to its distance to the central one, where the expected distance is controlled by ϕ. However, as argued by Boehmer et al. (2021b) and Boehmer, Faliszewski, and Kraiczy (2023) one disadvantage of the Mallows model is that choosing the dispersion parameter ϕ uniformly at random leads to a skewed dataset. That is why we use a normalized variant of the Mallows model $\mathcal{D}^{\succ^*,\text{norm-}\phi}_{\text{Mallows}}$ proposed by Boehmer et al. (2021b), which is parameterized by a normalized dispersion parameter norm-ϕ. The idea here is that preference orders sampled from $\mathcal{D}^{\succ^*,\text{norm-}\phi}_{\text{Mallows}}$ are at an expected swap distance of norm-$\phi \times \frac{n(n-1)}{4}$ from $\succ^*$. By setting norm-$\phi = 1$, we recover IC, whereas norm-$\phi = 0$ results in only $\succ^*$ being sampled, and norm-$\phi = 0.5$ results in preferences orders that fall in some sense exactly between the two. Specifically, sampling from $\mathcal{D}^{\succ^*,\text{norm-}\phi}_{\text{Mallows}}$, norm-$\phi$ is internally converted to a value ψ of the dispersion parameter such that the expected swap distance between $\succ^*$ and a sampled preference order from $\mathcal{D}^{\succ^*,\psi}_{\text{Mallows}}$ is norm-ϕ times $\frac{n(n-1)}{4}$. Subsequently a preference order from $\mathcal{D}^{\succ^*,\psi}_{\text{Mallows}}$ is drawn. Now, given a normalized dispersion parameter norm-$\phi \in [0, 1]$, to generate an SR instance, we draw $\succ^*$ uniformly at random from $\mathcal{L}(A)$. Afterwards, for each agent $a \in A$, we obtain its preferences by drawing a preference order from $\mathcal{D}^{\succ^*|_{A\setminus\{a\}},\text{norm-}\phi}_{\text{Mallows}}$.

Euclidean (Arkin, Bae, Efrat, Okamoto, Mitchell, & Polishchuk, 2009). Given some $d \in \mathbb{N}$, for each agent $a \in A$, we uniformly at random sample a point $\mathbf{p}^a$ from $[0, 1]^d$. Agent a ranks other agents increasingly by the Euclidean distance between their points, i.e., by $\ell_2(\mathbf{p}^a, \mathbf{p}^b)$ for $b \in A \setminus \{a\}$. The intuition is that each dimension represents some continuous property of the agents and agents prefer similar agents.

Reverse-Euclidean. Given some $p \in [0, 1]$ and $d \in \mathbb{N}$, we partition A into two sets $A_1 \uplus A_2$ with $|A_1| = \lfloor p \cdot |A| \rfloor$. Each agent corresponds to some uniformly at random sampled point $\mathbf{p}^a$ from $[0, 1]^d$ and ranks other agents according to their Euclidean distance. However, here an agent $a \in A_1$ ranks agents decreasingly, i.e., from the furthest to the closest one, by

model	parameter	number of instances
Impartial Culture (IC)	—	20
2-Impartial Culture (2-IC)	$p \in \{0.25, 0.5\}$	20 for each p
Mallows	norm-$\phi \in \{0.2, 0.4, 0.6, 0.8\}$	20 for each norm-ϕ
Mallows-MD	norm-$\phi \in \{0.2, 0.4, 0.6\}$	20 for each norm-ϕ
1D-Euclidean	—	20
2D-Euclidean	—	20
Reverse-2D-Euclidean	$p \in \{0.05, 0.15, 0.25\}$	20 for each p
Mallows-2D-Euclidean	norm-$\phi \in \{0.2, 0.4\}$	20 for each norm-ϕ
Expectations-2D-Euclidean	$\sigma \in \{0.2, 0.4\}$	20 for each σ
Fame-2D-Euclidean	$f \in \{0.2, 0.4\}$	20 for each f
Attributes	$d \in \{2, 5\}$	20 for each d

Table 1: Composition of our synthetic dataset with 460 instances (see Section 5.1).

their Euclidean distance to $\mathbf{p}^a$ and an agent $a \in A_2$ ranks agents increasingly, i.e., from the closest to the furthest one, by their Euclidean distance to $\mathbf{p}^a$. The intuition is similar to Euclidean, but a p-fraction of agents prefers agents that are different from them.

Mallows-Euclidean. Given a normalized dispersion parameter norm-$\phi \in [0, 1]$ and some $d \in \mathbb{N}$, we start by generating agents' intermediate preferences $(\succ_a)_{a \in A}$ according to the Euclidean model with d dimensions. Subsequently, for each $a \in A$, we obtain its final preferences by sampling a preference order from $\mathcal{D}^{\succ_a,\text{norm-}\phi}_{\text{Mallows}}$. The resulting instances are perturbed Euclidean instances.

Expectations-Euclidean. Given some $d \in \mathbb{N}$ and $\sigma \in \mathbb{R}^+$, for each agent $a \in A$, we sample one point $\mathbf{p}^a$ uniformly at random from $[0, 1]^d$. Subsequently, we sample a second point $\mathbf{q}^a$ from $[0, 1]^d$ using a d-dimensional Gaussian function with mean $\mathbf{p}^a$ and standard deviation σ. Agent a ranks the agents increasingly according to $\ell_2(\mathbf{p}^a, \mathbf{q}^b)$ for $b \in A \setminus \{a\}$. Again, agents are characterized by continuous attributes; however, their "ideal" points are not necessarily where they are, yet there is a certain correlation.

Fame-Euclidean. Given some $d \in \mathbb{N}$ and $f \in [0, 1]$, we sample for each agent $a \in A$ uniformly at random a point $\mathbf{p}^a \in [0, 1]^d$ and a number $f^a \in [0, f]$. Agent a ranks the other agents increasingly by $\ell_2(\mathbf{p}^a, \mathbf{p}^b) - f^b$ for $b \in A \setminus \{a\}$. The intuition is similar as for Euclidean, but some agents have a higher quality/fame f^a and are thus more attractive to everyone.

Attributes (Bhatnagar, Greenberg, & Randall, 2008). Given some $d \in \mathbb{N}$, for each agent $a \in A$ we uniformly at random sample $\mathbf{p}^a \in [0, 1]^d$ and $\mathbf{w}^a \in [0, 1]^d$. Agent a ranks the other agents decreasingly by the inner product of $\mathbf{w}^a$ and $\mathbf{p}^b$, i.e., by $\sum_{i \in [d]} \mathbf{w}^a_i \cdot \mathbf{p}^b_i$. The intuition is that there are different objective evaluation criteria and agents assign a different importance to them.

Mallows-MD. Given a normalized dispersion parameter norm-$\phi \in [0, 1]$, we start with an instance that realizes the mutual disagreement matrix MD^{2n} where for each $i \in [2n]$

agent a_i has preferences $a_{i+1} \succ_{a_i} a_{i+2} \succ_{a_i} \cdots \succ_{a_i} a_n \succ_{a_i} a_1 \succ_{a_i} a_2 \succ_{a_i} \cdots \succ_{a_i} a_{i-1}$. Subsequently, for each $a_i \in A$, we obtain its final preferences by sampling a preference order from the Mallows model $\mathcal{D}_{\text{Mallows}}^{\succ_{a_i},\text{norm-}\phi}$. The reason we consider this model is that it covers a part of the map that would otherwise remain uncovered.

Our dataset consists of 460 instances sampled from the above-described statistical cultures. That is, we sampled 20 instances for each of the following cultures: Impartial Culture, 2-IC with $p \in \{0.25, 0.5\}$, Mallows with norm-$\phi \in \{0.2, 0.4, 0.6, 0.8\}$, 1D/2D-Euclidean (with $d \in \{1, 2\}$), Reverse-Euclidean with $d = 2$ and $p \in \{0.05, 0.15, 0.25\}$, Mallows-Euclidean with $d = 2$ and norm-$\phi \in \{0.2, 0.4\}$, Expectations-Euclidean with $d = 2$ and $\sigma \in \{0.2, 0.4\}$, Fame-Euclidean with $d = 2$ and $f \in \{0.2, 0.4\}$, Attributes with $d \in \{2, 5\}$, and Mallows-MD with norm-$\phi \in \{0.2, 0.4, 0.6\}$. In addition, on our maps, we include the four extreme matrices. Our experiments presented in the following will provide evidence that our dataset covers the space of SR instances quite uniformly. We focus on 200 agents in the following (maps for, e.g., 50 and 100 agents look similar).

5.1.1 Drawing the Map

To draw a map of our dataset, we compute for each pair of instances/matrices their mutual attraction distance. Subsequently, we embed the instances as points in the two-dimensional Euclidean space. Our goal is that the Euclidean distance between two points reflects the mutual attraction distance between the two respective instances. To obtain the embedding, we use a variant of the force-directed algorithm of Kamada and Kawai (1989).[9] The general idea here is that we start with an arbitrary embedding of the instances, then we add an attractive force between each pair of instances whose strength reflects their mutual attraction distance and a repulsive force between each pair ensuring that there is a certain minimum distance between each two points. Subsequently, the instances move based on the applied forces until a minimal energy state is reached. We depict the map visualizing our dataset of 460 instances for 200 agents in Figure 2. In Appendix B, we also include maps for instances with 500 or 750 agents, which look very similar.

Quality of the Embedding. To correctly interpret the map, we stress that our embedding algorithm does not optimize a global objective function. Instead, the algorithm works in a decentralized fashion also aiming at producing a visually pleasant image. Consequently, the position of instances on the map can be different in different runs and certainly depend on which other instances are part of the map. To verify the quality of the embedding, we now want to analyze whether the two-dimensional visualization of our dataset as a map adequately reflects the mutual attraction distances between instances. We consider two different quality measures for the embedding which both use normalized distances, i.e., we divide them by the respective distance between mutual agreement and mutual disagreement. First we compute for each pair of instances $(\mathcal{I}, \mathcal{I}')$ its distortion which is defined as the maximum of (a) the normalized mutual attraction distance between $\mathcal{I}$ and $\mathcal{I}'$ divided by the normalized Euclidean distance between the points representing $\mathcal{I}$ and $\mathcal{I}'$ on the map and (b) the normalized Euclidean distance between the points representing $\mathcal{I}$ and $\mathcal{I}'$ on the

9. Szufa et al. (2020) and Boehmer et al. (2021b) used the closely related Fruchterman–Reingold algorithm; however, in our case the results provided by the Kamada-Kawai algorithm were visually more appealing.

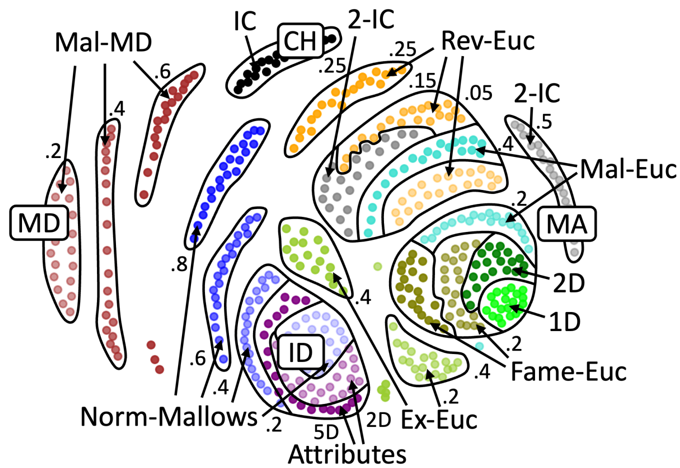


Figure 2: Map of 460 SR instances for 200 agents. Each instance is represented by a point. Roughly speaking, the closer two points are on the map, the more similar are the respective SR instances under the mutual attraction distance. Colors indicate the statistical culture that points were sampled from.

map divided by the normalized mutual attraction distance between $\mathcal{I}$ and $\mathcal{I}'$. The average distortion is 1.8, indicating that distances between instances are certainly not represented perfectly on the map. Nevertheless, this also underlines that the map creates a roughly correct picture of the space of SR instances (distances on the map are typically "off" only by a factor of two). However, we want to remark here that some error in the embedding is certainly to be expected because our space of SR instances is naturally too complex to be perfectly embedded into two-dimensional space. In Figure 3, we analyze which of the instances on the map are particularly challenging to embed and are thus misplaced. We do so by coloring each point on the map according to the average distortion[10] of all pairs involving this instance. We see that instances that fall into the middle of the map are particularly problematic and that instances close to MA and MD are embedded nearly perfectly.

Moreover, as a slightly simpler measure, we also consider for each pair of instances their normalized Euclidean distance on the map divided by their mutual attraction distance. We visualize the results as a histogram in Figure 4. We see that instances are mostly placed "too close to each other" and that for a majority of instances, the normalized Euclidean distance on the map is more than half of their mutual attraction distance. Overall, the results from this section draw a mixed picture: While the map is naturally not perfect, most distances are presented approximately accurately. In particular, as argued in the next section, the map groups instances sampled from the same culture together, implying that it is capable of detecting the shared underlying similarities between instances from the same culture. Moreover, in the experiments presented in the following, we will see that instances

10. Note that individual distortion values can be quite high, especially if two instances are very close to each other on the map. The largest singular distortion is 47.97 and is due to two IC elections, which are indeed embedded next to each other (while being not that similar).

