TSINGHUA SCIENCE AND TECHNOLOGY
ISSN 1007-0214 06/22 pp66–75
DOI: 10.26599/TST.2023.9010027
Volume 29, Number 1, February 2024

Robust Correlation Clustering Problem with Locally Bounded Disagreements

Sai Ji, Min Li, Mei Liang, and Zhenning Zhang*

Abstract: Min-max disagreements are an important generalization of the correlation clustering problem (CorCP). It can be defined as follows. Given a marked complete graph $G = (V, E)$, each edge in the graph is marked by a positive label "+" or a negative label "–" based on the similarity of the connected vertices. The goal is to find a clustering $\mathcal{C}$ of vertices V, so as to minimize the number of disagreements at the vertex with the most disagreements. Here, the disagreements are the positive cut edges and the negative non-cut edges produced by clustering $\mathcal{C}$. This paper considers two robust min-max disagreements: min-max disagreements with outliers and min-max disagreements with penalties. Given parameter $\delta \in (0, 1/14)$, we first provide a threshold-based iterative clustering algorithm based on LP-rounding technique, which is a $(1/\delta, 7/(1-14\delta))$-bi-criteria approximation algorithm for both the min-max disagreements with outliers and the min-max disagreements with outliers on one-sided complete bipartite graphs. Next, we verify that the above algorithm can achieve an approximation ratio of 21 for min-max disagreements with penalties when we set parameter $\delta = 1/21$.

Key words: min-max disagreements; outliers; penalties; approximation algorithm; LP-rounding

1 Introduction

Correlation clustering problem (CorCP) has many applications in machine learning, computer vision, data mining, and so on, and it has been widely studied in the past decade[1–6].

This CorCP was first introduced by Bansal et al.[7], and it is inspired from a document clustering problem. The input of the problem is a complete graph $G = (V, E)$ with N vertices, and each edge (p, q) is marked by a positive label "+" or a negative label "–" based on the similarity of vertex p and vertex q. The goal of the problem is to find a clustering $\mathcal{C}$ of set V, so as to minimize the number of positive cut edges and negative non-cut edges. Here, we can call each positive cut edge and each negative non-cut edge as a disagreement. In the CorCP, there is no limit on the number of clusters. When all the edges in the graph are positive, the optimal clustering of V contains only one cluster, and when all the edges are negative, the optimal clustering of V contains N clusters. Therefore, the number of clusters depends on the specific instance of the CorCP.

The CorCP is NP-hard, and people always use the combinational technique and LP-rounding technique to study the approximation algorithms rather than the exact algorithms of this problem[8–13]. The first constant approximation algorithm and the best approximation

A preliminary version of this paper was published in the proceedings of the 15th Annual International Conference on Combinatorial Optimization and Applications (COCOA), pp. 668–675, 2021.

• Sai Ji is with the Institute of Mathematics, Hebei University of Technology, Tianjin 300401, China. E-mail: jisai@hebut.edu.cn.
• Min Li is with the School of Mathematics and Statistics, Shandong Normal University, Jinan 250358, China. E-mail: liminemily@sdnu.edu.cn.
• Mei Liang is with the College of Statistics and Data Science, Beijing University of Technology, Beijing 100124, China. E-mail: chn_lmmei@emails.bjut.edu.cn.
• Zhenning Zhang is with the Department of Operations Research and Information Engineering, Beijing University of Technology, Beijing 100124, China. E-mail: zhangzhenning@bjut.edu.cn.
∗ To whom correspondence should be addressed.

Manuscript received: 2022-03-30; revised: 2023-01-05; accepted: 2023-03-06

algorithm for the CorCP are the 17 433-approximation algorithm given by Bansal et al.[7] and the 2.06-approximation algorithm given by Chawla et al.[14], respectively.

Besides the traditional CorCP, several interesting generalizations of the CorCP have also attracted the research interest of many scholars[15–23]. In this paper, we pay attention to three generalizations of the CorCP: min-max disagreements, CorCP with outliers, and CorCP with penalties.

