# Maximization of Minimum Weighted Hamming Distance between Set Pairs

**Tatsuya Matsuoka** TA.MATSUOKA@NEC.COM
*NEC Corporation*

**Shinji Ito** I-SHINJI@NEC.COM
*NEC Corporation*

**Editors:** Berrin Yanıkoğlu and Wray Buntine

## Abstract

*Finding diverse solutions* to combinatorial optimization problems is beneficial for a deeper understanding of complicated real-world problems and for simpler and more practical mathematical modeling. For this purpose, it is desirable that every solution is far away from one another and that solutions can be found in time not depending polynomially on the size of the family of feasible solutions. In this paper, we investigate the problem of finding diverse sets in the sense of maximizing the *minimum* of *weighted Hamming distance* between *set pairs*. Under a particular assumption, we provide an algorithm that gives diverse sets of almost $\mu/2$-approximation in expectation in the sense of maximization of the minimum of the expected value, where $\mu \in [0, 1]$ is a parameter on a subroutine. We further give a hardness result that any approximation ratio better than $2/3$ is impossible in polynomial time under the assumption of $\mathrm{P} \neq \mathrm{NP}$.

**Keywords:** Diverse solutions; Hamming distance; Multiplicative weight update

## 1. Introduction

Combinatorial optimization appears in various situations in daily life, such as working schedules of employees (Ernst et al., 2004), matchings (Edmonds, 1965), and facility location problems (Drezner and Hamacher, 2004). In many cases, modeling real-world problems by mathematical optimization precisely is quite difficult (e.g., Hanaka et al. (2022a,b)). For example, when a company decides on a working schedule for employees, it should consider various aspects, e.g., each employee's requests for a day off, their technical level, enough rest time, and the working schedule pattern prohibited or avoided in principle. Moreover, in real-world applications, personal relationships among employees may influence the quality of work, and it is quite difficult to deal with these kinds of matters enough when making a working schedule. Another example is the facility location problem. When considering the construction of new facilities, many aspects should be considered, such as distance from each house, an arrangement that is not biased, and feasibility constraints.

Owing to the above complexity of real-world problems, it is difficult to model them as mathematical formulations incorporating all detailed conditions and objectives. If the mathematical optimization problem does not reflect real-world conditions, then it is expected that an output (single solution) does not fulfill the user's wishes.

One approach to tackle the above issue is taking multiple diverse solutions (see, e.g., Hanaka et al. (2022a,b)). Finding diverse solutions has been studied in various fields such as , document summarization (Carbonell and Goldstein, 1998), ranking (Yang et al., 2019), data selection (Moumoulidou et al., 2021), search results (Gollapudi and Sharma, 2009; Drosou and Pitoura, 2010; Qin et al., 2012), matching (Hanaka et al., 2021; Fomin et al., 2023), and recommender systems (Kaminskas and Bridge, 2016; Kunaver and Požrl, 2017; Castells et al., 2021). In order to make decisions for real-world problems, diverse solutions are often helpful. By examining multiple diverse solutions, one can look out over diverse possibilities for solutions, and they may help find omitted constraints or components that should be considered in an objective function and recognize what is actually one would like to model as a mathematical optimization.

There have been existing works on problems of finding diverse sets. For the class of solutions, there have been mainly two types; taking diverse *elements* in a ground set (e.g., Erkut et al. (1994); Ravi et al. (1994)) and taking diverse *sets* in a given set family (e.g., Hanaka et al. (2022a,b)), where the set family itself may not be given, and this type includes the setting that the information of the set family is given as a membership oracle. For the class of objective function, which reflects the measure of diversity, two commonly considered forms are the *sum* of the distance between pairs (*max-sum* type) (e.g., Hanaka et al. (2022a,b)) and the *minimum* of the distance between pairs (*max-min* type) (e.g., Erkut et al. (1994); Fomin et al. (2023)). As this distance, *weighted Hamming distance* is a widely used one (e.g., Hanaka et al. (2022a,b)). More details will be given later in this section.

Our algorithm possesses both of the following characteristics: (i) time complexity which is in polynomial order in the size of the input ground set (not the size of the candidate set family), the reciprocal of a parameter that affects the property of the output of the algorithm, and the maximum value of the weight of elements, and (ii) the distance of the nearest pair being considered (our algorithm deals with the maximization of *minimum* weighted Hamming distance between *set pairs* using the expected value on the algorithm and applicability for problems with certain assumptions).

Let us introduce some notation for describing our problem setting precisely. Let $V$ be a finite set. For sets $X, Y \subseteq V$, let us denote the symmetric difference of $X$ and $Y$ by $X \triangle Y := (X \setminus Y) \cup (Y \setminus X)$. For $w \colon V \to \mathbb{R}_{\geq 0}$, *weighted Hamming distance* between $X$ and $Y$ is $\sum_{v \in X \triangle Y} w(v)$, and this is denoted by $d_w(X, Y)$. For $X \subseteq V$, we denote $\sum_{v \in X} w(v)$ by $w(X)$. Let $W$ stand for $\max_{v \in V} w(v)$. For a domain $\mathcal{D}$, we call an oracle a *$\mu$-approximation oracle* for the maximization problem on $\mathcal{D}$ if for any nonnegative function $f \colon \mathcal{D} \to \mathbb{R}_{\geq 0}$, it returns $D \in \mathcal{D}$ satisfying $f(D) \geq \mu \max_{D^* \in \mathcal{D}} f(D^*)$.

In this paper, when constructing an algorithm in Section 2, we assume the following: there exists a $\mu$-approximation oracle for finding $S \in \mathcal{S}$ maximizing $\sum_{i=1}^{k} \gamma_i d_w(S, S_i)$ for any $\gamma \in \Delta^k (:= \{\boldsymbol{x} \in \mathbb{R}^k \mid x_i \geq 0 \ (i = 1, \ldots, k), \ \sum_{i=1}^{k} x_i = 1\})$ and $S_1, \ldots, S_k \in \mathcal{S}$, which takes time $\theta$. (Note that by this assumption, for any integer $k'$ satisfying $1 \leq k' \leq k$, we can find $S \in \mathcal{S}$ with $\sum_{i=1}^{k'} \gamma_i d_w(S, S_i) \geq \mu \cdot \max_{S' \in \mathcal{S}} \sum_{i=1}^{k'} \gamma_i d_w(S', S_i)$ for any $\gamma \in \Delta^{k'}$ with time $\theta$ by the above oracle.) Section 2.3 introduces a specific example which shows that this assumption holds if two certain conditions (one on the emptyset and the other on weight maximization) hold. Our problem setting is the following.