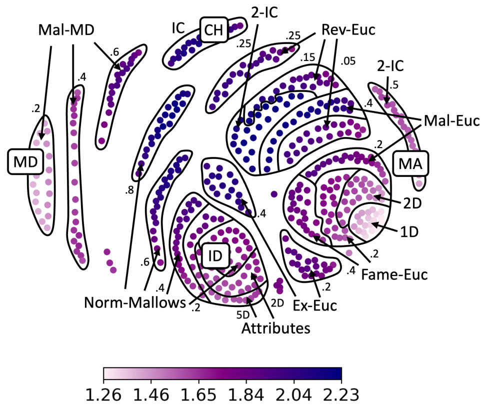


Figure 3: Average distortion for each instance on our map of SR instances for 200 agents.

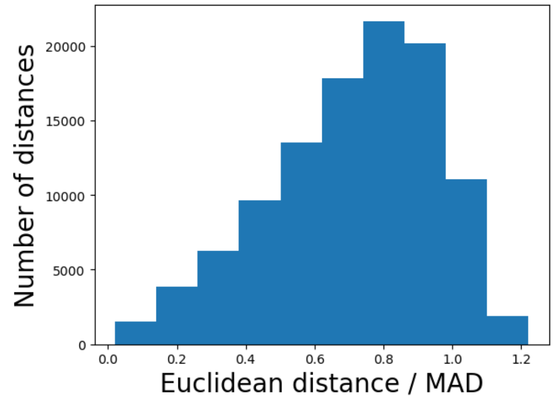


Figure 4: This histogram visualizes the normalized Euclidean distance of instance pairs on the map divided by their normalized mutual attraction distance.

close on the map oftentimes have similar properties to each other, which underlines that instances placed in the same region on the map are structurally similar and highlights the usefulness of the map as a visualization tool for experimental results.

5.2 Understanding the Map

We now take a closer look at the map of SR instances shown in Figure 2 and the average distance between instances sampled from different cultures shown in Figure 5. Our discussion will include an explanation of the different regions of the map and their properties, how to "navigate" the map as well as some additional arguments for why the grouping of instances on the map is meaning- and insightful.

Examining the map, what stands out is that for all cultures, instances sampled from this culture are placed close to each other on the map, resulting in an island-like structure. In fact, instances sampled from the same culture are usually close to each other under the mutual attraction distance (or at least closer to each other than to instances sampled from other cultures; see the diagonal line in Figure 5). While this is to be expected to a certain extent, this observation validates our approach in that the mutual attraction distance is seemingly able to identify the shared structure of instances sampled from the same statistical culture and in that our embedding algorithm can detect these clusters.

Moreover, interestingly, the different statistical cultures have different "variation", i.e., the average mutual attraction distance of two instances sampled from the same culture substantially differs for the different cultures. The Impartial Culture model has the highest variation with 0.59, while the Euclidean model for $d = 1$ has the lowest variation with 0.07 (see Figure 5). The value for Impartial Culture is quite remarkable, as it means that Impartial Culture instances are on average almost as far away from each other as, for example, ID is from the other extreme points. Because of the limitations of two-dimensional Euclidean space, this is not adequately represented on the map, as Impartial Culture instances are

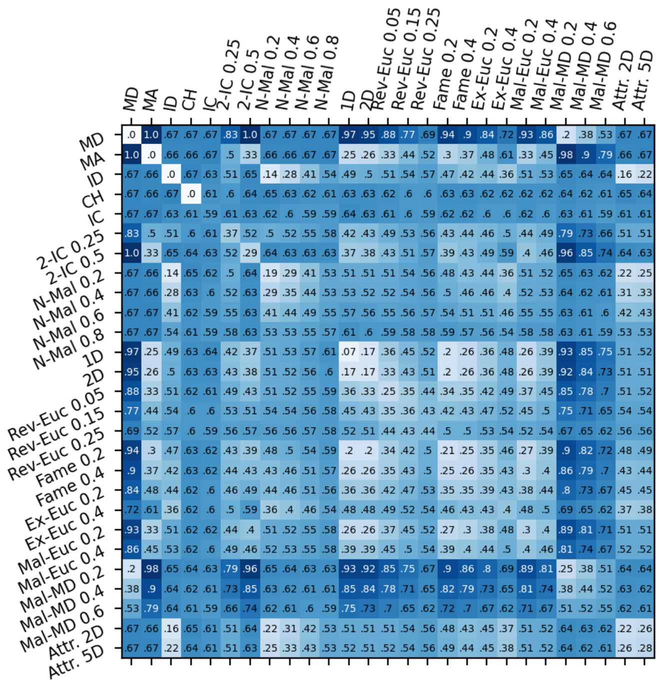
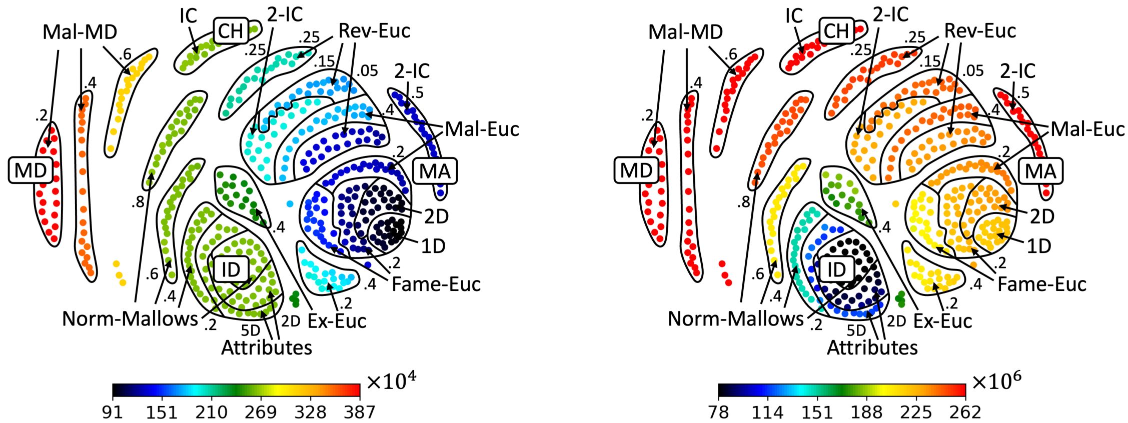


Figure 5: For each pair of statistical cultures, average mutual attraction distance between instances sampled from the two (normalized by the maximum possible distance between two SR instances). The first four lines/columns contain, for each statistical culture, the average distance of instances sampled from that culture to our four extreme matrices. The diagonal contains the average distance of two instances sampled from the same statistical culture.

still placed close to each other. The reason for this embedding is that Impartial Culture instances are all at a similar (even larger) distance from the other instances. In the following experiments, we observe that Impartial Culture instances nevertheless behave quite similarly to each other, justifying their placement next to each other on the map.

Taking a closer look at the map, we observe that our four extreme points fall into four different parts. On the right, we have the mutual agreement matrix MA. Accordingly, models for which mutual agreement is likely to appear all land on the right side of the map. These models are (i) the Euclidean model (where intuitively speaking agent a likes agent b if they are placed close to each other in the underlying space, making it also likely that b likes a), (ii) the Fame-Euclidean model for $f = 0.2$, the Mallows-Euclidean model for norm-$\phi = 0.2$, and the Reverse-Euclidean model for $p = 0.05$ (these three are basically all differently perturbed variants of Euclidean models with a "low" level of perturbation; as a result, they are also on average slightly further away from MA than Euclidean instances), and (iii) the 2-IC model for $p = 0.5$ (where we have some guaranteed level of mutual agreement because there are two equal-sized groups of agents and the agents from one group prefer each other to the agents from the other group).



(a) Mutuality

(b) Rank distortion

Figure 6: Map of 460 SR instances for 200 agents visualizing different quantities for each instance.

On the left, we have the mutual disagreement matrix MD with only instances from the Mallows-MD model being close to it (in general, it is to be expected that if we apply the Mallows model on top of some other model $\mathcal{X}$, then for small values of norm-ϕ the sampled instances are close to the ones from $\mathcal{X}$ but that they move further and further away towards Impartial Culture instances as norm-ϕ grows). That the mutual (dis)agreement matrices are at the two ends of the horizontal axis raises the question of whether the horizontal axis can be indeed interpreted as an indicator for the degree of mutuality in SR instances. This hypothesis gets strongly confirmed in Figure 6 (a) where we color the points on the map according to their mutuality value, which we define as the total difference between the mutual evaluations of agent pairs, i.e., $\sum_{a\in A}\sum_{i\in[|A|-1]}|\mathcal{MA}(a,i)-i|$. For MA this quantity is zero, whereas for MD it takes the maximum possible value. The nicely continuous shading in Figure 6 (a) indicates a strong correlation between the mutuality value of an instance and its x-coordinate on the map with instances that are close on the map having similar mutuality values. Moreover, the continuous coloring indicates that our dataset provides good and almost uniform coverage of the space of SR instances (at least in terms of their mutuality value).

Turning to the middle part of the map, the identity matrix ID can be found at the bottom. Close to ID are instances from cultures where agents' quality is "objective". Namely, Mallows model with norm-ϕ = 0.2 (where the preferences of agents are still often quite close to the central order) and the Attributes model with $d = 2$ (where each agent has two quality scores and the preferences of agents only differ in how they weight the quality scores). The chaos matrix CH is placed on the top of the map together with Impartial Culture instances. Mallows instances naturally form a continuous spectrum between Identity and Chaos. These observations give rise to the hypothesis that in instances placed at the bottom of the map most agents have similar preferences, while in instances placed at the top all agents have roughly the same popularity among the other agents and few agents are particularly (un)popular. To quantify this, we measure the rank distortion of an instance, i.e., for each agent we sum up the absolute difference between all pairs of entries in its mutual attraction vector $\sum_{a \in A} \sum_{i,j \in [|A|-1]} |\mathcal{MA}(a,i) - \mathcal{MA}(a,j)|$. Note that, for example, for an agent that is always ranked in the same position by all other agents this absolute difference is zero, whereas it is maximum for agents that are ranked exactly once in position j for each $j \in [2n-1]$ by the other agents. We show in Figure 6 (b) a map colored by the rank distortion of instances. The picture here is slightly different than for the horizontal axis (previously explained by the mutuality value) in that instances with the same y-coordinate might still have a very different rank distortion. In fact, what we rather see here is that the further a point is from ID on the map, the larger is its rank distortion and thus the higher is the disagreement concerning the quality of an agent (which is quite intuitive recalling that for both MA and MD the rank distortion is maximal, whereas for ID it is minimal).

6. Using the Map

To illustrate the usefulness of the map to evaluate experiments and to confirm our previous observation that instances that are close to each other on the map have indeed similar properties, we perform some example experiments.

6.1 Blocking Pairs and Stable Matchings

We start by analyzing various properties related to the number of pairs that block some matching. Specifically, we first compute for each SR instance the minimum number of blocking pairs over all matchings and then the average number of blocking pairs for a random matching. Lastly, we examine how many pairs block minimum-weight matchings. We visualize the results of our experiments in Figures 7 and 8.

6.1.1 Blocking Pair Minimizing Matching

Naturally, the most important question related to an SR instance is whether the instance admits a stable matching or not. Slightly more nuanced, it is also possible to ask for a matching minimizing the number of blocking pairs. As computing the minimum number of blocking pairs in an SR instance is NP-hard (Abraham, Biró, & Manlove, 2005), we solve this problem using an ILP. We visualize the results of this experiment on the map in Figure 7.

First, considering which instances admit a stable matching (green points on the map), we do not see a clear correlation with the instance's position on the map. This is also quite intuitive, given that whether an instance admits a stable matching might depend on some local configuration. Such configurations can naturally not be fully captured in the mutual attraction matrix. However, what is clearly visible is that for different cultures the probability of admitting a stable matching is quite different: On the one hand, instances sampled from the Euclidean, Fame-Euclidean, and Reverse-Euclidean models almost always admit a stable matching (for the Euclidean model this is even guaranteed). On the other hand, instances sampled from the Mallows-Euclidean and Expectations-Euclidean models only very rarely admit a stable matching. The drastic contrast between the Euclidean model

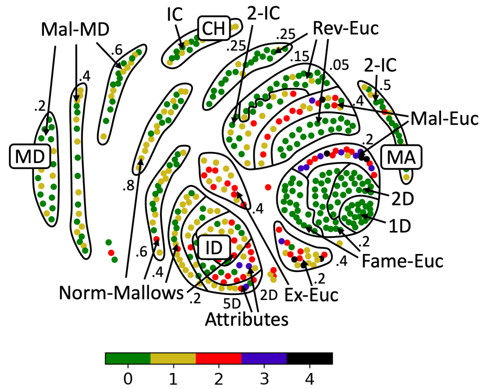


Figure 7: Map of 460 SR instances for 200 agents visualizing the minimum number of blocking pairs.

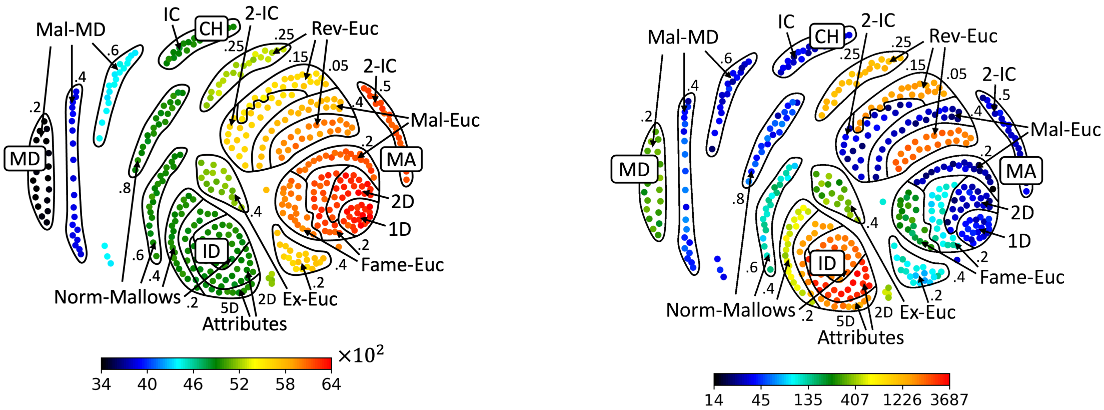
and the Mallows-Euclidean model with norm-$\phi = 0.2$ and between the Reverse-Euclidean and Expectations-Euclidean models is quite remarkable, as they are conceptually quite similar.

