Journal of the Operations Research Society of China (2019) 7:429–448
https://doi.org/10.1007/s40305-019-00260-1

Approximation Algorithms for Vertex Happiness

Yao Xu[1] · Yong Chen[2] · Peng Zhang[3] · Randy Goebel[1]

Received: 3 September 2018 / Revised: 10 April 2019 / Accepted: 1 July 2019 / Published online: 20 July 2019

Abstract
We investigate the maximum happy vertices (MHV) problem and its complement, the minimum unhappy vertices (MUHV) problem. In order to design better approximation algorithms, we introduce the supermodular and submodular multi-labeling (Sup- ML and Sub- ML) problems and show that MHV and MUHV are special cases of Sup- ML and Sub- ML, respectively, by rewriting the objective functions as set functions. The convex relaxation on the Lovász extension, originally presented for the submodular multi-partitioning problem, can be extended for the Sub- ML problem, thereby proving that Sub- ML (Sup- ML, respectively) can be approximated within a factor of $2 - 2/k$ ($2/k$, respectively), where k is the number of labels. These general results imply that MHV and MUHV can also be approximated within factors of $2/k$ and $2 - 2/k$, respectively, using the same approximation algorithms. For the MUHV problem, we also show that it is approximation-equivalent to the hypergraph multiway cut problem; thus, MUHV is Unique Games-hard to achieve a $(2 - 2/k - \varepsilon)$-approximation, for

This paper is dedicated to Professor Ding-Zhu Du in celebration of his 70th birthday.

This research was supported by the National Natural Science Foundation of China (Nos. 11771114, 11571252, and 61672323), the China Scholarship Council (No. 201508330054), the Natural Science Foundation of Shandong Province (No. ZR2016AM28), and the Natural Sciences and Engineering Research Council of Canada.

✉ Yao Xu
xu2@ualberta.ca

Yong Chen
chenyong@hdu.edu.cn

Peng Zhang
algzhang@sdu.edu.cn

Randy Goebel
rgoebel@ualberta.ca

1 Department of Computing Science, University of Alberta, Edmonton, AB T6G 2E8, Canada

2 Department of Mathematics, Hangzhou Dianzi University, Hangzhou 310018, China

3 School of Computer Science and Technology, Shandong University, Jinan 250101, China

any $\varepsilon > 0$. For the MHV problem, the $2/k$-approximation improves the previous best approximation ratio $\max\{1/k, 1/(\Delta + 1/g(\Delta))\}$, where Δ is the maximum vertex degree of the input graph and $g(\Delta) = (\sqrt{\Delta} + \sqrt{\Delta + 1})^2 \Delta > 4\Delta^2$. We also show that an existing LP relaxation for MHV is the same as the concave relaxation on the Lovász extension for Sup- ML; we then prove an upper bound of $2/k$ on the integrality gap of this LP relaxation, which suggests that the $2/k$-approximation is the best possible based on this LP relaxation. Lastly, we prove that it is Unique Games-hard to approximate the MHV problem within a factor of $\Omega(\log^2 k/k)$.

Keywords Vertex happiness · Multi-labeling · Submodular/supermodular set function · Approximation algorithm · Polynomial-time reduction · Integrality gap

Mathematics Subject Classification 68W25 · 90C05 · 90C27

1 Introduction

Inspired by the study on the homophyly governing the structures of large-scale networks such as social networks and citation networks, Zhang and Li [1] introduced a vertex coloring problem that can be defined as follows. One is given an undirected graph $G = (V, E)$ with a nonnegative weight $w(v)$ for each vertex $v \in V$, a color set $C = \{1, 2, \cdots, k\}$, and a partial vertex coloring function $c : V \mapsto C$. The goal is to color all the uncolored vertices such that the total weight of *happy* vertices is maximized. A vertex is *happy* if it shares the same color with all its neighbors in the coloring scheme. The problem is referred to as the *maximum happy vertices* (MHV) problem [1], and its complement the *minimum unhappy vertices* (MUHV) problem can be defined analogously to minimize the total weight of unhappy vertices, where a vertex is *unhappy* if its color is different from at least one of its neighbors.

In both the MHV and MUHV problems, there could be multiple vertices in the given graph pre-colored the same color. When only one vertex is pre-colored by the partial vertex coloring function c for each $i \in C$, we denote these two problems as the *restricted*-MHV and the *restricted*-MUHV problems, respectively. For MUHV, there is a polynomial-time reduction from the general MUHV problem to the restricted-MUHV problem, by creating a vertex for each $i \in C$ with a weight large enough, pre-coloring it with the color i, connecting it to all the vertices pre-colored i, and making them uncolored (see a detailed proof in Sect. 3). This reduction implies that the restricted-MUHV problem and the general MUHV problem are approximation-equivalent.

We remark that these two vertex coloring problems are in fact labeling problems, and we use "color" and "label" interchangeably; they are different from the classic *graph coloring* (Coloring) problem [2], in which a feasible vertex coloring scheme must assign different colors to any adjacent vertices. We also note that if no vertex is pre-colored i, for any i, then this color i can be removed without affecting the optimum; we therefore assume without loss of generality that every color is used in the given partial vertex coloring function c.

Given a graph $G = (V, E)$ with the vertex set V and the edge set E, for any subset $X \subseteq V$, define the *boundary* of X, denoted as $\partial(X)$, to be the subset of vertices of X, and each has at least one neighbor outside of X. Let $\iota(X) = X - \partial(X)$, which is called the *interior* of X. In a coloring scheme, let S_i denote the subset of all the vertices colored i; then, every vertex of $\partial(S_i)$ is unhappy since it has at least one neighbor not colored i, while all vertices of $\iota(S_i)$ are happy. We extend the vertex weight function to subsets of vertices, that is, $w(X) := \sum_{v \in X} w(v)$ for any $X \subseteq V$, and we define the set function $f(\cdot)$ as

$$f(X) := w(\partial(X)), \ \forall X \subseteq V. \tag{1.1}$$

A vertex coloring scheme one-to-one corresponds to a partition $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$ of the vertex set V, where each part S_i contains all the vertices colored i. This way, the MUHV problem can be cast as finding a partition $\mathcal{S}$ such that $f(\mathcal{S}) := \sum_{i=1}^{k} f(S_i)$ is minimized.

It is not hard to validate (see a detailed proof in Sect. 3) that the boundary $\partial(\cdot)$ of a vertex subset in the given graph $G = (V, E)$ has the following properties for any two subsets $X, Y \subseteq V$:

(i) $\partial(\varnothing) = \varnothing$;
(ii) $\partial(X \cap Y) \subseteq \partial(X) \cup \partial(Y)$;
(iii) $\partial(X \cup Y) \subseteq \partial(X) \cup \partial(Y)$;
(iv) $\partial(X \cap Y) \cap \partial(X \cup Y) \subseteq \partial(X) \cap \partial(Y)$.

Therefore, the set function $f : 2^V \to \mathbb{R}$ defined in Eq. (1.1) satisfies $f(X)+f(Y) \geqslant f(X \cap Y) + f(X \cup Y)$, for any two subsets $X, Y \subseteq V$ (see a detailed proof in Sect. 3). That is, $f(\cdot)$ is a *submodular* [3] function on the set V. This way, the MUHV problem can be cast as a special case of the following *submodular multi-labeling* (SUB- ML) problem:

Given a ground set V, a nonnegative submodular set function $f : 2^V \to \mathbb{R}_+$, with $f(\varnothing) = 0$, a set of labels $L = \{1, 2, \cdots, k\}$, and a partial labeling function $\ell : V \mapsto L$ which pre-assigns each label i to all the elements of a non-empty subset $T_i \subset V$, the goal of the SUB- ML problem is to find a partition $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$ of the ground set V to minimize $f(\mathcal{S}) = \sum_{i=1}^{k} f(S_i)$, where the part S_i is the subset of elements assigned with the label i.