**Problem 1** *Given a finite set $V$, a weight function $w\colon V \to \mathbb{R}_{\geq 0}$, and an integer $k \in \mathbb{Z}_{>0}$, and the information of $\mathcal{S} \subseteq 2^V$ is fixed but $\mathcal{S}$ itself is not explicitly given as an input, find $S_1, \dots, S_k$ with $S_i \in \mathcal{S}$ $(i = 1, \dots, k)$ maximizing* $\min_{1 \leq i < j \leq k} d_w(S_i, S_j)$.

Our contributions in this paper are as follows (details are explained in the following sections):

- We provide an algorithm which takes time polynomial in $|V|$, $1/\delta$, and $W$ whose output satisfies
$$\min_{1 \leq i < j \leq k} \mathbb{E} d_w(S_i, S_j) \geq \frac{\mu}{2}\Psi - \delta.$$
Here, $\mathbb{E}$ stands for the expectation over internal randomness of our algorithm (Algorithm 1), $\mu$ is a parameter on an oracle on the approximation of maximization of the weighted sum of weighted Hamming distances, $\Psi$ is the optimal value of our problem, and $\delta$ is an arbitrary input parameter. Also, we show that our algorithm runs in $\mathrm{O}\left(\frac{|V|^2 W^2 k \log k}{\delta^2}(\theta + |V|k)\right)$ time, where $\theta$ denotes the time complexity of the oracle approximately solving the corresponding maximization problem. *The main point of our contribution* is that we propose a framework combining MWU and Ravi et al. (1994)'s algorithm, giving diverse solutions satisfying the above inequality in time polynomial in $|V|$, $1/\delta$, and $W$ under a particular assumption on the maximization of the weighted sum of weighted Hamming distances. We give applications of our framework in Section 2.4.

- We show that our problem setting does not allow a polynomial-time approximation algorithm whose approximation ratio is better than 2/3 under the assumption of $\mathrm{P} \neq \mathrm{NP}$.

Our algorithm utilizes the framework of *multiplicative weight update* (MWU) method. The framework of MWU is well-studied and utilized in various fields (e.g., Arora et al. (2005); Bailey and Piliouras (2018)). For MWU, see, for example, Arora et al. (2012) as a survey.

### 1.1. Problem Categories

There have been many types of problems for finding diverse solutions. Here, we categorize common types of problems and clarify the position of our problem and results.

#### 1.1.1. Input and Solution

First, we categorize problems by the form of the *input* and the *output*. For a given ground set $V$, the following two types of problems have been commonly considered:

**Taking elements in a given set** Given a set $V$ and an integer $k \in \mathbb{Z}_{>0}$, find $v_1, \dots, v_k \in V$ with a required property.

**Taking sets in a given set family** Given a set $V$, the information of a set family $\mathcal{S} \subseteq 2^V$, and an integer $k \in \mathbb{Z}_{>0}$, find $S_1, \dots, S_k \in \mathcal{S}$ with a required property.

#### 1.1.2. Objective Function

Next, we categorize the type of the *objective function*. Although the purpose is to find diverse solutions in the following cases, mathematical formulations differ. Here, we write for the problem of taking sets in a given set family, but that of taking elements in a given set can be considered analogously. Here, we denote a distance by $d\colon \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$.

**Maximizing summation** Maximizing the sum of the distance between set pairs, that is, $\max_{S_1,\dots,S_k \in \mathcal{S}} \sum_{1 \leq i < j \leq k} d(S_i, S_j)$.

**Maximizing minimum** Maximizing the minimum distance between set pairs, that is, $\max_{S_1,\dots,S_k \in \mathcal{S}} \min_{1 \leq i < j \leq k} d(S_i, S_j)$.

#### 1.1.3. Distance

One key point for problems of taking diverse sets or elements is how to define the *distance* between pair of sets or elements. The following four types appear often.

**Arbitrary distance** An arbitrary function $d\colon \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$ which satisfies (i) $d(X, X) = 0$ holds for all $X \in \mathcal{S}$ and (ii) $d(X, Y) = d(Y, X)$ holds for all $X, Y \in \mathcal{S}$.

**Distance satisfying the triangle inequality** A distance $d\colon \mathcal{S} \times \mathcal{S} \to \mathbb{R}_{\geq 0}$ satisfying the above (i) and (ii) such that (iii) $d(X, Y) + d(Y, Z) \geq d(X, Z)$ holds for all $X, Y, Z \in \mathcal{S}$.

**Weighted Hamming distance**

$$d(S_i, S_j) := \sum_{e \in S_i \triangle S_j} w(e).$$

**(Unweighted) Hamming distance**

$$d(S_i, S_j) := |S_i \triangle S_j|.$$

This distance can be captured as the weighted Hamming distance with $w(v) = 1$ for all $v \in V$.

### 1.2. Why We Deal with This Problem Setting

**Maximizing summation and maximizing minimum** In our problem setting, we maximize the minimum weighted Hamming distance between set pairs. As written above, there are two types of commonly studied objective functions; summation maximization and minimum maximization. Although both problem settings have been well-studied, for obtaining diverse patterns of solutions, minimum maximization may be suitable for what the user wants in some cases. In Section 5 in Baste et al. (2019), they gave an example showing that maximizing summation might not output what one may expect. We shortly explain this Baste et al. (2019)'s example. For an even number $r$, given a path of $2r - 2$ vertices, they consider taking $r$ vertex covers each size of which is at most $r - 1$. Then they showed that for the case of $r = 6$, only two kinds of solutions appear (three copies of each solution). Baste et al. (2019) cites another example: taking points in a given square (Ulrich et al., 2010). In this example, if one wants to maximize the summation of pairwise distance, then the points are only on boundaries, and this may not be suitable for one's purpose.

**Set family and computational complexity** In our problem setting, we take *sets* $S_1, \ldots, S_k$ from a *set family* $\mathcal{S} \subseteq 2^V$, where $V$ is a given ground set. Although the set family version can be captured as a set version by considering $\mathcal{S}$ as a new ground set $V'$ and we take $S_1, \ldots, S_k \in V'$, what we would like to emphasize is that the time complexity of our algorithm depends on $|V|$, $1/\delta$, and $W$ in polynomial order and not on $|\mathcal{S}|$ (which can be $2^{|V|}$ in the worst case) in polynomial order.

**Weighted Hamming distance** *Hamming distance* is often used to describe the distance between two sets. *Weighted Hamming distance* is a natural quantitative extension of (unweighted) Hamming distance and has been used in various fields. A *distance with the triangle inequality* or an *arbitrary distance* are more general, but finding diverse solutions for these distances is often highly difficult. For example, for a maximizing minimum version of taking elements in a given set, if we consider an arbitrary distance, then constant factor approximation is impossible under $\mathrm{P} \neq \mathrm{NP}$ (Ravi et al., 1994). Even for the same problem with a distance with the triangle inequality, an approximation ratio of more than $1/2$ is NP-hard (Ravi et al., 1994).