Min-max disagreements are an important generalization of the CorCP, which was introduced by Puleo and Milenkovic[15]. The goal of CorCP is to optimize the number of global disagreements, while the goal of min-max disagreements is to optimize the number of local disagreements. Specifically, the goal of the min-max disagreements is to minimize the number of disagreements at the vertex with the most disagreements. Puleo and Milenkovic[15] first studied this problem and gave an LP-rounding based 48-approximation algorithm. Then, they provided a 10-approximation algorithm for the min-max disagreements on complete bipartite graphs, where the disagreements are measured on one side of the graph. In this paper, we refer to the latter problem simply as min-max disagreements on one-sided complete bipartite graphs. Later, Charikar et al.[16] provided an algorithm that achieves an approximation ratio of 7 for both min-max disagreements and min-max disagreements on one-sided complete bipartite graphs.

CorCP with outliers is a kind of robust CorCP, which was first introduced by Devvirt et al.[17] In this problem, in addition to a labeled complete graph, we are given an upper bound r for the number of vertices that can be un-clustered. We call each vertex that is not clustered an outlier and each vertex that is clustered a non-outlier. It is worth noting that once a vertex is selected as an outlier, it will not be clustered, so it will not produce any disagreements. Then, the goal of the CorCP with outliers is to find r outliers and a clustering of non-outliers, so as to minimize the number of disagreements generated by all the non-outliers. Devvrit et al.[17] proved that it is NP-hard to obtain any finite approximation factor of the CorCP with outliers unless the constraint on the number of outliers is not satisfied. Then, they provided a $(6, 6)$-bi-criteria approximation algorithm for the CorCP with outliers. In this paper, we define an algorithm as an (α, β)-bi-criteria approximation for the CorCP with outliers. If for any instance of this problem, the number of outliers output by the algorithm is no more than αr, the number of disagreements output by the algorithm shall not exceed β times of the number of disagreements in the optimal solution.

CorCP with penalties is one kind of robust CorCP, which was first introduced by Aboud and Rabani[23]. They provided a 9-approximation algorithm based on primal dual technique. Compared with CorCP with outliers, the number of un-clustered vertices has no limitation in CorCP with penalties, but each un-clustered vertex generates a penalty cost. The problem of the CorCP with penalties is described as follows. For a given marked complete graph, each vertex has a penalty cost. The goal is to find a set of un-clustered vertices, as well as a clustering of clustered vertices, so as to minimize the sum of the number of disagreements generated by clustered vertices and the penalty costs of the un-clustered vertices.

Since simple generalizations of the CorCP may not be able to accurately describe some practical problems, we combine min-max disagreements, CorCP with outliers, and CorCP with penalties to raise the following two problems: min-max disagreements with outliers and min-max disagreements with penalties. These two generalizations are described as follows. In the min-max disagreements with outliers, we are given a marked complete graph and an upper bound r of the number of outliers. The target is to find r outliers as well as a clustering of non-outliers, so as to minimize the number of disagreements at the non-outlier with the most disagreements. In the min-max disagreements with penalties, we are given a marked complete graph and an un-uniform penalty cost for each vertex. The purpose is to find a set of un-clustered vertices, as well as a clustering of clustered vertices, so as to minimize the maximum values of the disagreements at the clustered vertex and the largest penalty cost at the un-clustered vertex.

There are three main contributions of this paper.

(1) We introduce the min-max disagreements with outliers and provide a threshold-based iterative clustering algorithm based on Refs. [16, 17], which is a $(1/\delta, 7/(1-14\delta))$-bi-criteria approximation algorithm with the parameter $\delta \in (0, 1/14)$.

(2) We claim that the threshold-based iterative clustering algorithm is also suitable for the min-max disagreements with outliers on one-sided complete bipartite graphs, and has the same bi-criteria.