Conversely, given the graph $G = (V, E)$, we define another set function $g(\cdot)$ as

$$g(X) := w(\iota(X)), \ \forall X \subseteq V. \tag{1.2}$$

Then, $g(X) = w(X) - f(X)$ for any subset $X \subseteq V$, and consequently, $g(\cdot)$ is a *supermodular* [3] function on the set V. Thus, the MHV problem can be cast as finding a partition $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$ of the vertex set V such that $g(\mathcal{S}) = \sum_{i=1}^{k} g(S_i)$ is maximized, where each part S_i contains all the vertices colored i; it can also be cast as a special case of the *supermodular multi-labeling* (SUP- ML) problem that can be analogously defined.

1.1 Related Works

Classification problems have been formulated as cuts or partition or labeling or coloring, and have been widely studied for a very long time.

For the MHV problem, Zhang and Li [1] proved that it is polynomial-time solvable for $k = 2$ and it becomes NP-hard for $k \geqslant 3$; for $k \geqslant 3$, they presented two approximation algorithms: a greedy algorithm with an approximation ratio of $1/k$ and an $\Omega(1/\Delta^3)$-approximation based on a subset-growth technique, where Δ is the maximum vertex degree of the input graph. Later, Zhang et al. [4] presented an improved algorithm with an approximation ratio of $1/(\Delta + 1)$ based on a combination of randomized LP rounding techniques, which was further improved to $1/(\Delta+1/g(\Delta))$ with deeper analysis [5], where $g(\Delta) = (\sqrt{\Delta}+\sqrt{\Delta + 1})^2\Delta > 4\Delta^2$. Together, these imply that the current best approximation ratio for the MHV problem is $\max\{1/k, 1/\big(\Delta + 1/g(\Delta)\big)\}$. For the complementary MUHV problem, to the best of our knowledge, it hasn't been particularly studied in the literature.

Recall that the MHV and the MUHV problems are a special case of the Sup- ML and the Sub- ML problems, respectively. We again remind the readers that in an instance of these multi-labeling problems, each label is pre-assigned to at least one element and to multiple elements in general. A restricted version of the Sub- ML problem, when each label is pre-assigned to exactly one element, is the *submodular multiway partition* (Sub- MP) problem [6], which has been studied a great deal. The restricted-MUHV problem is a special case of the Sub- MP problem.

The Sub- MP problem was first studied by Zhao et al. [6], who presented a $(k - 1)$-approximation algorithm. Years later, Chekuri and Ene [7] proposed a convex relaxation for Sub- MP by using the Lovász extension, and they presented a 2-approximation based on this relaxation. This was further improved to a $(2 - 2/k)$-approximation shortly after by Ene et al. [8], which immediately gives a $(2-2/k)$-approximation for the restricted-MUHV and the general MUHV problems. On the inapproximability, Ene et al. [8] proved that any $(2-2/k-\varepsilon)$-approximation for Sub- MP requires exponentially many value queries for any $\varepsilon > 0$; otherwise, it implies NP = RP.

It is important to note that although the restricted-MUHV problem and the general MUHV problem are approximation-equivalent, we cannot simply conclude that the Sub- MP problem and the Sub- ML problem are also approximation-equivalent based on similar reduction proofs. The difference between Sub- MP and Sub- ML depends heavily on how the set function $f(\cdot)$ is defined; a change to the ground set V could alter the optimal solution value and a feasible solution value differently.

The Sub- MP problem includes many well-studied cut problems including the classic (edge-weighted) multiway cut (MC) problem [9], the node-weighted multiway cut (Node- MC) problem [10], and the *hypergraph multiway cut* (Hyp- MC) problem [11] as special cases. The classic MC problem is NP-hard for $k \geqslant 3$ even if all edges have unit weight [9], and there have been many approximation algorithms designed and analyzed for it [9,12–16]. Most of these approximation results are based on the *linear program* (LP) relaxation presented by Călinescu et al. [12], and the current best approximation ratio is 1.296 5 [16]. The Hyp- MC and the Node- MC problems are actually approximation-equivalent, and both of them admit a $(2-2/k)$-approximation [10,11]; on the negative side, they are proved more difficult to approximate that it is Unique Games-hard to achieve a $(2-2/k-\varepsilon)$-approximation, for any $\varepsilon > 0$ [8].

One can similarly define the complement of the SUB- MP problem, called the *supermodular multiway partition* (SUP- MP) problem. The restricted-MHV problem is then a special case of SUP- MP, which also includes the multiway uncut problem [17] as a special case, where the k terminals in the input graph can be considered as k elements each being pre-assigned with a distinct label. The multiway uncut problem seems only studied by Langberg et al. [17], who presented a 0.853 5-approximation based on an LP relaxation. When generalizing the multiway uncut problem to pre-assign multiple terminals in a part of the vertex partition, it becomes the maximum happy edges (MHE) problem, which is a special case of the SUP- ML problem. Zhang and Li [1] presented a $1/2$-approximation for the MHE problem on the basis of a simple division strategy; extending the LP relaxation for the multiway uncut, Zhang et al. [4] later improved the approximation ratio to $1/2 + \frac{\sqrt{2}}{4}h(k) \geqslant 0.853\,5$, where $h(k) \geqslant 1$ is a function in k.

More broadly, the multi-labeling problems can seem to be special cases of the *cost allocation* [18] (CA) problem, in which k different nonnegative set functions are given for evaluating the k parts of the partition separately; they are also closely related to the *optimal allocation* (OA) problem [19–23] in combinatorial auctions, where no elements are necessarily pre-assigned a label, but the set function (called utility function) is assumed to be monotone in general.

1.2 Our Contributions

Our target problems are the MHV and the MUHV problems, and we aim to design improved approximation algorithms for them and to prove the hardness results in approximability.

We first show that the convex relaxation on the Lovász extension for the SUB- MP problem [7] can be extended for the SUB- ML problem; therefore, the same approximation algorithm works for SUB- ML with a performance ratio of $(2-2/k)$. Analogously, we present the concave relaxation on the Lovász extension for the SUP- ML problem, thus showing that SUP- ML can be approximated within a factor of $2/k$. Therefore, the MUHV problem can be approximated within a factor of $(2-2/k)$ and the MHV problem can be approximated within a factor of $2/k$; the $2/k$-approximation for MHV improves the previous best ratio of $\max\{1/k, 1/\big(\Delta + 1/g(\Delta)\big)\}$, where Δ is the maximum vertex degree of the input graph and $g(\Delta) = (\sqrt{\Delta} + \sqrt{\Delta+1})^2\Delta > 4\Delta^2$.

For the MUHV problem, the $(2-2/k)$-approximation can also be obtained due to its approximation-equivalent to the restricted-MUHV problem, which is a special case of SUB- MP. We also prove that the MUHV problem is approximation-equivalent to the HYP- MC problem [11]; thus, MUHV is Unique Games-hard to approximate within a factor of $(2-2/k-\varepsilon)$, for any $\varepsilon > 0$. This hardness result gives another evidence that it is Unique Games-hard to achieve a $(2-2/k-\varepsilon)$-approximation for the general SUB- ML problem, for any $\varepsilon > 0$.

For the MHV problem, we show that the LP relaxation for the MHV problem presented in [4], called LP-MHV, is equivalent to the concave relaxation for the SUP-ML problem based on the Lovász extension to the set function $g(\cdot)$ defined in Eq. (1.2). We then prove an upper bound of $2/k$ on the integrality gap of LP-MHV and conclude that the $2/k$-approximation is the best possible based on LP-MHV; thus, the $2/k$-approximation is also the best possible for the SUP- ML problem based on the concave relaxation on the Lovász extension. On the inapproximability of MHV, we prove that it is Unique Games-hard to approximate within a factor of $\Omega(\log^2 k/k)$, by showing an approximation preserving reduction from the *maximum independent set* (MIS) problem [24]. This hardness result also gives another evidence that it is Unique Games-hard to achieve an $\Omega(\log^2 k/k)$-approximation for the general SUP- ML problem.