### 1.3. Related Works

Problems of taking diverse things have been well-investigated. Two standard objective functions for maximization are the sum of pairwise distances and the minimum of pairwise distances. For both objective functions, Ravi et al. (1994) gave algorithms and hardness results for a problem of taking elements in a given set. Chandra and Halldórsson (2001) gave several approximation algorithms and hardness results for diversification problems.

Max-min problems are also called *dispersion* problems, and heuristics have also been studied, see, e.g., Erkut et al. (1994). For max-min problems, Addanki et al. (2022) dealt with fair Max-Min diversification problem, which was introduced by Moumoulidou et al. (2021), the problem considering distinct categories and selecting the predetermined number of elements from each group. Akagi et al. (2018) proposed exact algorithm for the max-min $k$-dispersion problem which takes exponential time. Amano and Nakano (2020) gave a $1/4\sqrt{3}$-approximate algorithm for the max-min type problem, which considers not only the nearest point but also the second nearest point. Araki and Nakano (2022) dealt with a problem of max-min type dispersion on a line. Chen et al. (2019) gave an algorithm for a problem of an online version of a max-min type problem. Horiyama et al. (2021) gave algorithms for a problem of taking three points in max-min type. Kobayashi et al. (2021)'s topic is also a problem of taking three points in max-min type, but their work dealt with that on a convex polygon. Kobayashi et al. (2022) consider the same problem on a point set in a convex position. Singireddy and Basappa (2022) tackled a generalization of this problem: that of taking $k$ points in max-min type. Fomin et al. (2023) dealt with max-min type weighted problems on matroid bases and independent sets, showed their NP-hardness, and gave FPT algorithms for them.

For max-sum problems, Hanaka et al. (2021, 2022b) gave frameworks for finding diverse solutions. Hanaka et al. (2022a) dealt with a framework and applied the framework to some combinatorial problems. Gillenwater et al. (2015) defined submodular Hamming metrics, an extension of weighted Hamming distance, and gave an approximation algorithm for a problem with this metric.

---

**Algorithm 1** Proposed algorithm

Input $(V, w, \text{the information of } \mathcal{S}, k; \delta(>0))$
take $S_1 \in \mathcal{S}$ arbitrarily.
**for** $l = 2, \ldots, k$ **do**
  Set an output of Algorithm 2 to $S_l$.
**end for**
Output $S_1, \ldots, S_k$

---

We explain the difference and relation between our work and the work of Hanaka et al. (2022a). Hanaka et al. (2022a) gave a framework for a *max-sum* type problem. Their framework provides a *constant factor approximation ratio* for a problem for which *top-k enumeration* can be done in polynomial time. Note that for a ground set $V$, a membership oracle of $\mathcal{S} \subseteq 2^V$, and a weight function $w\colon V \to \mathbb{R}$, the procedure top-$k$ enumeration finds $S_1, \ldots, S_k \in \mathcal{S}$ such that for any $i \in [k]$ and any $S \in \mathcal{S} \setminus \{S_1, \ldots, S_k\}$, $w(S_i) \geq w(S)$ holds. On the other hand, the framework of our work deals with a *max-min* type problem. The output of our algorithm satisfies $\min_{1\leq i<j\leq k} \mathbb{E}d_w(S_i, S_j) \geq \frac{\mu}{2}\Psi - \delta$ and our algorithm can be used for problems for which $\mu$-approximation of maximization of the weighted sum of weighted Hamming distances can be done in time $\theta$.

### 1.4. Organization

The rest of this paper is organized as follows. In Section 2, we propose an algorithm for obtaining diverse solutions, analyze the time complexity of the algorithm, explain a special case on the oracle, and introduce applications of our algorithm. Section 3 gives a hardness result on Problem 1. We give some concluding remarks in Section 4.

## 2. Algorithm

For Problem 1, we propose an algorithm output of which satisfies a certain inequality on weighted Hamming distance between set pairs (details appear in Theorem 1 later.) Our algorithm as a whole is Algorithm 1. In Algorithm 1, Algorithm 2 is used as a subroutine repeatedly. Algorithm 2 utilizes the framework of MWU and decides the output $S_l$ probabilistically; for each $l$, candidates of $S_l$ are listed as $S^{(t)}$ ($t \in [1, T]$) (the same set can appear repeatedly) and we pick $S_l$ from them with probability in proportion to the number of appearances as $S^{(t)}$.

### 2.1. Approximation Performance Using Expected Value

For Problem 1, we provide the following result on the approximation ratio.

**Theorem 1** *Let $\Psi$ be the optimal value of Problem 1. Then, Algorithm 1 achieves the following inequality:*

$$\min_{1\leq i<j\leq k} \mathbb{E}d_w(S_i, S_j) \geq \frac{\mu}{2}\Psi - \delta.$$

*Here, $\mathbb{E}$ in this theorem denotes the expectation taken with respect to the internal randomness of Algorithm 1, and $\delta$ is an input parameter.*

**Algorithm 2** Step for taking $S_l$

Input $(V, w, \text{the information of } \mathcal{S}, l; \delta(>0); S_1, \ldots, S_{l-1})$
$W := \max_{v \in V} w(v),\ T := \max\left\{\left\lceil \frac{4|V|^2 W^2 \log(l-1)}{\delta^2} \right\rceil, 1\right\},\ \eta := \min\left\{\frac{1}{2}, \frac{\delta}{2|V|W}\right\}$
$\boldsymbol{\beta}^{(1)} = \mathbf{1}, \boldsymbol{\gamma}^{(1)} = \frac{1}{l-1}\mathbf{1}$
**for** $t = 1, \ldots, T-1$ **do**
  Take $S^{(t)} \in \mathcal{S}$ such that $\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(S^{(t)}, S_i)$ is larger than or equal to $\mu$ times the maximum.
  **for** $i = 1, \ldots, l-1$ **do**
    $\beta_i^{(t+1)} = \beta_i^{(t)} \left(1 - \eta \cdot \frac{d_w(S^{(t)}, S_i)}{|V|W}\right)$
  **end for**
  **for** $i = 1, \ldots, l-1$ **do**
    $\gamma_i^{(t+1)} = \beta_i^{(t+1)} / \sum_{i'=1}^{l-1} \beta_{i'}^{(t+1)}$
  **end for**
**end for**
Take $S^{(T)} \in \mathcal{S}$ such that $\sum_{i=1}^{l-1} \gamma_i^{(T)} d_w(S^{(T)}, S_i)$ is larger than or equal to $\mu$ times the maximum.
Output $S_l$ which is $S^{(t)}$ with the probability of frequency of appearances in rounds $t = 1, \ldots, T$.