However, moving to the minimum achievable number of blocking pairs, the picture of our dataset becomes more uniform: A large majority of the map (and cultures) solely consists of SR instances where the minimum number of blocking pairs is at most one (recall that all our experiments here are for $n = 200$ agents). Only in instances sampled from the Mallows-Euclidean and Expectations-Euclidean model (which only rarely admit a stable matching) is the minimum number of blocking pairs often two or more. Some further such instances can be found close to ID.

Overall, what we find here is that the minimum number of blocking pairs clearly depends on the model from which the respective SR instance was sampled, leading to a clustering of (very close to) stable instances on the map. However, there are also regions on the map exhibiting a mixed picture, for instance, the regions around Mutual Agreement and Identity; interestingly, it seems that while highly structured instances always admit a stable matching (like Euclidean instances where this is even guaranteed to be the case), slightly perturbing these instances leads to an increase in the minimum number of blocking pairs. Lastly, it is remarkable that a matching admitting at most four blocking pairs exists in all our 460 SR instances. This indicates that instances that are “far away” from stability are quite exceptional and motivates a future search for statistical cultures regularly producing such instances.

6.1.2 Expected Number of Blocking Pairs

Motivated by the fundamental importance of blocking pairs for stable matchings, we measure the expected number of blocking pairs for an arbitrary perfect matching. For this, for each instance, we sampled 100 perfect matchings uniformly at random from the set of all possible perfect matchings of agents and for each counted the number of blocking pairs. The observed averages are depicted in Figure 8 (a). As for the mutuality value, we get a nicely



(a) Average number of blocking pairs for a random perfect matching

(b) Number of blocking pairs for a minimum-weight matching (the scale is logarithmic)

Figure 8: For Section 6.1, maps of 460 SR instances for 200 agents visualizing different quantities for each instance.

continuous shading along the horizontal axis, which highlights a clear correlation between the mutuality value and the expected number of blocking pairs. This correlation is quite intuitive: If the mutual agreement in an instance is high, then agents in this instance are also more likely to form blocking pairs. If an agent a prefers an agent b to its current partner, then, because the mutuality is high, b also tends to like a and tends to prefer a to its current partner. If there is mutual disagreement, the situation is reversed: An agent prefers the agents that tend to dislike them to its current partner, which makes blocking pairs for random matchings less likely. This is also clearly visible in Figure 8 (a), as instances close to MA have a low expected number of blocking pairs, whereas for instances close to MD the expected number is much higher. Moreover, Figure 8 (a) again validates that instances that are close to each other on the map have similar properties and that our test dataset provides a good and uniform coverage of the space of SR instances.

6.1.3 Number of Blocking Pairs for Minimum-Weight Matching

We define the minimum-weight matching M in an instance as the perfect matching minimizing the summed rank that agents assign to their partner, i.e., M minimizes $\sum_{a\in A} \mathrm{pos}_{\succ_a}(M(a))$. If stability is not vital or if a stable matching might be too complicated to compute, a minimum-weight matching is a natural candidate matching to choose and might even serve as a heuristic for choosing a stable matching. Thus, it is an interesting question "how" stable such minimum-weight matchings are. We depict in Figure 8 (b) the number of pairs that block a minimum-weight matching for all instances from our dataset. Analyzing the results, we find that for almost all of our instances from the dataset, a minimum-weight matching is only blocked by a few pairs. There are two exceptions: Instances sampled from the Reverse-Euclidean model and instances close to identity. For both of these types of instances, the number of pairs blocking the minimum-weight matching is often above 1000. For Reverse-Euclidean, we see that these instances behave very differently than instances close to them sampled from different models. Slightly counterintuitively, for this model the higher p gets (i.e., the fraction of agents preferring agents further away), the lower the number of pairs that block a minimum-weight matching. For the identity region, we see that the minimum-weight matching is blocked by many pairs in all instances from this region. In general, what stands out from the map again is that instances sampled from the same model exhibit a very uniform behavior. Overall, the outlier behavior of Reverse-Euclidean (that we will see again in Figure 9 (a)) underlines that the map is not perfect and our mutual attraction distance and matrix (naturally) do not capture all facets of similarity. Nevertheless, the other results highlight the usefulness of the map as a visualization tool and an intuition provider.

6.2 Different Types of Stable Matchings

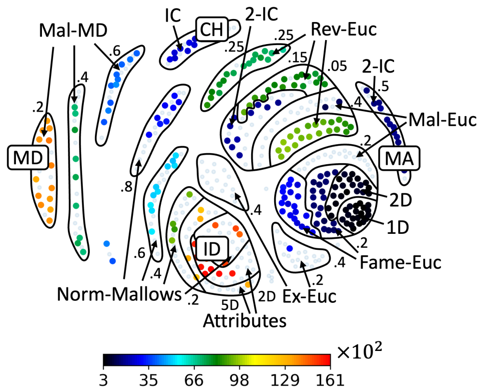
In this section, we analyze different types of stable matchings. Because multiple stable matchings in one instance can exist, formulating desirable properties for a stable matching, and computing the stable matching that performs best according to the desiderata is an active area of research (Manlove, 2013). In the following, we focus on different types of stable matchings related to agents' satisfaction.

6.2.1 Summed Rank Minimal Stable Matching

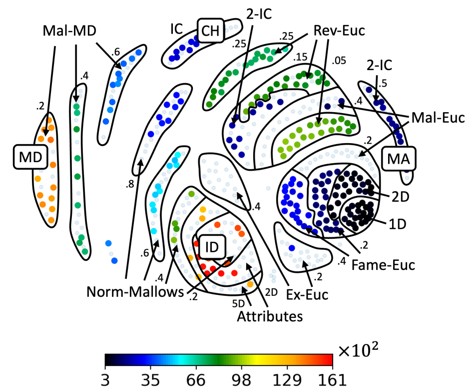
We start by analyzing summed rank minimal stable matchings, i.e., stable matchings M minimizing $\sum_{a\in A} \mathrm{pos}_{\succ_a}\big(M(a)\big)$ (these matchings are also sometimes called egalitarian matchings). Such a matching can also be interpreted as a stable matching maximizing the summed satisfaction of agents and is thus a natural candidate to pick if multiple stable matchings exist. However, computing it is NP-hard (Feder, 1992) and thus we resorted to an ILP. We visualize the quality of summed rank minimal stable matchings in Figure 9 (a) (we depict instances without a stable matching as transparent points). Focusing on instances that admit a stable matching, first observe that instances sampled from one culture again behave remarkably similarly. In addition, there is some, but certainly not a perfect correlation between the results and instances' position on the map: Ignoring Reverse-Euclidean instances which are a clear outlier here, if we move from chaos to mutual agreement, then the minimal summed rank decreases (as for perfect mutual agreement every agent can be matched to its top choice); in contrast, if we move from chaos to mutual disagreement or from chaos to identity, then the minimal summed rank increases monotonically. Remarkably, instances close to identity have a higher minimum summed rank than instances close to mutual disagreement, while for instances realizing the two matrices, the minimum summed rank for an instance with $2n$ agents is $2n^2$ for both of them.

6.2.2 Minimum Regret Stable Matching

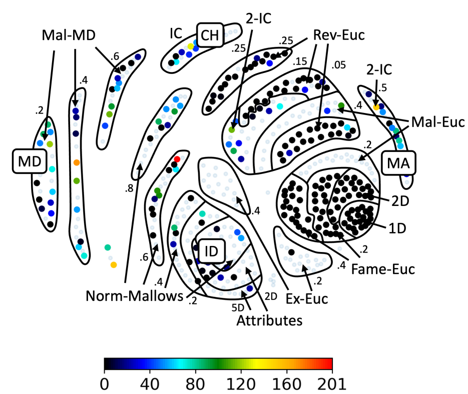
Moreover, we consider the stable matching that maximizes the satisfaction of the agent who is worst off, which is known as the minimum regret stable matching and can be computed in linear time (Gusfield & Irving, 1989). That is, we consider the maximum rank an agent assigns to its partner, i.e., $\max_{a\in A} \mathrm{pos}_{\succ_a}(M(a))$ in a stable matching M and want to minimize this value. It is another naturally attractive special stable matching. We show the results in Figure 10. Examining the map, the first remarkable observation is that



(a) Minimum summed rank of agents for partner in any stable matching (transparent points have no stable matching)



(b) Maximum summed rank of agents for partner in any stable matching (transparent points do not admit a stable matching)



(c) Difference between maximum and minimum summed rank of agents for partner in any stable matching (transparent points do not admit a stable matching)

Figure 9: For Section 6.2, maps of 460 SR instances for 200 agents visualizing different quantities for each instance.

instances that are close to Euclidean instances may behave very differently, even if they are sampled from the same statistical culture. A possible explanation for this is that in those instances stable matchings are often unique, leaving little flexibility to satisfy the worst-off agent. Moreover, again, some clear patterns on the map can be identified. For impartial culture instances and 2-IC instances the satisfaction of the worst-off agent is quite high. For instances close to mutual agreement the picture is quite mixed, while for instances close to identity, it is not possible to satisfy all agents adequately (this is quite intuitive because someone needs to be matched to the agents that are collectively considered to have a low quality in such instances). Moving from identity to mutual disagreement or chaos the situation of the worst-off agent monotonically improves.

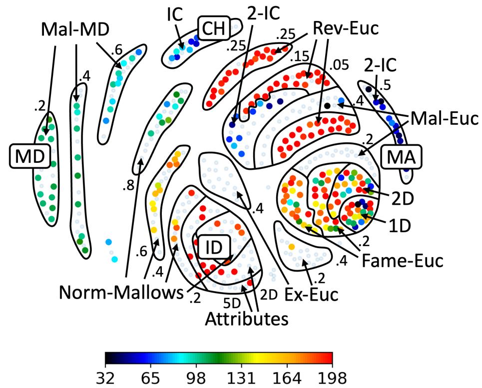


Figure 10: Maximum rank an agent assigns to its partner in a stable matching minimizing this value (transparent points do not admit a stable matching).

6.3 "Richness" of Set of Stable Matchings

Another question motivated by the fact that there might be multiple stable matchings in the same instance is to analyze how large the influence of the selected stable matching is. In other words, how much does it matter which matching is selected? We restrict our focus to the summed agent's satisfaction. For this, we consider the objective opposition to the summed rank minimal stable matching, i.e., we analyze the stable matching that maximizes the rank that agents assign to their partner (this is, in some sense, the worst stable matching that minimizes agents' satisfaction). We visualize the results in Figure 9 (b).[11]

Now, to quantify the "richness" of the set of stable matchings for an instance we use the difference between the maximum and minimum summed rank of agents for their partner in a stable matching as a proxy. We present the results in Figure 9 (c). Remarkably, this is the first of our maps where we see a very different behavior of instances sampled from the same culture. The reason for this might be that in most cases, the difference between the best and worst matching is small in comparison to the total satisfaction of the agents. This is particularly true for instances sampled from the Euclidean or similar models, which is quite intuitive, as in Euclidean instances there only exists a single stable matching.

6.4 Running Time Analysis

Lastly, to illustrate another possible application of the map, in Figure 11 we visualize the time our ILP, which we solved using Gurobi Optimization, LLC (2021), needed to find a summed rank minimal stable matching (from Section 6.2.1). Analyzing the results, again

11. While there is no simple pattern visible on the map, disregarding Reverse-Euclidean instances, there is a clear correlation between instance's behavior and their position on the map. The behavior can be nicely described by moving along the extreme matrices. Moving from mutual disagreement to chaos, the maximum summed rank monotonically decreases and if we move further towards identity (ignoring Reverse-Euclidean instances) it decreases even further. If we move from chaos to identity, then the maximum summed rank substantially increases, while moving from identity to mutual disagreement it first decreases and then increases again.

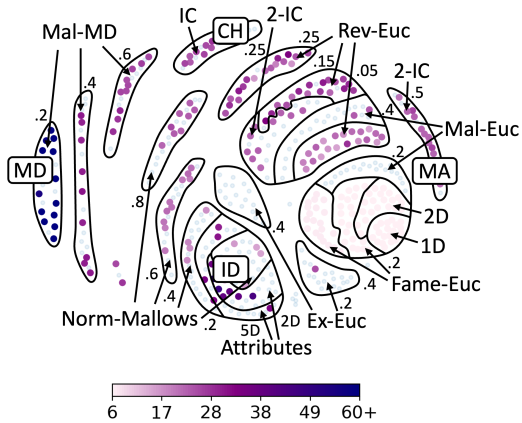


Figure 11: Seconds needed to compute summed rank minimal stable matching (transparent points have no stable matching).

instances from the same culture behave quite similarly to each other and the behavior of an instance is clearly connected to their position on the map. More specifically, instances from the Euclidean and Fame-Euclidean model seem to be particularly easy to solve, whereas instances close to ID and close to MD seem to be particularly challenging (possibly, because here the achievable minimum summed rank is quite high). Remarkably, for election-related problems typically Impartial Culture elections are the most challenging ones, and the more structure there is in an election, the easier it is to solve (Szufa et al., 2020). In sharp contrast to this, we observe that instances close to ID and MD, which are both heavily structured, are particularly challenging. We remark that naturally our observations on which instances are easy and which are hard are limited to the specific problem and solution method we considered.