1.3 Organization

The remainder of the paper is organized as follows: In Sect. 2, we introduce some basic notions such as the Lovász extension to a set function; we then present the relaxation based on the Lovász extension for the SUB- ML problem and a similar relaxation for the SUP- ML problem. We also present the approximation algorithm using the same randomized rounding technique for the SUB- MP problem in [8] and conclude that it is also a $(2 - 2/k)$-approximation for the SUB- ML problem and that it is a $2/k$-approximation for the SUP- ML problem. In Sect. 3, we study the MUHV problem which admits a $(2 - 2/k)$-approximation, and further show that it is approximation-equivalent to the HYP- MC problem; thus, MUHV is Unique Games-hard to approximate within a factor of $(2 - 2/k - \varepsilon)$, for any $\varepsilon > 0$. In Sect. 4, we study the MHV problem by first introducing the LP relaxation formulated in [4] and then showing its equivalence to the relaxation on the basis of the Lovász extension to the set function $f(\cdot)$ defined in Eq. (1.1), and proving an upper bound of $2/k$ on the integrality gap; lastly, we prove a hardness result for MHV that it is Unique Games-hard to achieve an $\Omega(\log^2 k/k)$-approximation. We conclude the paper in Sect. 5.

2 Preliminaries

Given a ground set $V = \{v_1, v_2, \cdots, v_n\}$, $y_j := y(v_j)$ is a real variable that maps the element v_j to the closed unit interval [0, 1]. For any nonnegative set function $f : 2^V \to \mathbb{R}_+$, its Lovász extension [3,25] is a function $\hat{f} : [0, 1]^V \to \mathbb{R}_+$ such that

$$\hat{f}(\vec{y}) = \sum_{j=1}^{n-1} (y_{\pi_j} - y_{\pi_{j+1}}) f(\{v_{\pi_1}, v_{\pi_2}, \cdots, v_{\pi_j}\}), \tag{2.1}$$

where $\vec{y} = (y_1, y_2, \cdots, y_n) \in [0, 1]^V$ and π is a permutation on $\{1, 2, \cdots, n\}$ such that $1 = y_{\pi_1} \geqslant y_{\pi_2} \geqslant \cdots \geqslant y_{\pi_n} = 0$. It has been proved by Lovász [3] that a set function is submodular (supermodular, respectively) if and only if its Lovász extension is convex (concave, respectively).

In the context of the SUB- ML problem with $f(\cdot)$ being the nonnegative submodular set function and $T_i \subset V$ being the non-empty subset of elements pre-labeled i, $i \in L = \{1, 2, \cdots, k\}$, we define a binary variable $y_j^i := y^i(v_j)$ for each pair of an element v_j and a label i, such that $y_j^i = 1$ if and only if the element v_j is labeled i. Next, y_j^i is relaxed to be a real variable in the closed unit interval [0, 1]. For each i, let $\vec{y}_i = (y_1^i, y_2^i, \cdots, y_n^i) \in [0, 1]^V$; let $\hat{f} : [0, 1]^V \to \mathbb{R}_+$ be the Lovász extension of $f(\cdot)$ as defined in Eq. (2.1). A relaxation that is based on the Lovász extension for SUB- ML can be written as follows:

$$\text{minimize} \quad \sum_{i=1}^{k} \hat{f}(\vec{y}_i) \tag{CP-Sub-ML}$$

$$\text{s.t.} \quad \sum_{i=1}^{k} y_j^i = 1, \quad \forall v_j \in V, \tag{2.2}$$

$$y_j^i = 1, \quad \forall v_j \in T_i, \; i \in L, \tag{2.3}$$

$$y_j^i \geqslant 0, \quad \forall v_j \in V, \; i \in L. \tag{2.4}$$

The submodularity of the function $f(\cdot)$ implies that (CP-Sub-ML) is a *convex program* (CP) and that it can be solved exactly in polynomial time [7].

In fact, such a relaxation that is based on the Lovász extension was proposed by Chekuri and Ene [7] for the SUB- MP problem, which is a special case of the SUB- ML problem in which $|T_i| = 1$ for every label i. We extend this relaxation for the SUB- ML problem with only one change that in the set of constraints (2.2) $y_j^i = 1$ holds for multiple elements v_j. We remark that one cannot reduce the SUB- ML problem to SUB- MP by cruelly contracting all the elements pre-labeled with the same label into a single element, which suggests incorrectly that all these pre-labeled elements were identical.

The following approximation algorithm RR first solves the convex program (CP–Sub-ML), followed by a randomized rounding scheme, which was applied to solve the SUB- MP problem in [8], to obtain a feasible solution to the SUB- ML problem.

Algorithm 1 RR

Step 1 Solve (CP-Sub-ML) to obtain an optimal fractional solution $\{y_j^i S \mid v_j \in V, i \in L\}$.
Step 2 Pick a parameter $\theta \in (\frac{1}{2}, 1]$ uniformly at random.
Step 3 Assign all elements of $S_i(\theta)$ the label i, for each $i \in L$.
Step 4 Pick a label i' from L uniformly at random, assign all elements of $R(\theta)$ the label i'.

Ene et al. [8] showed that RR is a $(2 - 2/k)$-approximation for SUB- MP. The algorithm uses a uniformly random variable θ in the interval $(\frac{1}{2}, 1]$ and defines the following $k+3$ sets:

$$\begin{aligned} S_i(\theta) &= \{v_j \mid y_j^i > \theta\}, \quad \text{for each } i \in L, \\ S(\theta) &= \bigcup_{i=1}^{k} S_i(\theta), \\ R(\theta) &= V - S(\theta), \\ Q(\theta) &= R(1-\theta). \end{aligned} \tag{2.5}$$

Then, it follows from the definition of Lovász extension in Eq. (2.1) that

$$\hat{f}(\vec{y}_i) = \sum_{j=1}^{n-1} (y_{\pi_j}^i - y_{\pi_{j+1}}^i) f(S_i(y_{\pi_{j+1}}^i)) = \int_0^1 f(S_i(\theta))\mathrm{d}\theta,$$

and thus, the optimal solution to (CP-Sub-ML) has a value

$$\text{OPT}(\text{CP} - \text{Sub} - \text{ML}) = \sum_{i=1}^{k} \hat{f}(\vec{y}_i) = \sum_{i=1}^{k} \int_0^1 f(S_i(\theta))\mathrm{d}\theta. \tag{2.6}$$

By proving the following three lemmas, Ene et al. [8] obtained an approximation ratio of $(2 - 2/k)$ for the algorithm for SUB- MP.

Lemma 2.1 (Lemma 2.2 in [8]) *The expected value of the solution returned by the Algorithm RR is*

$$\text{E[SOL(Sub-MP}, \theta)] = \left(2 - \frac{2}{k}\right) \sum_{i=1}^{k} \int_{\frac{1}{2}}^{1} f(S_i(\theta))\mathrm{d}\theta + \frac{2}{k} \sum_{i=1}^{k} \int_0^{\frac{1}{2}} f(S_i(\theta) \cup Q(\theta))\mathrm{d}\theta.$$

Lemma 2.2 (Lemma 2.5 in [8])

$$\sum_{i=1}^{k} \int_0^{\frac{1}{2}} f(S_i(\theta))\mathrm{d}\theta \geqslant \int_0^{\frac{1}{2}} f(Q(\theta))\mathrm{d}\theta.$$

Lemma 2.3 (Lemma 2.6 in [8])

$$\sum_{i=1}^{k} \int_0^{\frac{1}{2}} f(S_i(\theta))\mathrm{d}\theta \geqslant \sum_{i=1}^{k} \int_0^{\frac{1}{2}} f(S_i(\theta) \cup Q(\theta))\mathrm{d}\theta - (k-2) \int_0^{\frac{1}{2}} f(Q(\theta))\mathrm{d}\theta.$$

We observe that the performance analysis for the algorithm RR on the SUB- MP problem presented in [8] does not require $|\theta|$ for every label i. Therefore, the same analysis can also lead to the following theorem.