In order to prove Theorem 1, we use Lemmas 2 and 3.

**Lemma 2** *Output $S_l$ of Algorithm 2 satisfies*

$$\min_{i \in [l-1]} \mathbb{E}_{S_l} d_w(S_l, S_i) \geq \mu \cdot \max_{S \in \mathcal{S}} \min_{i \in [l-1]} d_w(S_i, S) - \delta.$$

The next lemma corresponds with Theorem 2 in Ravi et al. (1994) in the sense that we adapt this theorem in Ravi et al. (1994) to the case of maximization of the minimum of weighted Hamming distance between set pairs.

**Lemma 3 (Ravi et al. (1994))** *Let $\Psi$ be the optimal value of Problem 1 with the input $(V, w, \text{the information of } \mathcal{S}, k)$. Then, for arbitrary given $S_1, \ldots, S_{l-1} \in \mathcal{S}$, it holds that*

$$\max_{S \in \mathcal{S}} \min_{i \in [l-1]} d_w(S_i, S) \geq \frac{\Psi}{2}.$$

Now we write proofs of Lemmas 2 and 3. First, we write a proof of Lemma 2.
**Proof** Similar to the inequality (3.5) in Arora et al. (2012), the following inequality holds.

**Lemma 4** *For arbitrary $\delta > 0$, let $T = \max\left\{\left\lceil \frac{4|V|^2 W^2 \log(l-1)}{\delta^2} \right\rceil, 1\right\}$ and $\eta = \min\left\{\frac{1}{2}, \frac{\delta}{2|V|W}\right\}$. Then, for arbitrary $i^* \in [l-1]$, it holds that*

$$\frac{\sum_{t=1}^{T} \sum_{i=1}^{l-1} \gamma_i^{(t-1)} d_w(S^{(t)}, S_i)}{T} \leq \delta + \frac{\sum_{t=1}^{T} d_w(S^{(t)}, S_{i^*})}{T}.$$

Here we write a proof of Lemma 4.
**Proof** Let us denote

$$\Phi_t := \sum_{i=1}^{l-1} \beta_i^{(t)}.$$

Then, for $\Phi_{t-1}$ and $\Phi_t$, the following relation holds:

$$\begin{aligned}
\Phi_t = \sum_{i=1}^{l-1} \beta_i^{(t)} &= \sum_{i=1}^{l-1} \beta_i^{(t-1)} \left(1 - \eta \frac{d_w(S^{(t-1)}, S_i)}{|V|W}\right) \\
&= \Phi_{t-1} \sum_{i=1}^{l-1} \gamma_i^{(t-1)} \left(1 - \eta \frac{d_w(S^{(t-1)}, S_i)}{|V|W}\right) \\
&= \Phi_{t-1} \left(1 - \eta \sum_{i=1}^{l-1} \gamma_i^{(t-1)} \frac{d_w(S^{(t-1)}, S_i)}{|V|W}\right) \\
&\leq \Phi_{t-1} \exp\left(-\eta \sum_{i=1}^{l-1} \gamma_i^{(t-1)} \frac{d_w(S^{(t-1)}, S_i)}{|V|W}\right),
\end{aligned}$$

where the last inequality holds since $e^{-x} \geq 1 - x$ holds for all $x \in \mathbb{R}$. Let us denote $\beta_i^{(T+1)} := \beta_i^{(T)} \left(1 - \eta \cdot \frac{d_w(S^{(T)}, S_i)}{|V|W}\right)$ and $\gamma_i^{(T+1)} := \beta_i^{(T+1)} / \sum_{i'=1}^{l-1} \beta_{i'}^{(T+1)}$. Then,

$$\begin{aligned}
\Phi_{T+1} &\leq \Phi_1 \prod_{t=1}^{T} \exp\left(-\eta \sum_{i=1}^{l-1} \gamma_i^{(t)} \frac{d_w(S^{(t)}, S_i)}{|V|W}\right) \\
&= (l-1) \prod_{t=1}^{T} \exp\left(-\eta \sum_{i=1}^{l-1} \gamma_i^{(t)} \frac{d_w(S^{(t)}, S_i)}{|V|W}\right) \\
&= (l-1) \exp\left(-\eta \sum_{t=1}^{T} \sum_{i=1}^{l-1} \gamma_i^{(t)} \frac{d_w(S^{(t)}, S_i)}{|V|W}\right) \qquad (1)
\end{aligned}$$

holds. On the other hand, for arbitrary $i^* \in [l-1]$,

$$\Phi_{T+1} = \sum_{i=1}^{l-1} \beta_i^{(T+1)} \geq \beta_{i^*}^{(T+1)} = \prod_{t=1}^{T} \left(1 - \eta \frac{d_w(S^{(t)}, S_{i^*})}{|V|W}\right)$$

holds. Since $0 \leq \frac{d_w(S^{(t)}, S_{i^*})}{|V|W} \leq 1$ and $0 < \eta \leq 1/2$, $1 - \eta \frac{d_w(S^{(t)}, S_{i^*})}{|V|W} \geq (1-\eta)^{\frac{d_w(S^{(t)}, S_{i^*})}{|V|W}}$ holds for each $t \in [T]$. Thus,

$$\prod_{t=1}^{T} \left(1 - \eta \frac{d_w(S^{(t)}, S_{i^*})}{|V|W}\right) \geq (1-\eta)^{\sum_{t=1}^{T} \frac{d_w(S^{(t)}, S_{i^*})}{|V|W}} \qquad (2)$$

holds. Therefore, from (1) and (2),

$$\sum_{t=1}^{T} \frac{d_w(S^{(t)}, S_{i^*})}{|V|W} \log(1-\eta) \leq \log(l-1) - \eta \sum_{t=1}^{T} \sum_{i=1}^{l-1} \gamma_i^{(t-1)} \frac{d_w(S^{(t)}, S_i)}{|V|W}$$

holds. Hence,

$$\begin{aligned}\eta \sum_{t=1}^{T} \sum_{i=1}^{l-1} \gamma_i^{(t-1)} \frac{d_w(S^{(t)}, S_i)}{|V|W} &\leq \log(l-1) - \sum_{t=1}^{T} \frac{d_w(S^{(t)}, S_{i^*})}{|V|W} \log(1-\eta) \\ &= \log(l-1) + \sum_{t=1}^{T} \frac{d_w(S^{(t)}, S_{i^*})}{|V|W} \log \frac{1}{1-\eta} \\ &\leq \log(l-1) + \sum_{t=1}^{T} \frac{d_w(S^{(t)}, S_{i^*})}{|V|W} (\eta^2 + \eta)\end{aligned}$$