7. A Map of Stable Marriage Instances

The framework developed in this paper to draw a map of synthetic Stable Roommates (SR) instances can also be applied to different types of matching under preferences problems. In this section, we demonstrate how this can be done for the Stable Marriage (SM) problem, focusing on describing the adjustments necessary compared to the discussed SR setting.

Stable Marriage Instances. A SM instance $\mathcal{I}$ consists of a set A of agents partitioned into two sets U and W, which are traditionally referred to as *men* and *women*, respectively. We assume for simplicity that $|U| = |W|$. Each man $u \in U$ has a preference order $\succ_u \in \mathcal{L}(W)$ over all women and each woman $w \in W$ has a preference order $\succ_w \in \mathcal{L}(U)$ over all men. A *matching* is a subset M of man-woman pairs where each agent appears in at most one pair and a matching is *stable* if no man-woman pair exists preferring each other to their assigned partner. Note that a stable matching is guaranteed to exist in every SM instance (Gale & Shapley, 1962).

Mutual Attraction Distance between SM Instances. Let $\mathcal{I} = (U = \{u_1, \dots, u_n\}, W = \{w_1, \dots w_n\}, (\succ_u)_{u \in U}, (\succ_w)_{w \in W})$ be an SM instance. Recall that for some $i \in [2n-1]$ and $a \in U \cup W$, $\mathcal{MA}^{\mathcal{I}}(a,i)$ is the position of a in the preference order of the agent a' which is ranked in position i by a, i.e., $\mathcal{MA}^{\mathcal{I}}(a,i) = \mathrm{pos}_{\succ_{a'}}(a)$ where $a' := \mathrm{ag}_{\succ_a}(i)$. The mutual attraction vector of an agent $a \in U \cup W$ is $\mathcal{MA}^{\mathcal{I}}(a) = \big(\mathcal{MA}^{\mathcal{I}}(a,1), \dots, \mathcal{MA}^{\mathcal{I}}(a,n)\big)$.

For each instance $\mathcal{I}$, we define two *mutual attraction matrices*: $\mathcal{MA}^{\mathcal{I},U}$ and $\mathcal{MA}^{\mathcal{I},W}$. Matrix $\mathcal{MA}^{\mathcal{I},U}$ is the mutual attraction matrix of men, where the i-th row of $\mathcal{MA}^{\mathcal{I},U}$ is the vector $\mathcal{MA}^{\mathcal{I}}(u_i)$. Matrix $\mathcal{MA}^{\mathcal{I},W}$ is the mutual attraction matrix of women, where the i-th row of $\mathcal{MA}^{\mathcal{I},W}$ is the vector $\mathcal{MA}^{\mathcal{I}}(w_i)$. Thus, notably, an SM instance does not correspond to a single mutual attraction matrix but a pair of mutual attraction matrices.

The mutual attraction distance between two SM instances $\mathcal{I}$ with agents $U \cup W$ and $\mathcal{I}'$ with agents $U' \cup W'$ with $|U| = |W| = |U'| = |W'| = n$ is defined as:

$$\begin{aligned} &\min \Big(\mathrm{d}_{\mathrm{MAD}}(\mathcal{MA}^{\mathcal{I},U}, \mathcal{MA}^{\mathcal{I}',U'}) + \mathrm{d}_{\mathrm{MAD}}(\mathcal{MA}^{\mathcal{I},W}, \mathcal{MA}^{\mathcal{I}',W'}), \\ &\mathrm{d}_{\mathrm{MAD}}(\mathcal{MA}^{\mathcal{I},U}, \mathcal{MA}^{\mathcal{I}',W'}) + \mathrm{d}_{\mathrm{MAD}}(\mathcal{MA}^{\mathcal{I},W}, \mathcal{MA}^{\mathcal{I}',U'}) \Big), \end{aligned} \tag{5}$$

where for two $n \times n$ mutual attraction matrices A and B their mutual attraction distance $\mathrm{d}_{\mathrm{MAD}}(A,B)$ is still defined as: $\min_{\sigma \in \Pi([n],[n])} \sum_{i \in [n]} \ell_1\big(A_i, B_{\sigma(i)}\big)$. The idea behind Equation (5) is that we do not fix that the "women" in one instance are matched to the "women" in the other instance (as in applications of bipartite one-to-one matching problems the two sides are in some sense exchangeable). Accordingly, we compute the distance between the two instances for both possible mappings. The first alternative is to map men in $\mathcal{I}$ to men in $\mathcal{I}'$ and women in $\mathcal{I}$ to women in $\mathcal{I}'$. The second alternative is to map men in $\mathcal{I}$ to women in $\mathcal{I}'$ and women in $\mathcal{I}$ to men in $\mathcal{I}'$. Subsequently, we return the minimum.

Navigating the Space of SM Instances. Also for SM instances, it will prove useful to identify "canonical" pairs of extreme mutual attraction matrices. The first three extreme matrices identified for SR instances are still clearly relevant here: Identity here corresponds to the situation where all women have the same preferences over the men and all men have the same preferences over the women. This results in the following pair of matrices:

$$\left(\begin{bmatrix} 1 & \dots & 1 \\ 2 & \dots & 2 \\ & \vdots & \\ n & \dots & n \end{bmatrix}, \begin{bmatrix} 1 & \dots & 1 \\ 2 & \dots & 2 \\ & \vdots & \\ n & \dots & n \end{bmatrix} \right)$$

For mutual agreement, we still require that if agent a ranks agent b in position i, then b also ranks a in position i. This results in the following pair of matrices:

$$\left(\begin{bmatrix} 1 & 2 & \dots & n-1 & n \\ 1 & 2 & \dots & n-1 & n \\ & & \vdots & & \\ 1 & 2 & \dots & n-1 & n \end{bmatrix}, \begin{bmatrix} 1 & 2 & \dots & n-1 & n \\ 1 & 2 & \dots & n-1 & n \\ & & \vdots & & \\ 1 & 2 & \dots & n-1 & n \end{bmatrix} \right)$$

Notably, this pair of matrices is realizable. We can simply partition a complete bipartite graph with n vertices on each side into n perfect matchings $M_1, \dots, M_n$, where M_i gives the agent that is ranked in the position i in the agent's preferences. By Hall's theorem, such a partition into perfect matchings always exits.

For mutual disagreement, we analogously require that if a ranks agent b in position i, then b ranks a in position $n-i+1$. This results in the following pair of matrices:

$$\left(\begin{bmatrix} n & n-1 & \dots & 2 & 1 \\ n & n-1 & \dots & 2 & 1 \\ & & \vdots & & \\ n & n-1 & \dots & 2 & 1 \end{bmatrix}, \begin{bmatrix} n & n-1 & \dots & 2 & 1 \\ n & n-1 & \dots & 2 & 1 \\ & & \vdots & & \\ n & n-1 & \dots & 2 & 1 \end{bmatrix} \right)$$

One realization of this matrix pair is an SM instance where for $i \in [n]$ woman w_i has preferences $m_{i+1} \succ m_{i+2} \succ \cdots \succ m_n \succ m_1 \succ m_2 \succ \cdots \succ m_i$ and man m_i has preferences $w_i \succ w_{i+1} \succ \cdots \succ w_n \succ w_1 \succ w_2 \succ \cdots \succ w_{i-1}$.

Our fourth extreme matrix, which is the chaos matrix, has no naturally defined analog in the SM setting, which is why we omit it.[12] Moreover, determining the maximum distance between two realizable matrix pairs remains an open problem (in our experiments, the mutual agreement and mutual disagreement pairs are furthest away).

Creating and Drawing the Map. To create our map of SM instances, we sample from statistical cultures similar to the ones for SR. Notably, to maintain focus, we assume that the preferences of both women and men are generated using the same statistical culture (of course, in principle it would also be possible to combine different cultures, but we focus on the simpler case here).

We now describe how to adapt the cultures for SR instances to the bipartite SM setting and refer to Section 5.1 for the full descriptions. For the *Impartial Culture* and *Mallows* models, using the described procedure for SR instances, we sample for each woman $w \in W$ a preference order from $\mathcal{L}(U)$ and independently for each man $u \in U$ a preference order from $\mathcal{L}(W)$. For the *2-IC* model, given some $p \in [0, 0.5]$, we partition U into two sets $U_1 \uplus U_2$ with $|U_1| = \lfloor p \cdot |U| \rfloor$ and we partition W into two sets $W_1 \uplus W_2$ with $|W_1| = \lfloor p \cdot |W| \rfloor$. Each man $u \in U$, respectively woman $w \in W$, samples one preference order $\succ$ from $\mathcal{L}(W_1)$, respectively $\mathcal{L}(U_1)$, and one order $\succ'$ from $\mathcal{L}(W_2)$, respectively $\mathcal{L}(U_2)$. If $a \in U_1 \cup W_1$, then a's preferences start with $\succ$ followed by $\succ'$. If $a \in U_2 \cup W_2$, then a''s preferences start with $\succ'$ followed by $\succ$. For the *Euclidean*, *Mallows-Euclidean*, *Expectations-Euclidean*, *Fame-Euclidean*, and *Attributes* models, we sample for each agent a point and a vector as described for the respective model for SR. Subsequently, a woman ranks all men according to their "distance" (as defined for the respective model) and a man ranks all women according to their distance. For the *Reverse-Euclidean* model, given some $p \in [0, 1]$, we partition U into two sets $U_1 \uplus U_2$ with $|U_1| = \lfloor p \cdot |U| \rfloor$ and W into two sets $W_1 \uplus W_2$ with $|W_1| = \lfloor p \cdot |W| \rfloor$. We sample the agents' preferences as in the Euclidean model and subsequently reverse the preferences of all agents in $U_2 \cup W_2$. Lastly, for *Mallows-MD*, we start with an SM instance realizing MD, where for $i \in [n]$ woman w_i has preferences $m_{i+1} \succ_{w_i} m_{i+2} \succ_{w_i} \cdots \succ_{w_i}$

12. Note that intuitively taking one matrix from the mutual agreement pair and one matrix from the mutual disagreement pair could be a viable fourth extreme point. However, it is easy to see that the resulting (and similar) matrix pairs are not realizable.

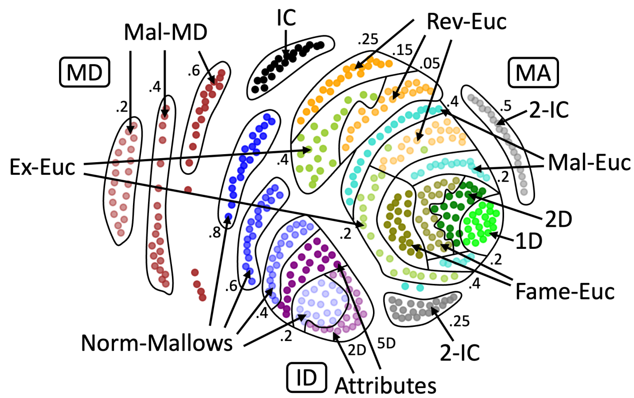


Figure 12: Map of 460 SM instances. Each instance is represented by a point. Roughly speaking, the closer two points are on the map, the more similar the respective SM instances under the mutual attraction distance. The color of a point indicates the statistical culture the respective instance was sampled from.

$m_n \succ_{w_i} m_1 \succ_{w_i} m_2 \succ_{w_i} \cdots \succ_{w_i} m_i$ and man m_i has preferences $w_i \succ_{m_i} w_{i+1} \succ_{m_i} \cdots \succ_{m_i} w_n \succ_{m_i} w_1 \succ_{m_i} w_2 \succ_{m_i} \cdots \succ_{m_i} w_{i-1}$. As for SR, for each agent $a \in U \cup W$, we obtain its final preferences by sampling a preference order from $\mathcal{M}_{\text{norm-}\phi,n,\succ_a}$.

As for SR, our dataset consists of 460 SM instances sampled from the above-described statistical cultures, where we use the same parameter configurations as for SR (as described in Section 5.1; see also Table 1). In addition, on our maps, we include the three extreme matrix pairs described previously.

Moreover, as for SR, to draw the map, we first compute the mutual attraction distance of each pair of instances and subsequently embed them into the two-dimensional Euclidean space using a variant of the force-directed Kamada-Kawai algorithm (Kamada & Kawai, 1989). We depict the map visualizing our dataset of 460 instances for 100 men and 100 women in Figure 12. Overall, the map of SM instances from Figure 12 is very similar to the map of SR instances from Figure 2, where the only cultures that are placed slightly differently in the two maps are 2-IC and Expectations-Euclidean.

Furthermore, we depict in Figure 13 the average distance between the different statistical cultures. Unsurprisingly, the general picture in Figure 13 is very similar as in Figure 5 for SR, with Expectations-Euclidean being the culture that produces the largest differences between SR and SM (which is then also reflected on the map, as this culture is placed differently in the two maps).

Using the Map. To showcase possible use cases of our map of SM instances we repeat some of the experiments that we conducted for SR in Section 6.