Theorem 2.4 *Algorithm RR is a* $(2 - 2/k)$*-approximation for the* SUB- ML *problem.*

Replacing the submodular function $f(\cdot)$ by a supermodular function $g(\cdot)$ and inverting the minimization to the maximization and letting $\hat{g} : [0, 1]^V \to \mathbb{R}_+$ be the Lovász extension of $g(\cdot)$, a relaxation based on the Lovász extension for the SUP- ML problem can be written as follows:

$$\text{maximize} \quad \sum_{i=1}^{k} \hat{g}(\vec{y}_i), \tag{CP-Sup-ML}$$

$$\text{s.t.} \quad \sum_{i=1}^{k} y_j^i = 1, \quad \forall v_j \in V, \tag{2.7}$$

$$y_j^i = 1, \quad \forall v_j \in T_i, \; i \in L, \tag{2.8}$$

$$y_j^i \geqslant 0, \quad \forall v_j \in V, \; i \in L. \tag{2.9}$$

(CP-Sup-ML) is a *concave program* and can similarly be solved in polynomial time. Using an analogous argument as the proof of Theorem 2.4, we can obtain the following corollary on the SUP- ML problem (Eq. (2.6) and Lemma 2.1 still hold for a supermodular set function, and in Lemma 2.3, the only difference is that the "$\geqslant$" sign of the inequality would be "$\leqslant$" for a supermodular set function).

Corollary 2.5 *Algorithm RR is a* $2/k$*-approximation for* SUP*-ML.*

3 The Minimum Unhappy Vertices (MUHV) Problem

Recall that the MUHV problem can be cast as finding a partition $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$ of the vertex set V such that $f(\mathcal{S}) = \sum_{i=1}^{k} f(S_i)$ is minimized, where the set function $f(\cdot)$ is defined in Eq. (1.1) and S_i is the subset of vertices colored i, for each i.

First, we prove the following two lemmas.

Lemma 3.1 *Given a graph* $G = (V, E)$*, the boundary* $\partial : 2^V \mapsto \mathbb{R}$ *has the following properties for any two subsets* $X, Y \subseteq V$:

(i) $\partial(\varnothing) = \varnothing$;
(ii) $\partial(X \cap Y) \subseteq \partial(X) \cup \partial(Y)$;
(iii) $\partial(X \cup Y) \subseteq \partial(X) \cup \partial(Y)$;
(iv) $\partial(X \cap Y) \cap \partial(X \cup Y) \subseteq \partial(X) \cap \partial(Y)$.

Proof Recall that for any $X \subseteq V$, $\partial(X)$ is the subset of vertices of X each has at least one neighbor outside of X. It follows that $\partial(\varnothing) = \varnothing$.

Next, for any $v \in \partial(X \cap Y)$, $v \in X \cap Y$ and v has a neighbor $u \notin X \cap Y$. That is, u is either outside of X or outside of Y. If u is outside of X, then $v \in \partial(X)$; otherwise, $v \in \partial(Y)$. Therefore, $\partial(X \cap Y) \subseteq \partial(X) \cup \partial(Y)$.

For any $v \in \partial(X \cup Y)$, $v \in X \cup Y$ and v has a neighbor $u \notin X \cup Y$. If $v \in X$, then $v \in \partial(X)$; otherwise, $v \in \partial(Y)$. Therefore, $\partial(X \cup Y) \subseteq \partial(X) \cup \partial(Y)$.

Lastly, from the last paragraph, if $v \in \partial(X \cap Y) \cap \partial(X \cup Y)$, then $v \in X \cap Y$ and v has a neighbor $u \notin X \cup Y$. These imply that $v \in \partial(X)$ and $v \in \partial(Y)$, *i.e.*, $v \in \partial(X) \cap \partial(Y)$. Therefore, $\partial(X \cap Y) \cap \partial(X \cup Y) \subseteq \partial(X) \cap \partial(Y)$.

Lemma 3.2 *Given a graph $G = (V, E)$, the set function $f(X) := w(\partial(X))$ defined in Eq.* (1.1) *satisfies $f(X) + f(Y) \geqslant f(X \cap Y) + f(X \cup Y)$, for any two subsets $X, Y \subseteq V$.*

Proof According to Lemma 3.1, the boundary $\partial : 2^V \mapsto \mathbb{R}$ satisfies

(ii) $\partial(X \cap Y) \subseteq \partial(X) \cup \partial(Y)$;
(iii) $\partial(X \cup Y) \subseteq \partial(X) \cup \partial(Y)$.

Therefore, $\partial(X \cap Y) \cup \partial(X \cup Y) \subseteq \partial(X) \cup \partial(Y)$ also holds. Furthermore, the boundary $\partial : 2^V \mapsto \mathbb{R}$ also satisfies

(iv) $\partial(X \cap Y) \cap \partial(X \cup Y) \subseteq \partial(X) \cap \partial(Y)$.

We thus conclude that

$$\begin{aligned} & w(\partial(X \cap Y) \cup \partial(X \cup Y)) + w(\partial(X \cap Y) \cap \partial(X \cup Y)) \\ \leqslant \; & w(\partial(X) \cup \partial(Y)) + w(\partial(X) \cap \partial(Y)), \end{aligned}$$

which is exactly

$$f(X) + f(Y) \geqslant f(X \cap Y) + f(X \cup Y).$$

Lemma 3.2 implies the submodularity of the function $f(\cdot)$ defined in Eq. (1.1); thus, we have:

Lemma 3.3 *The set function $f(\cdot)$ defined in Eq.* (1.1) *is submodular.*

Therefore, the MUHV problem is a special case of the SUB- ML problem, and the following theorem immediately follows according to Theorem 2.4.

Theorem 3.4 *There exists a $(2 - 2/k)$-approximation algorithm for the MUHV problem.*

On the other hand, we can prove that the general MUHV problem and the restricted-MUHV problem are approximation-equivalent by showing the following Lemma 3.5. Then, Theorem 3.4 also follows the fact that the restricted-MUHV problem can be cast as a special case of the SUB- MP problem, which admits a $(2-2/k)$-approximation [8].

Lemma 3.5 *If the restricted-*MUHV *problem admits a ρ-approximation algorithm, then the general* MUHV *problem also admits a ρ-approximation algorithm.*

Proof We prove this lemma by constructing a polynomial-time reduction from the general MUHV problem to the restricted-MUHV problem.

Given an instance $\mathcal{I} = (G = (V, E), w(\cdot), C = \{1, 2, \cdots, k\}, c)$ of the general MUHV problem, we construct an instance of $\mathcal{I}' = (G' = (V', E'), w'(\cdot), C = \{1, 2, \cdots, k\}, c')$ of restricted-MUHV as follows:

- For each color $i \in C$, we create a vertex t_i and connect t_i to all the vertices $v \in V$ with $c(v) = i$;
- Let $V' = V \cup T$, where $T = \{t_1, t_2, \cdots, t_k\}$, and $E' = E \cup \bigcup_{i=1}^{k} \{(t_i, v) | c(v) = i\}$;
- For each $v \in V$, let $w'(v) = w(v)$; for each $t_i \in T$, let $w(t_i) = W = \rho \cdot w(V) + 1$;
- Let $C = \{1, 2, \cdots, k\}$ still be the color set, and the partial coloring function c' only pre-colors the vertices in T with $c'(t_i) = i$, for $i \in C$.

Let $\text{OPT}(\mathcal{I})$ be the total weight of the optimal set of unhappy vertices in G; let $\text{OPT}(\mathcal{I}')$ be the total weight of the optimal set of unhappy vertices in G'.

For any coloring function c^* that completes the given partial coloring function c for G, we can apply the same function c^* to color all the uncolored vertices in G'. Then for each $t_i \in T$, $c^*(v) = i$ for any $v \in N(t_i)$, t_i must be happy. Thus, for any vertex in G, its happiness must be identical to the corresponding vertex in G', and they share the same weight. Therefore, under this coloring scheme, the set of unhappy vertices in G' has the same weight as the set of unhappy vertices in G. This also indicates that $\text{OPT}(\mathcal{I}') \leqslant \text{OPT}(\mathcal{I}) \leqslant w(V)$.