holds. Here, the last inequality holds by $0 < \eta \leq 1/2$. Therefore, we obtain

$$\begin{aligned}\sum_{t=1}^{T} \sum_{i=1}^{l-1} \gamma_i^{(t-1)} d_w(S^{(t)}, S_i) - \sum_{t=1}^{T} d_w(S^{(t)}, S_{i^*}) &\leq \frac{|V|W \log(l-1)}{\eta} + \eta \sum_{t=1}^{T} d_w(S^{(t)}, S_{i^*}) \\ &\leq \frac{|V|W \log(l-1)}{\eta} + \eta |V| W T,\end{aligned}$$

where we use $d_w(S^{(t)}, S_i) \leq |V|W$ for any $t \in [T]$ and $i \in [l-1]$. If $\delta \leq |V|W$, when $T = \max\left\{\left\lceil \frac{4|V|^2 W^2 \log(l-1)}{\delta^2} \right\rceil, 1\right\}$, $\eta = \min\left\{\frac{1}{2}, \frac{\delta}{2|V|W}\right\}$,

$$\frac{\sum_{t=1}^{T} \sum_{i=1}^{l-1} \gamma_i^{(t-1)} d_w(S^{(t)}, S_i)}{T} \leq \delta + \frac{\sum_{t=1}^{T} d_w(S^{(t)}, S_{i^*})}{T} \tag{3}$$

holds. On the other hand, since $0 \leq d_w(S^{(t)}, S_i) \leq |V|W$ holds for arbitrary $t$ and $i$, the inequality (3) also holds under the condition of $\delta > |V|W$. ■

Now we move back to the proof of Lemma 2. From von Neumann (1928), the following equality holds:

$$\max_{\boldsymbol{S} \in \mathrm{conv}(\mathcal{S})} \min_{\boldsymbol{\gamma} \in \Delta^{l-1}} \sum_{i=1}^{l-1} \gamma_i d_w(S_i, \boldsymbol{S}) = \min_{\boldsymbol{\gamma} \in \Delta^{l-1}} \max_{\boldsymbol{S} \in \mathrm{conv}(\mathcal{S})} \sum_{i=1}^{l-1} \gamma_i d_w(S_i, \boldsymbol{S}), \tag{4}$$

where $\mathrm{conv}(\mathcal{S})$ denotes the convex hull of $\mathcal{S}$, and for a decomposition of $\boldsymbol{S}$ to a weighted sum of elements in $\mathcal{S}$, that is, $\boldsymbol{S} = \sum_{S' \in \mathcal{S}' \subseteq \mathcal{S}} \alpha_{S'} S'$, $d_w(S_i, \boldsymbol{S}) := \sum_{S' \in \mathcal{S}'} \alpha_{S'} d_w(S_i, S')$. (Note that since a problem we deal with in Algorithm 2, that is, maximizing $\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(S^{(t)}, S_i)$ has an optimal solution in $\mathcal{S}$ and thus the fact that $\boldsymbol{S}$ in (4) is in $\mathrm{conv}(\mathcal{S})$ does not cause trouble in our case.) Let us denote this value of (4) by $\lambda^*$. For an arbitrary $\boldsymbol{\gamma}$, the following inequality holds:

$$\mu \lambda^* \leq \frac{\sum_{t=1}^{T} (\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(S^{(t)}, S_i))}{T} \leq \delta + \min_{i \in [l-1]} \left\{ \frac{\sum_{t=1}^{T} d_w(S^{(t)}, S_i)}{T} \right\},$$

where the former inequality holds by the definition of $\lambda^*$ (the fact that $\lambda^*$ is equal to the right-hand side of (4)) and that $S^{(t)}$ attains $\mu$ times the maximum value, and the latter inequality holds by Lemma 4. Hence,

$$\mu\lambda^* - \delta \leq \min_{i\in[l-1]} \left\{ \frac{\sum_{t=1}^{T} d_w(S^{(t)}, S_i)}{T} \right\} = \min_{i\in[l-1]} \mathbb{E}_t d_w(S^{(t)}, S_i)$$

holds. Therefore, it holds that

$$\min_{i\in[l-1]} \mathbb{E}_{S_l} d_w(S_l, S_i) \geq \mu \cdot \max_{S\in\mathcal{S}} \min_{i\in[l-1]} d_w(S_i, S) - \delta.$$

■

In order to make this paper self-contained, we write a proof of Lemma 3. The following proof of Lemma 3 is the same kind as that of Theorem 2 in Ravi et al. (1994), and Lemma 5 corresponds with Part b of Claim 1 in Ravi et al. (1994).

**Proof** Let us denote an optimal solution of Problem 1 by $S_1^*, \ldots, S_k^*$ and

$$\Psi := \min_{1\leq i<j\leq k} d_w(S_i^*, S_j^*).$$

For each $S_i^*$ $(i = 1, \ldots, k)$, let us denote

$$\mathcal{C}_i^* = \left\{ S \in \mathcal{S} \;\middle|\; d_w(S_i^*, S) < \frac{\Psi}{2} \right\}.$$

Since $S_i^* \in \mathcal{C}_i^*$ $(i = 1, \ldots, k)$, each $\mathcal{C}_i^*$ is not empty. The following lemma holds.

**Lemma 5 (Ravi et al. (1994))** *For arbitrary $i$ and $j$ with $i \neq j$, $\mathcal{C}_i^* \cap \mathcal{C}_j^* = \emptyset$ holds.*

For completeness, we write a proof of Lemma 5.

**Proof** Suppose that there exist $i$ and $j$ with $i \neq j$ satisfying $\mathcal{C}_i^* \cap \mathcal{C}_j^* \neq \emptyset$. Suppose $S \in \mathcal{C}_i^* \cap \mathcal{C}_j^*$. Then, by the definition of $\mathcal{C}_i^*$, $d_w(S_i^*, S) < \Psi/2$ holds. In the same way, by the definition of $\mathcal{C}_j^*$, $d_w(S_j^*, S) < \Psi/2$ holds. Since $S_i^*$ and $S_j^*$ are elements in the optimal solution, $d_w(S_i^*, S_j^*) \geq \Psi$ holds. Therefore, it holds that $d_w(S_i^*, S) + d_w(S_j^*, S) < d_w(S_i^*, S_j^*)$, and this inequality contradicts the triangle inequality. ■

We move back to the proof of Lemma 3. By Lemma 5, when the algorithm adds a new set, the number of $\mathcal{C}_i^*$ including the set is at most one. Thus, at step $l$, the number of $\mathcal{C}_i^*$ remained not taken a set inside is at least $k - l - 1$. Therefore, the optimal value at step $l$ in the algorithm is larger than or equal to $\Psi/2$. ■

Finally, we prove Theorem 1.