Specifically, analogous to Figure 8 (a), in Figure 14 (a), we visualize the average number of blocking pairs for a random perfect matching in our SM instances (by sampling 100 perfect matchings and taking the average of the number of pairs blocking them). The

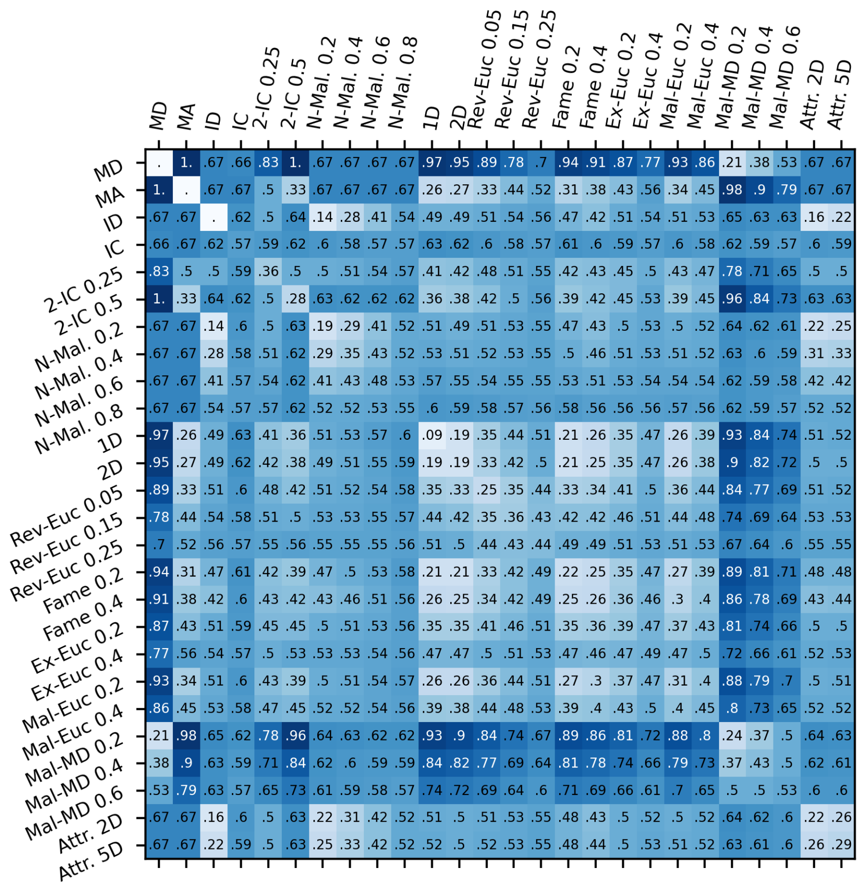
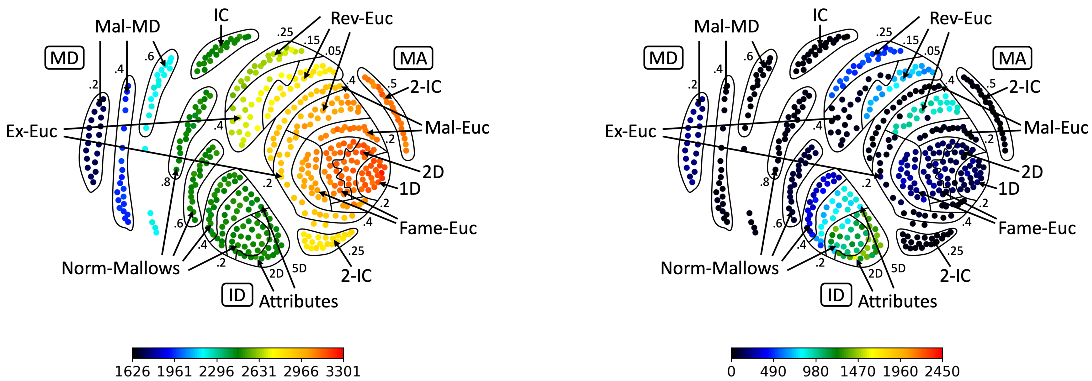


Figure 13: For each pair of statistical cultures for sampling SM instances, average normalized mutual attraction distance between instances sampled from the two. The first three lines/columns contain, for each statistical culture, the average distance of instances sampled from the culture to our three extreme matrix pairs. The diagonal contains the average distance of two instances sampled from the same statistical culture.

results are very similar as for SR and in particular the average number of blocking pairs strongly correlates with the position of instances on the map.

Moreover, analogous to Section 6.1.3, in Figure 14 (b), we show the number of blocking pairs for a matching minimizing the summed rank agents have for their partner. Remarkably, for a large majority of instances, this matching is quite close to being stable. The general picture here is again very similar as for SR; in particular, for both SM and SR the different statistical cultures produce instances with very similar properties.

Next, we consider the stable matching that minimizes/maximizes the summed rank that agents have for their partner (as in Section 6.2.1). We show the summed rank that agents have for their partner in Figure 15 (a) for the summed rank minimal matching and in Figure 15 (b) for the summed rank maximal matching. The general picture here looks again very similar as for SR. In particular, for certain regions on the map instances falling



(a) Average number of blocking pairs for a perfect matching

(b) Number of blocking pairs for a minimum-weight matching (the scale is logarithmic)

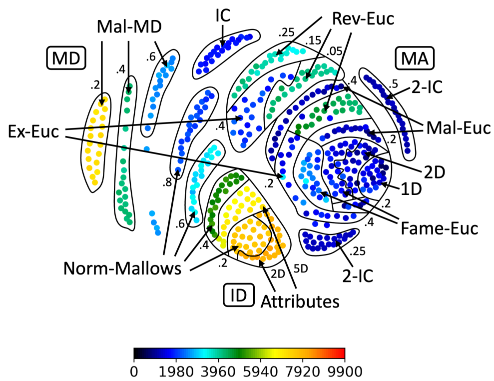
Figure 14: Map of 460 SM instances for 100 men and 100 women visualizing different quantities for each instance.

in this region show a uniform behavior. Moreover, in Figure 15 (c), we show the difference between the summed rank agents have for their partner in the stable matching maximizing and minimizing this value. Comparing this map to the respective map for SR, what stands out is that for SM for some instances there is a larger difference between the summed rank minimal and summed rank maximal matching than for SR, indicating that the space of stable matchings for some of the sampled SM instances is "richer". Nevertheless, still for most of our SM instances there is only little difference between the summed rank minimal and maximal matching; this holds in particular for most of the instances from the bottom-right region of the map.

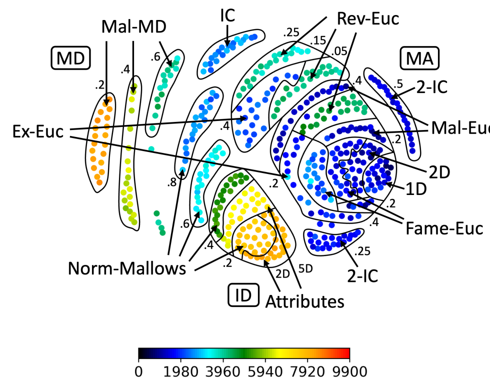
Lastly, analogous to Section 6.2.2, in Figure 16, we depict the maximum rank an agent has for its partner in a stable matching minimizing this value. Notably, here, the results for SM differ from the results for SR. In particular, for SM, there are more instances where some agent is always matched to its almost least preferred agent than for SR (this contrast is most clear for 1D- and 2D-Euclidean instances). Overall, ignoring 2-IC, in Figure 16, a split of the map for SM instances is visible where in instances from the bottom right part some agent is matched to one of its least preferred agents in every stable matching, whereas in instances from the top left part of the map, in some stable matching even the worst-off agent is matched to a partner that is not in the bottom 20% of its preferences.

8. Discussion

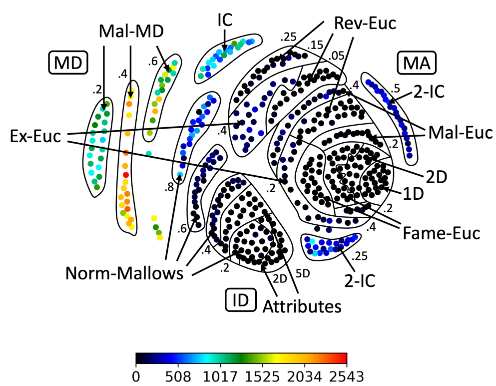
Contributing to the toolbox for experiments for stable matching problems, we have introduced the polynomial-time computable mutual attraction distance and analyzed its properties as well as the space it induces. As a second step, we have described a variety of statistical cultures for generating synthetic stable matching instances. In our following experiments, we have provided evidence that the instances produced by the different cultures



(a) Minimum summed rank of agents for partner in any stable matching



(b) Maximum summed rank of agents for partner in any stable matching



(c) Difference between maximum and minimum summed rank of agents for partner in any stable matching

Figure 15: Map of 460 SM instances for 100 men and 100 women visualizing different quantities for each instance.

behave quite differently from each other, which makes our collection of cultures well-suited to create diverse easily customizable test datasets. One specific application of these two contributions is our map of stable matching instances. We have verified that the produced map is meaningful in the sense that it groups instances with similar properties together, and have provided intuitive interpretations of the different regions on the map.

To demonstrate the capabilities of the map and our test dataset, we have conducted some example experiments. Overall, our experimental results underline the importance of using diverse test data. Among others, we have observed that sampling preferences uniformly at random results in instances that behave very similarly (and often very differently than instances sampled from other models) and that such instances only cover a small part of

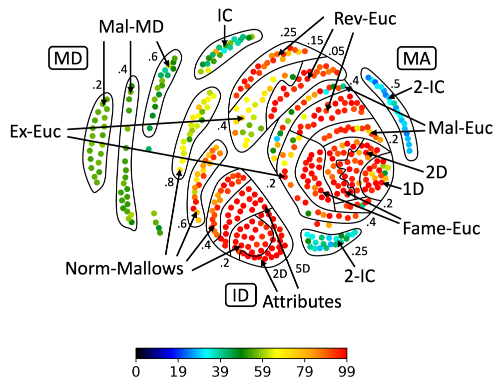


Figure 16: Map of 460 SM instances for 100 men and 100 women visualizing the maximum rank an agent assigns to its partner in a stable matching minimizing this value.

the space of instances. Overall, this questions the common practice of only examining preferences sampled uniformly at random in experiments, as it is quite unclear whether the results of these experiments generalize.[13] Specifically, our results presented in Section 6 (and Section 7) demonstrate the insufficiency of only using uniformly at random sampled preferences in SM and SR instances when analyzing properties of specific stable matchings and performances of algorithms. For instance, assume that one would be interested in analyzing the maximum total satisfaction of agents in a stable matching in an SR instance (cf. Figure 9 (a)) and only examine instances generated from the Impartial Culture model. Then, one's conclusion would probably be along the following lines: "The maximum total satisfaction of agents in a stable matching is similar in all instances and generally quite high, i.e., in an SR instance with 200 agents the average rank that agents have for their partner is close to 15." However, this conclusion is rendered invalid as soon as one moves beyond Impartial Culture instances. As we observe in Section 6.2.1, the maximum total satisfaction of agents highly varies based on the instance, with some instances admitting matchings where everyone is almost perfectly satisfied while others only admit matching where the total satisfaction of agents is quite low. For instance, in SR instances with 200 agents sampled from the Mal-MD model with parameter 0.2 the average rank that agents have for their partner is never below 60, a behavior that would remain unseen when only examining Impartial Culture instances. This is slightly worrisome given that in the past several papers have analyzed properties of different types of stable matchings (Cooper & Manlove, 2019, 2020) and conducted performance evaluations of algorithms (Genc et al., 2017; Cooper & Manlove, 2020; Erdem et al., 2020) in SM and SR instances only using random preferences.

13. Further, it is also unclear why instances with uniformly at random sampled preferences are particularly practically useful. Quite the contrary, there is some evidence that preferences, in reality, are often not drawn uniformly at random: In the general setting of agents ranking different alternatives, Boehmer et al. (2021b) and Boehmer and Schaar (2023) analyzed real-world preference data from a variety of sources observing that only few of these match a random preference sampling.

Moreover, our experiments and analysis of the map also reveal that instances that have a small mutual attraction distance (and thus are close to each other on the map) tend to have similar properties. This underlines the usefulness of the mutual attraction distance measure to assess the similarity of instances. Furthermore, it highlights the capabilities of the map as a non-aggregate visualization tool: Instead of presenting experimental results by listing different (sometimes non-robust) statistical quantities, on the map, we can depict the results on an instance level, thereby showing the full picture. Using this, it is often possible to identify general high-level trends and typical behavior of instances from different parts of the space. In a similar vein, the map also supports the informed planning of more focused follow-up experiments, by looking for parts on the map that show an interesting behavior and analyzing the respective cultures in more detail. To use the map for this purpose, a meaningful placement of the instances on the map which groups similar instances together is vital. In fact, the maps shown in this paper provide some initial clear evidence that this is indeed the case, which also justifies the usage of the mutual attraction distance as a practically useful and sufficient distance measure.

For future work, from a theoretical perspective, it would be interesting to extend our theoretical analysis of the space of SR instances to SM instances, and to analyze the theoretical properties of the space of SM and SR instances induced by other distance measures such as the swap or Spearman distance. From a more practical perspective, it would be very valuable to see where real-world instances lie on the map, which first of all requires the collection of such data. Moreover, while our experiments indicate that our collected dataset is quite diverse, it is also clear that not all possible behaviors will be seen in one of its instances. For instance, our dataset does not contain any instances where every matching is "far away" from being stable, indicating that there are certain "holes" in the space of stable matching instances covered by our dataset. Designing cultures that generate instances falling into these holes and thereby further extending the diversity of the dataset would be a valuable addition. Lastly, applying our framework to other types of stable matching problems is another interesting direction for the future.