(3) We introduce the threshold-based iterative clustering algorithm to overcome the problem of min-

max disagreements with penalties, and the algorithm is guaranteed by 21 with $\delta = 1/21$.

The rest of this paper is structured as follows. In Section 2, we describe the definition and two formulations for the min-max disagreements with outliers. In Section 3, we design a threshold-based iterative clustering algorithm for the min-max disagreements with outliers. In Sections 4 and 5, we state that the threshold-based iterative clustering algorithm is both effective for min-max disagreements with outliers on one-sided complete bipartite graphs and min-max disagreements with penalties. Finally, some conclusions are provided in Section 6.

2 Min-Max Disagreements

Firstly, we introduce some notations and definitions used throughout our paper.

For each positive integer m, let $[m]$ be the set of all the positive integers less than or equal to m. $G = (V, E)$ is a given marked graph, where V is the set of vertexes and E is the set of edges. Let $E^+ := \{(p,q) \in E : (p,q) \text{ is a positive edge}\}$, and $E^- := E \backslash E^+$. For a set $S \subseteq V$, the clustering is denoted by $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}, m \in [|S|]$. The disagreement of a vertex $q \in C_i, i \in [m]$ is the positive cut edges $\{(p,q) \in E^+ : p \in S \backslash C_i\}$ and the negative non-cut edges $\{(p,q) \in E^- : p \in C_i\}$. We use the following symbols to denote the numbers of these kinds of edges, that is,

$$\begin{aligned} \text{dis}_q^+(\mathcal{C}, S) &= |\{(p,q) \in E^+ : p \in S \backslash C_i\}|, \\ \text{dis}_q^-(\mathcal{C}, S) &= |\{(p,q) \in E^- : p \in C_i\}|, \\ \text{dis}_q(\mathcal{C}, S) &= \text{dis}_q^+(\mathcal{C}, S) + \text{dis}_q^-(\mathcal{C}, S). \end{aligned}$$

In the following, we give the descriptions of the problems we considered.

Definition 1 (Min-max disagreements) Given a marked complete graph $G = (V, E)$ with $N = |V|$ vertices, the goal of the problem is to find a clustering $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}, m \in [N]$ of V, such that

$$\max_{q \in C_i, i \in [m]} \text{dis}_q(\mathcal{C}, V)$$

is minimized.

Definition 2 (Min-max disagreements with outliers) Given an instance $\mathcal{I} = \{G = (V, E), r\}$, where G is a marked complete graph with N vertices and an upper bound r on the number of outliers, the target is to find a set O with r outliers as well as a clustering $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}, m \in [N - r]$ of $V \backslash O$, such that

$$\max_{q \in C_i, i \in [m]} \text{dis}_q(\mathcal{C}, V \backslash O)$$

is minimized.

Definition 3 (Bi-criteria approximation algorithm for the min-max disagreements with outliers) Algorithm A is an (α, β)-bi-criteria approximation algorithm for the min-max disagreements with outliers. If for any instance $\mathcal{I} = \{G = (V, E), r\}$ of the min-max disagreements with outliers, algorithm A can return a set O of outliers and a clustering $\mathcal{C}$ of $V \backslash O$ which satisfy:

(1) $|O| \leqslant \alpha r$;

(2) $\max_{q \in V \backslash O} \text{dis}_q(\mathcal{C}, V \backslash O) \leqslant \beta \max_{q \in V \backslash O^*} \text{dis}_q(\mathcal{C}^*, V \backslash O^*)$, where O^* and $\mathcal{C}^*$ are the set of outliers and the clustering of non-outliers returned by the optimal algorithm, respectively.

3 Min-Max Disagreements with Outliers

In this section, we present our threshold-based iterative clustering algorithm for the min-max disagreements with outliers based on Refs. [16, 17]. Two formulations of the problem are given in Section 3.1. The approximation algorithm and the theoretical analysis are shown in Sections 3.2 and 3.3, respectively.