**Proof** By combining Lemmas 2 and 3, for each step taking $S_l$,

$$\min_{i\in[l-1]} \mathbb{E}_{S_l} d_w(S_l, S_i) \geq \mu \cdot \max_{S\in\mathcal{S}} \min_{i\in[l-1]} d_w(S_i, S) - \delta \geq \frac{\mu}{2}\Psi - \delta$$

holds. This means that

$$\mathbb{E}_{S_l} d_w(S_l, S_i) \geq \frac{\mu}{2}\Psi - \delta$$

holds for arbitrary $l$ and $i$, and thus it holds that $\min_{1\leq i<j\leq k} \mathbb{E} d_w(S_i, S_j) \geq \frac{\mu}{2}\Psi - \delta$. ■

### 2.2. Time Complexity

We analyze the total time complexity of Algorithm 1. First, we see the subroutine, Algorithm 2 for a fixed $l$. Here, $\theta$ denotes the time complexity of the oracle of the procedure "Take $S^{(t)} \in \mathcal{S}$ such that $\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(S^{(t)}, S_i)$ is lagrer than or equal to $\mu$ times the maximum." (Note that $1 \leq l-1 \leq k$.) We repeat solving weight-maximization problem approximately (time complexity: $\theta$) and updating $\boldsymbol{\beta}^{(t)}$ and $\boldsymbol{\gamma}^{(t)}$ (time complexity: $\mathrm{O}(l|V|)$) for $T-1$ times and find $S^{(T)}$ (time complexity: $\theta$). Since $T = \max\left\{\left\lceil \frac{4|V|^2W^2\log(l-1)}{\delta^2} \right\rceil, 1\right\}$, time complexity of Algorithm 2 (round $l$) is $O\left(\frac{|V|^2W^2\log l}{\delta^2}(\theta + l|V|)\right)$. Taking the sum from $l=2$ to $k$ and also summing time complexity for taking $S_1$, by $\int x \log x \mathrm{d}x = \frac{1}{4}x^2(2\log x - 1) + C$, total time complexity of Algorithm 1 is

$$O\left(\frac{|V|^2W^2k\log k}{\delta^2}(\theta + |V|k)\right).$$

### 2.3. Special Case: Reduction to Weight Maximization for the Oracle

In the above, we assume that there exists a $\mu$-approximation oracle for maximizing the function $\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(S^{(t)}, S_i)$ in time $\theta$. We show that for a certain class, $S^{(t)}$ can be calculated in polynomial time concretely.

**Theorem 6** *Suppose the following two conditions hold for $\mathcal{S} \subseteq 2^V$:*

**(i)** $\emptyset \in \mathcal{S}$ *and*

**(ii)** *for any $w' \colon V \to \mathbb{R}$, we can find $S' \in \mathcal{S}$ such that $\sum_{e \in S'} w'(e)$ is larger than or equal to $\mu$ times the maximum in polynomial time.*

*Then, we can find $S^{(t)} \in \mathcal{S}$ such that $\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(S^{(t)}, S_i)$ is larger than or equal to $\mu$ times the maximum in polynomial time.*

**Proof** Let

$$w''(i,e) = \begin{cases} w(e) & (e \notin S_i) \\ -w(e) & (e \in S_i) \end{cases}$$

and let us denote $\tilde{w}(e) = \sum_{i=1}^{l-1} \gamma_i^{(t)} w''(i,e)$. Then, it holds that

$$\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(S, S_i) = \tilde{w}(S) + \sum_{i=1}^{l-1} \gamma_i^{(t)} w(S_i).$$

Let $S^* \in \mathcal{S}$ be the maximizer of $\sum_{i=1}^{l-1} \gamma_i^{(t)} d_w(\cdot, S_i)$. Since the value $\sum_{i=1}^{l-1} \gamma_i^{(t)} w(S_i)$ does not depend on $S^*$, the maximizer of $w'(\cdot)$ is also $S^*$. Since we assume (i) and it holds that $w'(\emptyset) = 0$, $\tilde{w}(S^*) \geq \tilde{w}(\emptyset) = 0$. By the assumption (ii), we can find $\hat{S} \in \mathcal{S}$ such that $\tilde{w}(\hat{S}) \geq \mu \tilde{w}(S^*)$ holds by the oracle. Then, it holds that

$$\tilde{w}(\hat{S}) + \sum_{i=1}^{l-1} \gamma_i^{(t)} w(S_i) \geq \mu \tilde{w}(S^*) + \sum_{i=1}^{l-1} \gamma_i^{(t)} w(S_i) \geq \mu\left(\tilde{w}(S^*) + \sum_{i=1}^{l-1} \gamma_i^{(t)} w(S_i)\right).$$

Thus, $\hat{S}$ satisfies the statement in the theorem, and we can find $\hat{S}$ in polynomial time. ■

**Remark 7** *Assumptions (i) and (ii) in Theorem 6 are one pair forming a sufficient condition, and actually, these can be loosened. For example, (i) can be replaced with "there exists $S \in \mathcal{S}$ with $\tilde{w}(S) \geq 0$".*

### 2.4. Applications

#### 2.4.1. MATCHING

Matching is one of the most fundamental structures in graphs, and it has been well-investigated in various fields. Let $G = (V, E)$ be an undirected graph and $w \colon E \to \mathbb{R}$ be a weight function on edges. A set $M \subseteq E$ is called a *matching* if any distinct pair of elements in $M$ do not share endpoints. For a given integer $r$, the following problem is considered.

**Problem 2 (Maximum weight matching)** *Given an undirected graph $G = (V, E)$, a weight function $w \colon E \to \mathbb{R}$, and an integer $r \in \mathbb{Z}_{>0}$, find a maximum weight matching of size less than or equal to $r$ in $G$ with respect to edge weight $w$.*

A diverse version of weighted matchings is formulated as follows.

**Problem 3** *Given an undirected graph $G = (V, E)$, a weight function $w \colon E \to \mathbb{R}$, and integers $k, r \in \mathbb{Z}_{>0}$, find matchings $M_1, \ldots, M_k \subseteq E$ of size less than or equal to $r$ maximizing $\min_{1 \leq i < j \leq k} d_w(M_i, M_j)$.*

Note that for the problem similar to Problem 3 whose feasible solutions $M_1, \ldots, M_k$ are of size $r$ and whose objective function is $\sum_{1 \leq i < j \leq k} d_w(M_i, M_j)$, Hanaka et al. (2022a) gave a $\max\{1 - 2/k, 1/2\}$-approximate algorithm, but our problem setting is max-min type.