If the restricted-MUHV admits a ρ-approximation algorithm, then we can always find in polynomial time a coloring function c'^* that colors all the uncolored vertices in G', which makes $R \subseteq V'$ the set of unhappy vertices in G' and $w(R) \leqslant \rho \cdot \text{OPT}(\mathcal{I}') \leqslant \rho \cdot w(V)$. Under this coloring scheme, we must have $t_i \notin R$ for every $t_i \in T$, that is to say vertices in T must be all happy. Assume for the sake of contradiction that t_i is unhappy for some $t_i \in T$, then we have $w(R) \geqslant W = \rho \cdot w(V) + 1 > \rho \cdot w(V)$, a contradiction. Then, by applying the same function c'^* to color all the corresponding uncolored vertices in G, every vertex in G has the same happiness as the corresponding vertex in G', and they share the same weight. Therefore, under this coloring scheme, the corresponding set R in G is also the set of unhappy vertices in G.

In summary, the general MUHV problem is polynomial-time reducible to the restricted-MUHV problem, and if there exists a ρ-approximation for the restricted-MUHV problem, then the general MUHV problem also admits a ρ-approximation algorithm.

Next, we prove that the restricted-MUHV problem and the HYP- MC problem are approximation-equivalent; thus, MUHV and HYP- MC are also approximation-equivalent.

Given a hypergraph $H = (V_H, E_H)$ with a nonnegative weight $w(e)$ for each hyperedge $e \in E_H$ and a set of k terminals $T = \{t_1, t_2, \cdots, t_k\} \subseteq V$, the HYP- MC problem asks to remove a minimum-weight set of hyperedges so that every pair of terminals is disconnected.

Lemma 3.6 *There is an approximation preserving reduction from the restricted-MUHV problem to the HYP-MC problem.*

Proof Given an instance $\mathcal{I} = (G = (V, E), w(\cdot), C = \{1, 2, \cdots, k\}, c)$ of the restricted-MUHV problem, we construct an instance $\mathcal{I}' = (H = (V, E_H), w'(\cdot), T = \{t_1, t_2, \cdots, t_k\})$ of the HYP- MC problem as follows:

- Let the vertex set be V; for each $i \in C$, let v_i, which is pre-colored i, be a terminal t_i; let $T = \{t_1, t_2, \cdots, t_k\}$ be the terminal set;

- For each $v \in V$, we create a hyperedge $e_v = N[v]$ and add it to the hyperedge set E_H, where $N[v] = \{v\} \cup N(v)$ is the set of all the neighbors of v in G along with v itself;
- For each hyperedge $e_v \in E_H$, let $w'(e_v) = w(v)$.

Each vertex in G corresponds one-to-one to a hyperedge in H and shares the same weight.

Consider a simple path P connecting two terminals t_i and t_j in the constructed hypergraph H. Every two consecutive vertices on P must belong to a common hyperedge; thus, the path P corresponds one-to-one to a simple path connecting the two vertices t_i and t_j in G, which we also denote as P without any ambiguity. For any coloring function c^* that completes the given partial coloring function c for G, we have $c^*(t_i) = i$ for each $i = \{1, 2, \cdots, k\}$. It follows that any simple path P connecting t_i and t_j must contain at least one vertex $v \in V$ such that its preceding vertex or its succeeding vertex is colored differently from v itself. The vertex v is thus unhappy under the coloring scheme c^*. Then, in the hypergraph H, removing the corresponding hyperedge e_v breaks the path P, thus disconnecting t_i and t_j via the path P. Therefore, removing all the hyperedges, the corresponding vertices of which in the graph G are unhappy disconnects of all pairs of terminals in H, and the total weight of the removed hyperedges is equivalent to the total weight of the unhappy vertices in G.

Conversely, consider a subset E_H^* of hyperedges in the hypergraph $H = (V_H, E_H)$, the removal of which disconnects of which all pairs of terminals. Let V^i and E_H^i denote the subsets of vertices and hyperedges in the connected component of the remainder hypergraph $(V, E_H - E_H^*)$ that contains the terminal t_i, for each $i = 1, 2, \cdots, k$. We complete the partial coloring function c by coloring all vertices of the corresponding vertex set V^i in G with the color i, for $i = 1, 2, \cdots, k$, and coloring all other remaining vertices of V with the color 1. Clearly, all the vertices corresponding to the hyperedges of $E_H - E_H^*$ are happy. Thus, the total weight of unhappy vertices under this coloring scheme is no more than $w(E_H^*) := \sum_{e \in E_H^*} w(e)$.

In summary, the restricted-MUHV problem is polynomial-time reducible to the Hyp- MC problem, and our reduction preserves the value of any feasible solution and consequently preserves the approximation ratio.

We note that due to the $(2-2/k)$-approximation for the Hyp- MC problem [10,11], Theorem 3.4 can also be proved according to Lemma 3.6.

Lemma 3.7 *There is an approximation preserving reduction from the Hyp-MC problem to the restricted-MUHV problem.*

Proof Given an instance $\mathcal{I} = (H = (V_H, E_H), w(\cdot), T = \{t_1, t_2, \cdots, t_k\})$ of the Hyp- MC problem, we construct an instance $\mathcal{I}' = (G = (V, E), w'(\cdot), C = \{1, 2, \cdots, k\}, c)$ of the restricted-MUHV problem as follows:

- For each hyperedge $e \in E_H$, we create a vertex v_e;
- Let the vertex set be $V = V_H \cup V_E$, where $V_E = \{v_e \mid e \in E_H\}$ and call $T = \{t_1, t_2, \cdots, t_k\} \subseteq V$ the terminal set;
- For each vertex $v \in V_H$, let $w'(v) = 0$; for each vertex $v_e \in V_E$, let $w'(v_e) = w(e)$;

– For each vertex $v_e \in V_E$, it is adjacent to every vertex of e; let the edge set be $E = \{\{v_e, v\} \mid e \in E_H, v \in e\}$;
– Let the color set be $C = \{1, 2, \cdots, k\}$, and let the partial coloring function $c : V \mapsto C$ pre-color the terminal t_i with i, for each $i \in C$.

We note that the graph G is actually bipartite, and the two parts of vertices are V_H and V_E.

Consider a simple path P connecting two terminals t_i and t_j in the hypergraph H. Every two consecutive vertices on P must belong to a common hyperedge; thus, the path P corresponds one-to-one to a simple path connecting the two vertices t_i and t_j in G, which we also denote as P without any ambiguity. For any coloring function c^* that completes the given partial coloring function c for G, we have $c^*(t_i) = i$ for each $i = \{1, 2, \cdots, k\}$. It follows that any simple path P connecting t_i and t_j must contain at least one vertex $v_e \in V_E$ such that its preceding vertex and its succeeding vertex, both in V_H, are colored differently. The vertex v_e is thus unhappy under the coloring scheme c^*. Then, in the hypergraph H, removing the corresponding hyperedge e breaks the path P, thus disconnecting t_i and t_j via the path P. Therefore, removing all the hyperedges, the corresponding vertices of which in the graph G are unhappy disconnects all pairs of terminals, and the total weight of the removed hyperedges is equivalent to the total weight of the unhappy vertices in G.

Conversely, consider a subset E_H^* of hyperedges in the hypergraph $H = (V_H, E_H)$, the removal of which disconnects all pairs of terminals. Let V_H^i and E_H^i denote the subsets of vertices and hyperedges in the connected component of the remainder hypergraph $(V_H, E_H - E_H^*)$ that contains the terminal t_i, for each $i = 1, 2, \cdots, k$. Denote the vertex subsets in the constructed graph $G = (V, E)$ corresponding to V_H^i and E_H^i as V_H^i and V_E^i, respectively, for $i = 1, 2, \cdots, k$. We complete the partial coloring function c by coloring all vertices of $V_H^i \cup V_E^i$ with the color i, for $i = 1, 2, \cdots, k$, and coloring all the other remaining vertices of V with the color 1. Clearly, all the vertices of $\{v_e \mid e \in E_H - E_H^*\}$ are happy. Because every vertex of V_H has weight 0 (such that we may ignore its happiness), we conclude that the total weight of unhappy vertices under this coloring scheme is no more than $w(E_H^*) := \sum_{e \in E_H^*} w(e)$.