3.1 Definition and formulation of the min-max disagreements with outliers

In order to give an integer programming for the min-max disagreements with outliers, we introduce variables X_{pq} and Z_{pq} for each edge (p,q) and a variable Y_q for each vertex q. These variables are explained as follows.

(1) For each $(p,q) \in E$, let X_{pq} indicate whether (p,q) is a cut edge. If $X_{pq} = 1$, then (p,q) is a cut edge. Otherwise, (p,q) is a non-cut edge.

(2) For each $q \in V$, let Y_q indicate whether vertex q is an outlier. If $Y_q = 1$, then q is an outlier. Otherwise, q is a non-outlier.

(3) For each $(p,q) \in E$, let Z_{pq} indicate whether edge (p,q) is a disagreement. If $Z_{pq} = 1$, then edge (p,q) is a disagreement. Otherwise, edge (p,q) is not a disagreement.

Based on above three types of variables, we can formulate the min-max disagreements with outliers as Formula (1).

$$\begin{aligned} \min \quad & \max_{q \in V} \sum_{p \in V} Z_{pq}, \\ \text{s.t.,} \quad & X_{pq} + X_{qw} \geqslant X_{pw}, \quad \forall p, q, w \in V, \\ & Y_p + Y_q + Z_{pq} \geqslant 1 - X_{pq}, \ \forall (p,q) \in E^-, \\ & Y_p + Y_q + Z_{pq} \geqslant X_{pq}, \ \forall (p,q) \in E^+, \\ & \sum_{q \in V} Y_q = r, \\ & X_{qq} = 0, \quad \forall q \in V, \\ & X_{pq}, Z_{pq}, Y_p \in \{0, 1\}, \quad \forall p, q \in V \end{aligned} \tag{1}$$

The objective value is the number of positive cut edges and negative non-cut edges for a vertex q. There are five kinds of constraints in Formula (1). The first constraint guarantees that the output is a reasonable solution of the CorCP. The next two constraints give the condition for an edge to be a disagreement. The fourth constraint ensures that there are exactly r outliers output by Formula (1). The last constraint ensures that each vertex will only be in one cluster. By relaxing the each variable to interval [0, 1], we obtain the following LP relaxation of Formula (1):

$$
\begin{aligned}
\min \quad & \max_{q\in V} \sum_{p\in V} Z_{pq}, & \\
\text{s.t.,} \quad & X_{pq} + X_{qw} \geqslant X_{pw}, & \forall p,q,w \in V, \\
& Y_p + Y_q + Z_{pq} \geqslant 1 - X_{pq}, & (p,q) \in E^-, \\
& Y_p + Y_q + Z_{pq} \geqslant X_{pq}, & \forall (p,q) \in E^+, \quad (2) \\
& \sum_{q\in V} Y_q = r, & \\
& X_{qq} = 0, & \forall q \in V, \\
& X_{pq}, Z_{pq}, Y_p \in [0,1], & \forall p,q \in V
\end{aligned}
$$

3.2 Algorithm

In this subsection, we provide a threshold-based iterative clustering algorithm for the min-max disagreements with outliers. All steps of the algorithm are based on the fractional optimal solution (X^*, Y^*, Z^*) of Formula (2). Algorithm 1 has two main processes. The first process is finding outliers (Line 3). The second is clustering non-outliers, which is the core of the algorithm (Lines 4–11).

In the first process, we find a set of outliers based on a parameter δ as well as the value Y^* of each vertex. For each $q \in V$, we make q as an outlier if $Y_q^* \geqslant \delta$. The number of outliers may not satisfy the fourth constraint of Formula (2), but we can prove that the number of outliers does not exceed a constant multiple of r.

In the second process, we find a clustering of the non-outliers based on the value of X^*. This process is an iterative process. In each iteration, we first construct a set of neighbors for each un-clustered non-outlier based on a given parameter $1/7$. Then, we specify the vertex with the most neighbors as the center. Finally, we construct a cluster for the center based on a parameter $3/7$. We repeat the iterative process until all the non-outliers are clustered.