By Edmonds (1965), maximum weight matching of size $r$ can be found in polynomial time. Thus, from this Edmonds (1965)'s result and Theorem 1, the following corollary is obtained.

**Corollary 8** *For Problem 3, by applying Algorithm 1, the output $M_1, \ldots, M_k$ satisfies*

$$\min_{1 \leq i < j \leq k} \mathbb{E} d_w(M_i, M_j) \geq \frac{\Psi_{\mathrm{DM}}}{2} - \delta,$$

*where $\Psi_{\mathrm{DM}}$ is the optimal value of Problem 3 and $\delta$ is a parameter in Algorithm 1 which the user determines.*

#### 2.4.2. MATROID

A pair $\mathcal{M} = (E, \mathcal{I})$ of a ground set $E$ and a set family $\mathcal{I} \subseteq 2^E$ is called a *matroid* if the following three conditions hold: (i) $\emptyset \in \mathcal{I}$, (ii) $Y \in \mathcal{I} \Longrightarrow X \in \mathcal{I}$ holds for all $X \subseteq Y \subseteq E$, and (iii) there exists $e \in Y \setminus X$ such that $X \cup \{e\}$ holds for all $X, Y \in \mathcal{I}$ with $|X| < |Y|$. The family $\mathcal{B}$ of maximal elements of $\mathcal{I}$ is called a base family of $\mathcal{M}$, and each element in $\mathcal{B}$ is called a base. The concept of matroids includes many fundamental concepts, and thus it is crucial to analyze problems on matroids; e.g., if a matroid is a graphic matroid, which corresponds to a certain graph, each base of the matroid corresponds to a spanning tree in the graph.

In Fomin et al. (2023), Weighted Diverse Bases problem was considered. Here, we state an optimization version of a similar problem (the above problem in Fomin et al. (2023) is a decision problem).

**Problem 4** *Given a matroid $\mathcal{M} = (E, \mathcal{I})$, a weight function $w \colon E \to \mathbb{R}$, and an integer $k \in \mathbb{Z}_{>0}$, find independent sets $I_1, \ldots, I_k \in \mathcal{I}$ of $\mathcal{M}$ maximizing $\min_{1 \le i < j \le k} d_w(I_i, I_j)$.*

For the version of Problem 4 whose objective function is $\sum_{1 \le i < j \le k} d_w(B_i, B_j)$ where $B_1, \ldots, B_k$ are bases, Hanaka et al. (2022b) gave a polynomial-time algorithm, but again, our problem setting is different.

By Rado (1957), for a given matroid $\mathcal{M} = (E, \mathcal{I})$ and a weight function $w \colon E \to \mathbb{R}$, a maximum-weight base can be found in polynomial time. Thus, from this Rado (1957)'s result and Theorem 1, we obtain the following corollary on Problem 4.

**Corollary 9** *For Problem 4, by applying Algorithm 1, the output $I_1, \ldots, I_k$ satisfies*

$$\min_{1 \le i < j \le k} \mathbb{E} d_w(I_i, I_j) \ge \frac{\Psi_{\mathrm{DI}}}{2} - \delta,$$

*where $\Psi_{\mathrm{DI}}$ is the optimal value of Problem 4 and $\delta$ is a parameter in Algorithm 1 which the user determines.*

## 3. Hardness on Approximation

For the hardness of Problem 1, we show the following statement on the approximation ratio.

**Theorem 10** *Under the assumption of $\mathrm{P} \neq \mathrm{NP}$, for any $\epsilon > 0$, there does not exist $(2/3 + \epsilon)$-approximation algorithm for Problem 1.*

**Proof** We utilize 3D MATCHING problem for reduction.

**Problem 5 (3D MATCHING)** *Given a hypergraph $\mathcal{H} = (V_1, V_2, V_3; \mathcal{E})$ with $\mathcal{E} \subseteq V_1 \times V_2 \times V_3$ and an integer $k'$, find $\mathcal{M} \subseteq \mathcal{E}$ such that $|\mathcal{M}| = k'$ and that arbitrary pair of distinct elements in $\mathcal{M}$ are disjoint.*

Each solution of Problem 5 is called a *3-dimensional matching*. By Karp (1972), Problem 5 is NP-complete. Let us consider that an arbitrary input $(V_1, V_2, V_3; \mathcal{E}; k')$ of Problem 5 is given. Then, we reduce this instance to an instance of Problem 1 as follows: $V := V_1 \cup V_2 \cup V_3$, $\mathcal{S} := \{\{v_1, v_2, v_3\} \mid (v_1, v_2, v_3) \in \mathcal{E}\}$, $k := k'$, and $w(v) := 1$ $(\forall v \in V)$. Then, the relation between the instance $(V, \mathcal{S}, k, w)$ (hereafter we call this instance (A)) of Problem 1 and the original instance $(V_1, V_2, V_3; \mathcal{E}; k')$ (hereafter we call this instance (B)) of Problem 5 is

- the optimal value of (A) is six $\iff$ (B) has a 3-dimensional matching of size $k'$, and
- the optimal value of (A) is less than or equal to four $\iff$ (B) does not have a 3-dimensional matching of size $k'$.

If there exists $(2/3+\epsilon)$-approximation algorithm for Problem 1, by applying the algorithm to (A), if the function value of the output solution is larger than four, (B) has a 3-dimensional matching with size $k'$, and if the function value of the output solution is less than or equal to four, (B) does not have a 3-dimensional matching with size $k'$. Thus, it means Problem 5, which is NP-complete, is solved in polynomial time. Therefore, under the assumption of $\mathrm{P} \neq \mathrm{NP}$, there does not exist $(2/3+\epsilon)$-approximation algorithm for Problem 1. ■

## 4. Concluding Remarks

In this paper, we give an algorithm for the problem of finding diverse sets in the sense of maximization of the minimum weighted Hamming distance between set pairs. Also, we gave a hardness result on the approximation ratio.

As future works, both better algorithms and hardness directions can be considered. For the former, better approximation algorithms are hoped, which may be an algorithm such that the expected value of the minimum distance is lower bounded, or the left-hand side is the same as our result but the right-hand side is larger than our result $(\mu/2)\Psi - \delta$. For the latter, since our hardness result may not be tight, we think further investigation on the hardness of our problem is a worthwhile research direction.

## References


Raghavendra Addanki, Andrew McGregor, Alexandra Meliou, and Zafeiria Moumoulidou. Improved approximation and scalability for fair max-min diversification. In *ICDT*, pages 7:1–7:21, 2022.