In summary, the Hyp- MC problem is polynomial-time reducible to the restricted-MUHV problem, and our reduction preserves the value of any feasible solution and consequently preserves the approximation ratio.

Ene et al. [8] proved that achieving a $(2 - 2/k - \varepsilon)$-approximation for Hyp- MC is NP-hard, for any $\varepsilon > 0$, assuming the Unique Games Conjecture. According to Lemma 3.7, we have Theorem 3.8; due to MUHV being a special case of Sub- ML, Corollary 3.9 immediately follows.

Theorem 3.8 *No $(2 - 2/k - \varepsilon)$-approximation algorithm for the restricted-MUHV or the general MUHV problem exists, for any $\varepsilon > 0$, assuming the Unique Games Conjecture.*

Corollary 3.9 *No $(2 - 2/k - \varepsilon)$-approximation algorithm for the Sup-ML problem exists, for any $\varepsilon > 0$, assuming the Unique Games Conjecture.*

4 The Maximum Happy Vertices (MHV) Problem

4.1 A 2/k-Approximation for MHV

Recall that the MHV problem can be cast as finding a partition $\mathcal{S} = \{S_1, S_2, \cdots, S_k\}$ of the vertex set V such that $g(\mathcal{S}) = \sum_{i=1}^{k} g(S_i)$ is maximized, where the set function $g(\cdot)$ is defined in Eq. (1.2) and S_i is the subset of vertices colored i, for each i. The following lemma can be proved analogously to Lemma 3.3.

Lemma 4.1 *The set function $g(\cdot)$ defined in Eq. (1.2) is supermodular.*

Therefore, the MHV problem is a special case of the SUP- ML problem, and the following theorem immediately follows according to Corollary 2.5:

Theorem 4.2 *Algorithm RR is a $2/k$-approximation for the MHV problem, which is a special case of the SUP-ML problem.*

The following LP relaxation for the MHV problem (LP-MHV) on a given graph $G = (V, E)$ was formulated in [4], where $V = \{v_1, v_2, \cdots, v_n\}$, $w_j = w(v_j)$ denotes the weight of the vertex v_j, C is the color set, $c(v_j) = i$ if the vertex v_j is pre-colored i, a binary variable $y_j^i := y^i(v_j)$ denotes whether the vertex v_j is colored i, z_j^i indicates whether the vertex v_j is happy by color i, z_j indicates whether the vertex v_j is happy, and $N[v_j]$ is the closed neighborhood of the vertex v_j.

$$\text{maximize} \quad \sum_{j=1}^{n} w_j z_j, \tag{LP-MHV}$$

$$\text{s.t.} \quad \sum_{i=1}^{k} y_j^i = 1, \quad \forall v_j \in V, \tag{4.1}$$

$$y_j^i = 1, \quad \forall v_j \in V,\ \forall i \in C \text{ s.t. } c(v_j) = i, \tag{4.2}$$

$$z_j^i = \min_{v_h \in N[v_j]} \{y_h^i\}, \quad \forall v_j \in V,\ \forall i \in C, \tag{4.3}$$

$$z_j = \sum_{i=1}^{k} z_j^i, \quad \forall v_j \in V, \tag{4.4}$$

$$z_j,\ z_j^i,\ y_j^i \geqslant 0, \quad \forall v_j \in V,\ \forall i \in C. \tag{4.5}$$

For each color i, since there is at least one vertex pre-colored i and at least one vertex pre-colored another color (due to $k \geqslant 2$), we let $\vec{y}_i = (y_1^i, y_2^i, \cdots, y_n^i)$ and π be the permutation for $\vec{y}_i$ such that $1 = y_{\pi_1}^i \geqslant y_{\pi_2}^i \geqslant \cdots \geqslant y_{\pi_n}^i = 0$. In the concave relaxation (CP-Sup-ML) based on the Lovász extension for SUP- ML, when we set the supermodular set function g as in Eq. (1.2), the objective function of

(CP-Sup-ML) becomes

$$\sum_{i=1}^{k} \hat{g}(\vec{y}_i) = \sum_{i=1}^{k} \sum_{j=1}^{n-1} \left(y^i_{\pi_j} - y^i_{\pi_{j+1}} \right) g(\{v_{\pi_1}, v_{\pi_2}, \cdots, v_{\pi_j}\})$$
$$= \sum_{i=1}^{k} \sum_{j=1}^{n-1} \left(y^i_{\pi_j} - y^i_{\pi_{j+1}} \right) \sum_{v_h \in i(\{v_{\pi_1}, v_{\pi_2}, \cdots, v_{\pi_j}\})} w_h. \tag{4.6}$$

For each vertex $v_p \in V$, let v_q denote its neighbor that appears the last in the permutation $(v_{\pi_1}, v_{\pi_2}, \cdots, v_{\pi_n})$. Assume $p = \pi_{j_1}$ and $q = \pi_{j_2}$. Clearly, $v_p \in i(\{v_{\pi_1}, v_{\pi_2}, \cdots, v_{\pi_j}\})$ if and only if $p, q \in \{\pi_1, \pi_2, \cdots, \pi_j\}$, that is to say, we must have $j_1, j_2 \leqslant j$. It follows that for the vertex $v_p \in V$, the coefficient of w_p in Eq. (4.6) is

$$\sum_{i=1}^{k} \sum_{j=\max\{j_1, j_2\}}^{n} \left(y^i_{\pi_j} - y^i_{\pi_{j+1}} \right) = \sum_{i=1}^{k} z^i_p = z_p,$$

where the last two equalities hold due to Constraints (4.3, 4.4) of (LP-MHV). This shows that by setting the supermodular set function g as defined in Eq. (1.2), (CP–Sup-ML) is the same as (LP-MHV). Therefore, we have the following theorem.

Theorem 4.3 *The LP relaxation for the MHV problem (LP-MHV) is the same as the relaxation based on the Lovász extension for the* Sup*-ML problem CP-Sub-ML, when the MHV problem is cast into the* Sup*-ML problem.*

Theorem 4.4 *The integrality gap of (LP-MHV) has an upper bound of* $2/k$.

Proof We prove this theorem by constructing an instance $\mathcal{I} = (G = (V, E), w(\cdot), C = \{1, 2, \cdots, k\}, c)$ of the MHV problem.

- Let $T = \{t_1, t_2, \cdots, t_k\}$ be a set of k pre-colored vertices, called *terminals*; all terminals have the same weight $w_t \geqslant 0$, and the terminal t_i is pre-colored i, *i.e.*, $c(t_i) = i$.
- Associated with each pair of distinct terminals t_i and t_j, $i < j$, there is a vertex $b_{\{ij\}}$. Let $V_b = \{b_{\{ij\}} \mid i < j\}$, then $|V_b| = \binom{k}{2}$; all vertices of V_b have the same weight $w_b \geqslant 0$, and none of them is pre-colored.
- The vertex set $V = T \cup V_b$; the edge set $E = \{\{t_i, b_{\{ij\}}\}, \{t_j, b_{\{ij\}}\} \mid i < j\}$. Clearly, $|V| = k + \binom{k}{2}$ and $|E| = 2\binom{k}{2}$.