3.3 Theoretical analysis of the bi-criteria ratio

Let O be the set of outliers and $C = \{q_1^*, q_2^*, \ldots, q_m^*\}$ be the center set output by Algorithm 1. Moreover, let $\mathcal{C} = \{C_{q_1^*}, C_{q_2^*}, \ldots, C_{q_m^*}\}$ be the clustering of $V \backslash O$. From Algorithm 1 and the constraints of Formula (2), we have Property 1.

Algorithm 1 Threshold-based iterative clustering algorithm

Input: Given an instance $\mathcal{I} = \{G = (V,E), r\}$ and a parameter $\delta \in (0, 1/14)$.

Output: A set O of outliers and a clustering of $V \backslash O$.

1: Let $\mathcal{U} := V, C := \varnothing, O := \varnothing$, and $t := 1$.
2: Obtain the optimal fractional solution (X^*, Y^*, Z^*) of $\mathcal{I}$ by solving Programming (2).
3: Update $O := \{q \in V : Y_q^* \geqslant \delta\}$ and $\mathcal{U} := \mathcal{U} \backslash O$
4: **while** $\mathcal{U} \neq \varnothing$ **do**
5: **for** each $q \in \mathcal{U}$ **do**
6: Denote $C_q^*(t) := \{p \in \mathcal{U} : X_{pq}^* \leqslant 1/7\}$.
7: **end for**
8: Choose vertex
$$q_t^* := \underset{q\in\mathcal{U}}{\operatorname{argmax}} |C_q^*|.$$
9: Make $C_{q_t^*} := \{p \in \mathcal{U} : X_{pq^*}^* \leqslant 3/7\}$ as a cluster.
10: Update $\mathcal{U} := \mathcal{U} - C_{q_t^*}, C := C \cup \{q_t^*\}$, and $t := t+1$.
11: **end while**
12: **return** O and clustering $\mathcal{C} = \{C_{q_1^*}, C_{q_2^*}, \ldots, C_{q_t^*}\}$.

Property 1 The solutions returned by Algorithm 1 satisfies:

(1) For each $q \in O$, we have $Y_q^* \geqslant \delta$;

(2) For each $q \in V \backslash O$, we have $Y_q^* < \delta$;

(3) For each $(p,q) \in E^+, p, q \in V \backslash O$, we have $Z_{pq}^* \geqslant X_{pq}^* - 2\delta$;

(4) For each $(p,q) \in E^-, p, q \in V \backslash O$, we have $Z_{pq}^* \geqslant 1 - X_{pq}^* - 2\delta$.

Lemma 1 The size of set O is no more than r/δ.

Proof From the forth constraint of Programming (2), we have

$$\sum_{q\in V} Y_q^* = r.$$

Moreover, from Property 1-(1), we can obtain that

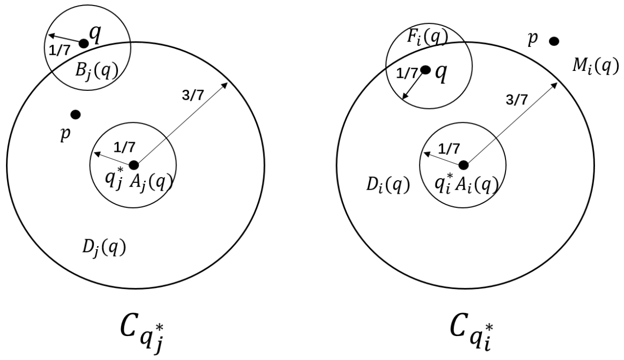
$$|O| \leqslant \sum_{q\in O} \frac{1}{\delta} Y_q^* \leqslant \frac{1}{\delta} \sum_{q\in V} Y_q^* = \frac{r}{\delta}.$$