Toshihiro Akagi, Tetsuya Araki, Takashi Horiyama, Shin-ichi Nakano, Yoshio Okamoto, Yota Otachi, Toshiki Saitoh, Ryuhei Uehara, Takeaki Uno, and Kunihiro Wasa. Exact algorithms for the max-min dispersion problem. In *FAW*, pages 263–272, 2018.

Kazuyuki Amano and Shin-ichi Nakano. An approximation algorithm for the 2-dispersion problem. *IEICE Trans. Inf.& Syst.*, 103(3):506–508, 2020.

Tetsuya Araki and Shin-ichi Nakano. Max–min dispersion on a line. *J. Comb. Optim.*, 44: 1824–1830, 2022.

Sanjeev Arora, Elad Hazan, and Satyen Kale. Fast algorithms for approximate semidefinite programming using the multiplicative weights update method. In *FOCS*, pages 339–348, 2005.

Sanjeev Arora, Elad Hazan, and Satyen Kale. The multiplicative weights update method: a meta-algorithm and applications. *Theory Comput.*, 8(1):121–164, 2012.

James P Bailey and Georgios Piliouras. Multiplicative weights update in zero-sum games. In *EC*, pages 321–338, 2018.

Julien Baste, Lars Jaffke, Tomáš Masařík, Geevarghese Philip, and Günter Rote. FPT algorithms for diverse collections of hitting sets. *Algorithms*, 12(12):254, 2019.

Jaime Carbonell and Jade Goldstein. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In *SIGIR*, pages 335–336, 1998.

Pablo Castells, Neil Hurley, and Saul Vargas. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*, pages 603–646. Springer, 2021.

Barun Chandra and Magnús M Halldórsson. Approximation algorithms for dispersion problems. *Journal of Algorithms*, 38(2):438–465, 2001.

Jing Chen, Bo Li, and Yingkai Li. Efficient approximations for the online dispersion problem. *SIAM J. Comput.*, 48(2):373–416, 2019.

Zvi Drezner and Horst W Hamacher. *Facility Location: Applications and Theory*. Springer Science & Business Media, 2004.

Marina Drosou and Evaggelia Pitoura. Search result diversification. *ACM SIGMOD Record*, 39(1):41–47, 2010.

Jack Edmonds. Paths, trees, and flowers. *Can. J. Math.*, 17:449–467, 1965.

Erhan Erkut, Yilmaz Ülküsal, and Oktay Yenicerioğlu. A comparison of p-dispersion heuristics. *Computers & Operations Research*, 21(10):1103–1113, 1994.

Andreas T Ernst, Houyuan Jiang, Mohan Krishnamoorthy, and David Sier. Staff scheduling and rostering: A review of applications, methods and models. *Eur. J. Oper. Res.*, 153 (1):3–27, 2004.

Fedor V Fomin, Petr A Golovach, Fahad Panolan, Geevarghese Philip, and Saket Saurabh. Diverse collections in matroids and graphs. *Math. Program.*, pages 1–33, 2023.

Jennifer A Gillenwater, Rishabh K Iyer, Bethany Lusch, Rahul Kidambi, and Jeff A Bilmes. Submodular Hamming metrics. In *NIPS*, pages 3141–3149, 2015.

Sreenivas Gollapudi and Aneesh Sharma. An axiomatic approach for result diversification. In *WWW*, pages 381–390, 2009.

Tesshu Hanaka, Yasuaki Kobayashi, Kazuhiro Kurita, and Yota Otachi. Finding diverse trees, paths, and more. In *AAAI*, pages 3778–3786, 2021.

Tesshu Hanaka, Masashi Kiyomi, Yasuaki Kobayashi, Yusuke Kobayashi, Kazuhiro Kurita, and Yota Otachi. A framework to design approximation algorithms for finding diverse solutions in combinatorial problems. *arXiv preprint arXiv:2201.08940*, 2022a.

Tesshu Hanaka, Yasuaki Kobayashi, Kazuhiro Kurita, See Woo Lee, and Yota Otachi. Computing diverse shortest paths efficiently: a theoretical and experimental study. In *AAAI*, pages 3758–3766, 2022b.

Takashi Horiyama, Shin-ichi Nakano, Toshiki Saitoh, Koki Suetsugu, Akira Suzuki, Ryuhei Uehara, Takeaki Uno, and Kunihiro Wasa. Max-min 3-dispersion problems. *IEICE Trans. Fundamentals*, 104(9):1101–1107, 2021.

Marius Kaminskas and Derek Bridge. Diversity, serendipity, novelty, and coverage: a survey and empirical analysis of beyond-accuracy objectives in recommender systems. *ACM Trans. Interact. Intell. Syst.*, 7(1):1–42, 2016.

Richard M Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Springer, 1972.

Yasuaki Kobayashi, Shin-ichi Nakano, Kei Uchizawa, Takeaki Uno, Yutaro Yamaguchi, and Katsuhisa Yamanaka. Max-min 3-dispersion on a convex polygon. In *EuroCG*, 2021.

Yasuaki Kobayashi, Shin-Ichi Nakano, Kei Uchizawa, Takeaki Uno, Yutaro Yamaguchi, and Katsuhisa Yamanaka. An $O(n^2)$-time algorithm for computing a max-min 3-dispersion on a point set in convex position. *IEICE Trans. Inf.& Syst.*, 105(3):503–507, 2022.

Matevž Kunaver and Tomaž Požrl. Diversity in recommender systems–a survey. *Knowledge-Based Systems*, 123:154–162, 2017.

Zafeiria Moumoulidou, Andrew McGregor, and Alexandra Meliou. Diverse data selection under fairness constraints. In *ICDT*, pages 13:1–13:25, 2021.

Lu Qin, Jeffrey Xu Yu, and Lijun Chang. Diversifying top-k results. *Proceedings of the VLDB Endowment*, 5(11), 2012.

Richard Rado. Note on independence functions. *Proceedings of the London Mathematical Society*, 3(1):300–320, 1957.

Sekharipuram S Ravi, Daniel J Rosenkrantz, and Giri Kumar Tayi. Heuristic and special case algorithms for dispersion problems. *Oper. Res.*, 42(2):299–310, 1994.

Vishwanath R Singireddy and Manjanna Basappa. Max-min $k$-dispersion on a convex polygon. *arXiv preprint arXiv:2205.02021*, 2022.

Tamara Ulrich, Johannes Bader, and Lothar Thiele. Defining and optimizing indicator-based diversity measures in multiobjective search. In *PPSN*, pages 707–717, 2010.

J von Neumann. Zur theorie der gesellschaftsspiele. *Mathematische Annalen*, 100:295–320, 1928.

Ke Yang, Vasilis Gkatzelis, and Julia Stoyanovich. Balanced ranking with diversity constraints. In *IJCAI*, pages 6035–6042, 2019.