Acknowledgments

NB was supported by the DFG project ComSoc-MPMS (NI 369/22). KH was supported by the DFG project FPTinP (NI 369/16). This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 101002854). A short version of the paper appears in the proceedings of the *2023 International Conference on Autonomous Agents and Multiagent Systems* (AAMAS 2023), pages 1003–1011 and won the Pragnesh Jay Modi Best Student Paper Award at AAMAS 2023. The paper has also been presented at the 6th International Workshop on Matching Under Preferences (MATCH-UP 2022). We are grateful to the anonymous *MATCH-UP* 2022 and *AAMAS* 2023 reviewers for their thoughtful, constructive, and helpful comments.

Appendix A. Additional Material for Section 4

In this section, we present all missing proofs from Section 4.

A.1 Identity

Given $n \in \mathbb{N}$, the $2n \times (2n-1)$-matrix ID^{2n} can be written as follows:

$$\begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 & 1 \\ 1 & 2 & 2 & 2 & \dots & 2 & 2 \\ 2 & 2 & 3 & 3 & \dots & 3 & 3 \\ 3 & 3 & 3 & 4 & \dots & 4 & 4 \\ & & & & \vdots & & \\ 2n-2 & 2n-2 & 2n-2 & 2n-2 & \dots & 2n-2 & 2n-1 \\ 2n-1 & 2n-1 & 2n-1 & 2n-1 & \dots & 2n-1 & 2n-1 \end{bmatrix}$$

Proposition 24. *For every $n \in \mathbb{N}$, an SR instance $\mathcal{I}$ is a realization of ID^{2n} if and only if the preferences in $\mathcal{I}$ are derived from a master list. In particular, the realization of ID^{2n} is unique.*

Proof. ($\Rightarrow$): Let $\mathcal{I}^*$ be a realization of ID^{2n}. We refer to the agents from $\mathcal{I}^*$ as $a_1, a_2, \dots, a_{2n}$ and assume without loss of generality that the i-th row of ID^{2n} belongs to a_i for every $i \in [2n]$. We show by induction on i that agent a_i is the i-th agent in the preferences of a_j for $j > i$ and the $(i-1)$-th agent in the preferences of a_j for $j < i$ (which is equivalent to $\mathcal{I}^*$ being derived from the master list $a_1 \succ a_2 \succ \dots \succ a_{2n}$). For $i = 0$, there is nothing to show (as agent a_0 does not exist). So fix $i > 0$. The i-th row of ID^{2n} contains the entry "$i-1$" precisely $i-1$ times and the entry i precisely $2n-i$ times. By the induction hypothesis, the preferences of a_j for $j > i$ start with $a_1 \succ a_2 \succ \dots \succ a_{i-1}$, so a_i cannot be the $(i-1)$-th agent in the preferences of a_j. Consequently, a_i is the i-th agent in the preferences of a_j for every $j > i$. It follows that a_i is the $(i-1)$-th agent in the preferences of all remaining agents, that is, of agent a_j for every $j < i$ (because the $2n-i$ entries of value i in the i-th row belong to agents $a_{j'}$ for $j' > i$).

($\Leftarrow$): Let $\mathcal{I}$ be derived from the master list $a_1 \succ a_2 \succ \dots \succ a_{2n}$. Then a_i is the $(i-1)$-th agent in the preferences of a_j for $j < i$ and the i-th agent in the preferences of a_j for $j > i$. It follows that the mutual attraction matrix R of $\mathcal{I}$ fulfills that $R[i,j] = \begin{cases} i & j \geq i \\ i-1 & j < i \end{cases}$, i.e., we have $R = \mathrm{ID}^{2n}$. ∎

A.2 Mutual Agreement

Given $n \in \mathbb{N}$, the $2n \times (2n-1)$-matrix MA^{2n} can be written as follows:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & \dots & 2n-2 & 2n-1 \\ 1 & 2 & 3 & 4 & \dots & 2n-2 & 2n-1 \\ 1 & 2 & 3 & 4 & \dots & 2n-2 & 2n-1 \\ & & & & \vdots & & \\ 1 & 2 & 3 & 4 & \dots & 2n-2 & 2n-1 \\ 1 & 2 & 3 & 4 & \dots & 2n-2 & 2n-1 \end{bmatrix}$$

Proposition 25. *For every $n \in \mathbb{N}$, there is a bijection between realizations of MA^{2n} and the set of Round-Robin tournaments. In particular, there are several non-isomorphic realizations of MA^{2n} for $n = 4$.*

Proof. We prove the statement by giving an injective function from the realizations of MA^{2n} to the set of Round-Robin tournaments as well as an injective function from Round-Robin tournaments to realizations of MA^{2n}.

Formally, we interpret Round-Robin tournaments as colored 1-factorizations of the complete graph K_{2n}.

Definition 26. *A* colored 1-factorization *of the complete graph K_{2n} is a function $f : E(K_{2n}) \to [2n-1]$ such that $f^{-1}(j)$ is a perfect matching for every $j \in [2n-1]$.*

Accordingly, in the following, we speak of colored 1-factorizations of K_{2n}. Further, we identify the vertices of K_{2n} with the agents of a realization of MA^{2n}.

We start by describing an injective function from the realizations of MA^{2n} to the set of colored 1-factorizations of K_{2n}. Let $\mathcal{I}$ be a realization of MA^{2n}. We create a colored 1-factorization f of K_{2n} as follows (where we identify the vertices of K_{2n} with the agents of $\mathcal{I}$). For every $j \in [2n-1]$, we set $f(\{a, a'\}) = j$ if and only if a is the j-th agent in the preferences of a'. Since $\mathrm{MA}^{2n}[i,j] = j$ for every $j \in [2n-1]$, this is indeed a colored 1-factorization. This function from the realizations of MA^{2n} to the colored 1-factorizations of K_{2n} is clearly injective.

Next, we give an injective function from the set of colored 1-factorizations to the realizations of MA^{2n}. Given a colored 1-factorization on K_{2n}, we create preferences for each vertex as follows: The j-th vertex in the preferences of some vertex v is the vertex w such that $f(\{v,w\}) = j$. Since we have a colored 1-factorization, these are well-defined and feasible preferences. Let $\mathcal{I}$ be the resulting SM instance. The function is clearly injective. It remains to show that the mutual attraction matrix R of $\mathcal{I}$ is MA^{2n}. For every vertex v with vertex w at the j-th position in its preferences, we have $f(\{v,w\}) = j$, implying that also w has v in the j-th position in its preferences. Consequently, we have $R[i,j] = j$ for every $i \in [2n]$ and $j \in [2n-1]$, i.e., $R = \mathrm{MA}^{2n}$ for $n \geq 4$.

We remark that in general, there may be multiple non-isomorphic 1-factorizations (Dickson & Safford, 1906) and thus also multiple non-isomorphic realizations of MA^{2n}. □

A.3 Mutual Disagreement

Given $n \in \mathbb{N}$, the $2n \times (2n-1)$-matrix MD^{2n} is defined as follows:

$$\begin{bmatrix} 2n-1 & 2n-2 & 2n-3 & 2n-4 & \dots & 2 & 1 \\ 2n-1 & 2n-2 & 2n-3 & 2n-4 & \dots & 2 & 1 \\ 2n-1 & 2n-2 & 2n-3 & 2n-4 & \dots & 2 & 1 \\ & & & & \vdots & & \\ 2n-1 & 2n-2 & 2n-3 & 2n-4 & \dots & 2 & 1 \\ 2n-1 & 2n-2 & 2n-3 & 2n-4 & \dots & 2 & 1 \end{bmatrix}$$

Proposition 27. *For every $n \in \mathbb{N}$, matrix MD^{2n} is realizable. For $n = 3$, matrix MD^{2n} has multiple non-isomorphic realizations.*

Proof. We start by proving the first part. The following preferences on agents $a_1, \dots, a_{2n}$ are a realization for MD^{2n}:

$$a_i : a_{i+1} \succ a_{i+2} \succ \dots \succ a_{2n} \succ a_1 \succ a_2 \succ \dots \succ a_{i-1}\,.$$

It remains to show that this is indeed a realization of MD^{2n}. So fix $i \in [2n]$ and $j \in [2n-1]$. For the rest of the proof, all indices are taken modulo $2n$. The j-th position in the preferences of a_i is a_{i+j}. In the preferences of a_{i+j}, agent a_i is at the $(2n-j)$-th position. Consequently, the preference profile realizes MD^{2n}.

We now turn to the second part. For $n = 3$, the following realization is not isomorphic to the one described above, but its mutual agreement matrix is MD^{2n}:

$$\begin{aligned}
a_1 &: a_2 \succ a_3 \succ a_4 \succ a_5 \succ a_6 \\
a_2 &: a_4 \succ a_6 \succ a_5 \succ a_3 \succ a_1 \\
a_3 &: a_5 \succ a_2 \succ a_6 \succ a_1 \succ a_4 \\
a_4 &: a_3 \succ a_5 \succ a_1 \succ a_6 \succ a_2 \\
a_5 &: a_6 \succ a_1 \succ a_2 \succ a_4 \succ a_3 \\
a_6 &: a_1 \succ a_4 \succ a_3 \succ a_2 \succ a_5
\end{aligned}$$ ☐

A.4 Chaos

Given $n \in \mathbb{N}$, the $2n \times (2n-1)$-matrix CH^{2n} can be written as follows:

$$\begin{bmatrix}
1 & 2 & 3 & 4 & \dots & 2n-2 & 2n-1 \\
1 & n+1 & 2 & n+2 & \dots & 2n-1 & n \\
2 & n+2 & 3 & n+3 & \dots & 1 & n+1 \\
3 & n+3 & 4 & n+4 & \dots & 2 & n+2 \\
4 & n+4 & 5 & n+5 & \dots & 3 & n+3 \\
 & & & & \vdots & & \\
2n-2 & n-1 & 2n-1 & n & \dots & 2n-3 & n-2 \\
2n-1 & n & 1 & n+1 & \dots & 2n-2 & n-1
\end{bmatrix}$$

Proposition 28. *For every $n \in \mathbb{N}$ such that $2n-1$ is not divisible by 3, matrix CH^{2n} is realizable, and the realization is unique.*

Proof. We remark that for each $\ell \in [n]$ and $i \in [2, 2n]$ it holds that $\mathrm{CH}^{2n}[i][2\ell-1] = i+\ell-2 \mod 2n-1$. Further, for each $\ell \in [n-1]$ it holds that $\mathrm{CH}^{2n}[i][2\ell] = i+\ell+n-2 \mod 2n-1$. First, we show that $n-1$, n, and $n+1$ are coprime to $2n-1$ (assuming that $2n-1$ is not divisible by 3). Any integer dividing $n-1$ and $2n-1$ also divides $2n-1-2\cdot(n-1) = 1$. Any integer dividing n and $2n-1$ also divides $2\cdot n-(2n-1) = 1$. Any integer dividing $n+1$ and $2n-1$ also divides $2\cdot(n+1)-(2n-1) = 3$. Since 3 is prime and does not divide $2n-1$, it follows that $n+1$ and $2n-1$ are coprime.

Since $2n-1$ and n are coprime, each row of CH^{2n} contains each number from $[2n-1]$ exactly once. Since the entries in the first column of rows 2 to $2n$ are $1, 2, \dots, 2n-1$, it follows that each number from $[2n-1]$ is contained exactly once in each row and once in each column of the submatrix arising from deleting the first row. Consequently, for every $j \in [2n-1]$, there exists exactly one $i \in \{2, 3, 4, \dots, 2n\}$ with $\mathrm{CH}^{2n}[i,j] = j$.

Fix $i > 1$ and let $j \in [2n-1]$. Let $j^* := \mathrm{CH}^{2n}[i][j]$ and assume that $j^* \neq j$. We claim that there exists exactly one $i^* \in \{2, 3, \dots, 2n\}$ such that $\mathrm{CH}^{2n}[i^*][j^*] = j$ and that $i^* \neq i$. Existence and uniqueness of i^* follows as j is contained once in each column (so in particular also in column j^*). It remains to show that $i^* \neq i$. So assume towards a contradiction that $i^* = i$. We have $j^* = i + nj - n - 1 \mod 2n - 1$. Since $j^* \neq j$, it follows that

$$i + (n-1) \cdot j - n - 1 \neq 0 \mod 2n - 1. \tag{6}$$

Since $\mathrm{CH}^{2n}[i][j^*] = j$, we have $i + n(i + nj - n - 1) - n - 1 = j \mod 2n - 1$ which (using the identity $(n^2 - 1)j = (n+1)(n-1)j$) is equivalent to

$$(n+1) \cdot (i + (n-1) \cdot j - n - 1) = 0 \mod 2n - 1 \tag{7}$$

Since $n+1$ and $2n-1$ are coprime, Equation (7) implies that $i + (n-1) \cdot j - n - 1 = 0 \mod 2n - 1$, contradicting Equation (6).

In the realization of CH^{2n}, agent a_i ranks at position j with $\mathrm{CH}^{2n}[i][j] = j^*$ the agent i^* with $\mathrm{CH}^{2n}[i^*][j^*] = j$. The resulting preference profile clearly realizes CH^{2n}, so it remains to show that it is indeed a preference profile, i.e., each agent appears exactly once in the preferences of some other agent. It suffices to show that no agent a_{i^*} appears twice in the preferences of an agent a_i. So assume towards a contradiction that a_{i^*} appears at positions j_1 and j_2 (where $j_1 \neq j_2$) in the preference of a_i. Thus, we have $j_1 = i^* + ni + n^2 j_1 - n^2 - 2n - 1 \mod 2n - 1$ and $j_2 = i^* + ni + n^2 j_2 - n^2 - 2n - 1 \mod 2n - 1$. Consequently, we have $(n^2 - 1) \cdot j_1 = i^* + ni - n^2 - 2n - 1 = (n^2 - 1) j_2 \mod 2n - 1$. It follows that $(n-1)(n+1) \cdot (j_1 - j_2) = 0 \mod 2n - 1$. Since $n-1$ and $2n-1$ as well as $n+1$ and $2n-1$ are coprime, it follows that $j_1 = j_2 \mod 2n - 1$, a contradiction.