Let c^* denote a coloring function that completes the given partial coloring function c, that is to say, c^* assigns a color for each vertex, and it assigns the color i to the terminal t_i, for each $i \in C$. Then,

- All vertices of V_b must be unhappy, since the vertex $b_{\{ij\}}$ is adjacent to two terminals t_i and t_j colored with distinct colors;
- The terminal t_i is adjacent to $k-1$ vertices $\{b_{\{ij\}} \mid j \neq i\}$, while the vertex $b_{\{ij\}}$ is adjacent to the terminals t_i and t_j; it follows that if t_i is happy, then all vertices of $\{b_{\{ij\}} \mid j \neq i\}$ are colored i, and subsequently, none of the other terminals can be happy; in other words, at most one of the k terminals can be happy, regardless of what the coloring function c^* is.

Let OPT(MHV) denote the value of an optimal solution to the constructed instance $\mathcal{I}$; we obtain

$$\text{OPT(MHV)} \leqslant w_t. \tag{4.7}$$

Consider the following fractional feasible solution to the instance $\mathcal{I}$ in the LP relaxation (LP-MHV),

- For each terminal $t_i \in T$, $y^i(t_i) = 1$ and $y^j(t_i) = 0$ for all $j \neq i$;
- For each vertex $b_{\{ij\}} \in V_b$, $y^i(b_{\{ij\}}) = y^j(b_{\{ij\}}) = 1/2$ and $y^\ell(b_{\{ij\}}) = 0$ for all $\ell \neq i, j$;
- For each terminal $t_i \in T$, we set $z^i(t_i) = y^i(b_{\{ij\}}) = 1/2$, $z^j(t_i) = 0$ for all $j \neq i$, and $z(t_i) = \sum_{\ell=1}^{k} z^\ell(t_i) = 1/2$;
- For each vertex $b_{\{ij\}} \in V_b$, we set $z^\ell(b_{\{ij\}}) = 0$ for all $\ell \in C$, and $z(b_{\{ij\}}) = 0$.

Let OPT(LP − MHV) denote the optimum of the instance $\mathcal{I}$ in the LP relaxation (LP-MHV). It is greater than or equal to the value of the above fractional feasible solution, that is,

$$\text{OPT(LP} - \text{MHV)} \geqslant \frac{1}{2} k w_t. \tag{4.8}$$

Combining Eqs. (4.7) and (4.8), it gives an upper bound on the integrality gap of the LP relaxation (LP-MHV):

$$\frac{\text{OPT(MHV)}}{\text{OPT(LP} - \text{MHV)}} \leqslant \frac{1}{\frac{1}{2}k} = \frac{2}{k}.$$

Theorems 4.2 and 4.4 together imply that the $2/k$-approximation Algorithm RR for the MHV problem is the best possible based on the LP relaxation (LP-MHV). Furthermore, the following corollary immediately follows.

Corollary 4.5 *The $2/k$-approximation algorithm RR for the Sup-ML problem is the best possible based on the concave relaxation on the Lovász extension (CP-Sup-ML).*

4.2 A Hardness Result for MHV

In this section, we show a hardness result on approximating the MHV problem by a reduction from the MIS problem, in which we are given an undirected graph $G = (V, E)$ with a nonnegative weight $w(v)$ for each vertex $v \in V$. The goal is to find a maximum-weight *independent set* $I \subseteq V$, *i.e.*, a subset of pairwise non-adjacent vertices. We also note that if there is a connected component in G that is exactly a clique, then the maximum-weight vertex in the clique is the optimal solution to the MIS problem on this connected component, *i.e.*, MIS is linear-time solvable on a clique. Thus, we assume without loss of generality that the input graph of the MIS problem does not contain a connected component being a clique.

We observe that any graph G with maximum degree of $\Delta \geqslant 3$ can also be viewed as a Δ-partite graph if G contains no clique of size $\Delta + 1$, by solving the classic COLORING problem, which is to color all the vertices in the given graph such that no two adjacent vertices have the same color. In other words, the COLORING problem asks to partition the vertex set into subsets of independent sets, and the number of the subsets of independent sets is then equivalent to the number of colors required to color all the vertices. From Brooks' theorem [26], along with a simplified proof presented by Lovász [27], one can solve the COLORING problem on G by using at most Δ colors in polynomial time.

Given an instance $\mathcal{I} = (G = (V, E), w(\cdot))$ of MIS, where G is a k-partite graph, with $k \geqslant 3$ and $V_1, V_2, \cdots, V_k$ being the k parts of the vertex set V, we construct an instance $\mathcal{I}' = (G' = (V', E'), w'(\cdot), C = \{1, 2, \cdots, k\}, c)$ of MHV as follows:

- For each edge $e = (u, v) \in E$, we break it into two edges (u, x_e) and (v, x_e) by creating an additional vertex x_e;
- Let $V' = V \cup X$, where $X = \{x_e | e \in E\}$, and $E' = \{(u, x_e), (v, x_e) | e = (u, v) \in E\}$;
- For each vertex $v \in V$, let $w'(v) = w(v)$; for each vertex $x_e \in X$, let $w'(x_e) = 0$;
- Let the color set be $C = \{1, 2, \cdots, k\}$ and let the partial coloring function $c : V' \mapsto C$ pre-color each vertex $v_i \in V_i$ with i, for $i = 1, 2, \cdots, k$.

We note that in the graph G', only the vertices in X are uncolored; all the neighbors of any vertex in V are in X; each vertex $x_e \in X$ has exactly two neighbors u and v, which correspond to the two endpoints of $e = (u, v) \in E$, and $c(u) \neq c(v)$ since G is k-partite; thus, x_e must be unhappy.

Consider an independent set $I \subseteq V$ of G. For any two vertices $u, v \in I$ in the graph G', they do not share any neighbor, *i.e.*, $N(u) \cap N(v) = \varnothing$. Assume for the sake of contradiction that there exists some $x \in N(u) \cap N(v)$, then $x \in X$ and $N(x) = \{u, v\}$, indicating that $(u, v) \in E$ in graph G, which contradicts to I being an independent set of G. Then in graph G', for any $v \in I$, we color every vertex in $N(v)$ with $c(v)$; for any x_e of the remaining uncolored vertices in X, with $N(x_e) = \{v_i, v_j\}$, where $c(v_i) = i$ and $c(v_j) = j$, we color x_e with any color in $C - \{i, j\}$. This is a feasible coloring scheme for G', which makes all the vertices in I happy and all the vertices in $V' - I$ unhappy in G'. Since in the constructed instance, the weights of all vertices in V are unchanged, the total weight of I is also unchanged.

Conversely, consider a feasible coloring scheme for G', which makes all the vertices in $S \subseteq V'$ happy and the remaining vertices unhappy. Then, $S \subseteq V$ and for any two vertices $u, v \in S$, either $u, v \in V_i$ for some $i \in C$ or $u \in V_i$ and $v \in V_j$ for two distinct $i, j \in C$. In both cases, we can conclude $(u, v) \notin E$ in G. The first case is straightforward; for the second case, assume for the sake of contradiction that $(u, v) \in E$ in G, then u and v cannot be both happy in G' since $c(u) \neq c(v)$ and they share a common neighbor. Therefore, S is also an independent set in G. Still, since in the constructed instance, the weights of all vertices in V are unchanged, the total weight of S is also unchanged.

Therefore, any feasible solution to the given instance $\mathcal{I}$ of MIS corresponds one to one to a feasible solution to the constructed instance $\mathcal{I}'$ of MHV, and the two solutions have exactly the same value.

In summary, the MIS problem is polynomial-time reducible to the MHV problem, and our reduction preserves the value of any feasible solution and consequently preserves the approximation ratio. Austrin [28] proved that MIS is Unique Games-hard to approximate within a factor of $\Omega(\log^2 \Delta/\Delta)$, where Δ is the maximum vertex degree of the input graph; thus, we have Theorem 4.6. Since MHV is a special case of Sup- ML, Corollary 4.7 immediately follows.

Theorem 4.6 *The MHV problem is Unique Games-hard to approximate within a factor of* $\Omega(\log^2 k/k)$.

Corollary 4.7 *The Sup-ML problem is Unique Games-hard to approximate within a factor of* $\Omega(\log^2 k/k)$.