The uniqueness of the realization is easy to see as agent a_i must rank at position j an agent a_{i^*} with $\mathrm{CH}^{2n}[i^*][\mathrm{CH}^{2n}[i][j]] = j$. □

A.5 Proof of Proposition 23

This section is devoted to proving the following statement:

Proposition 29. *For each $n \in \mathbb{N}$, we have*

1. $\mathrm{d}_{\mathrm{MAD}}(MA^{2n}, MD^{2n}) = 4 \cdot (n-1) \cdot n^2$,
2. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, MA^{2n}) = \mathrm{d}_{\mathrm{MAD}}(MA^{2n}, CH^{2n}) = \frac{8}{3}n^3 - 4n^2 + \frac{4}{3}n$,
3. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, MD^{2n}) = \mathrm{d}_{\mathrm{MAD}}(MD^{2n}, CH^{2n}) = \frac{8}{3}n^3 - 2n^2 - \frac{2}{3}n$,
4. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, CH^{2n}) = \frac{8}{3}n^3 \pm O(n^2)$.

Lemma 30. $\mathrm{d}_{\mathrm{MAD}}(MA^{2n}, MD^{2n}) = 4 \cdot (n-1) \cdot n^2$.

Proof. Independent of the mapping of agents, we get the following distance per row:

$$\begin{aligned}\sum_{j=1}^{2n-1} |j - (2n-j)| = \sum_{j=1}^{2n-1} |2j - 2n| &= \sum_{j=1}^{n-1} (2n - 2j) + \sum_{j=n+1}^{2n-1} (2j - 2n) \\ &= 2n \cdot (n-1) - 2\sum_{j=1}^{n-1} j + 2\sum_{j=1}^{n-1} j = 2n \cdot (n-1)\,.\end{aligned}$$

Summing over all $2n$ rows proves the lemma. □

Note that Proposition 16 and Lemma 30 imply that the distance between MA^{2n} and MD^{2n} is the maximum possible distance between any two realizable matrices.

Lemma 31. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, MA^{2n}) = \frac{8}{3}n^3 - 4n^2 + \frac{4}{3}n$.

Proof. For the i-th row of ID^{2n}, the distance to any row from MA^{2n} is

$$\begin{aligned}\sum_{j=1}^{i-1}((i-1)-j)+\sum_{j=i}^{2n-1}(j-i) &= (i-1)^2 - \sum_{j=1}^{i-1} j + \sum_{j=1}^{2n-1-i} j \\ &= (i-1)^2 - 0.5\cdot(i-1)\cdot i + 0.5\cdot(2n-1-i)\cdot(2n-i)\end{aligned}$$

This is a polynomial of second degree in i and n. Thus, summing over all rows (i.e., from $i = 1$ to $2n$) yields a polynomial of third degree in n. This is uniquely determined by any four points, e.g., by $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2\cdot 1}, \mathcal{MA}^{2\cdot 1}) = 0$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2\cdot 2}, \mathcal{MA}^{2\cdot 2}) = 8$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2\cdot 3}, \mathcal{MA}^{2\cdot 3}) = 32$, and $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2\cdot 4}, \mathcal{MA}^{2\cdot 4}) = 112$. Consequently, we have $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2n}, \mathrm{MA}^{2n}) = \frac{8}{3}n^3 - 4n^2 + \frac{4}{3}n$. □

Lemma 32. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, MD^{2n}) = \frac{8}{3}n^3 - 2n^2 - \frac{2}{3}n$.

Proof. For any $i \le n$, the distance of the i-th row of ID^{2n} to any row from MD^{2n} is

$$\sum_{j=1}^{i-1}(2n-j-(i-1)) + \sum_{j=i}^{2n-i}(2n-j-i) + \sum_{j=2n-i+1}^{2n-1} i-(2n-j)$$

This is a polynomial of second degree in n and i.

For any $i > n$, the distance of the i-th row of ID^{2n} to any row from MD^{2n} is

$$\begin{aligned}&\sum_{j=1}^{2n-i}(2n-j-(i-1)) + \sum_{j=2n-i+1}^{i-1}(i-1-(2n-j)) + \sum_{j=i}^{2n-1} i-(2n-j) \\ &\quad= \sum_{j=1}^{2n-i}(2n-j-(i-1)) + \sum_{j=2n-i+1}^{2n-1} i-(2n-j)-(i-(2n-i+1))\end{aligned}$$

Again, this is a polynomial of second degree in n and i.

Consequently, summing up the distance over all rows (i.e., summing over i from 1 to $2n$) yields a polynomial of third degree in n. A polynomial of third degree is uniquely characterized by any four points on the polynomial. Using $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^2, \mathrm{MD}^2) = 0$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^4, \mathrm{MD}^4) = 12$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^6, \mathrm{MD}^6) = 52$, and $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^8, \mathrm{MD}^8) = 136$, we get that $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2n}, \mathrm{MD}^{2n}) = \frac{8}{3}n^3 - 2n^2 - \frac{2}{3}n$. □

Lemma 33. $\mathrm{d}_{\mathrm{MAD}}(MA^{2n}, CH^{2n}) = \frac{8}{3}n^3 - 4n^2 + \frac{4}{3}n$.

Proof. As every row from MA^{2n} is identical, the mapping between the rows is irrelevant. The first row of CH^{2n} is identical to any row of MA^{2n} and thus does not contribute to the mutual attraction distance. Fix some $i > 1$. Let $j^*_{\text{odd}} := \min\{n, 2n+1-i\}$ and $j^*_{\text{even}} := \min\{n-1, n+1-i\}$. For $i > 1$, the i-th row of CH^{2n} contributes

$$
\begin{aligned}
\sum_{j=1}^{2n-1} |j - \mathrm{CH}^{2n}[i][j]| &= \sum_{\ell=1}^{n} |\mathrm{CH}^{2n}[i][2\ell-1] - (2\ell-1)| + \sum_{\ell=1}^{n-1} |\mathrm{CH}^{2n}[i][2\ell] - 2\ell| \\
&= \sum_{\ell=1}^{j^*_{\text{odd}}} |i+\ell-2-2\ell+1| + \sum_{\ell=j^*_{\text{odd}}}^{n} |i+\ell-2-(2n-1)-2\ell+1| \\
&+ \sum_{\ell=1}^{j^*_{\text{even}}} |i+\ell+n-2-2\ell| + \sum_{\ell=j^*_{\text{even}}}^{n} |i+\ell+n-2-(2n-1)-2\ell| \\
&= \sum_{\ell=1}^{j^*_{\text{odd}}} |i-\ell-1| + \sum_{\ell=j^*_{\text{odd}}}^{n} |i-\ell-2n| \\
&+ \sum_{\ell=1}^{j^*_{\text{even}}} |i-\ell+n-2| + \sum_{\ell=j^*_{\text{even}}}^{n} |i-\ell-n-1|
\end{aligned}
$$

One easily verifies that this is a polynomial of second degree in n and i. Consequently, summing up the distance over all rows (i.e., summing over i from 1 to $2n$) yields a polynomial of third degree in n. A polynomial of third degree is uniquely characterized by any four points on the polynomial. Using $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 1}, \mathrm{MA}^{2\cdot 1}) = 0$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 3}, \mathrm{MA}^{2\cdot 3}) = 40$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 4}, \mathrm{MA}^{2\cdot 4}) = 112$, and $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 6}, \mathrm{MA}^{2\cdot 6}) = 448$, we get that $\mathrm{d}_{\mathrm{MAD}}(\mathrm{MA}^{2n}, \mathrm{CH}^{2n}) = \frac{8}{3}n^3 - 4n^2 + \frac{4}{3}n$. $\square$

Lemma 34. $\mathrm{d}_{\mathrm{MAD}}(MD^{2n}, CH^{2n}) = \frac{8}{3}n^3 - 2n^2 - \frac{2}{3}n$.

Proof. As every row from MD^{2n} is identical, the mapping between the rows is irrelevant. The first row of CH^{2n} is identical to any row of MD^{2n} and thus does not contribute to the distance. Fix some $i > 1$. Let $j^*_{\text{odd}} := \min\{n, 2n+1-i\}$ and $j^*_{\text{even}} := \min\{n-1, n+1-i\}$.

For $i > 1$, the i-th row of CH^{2n} contributes

$$\begin{aligned}
\sum_{j=1}^{2n-1} |2n - j - \mathrm{CH}^{2n}[i][j]| = &\sum_{\ell=1}^{n} |\mathrm{CH}^{2n}[i][2\ell-1] - (2n-(2\ell-1))| \\
&+ \sum_{\ell=1}^{n-1} |\mathrm{CH}^{2n}[i][2\ell] - (2n-2\ell)| \\
&= \sum_{\ell=1}^{j^*_{\text{odd}}} |i+\ell-2-2n+2\ell-1| + \sum_{\ell=j^*_{\text{odd}}}^{n} |i+\ell-2-(2n-1)-2n+2\ell-1| \\
&+ \sum_{\ell=1}^{j^*_{\text{even}}} |i+\ell+n-2-2n+2\ell| + \sum_{\ell=j^*_{\text{even}}}^{n} |i+\ell+n-2-(2n-1)-2n+2\ell| \\
&= \sum_{\ell=1}^{j^*_{\text{odd}}} |i+3\ell-3-2n| + \sum_{\ell=j^*_{\text{odd}}}^{n} |i+3\ell-4n-2| \\
&+ \sum_{\ell=1}^{j^*_{\text{even}}} |i+3\ell-n-2| + \sum_{\ell=j^*_{\text{even}}}^{n} |i+3\ell-3n-1|
\end{aligned}$$

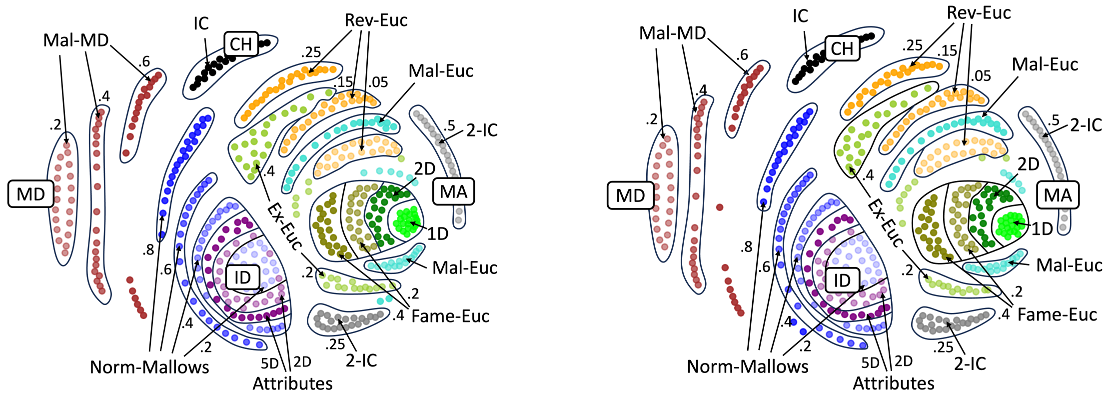
One easily verifies that this is a polynomial of second degree in n and i. Consequently, summing up the distance over all rows yields a polynomial of third degree in n. A polynomial of third degree is uniquely characterized by any four points on the polynomial. Using $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 1}, \mathrm{MD}^{2\cdot 1}) = 0$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 3}, \mathrm{MD}^{2\cdot 3}) = 52$, $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 4}, \mathrm{MD}^{2\cdot 4}) = 136$, and $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2\cdot 6}, \mathrm{MD}^{2\cdot 6}) = 500$, we get that $\mathrm{d}_{\mathrm{MAD}}(\mathrm{CH}^{2n}, \mathrm{MD}^{2n}) = \frac{8}{3}n^3 - 2n^2 - \frac{2}{3}n$. □

Lemma 35. $\mathrm{d}_{\mathrm{MAD}}(ID^{2n}, CH^{2n}) = \frac{8}{3}n^3 \pm O(n^2)$

Proof. We only prove the leading term. In order to do so, we use a (non-realizable) $2n \cdot (2n-1)$-matrix ID^* which is very similar to ID^{2n} but has an even simpler structure: The i-th row of ID^* consists solely of i's, i.e., $\mathrm{ID}^*[i][j] = i$ for every $j \in [2n-1]$. We have $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2n}, \mathrm{ID}^*) = O(n^2)$ as $\mathrm{ID}^*[i][j] - \mathrm{ID}^{2n}[i][j] \in \{0,1\}$ for every $i \in [2n]$ and $j \in [2n-1]$. Since every row of ID^* contains only one number and each row of MA^{2n} as well as CH^{2n} contain each number from $[2n-1]$ exactly once, it follows that $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^*, \mathrm{MA}^{2n}) = \mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^*, \mathrm{CH}^{2n})$. Thus, we have

$$\begin{aligned}
\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2n}, \mathrm{CH}^{2n}) &= \mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^*, \mathrm{CH}^{2n}) \pm O(n^2) \\
&= \mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^*, \mathrm{MA}^{2n}) \pm O(n^2) \\
&= \mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2n}, \mathrm{MA}^{2n}) \pm O(n^2) \\
&= \frac{8}{3}n^3 \pm O(n^2)
\end{aligned}$$ □

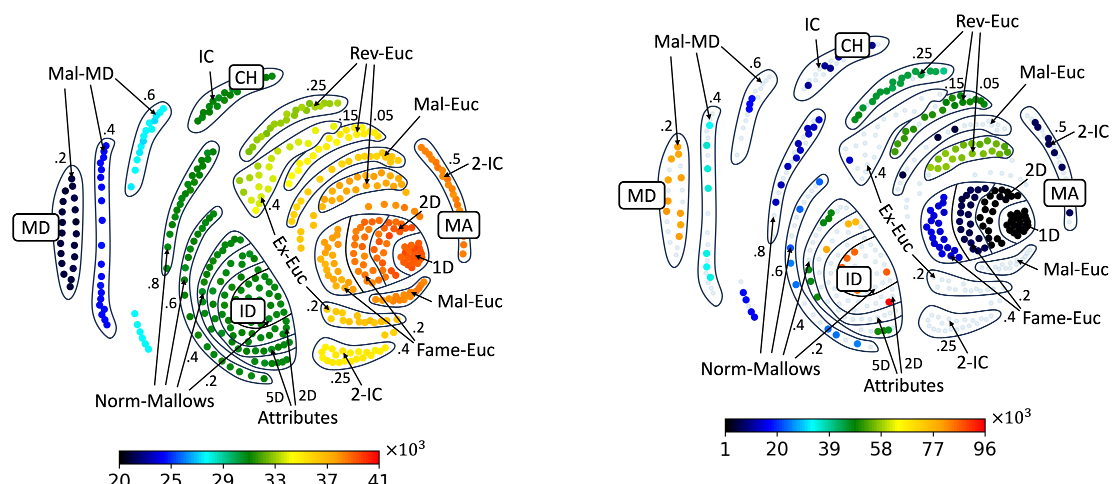
We conjecture that $\mathrm{d}_{\mathrm{MAD}}(\mathrm{ID}^{2n}, \mathrm{CH}^{2n}) = \frac{8}{3}n^3 - 3n^2 - \frac{5}{3}n + 2$ (we verified this conjecture for different values of n with the help of a computer).



(a) 500 agents

(b) 750 agents

Figure 17: Maps of 460 SR instances.



(a) Average number of blocking pairs for a perfect matching

(b) Minimum summed rank of agents for partner in any stable matching (transparent points have no stable matching)

Figure 18: Map of 460 SR instances for 500 agents visualizing different quantities for each instance.

Appendix B. Maps of Roommates for Different Agent Numbers

In the main body, we have focused on SR instances with 200 agents. Here, we present some maps for instances with 500 and 750 agents. It turns out that the maps for 500 agents (Figure 17 (a)) and 750 agents (Figure 17 (b)) are very similar to the map for 200 agents presented in the main body (Figure 2). The main difference is that the placement of the 2-IC instances with $p = 0.25$ and of the Expectations-Euclidean instances changes. These instances still appear as one cluster but in a slightly different position as in the map for 200 agents. However, their placement in the maps for 500 and 750 agents is very similar, making these two maps almost indistinguishable. We also reran our experiments for 500 agents, observing very similar high-level trends as reported in our experiments for 200 agents. As examples, we show in Figure 18 (a) the map for 500 agents visualizing the average number of blocking pairs for a perfect matching and in Figure 18 (b) the map visualizing the value of the summed rank minimal matching.

References

Abraham, D. J., Biró, P., & Manlove, D. F. (2005). "Almost stable" matchings in the roommates problem. In *Proceedings of the Third International Workshop on Approximation and Online Algorithms (WAOA '05)*, pp. 1–14. Springer.

Abraham, D. J., Levavi, A., Manlove, D. F., & O'Malley, G. (2008). The stable roommates problem with globally ranked pairs. *Internet Math.*, *5*(4), 493–515.

Agarwal, I., & Cole, R. (2023). Stable matching: Choosing which proposals to make. In *Proceedings of the 50th International Colloquium on Automata, Languages, and Programming (ICALP '23)*, pp. 8:1–8:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Arkin, E. M., Bae, S. W., Efrat, A., Okamoto, K., Mitchell, J. S. B., & Polishchuk, V. (2009). Geometric stable roommates. *Inf. Process. Lett.*, *109*(4), 219–224.

Bhatnagar, N., Greenberg, S., & Randall, D. (2008). Sampling stable marriages: why spouse-swapping won't work. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '08)*, pp. 1223–1232. SIAM.

Biedl, T. C., Brandenburg, F., & Deng, X. (2005). Crossings and permutations. In *Proceedings of the 13th International Symposium (GD '05)*, pp. 1–12. Springer.

Boehmer, N. (2023). *Application-Oriented Collective Decision Making Experimental Toolbox and Dynamic Environments.* Ph.D. thesis, TU Berlin.

Boehmer, N., Bredereck, R., Elkind, E., Faliszewski, P., & Szufa, S. (2022). Expected frequency matrices of elections: Computation, geometry, and preference learning. In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS '22)*.

Boehmer, N., Bredereck, R., Faliszewski, P., & Niedermeier, R. (2021a). Winner robustness via swap- and shift-bribery: Parameterized counting complexity and experiments. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI '21)*, pp. 52–58. ijcai.org.

Boehmer, N., Bredereck, R., Faliszewski, P., Niedermeier, R., & Szufa, S. (2021b). Putting a compass on the map of elections. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence (IJCAI '21)*, pp. 59–65. ijcai.org.

Boehmer, N., Faliszewski, P., & Kraiczy, S. (2023). Properties of the Mallows model depending on the number of alternatives: A warning for an experimentalist. In *Proceedings of* *the 40th International Conference on Machine Learning (ICML '23)*, pp. 2689–2711. PMLR.

Boehmer, N., Faliszewski, P., Niedermeier, R., Szufa, S., & Was, T. (2022a). Understanding distance measures among elections. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI '22)*, pp. 102–108. ijcai.org.

Boehmer, N., Heeger, K., & Niedermeier, R. (2022b). Theory of and experiments on minimally invasive stability preservation in changing two-sided matching markets. In *Proceedings of the Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI '22)*, pp. 4851–4858. AAAI Press.

Boehmer, N., & Schaar, N. (2023). Collecting, classifying, analyzing, and using real-world ranking data. In *Proceedings of the 2023 International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*, pp. 1706–1715. ACM.

Bredereck, R., Heeger, K., Knop, D., & Niedermeier, R. (2020). Multidimensional stable roommates with master list. In *Proceedings of the 16th International Conferenc of the Web and Internet Economics (WINE '20)*, pp. 59–73. Springer.

Brilliantova, A., & Hosseini, H. (2022). Fair stable matching meets correlated preferences. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS '22)*, pp. 190–198. IFAAMAS.

Cooper, F., & Manlove, D. F. (2019). Two-sided profile-based optimality in the stable marriage problem. *preprint*, *arXiv:1905.06626*.

Cooper, F., & Manlove, D. F. (2020). Algorithms for new types of fair stable matchings. In *Proceedings of the 18th International Symposium on Experimental Algorithms (SEA '20)*, pp. 20:1–20:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik.

Cui, L., & Jia, W. (2013). Cyclic stable matching for three-sided networking services. *Comput. Networks*, *57*(1), 351–363.

Delorme, M., García, S., Gondzio, J., Kalcsics, J., Manlove, D. F., & Pettersson, W. (2019). Mathematical models for stable matching problems with ties and incomplete lists. *Eur. J. Oper. Res.*, *277*(2), 426–441.

Diaconis, P., & Graham, R. L. (1977). Spearman's footrule as a measure of disarray. *J. R. Stat. Soc. Series B Stat. Methodol.*, *39*(2), 262–268.

Dickson, L. E., & Safford, F. H. (1906). Solutions of Problems: Group Theory: 8. *Amer. Math. Monthly*, *13*(6-7), 150–151.

Dwork, C., Kumar, R., Naor, M., & Sivakumar, D. (2001). Rank aggregation methods for the web. In *Proceedings of the Tenth International World Wide Web Conference (WWW '10)*, pp. 613–622. ACM.

Erdem, E., Fidan, M., Manlove, D. F., & Prosser, P. (2020). A general framework for stable roommates problems using answer set programming. *Theory Pract. Log. Program.*, *20*(6), 911–925.

Faliszewski, P., Skowron, P., Slinko, A., Szufa, S., & Talmon, N. (2019). How similar are two elections?. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI '19)*, pp. 1909–1916. AAAI Press.

Feder, T. (1992). A new fixed point approach for stable networks and stable marriages. *J. Comput. Syst. Sci.*, *45*(2), 233–284.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *Am. Math. Mon.*, *69*, 9–15.

Genc, B., Siala, M., O'Sullivan, B., & Simonin, G. (2017). Finding robust solutions to stable marriage. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI '17)*, pp. 631–637. ijcai.org.

Genc, B., Siala, M., Simonin, G., & O'Sullivan, B. (2019). An approach to robustness in the stable roommates problem and its comparison with the stable marriage problem. In *Proceedings of the 16th International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR 2019)*, pp. 320–336. Springer.

Gurobi Optimization, LLC (2021). Gurobi Optimizer Reference Manual. `https://www.gurobi.com`. Accessed: 2022-04-25.

Gusfield, D., & Irving, R. W. (1989). *The Stable marriage problem - structure and algorithms*. Foundations of computing series. MIT Press.

Harary, F., & Moser, L. (1966). The theory of round robin tournaments. *Am. Math. Mon.*, *73*(3), 231–246.

Holyer, I. (1981). The NP-completeness of edge-coloring. *SIAM J. Comput.*, *10*(4), 718–720.

Irving, R. W., & Manlove, D. F. (2009). Finding large stable matchings. *ACM J. Exp. Algorithmics*, *14*, 1.2–1.30.

Irving, R. W., Manlove, D. F., & Scott, S. (2008). The stable marriage problem with master preference lists. *Discret. Appl. Math.*, *156*(15), 2959–2977.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Inf. Process. Lett.*, *31*(1), 7–15.

Kamiyama, N. (2019). Many-to-many stable matchings with ties, master preference lists, and matroid constraints. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '19)*, pp. 583–591. IFAAMAS.

Knuth, D. E. (1976). *Mariages stables et leurs relations avec d'autres problèmes combinatoires*. Les Presses de l'Université de Montréal, Montreal, Que. Introduction à l'analyse mathématique des algorithmes, Collection de la Chaire Aisenstadt.

Kwanashie, A., & Manlove, D. F. (2013). An integer programming approach to the hospitals/residents problem with ties. In *Selected Papers of the International Conference on Operations Research (OR '13)*, pp. 263–269. Springer.

Mallows, C. (1957). Non-null ranking models. *Biometrica*, *44*, 114–130.

Manlove, D. F. (2013). *Algorithmics of Matching Under Preferences*, Vol. 2 of *Series on Theoretical Computer Science*. WorldScientific.

Manlove, D. F., Milne, D., & Olaosebikan, S. (2022). Student-project allocation with preferences over projects: Algorithmic and experimental results. *Discret. Appl. Math.*, *308*, 220–234.

Manne, F., Naim, M., Lerring, H., & Halappanavar, M. (2016). On stable marriages and greedy matchings. In *Proceedings of the Seventh SIAM Workshop on Combinatorial Scientific Computing (CSC '16)*, pp. 92–101. SIAM.

Mertens, S. (2015). Stable roommates problem with random preferences. *J. Stat. Mech. Theory Exp.*, *2015*(1), P01020.

O'Malley, G. (2007). *Algorithmic aspects of stable matching problems*. Ph.D. thesis, University of Glasgow.

Ostrovsky, R., & Rosenbaum, W. (2015). Fast distributed almost stable matchings. In *Proceedings of the 2015 ACM Symposium on Principles of Distributed Computing (PODC '15)*, pp. 101–108. ACM.

Pettersson, W., Delorme, M., García, S., Gondzio, J., Kalcsics, J., & Manlove, D. F. (2021). Improving solution times for stable matching problems through preprocessing. *Comput. Oper. Res.*, *128*, 105128.

Podhradsky, A. (2010). Stable marriage problem algorithms. Master's thesis, Masaryk University, Faculty of Informatics.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesth. Analg.*, *126*(5), 1763–1768.

Siala, M., & O'Sullivan, B. (2017). Rotation-based formulation for stable matching. In *Proceedings of the 23rd International Conference on Principles and Practice of Constraint Programming (CP '17)*, pp. 262–277. Springer.

Szufa, S., Faliszewski, P., Skowron, P., Slinko, A., & 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)*, pp. 1341–1349. IFAAMAS.

Szufa, S., Faliszewski, P., Janeczko, L., Lackner, M., Slinko, A., Sornat, K., & Talmon, N. (2022). How to sample approval elections?. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence (IJCAI '22)*, pp. 496–502. ijcai.org.

Teo, C., & Sethuraman, J. (2000). On a cutting plane heuristic for the stable roommates problem and its applications. *Eur. J. Oper. Res.*, *123*(1), 195–205.

Tziavelis, N., Giannakopoulos, I., Johansen, R. Q., Doka, K., Koziris, N., & Karras, P. (2020). Fair procedures for fair stable marriage outcomes. In *Proceedings of the Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI '20)*, pp. 7269–7276. AAAI Press.

Woeginger, G. J. (2003). Banks winners in tournaments are difficult to recognize. *Soc. Choice Welf.*, *20*(3), 523–528.