5 Conclusions

We studied the MHV problem and its complement, the MUHV problem. We first showed that by rewriting the objective functions as set functions, the MHV and MUHV problems are actually a special case of the supermodular and submodular multi-labeling (Sup- ML and Sub- ML) problems, respectively. We next showed that the convex relaxation on the Lovász extension, presented by Chekuri and Ene for the submodular multi-partitioning (Sub- MP) problem [7], can be extended for the Sub- ML problem, thus proving that the Sub- ML (Sup- ML, respectively) can be approximated within a factor of $2 - 2/k$ ($2/k$, respectively). These general results imply that the MHV and the MUHV problems can also be approximated within $2/k$ and $2 - 2/k$, respectively, using the same approximation algorithms.

For MUHV, we showed that it is approximation-equivalent to the Hyp- MC problem; thus, it is Unique Games-hard to achieve a $(2 - 2/k - \varepsilon)$-approximation for MUHV, for any $\varepsilon > 0$. This hardness result also gives another evidence that it is Unique Games-hard to achieve a $(2 - 2/k - \varepsilon)$-approximation for the general Sub-ML problem, for any $\varepsilon > 0$.

For MHV, the $2/k$-approximation improves the previous best approximation ratio $\max\{1/k, 1/\big(\Delta + 1/g(\Delta)\big)\}$ [1,5] to $\max\{2/k, 1/\big(\Delta + 1/g(\Delta)\big)\}$, where Δ is the maximum vertex degree of the input graph and $g(\Delta) = (\sqrt{\Delta} + \sqrt{\Delta + 1})^2\Delta > 4\Delta^2$. We also showed that the LP relaxation for MHV presented by Zhang et al. [4] is the same as the concave relaxation on the Lovász extension for the SUP- ML problem; we then proved an upper bound of $2/k$ on the integrality gap of this LP relaxation. These suggest that the $2/k$-approximation algorithm is the best possible based on this LP relaxation; thus, the $2/k$-approximation algorithm is also the best possible based on the concave relaxation on the Lovász extension for the SUP- ML problem. Further, we proved that it is Unique Games-hard to approximate the MHV problem within a factor of $\Omega(\log^2 k/k)$, by a reduction from MIS, which also presents another evidence that the general SUP- ML problem is Unique Games-hard to approximate within a factor of $\Omega(\log^2 k/k)$. A possible future work would be to see if the approximation ratio of $\max\{2/k, 1/\big(\Delta + 1/g(\Delta)\big)\}$ for the MHV problem can be further improved.

Acknowledgements We want to thank Dr. Guo-Hui Lin and the anonymous reviewers of this paper for their valuable comments and suggestions that helped to improve the presentation

References

[1] Zhang, P., Li, A.: Algorithmic aspects of homophyly of networks. Theor. Comput. Sci. **593**, 117–131 (2015)
[2] Harary, F.: Graph Theory. Addison-Wesley, Reading (1969)
[3] Lovász, L.: Submodular functions and convexity. In: Bachem, A., Grötschel, M., Korte, B. (eds.) Mathematical Programming the State of the Art, pp. 235–257. Springer, Berlin (1983)
[4] Zhang, P., Jiang, T., Li, A.: Improved approximation algorithms for the maximum happy vertices and edges problems. In: Wang, L., Zhu, D. (eds.) International Computing and Combinatorics Conference, pp. 159–170. Springer, Berlin (2015)
[5] Zhang, P., Xu, Y., Jiang, T., Li, A., Lin, G., Miyano, E.: Improved approximation algorithms for the maximum happy vertices and edges problems. Algorithmica **80**(5), 1412–1438 (2017)
[6] Zhao, L., Nagamochi, H., Ibaraki, T.: Greedy splitting algorithms for approximating multiway partition problems. Math. Prog. **102**, 167–183 (2005)
[7] Chekuri, C., Ene, A.: Approximation algorithms for submodular multiway partition. In: IEEE 52nd Annual Symposium on Foundations of Computer Science (FOCS), 2011, IEEE, pp. 807–816 (2011a)
[8] Ene, A., Vondrák, J., Wu, Y.: Local distribution and the symmetry gap: Approximability of multiway partitioning problems. In: Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM, pp. 306–325 (2013)
[9] Dahlhaus, E., Johnson, D.S., Papadimitriou, C.H., Seymour, P.D., Yannakakis, M.: The complexity of multiterminal cuts. SIAM J. Comput. **23**, 864–894 (1994)
[10] Garg, N., Vazirani, V.V., Yannakakis, M.: Multiway cuts in node weighted graphs. J. Algorithms **50**, 49–61 (2004)
[11] Okumoto, K., Fukunaga, T., Nagamochi, H.: Divide-and-conquer algorithms for partitioning hypergraphs and submodular systems. Algorithmica **62**, 787–806 (2012)
[12] Călinescu, G., Karloff, H., Rabani, Y.: An improved approximation algorithm for multiway cut. In: Proceedings of the thirtieth annual ACM symposium on Theory of computing, ACM, pp. 48–52 (1998)
[13] Freund, A., Karloff, H.: A lower bound of $8/(7 + \frac{1}{k-1})$ on the integrality ratio of the Călinescu–Karloff–Rabani relaxation for multiway cut. Inf. Process. Lett. **75**, 43–50 (2000)
[14] Karger, D.R., Klein, P., Stein, C., Thorup, M., Young, N.E.: Rounding algorithms for a geometric embedding of minimum multiway cut. Math. Oper. Res. **29**, 436–461 (2004)
[15] Buchbinder, N., Naor, J.S., Schwartz, R.: Simplex partitioning via exponential clocks and the multiway cut problem. In: Proceedings of the forty-fifth annual ACM symposium on Theory of computing, ACM, pp. 535–544 (2013)
[16] Sharma, A., Vondrák, J.: Multiway cut, pairwise realizable distributions, and descending thresholds. In: Proceedings of the 46th Annual ACM Symposium on Theory of Computing, ACM, pp. 724–733 (2014)

[17] Langberg, M., Rabani. Y., Swamy, C.: Approximation algorithms for graph homomorphism problems. In: Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, pp. 176–187. Springer, Berlin (2006)
[18] Chekuri, C., Ene, A.: Submodular cost allocation problem and applications. In: Automata, Languages and Programming, pp 354–366. Springer, Berlin (2011)
[19] Lehmann, B., Lehmann, D., Nisan, N.: Combinatorial auctions with decreasing marginal utilities. In: Proceedings of the 3rd ACM conference on Electronic Commerce, ACM, pp. 18–28 (2001)
[20] Feige, U., Vondrak, J.: The allocation problem with submodular utility functions. In: Proceedings of the 47th Annual IEEE Symposium on Foundations of Computer Science (2006)
[21] Dobzinski, S., Schapira, M.: An improved approximation algorithm for combinatorial auctions with submodular bidders. In: Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithm, Society for Industrial and Applied Mathematics, pp 1064–1073 (2006)
[22] Feige, U.: On maximizing welfare when utility functions are subadditive. SIAM J. Comput. **39**, 122–142 (2009)
[23] Khot, S., Lipton, R.J., Markakis, E., Mehta, A.: Inapproximability results for combinatorial auctions with submodular utility functions. Algorithmica **52**, 3–18 (2008)
[24] Garey, M.R., Johnson, D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. W. H. Freeman and Company, San Francisco (1979)
[25] Vondrák, J.: Symmetry and approximability of submodular maximization problems. SIAM J. Comput. **42**, 265–304 (2013)
[26] Brooks, R.L.: On colouring the nodes of a network. Math. Proc. Camb. Philos. Soc. **37**, 194–197 (1941)
[27] Lovász, L.: Three short proofs in graph theory. J. Comb. Theory Ser. B **19**(3), 269–271 (1975)
[28] Austrin, P., Khot, S., Safra, M.: Inapproximability of vertex cover and independent set in bounded degree graphs. In: 24th Annual IEEE Conference on Computational Complexity, pp 74–80 (2009)