The lemma is concluded. ■

Next, we analyze the value of $\mathrm{dis}_q(\mathcal{C}, V\backslash O)$ for each $q \in C_{q_i^*}, i \in [m]$. As shown in Fig. 1, for each vertex $q \in C_{q_i^*}, i \in [m]$, and $j \in [i-1]$, denote

$$
\begin{aligned}
A_j(q) &= C_{q_j^*}^*(j), \\
B_j(q) &= C_{q_j^*} \cap C_q^*(j), \\
D_j(q) &= C_{q_j^*} \backslash \left(A_j(q) \cup B_j(q)\right).
\end{aligned}
$$

Similar, for each $q \in C_{q_i^*}, i \in [m]$, denote



(a) $p \in C_{q_j^*},\ j \in [i-1]$ (b) $p \in C_{q_j^*},\ j \in [m]\backslash[i]$

Fig. 1 Relationship between vertex *q* and several special sets.

$$\begin{aligned}
A_i(q) &= C_{q_i^*}^*(i), \\
D_i(q) &= C_{q_i^*} - \left(A_i(q) \cup C_q^*(i)\right), \\
F_i(q) &= C_q^*(i) - C_{q_i^*}, \\
M_i(q) &= \cup_{t \in [m]\backslash[i]} C_{q_t^*} - F_i(q).
\end{aligned}$$

Next, we analyze the number of disagreements at each vertex $q \in C_{q_i^*}, i \in [m]$ based on the above sets.

3.3.1 Positive cut edges

In this subsection, we analyze the number of disagreements generated by positive cut edges at each vertex $q \in C_{q_i^*}, i \in [m]$.

Lemma 2 For each vertex $q \in C_{q_i^*}, i \in [m]$, the number of positive cut edge $(p,q), p \in M_i(q) \cup D_j(q),\ j \in [i-1]$, is no more than

$$\frac{7}{1-14\delta} \sum_{\substack{(p,q) \in E^+ \\ p \in M_i(q) \cup D_j(q) \\ j \in [i-1]}} Z_{pq}^*.$$

Proof From the construction of sets $D_j, j \in [i-1]$, and $M_i(q)$, for each cut edge $(p,q) \in E^+, p \in M_i(q) \cup D_j(q), j \in [i-1]$, we have

$$X_{pq}^* \geqslant \frac{1}{7}.$$

Moreover, from Property 1-(3), we can obtain

$$Z_{pq}^* \geqslant X_{pq}^* - 2\delta \geqslant \frac{1}{7} - 2\delta,$$

which indicates that the disagreement generated by (p,q) is no more than

$$\frac{7}{1-14\delta} Z_{pq}^*.$$

We sum all the $p \in M_i(q) \cup D_j(q), j \in [i-1]$, and we have that the number of positive cut edge $(p,q), p \in M_i(q) \cup D_j(q),\ j \in [i-1]$ is no more than

$$\frac{7}{1-14\delta} \sum_{\substack{(p,q) \in E^+ \\ p \in M_i(q) \cup D_j(q) \\ j \in [i-1]}} Z_{pq}^*.$$

The lemma is concluded. ■

Lemma 3 For each vertex $q \in C_{q_i^*}, i \in [m]$, the number of positive cut edge $(p,q), p \in A_j(q) \cup B_j(q), j \in [i-1]$ is no more than

$$\sum_{\substack{(p,q) \in E^+ \\ p \in A_j(q) \\ j \in [i-1]}} \frac{7}{1-7\delta} Z_{pq}^* + \sum_{\substack{(p,q) \in E^- \\ p \in A_j(q) \\ j \in [i-1]}} \frac{7}{2-14\delta} Z_{pq}^*.$$

Proof Given a vertex $q \in C_{q_i^*}, i \in [m]$ and any $j \in [i-1]$, for each positive edge $(p,q), p \in A_j(q)$, from Property 1-(3) we receive that

$$\begin{aligned}
Z_{pq}^* \geqslant\ & X_{pq}^* - 2\delta \geqslant \\
& X_{qq_j^*}^* - X_{pq_j^*}^* - 2\delta \geqslant \\
& \frac{2}{7} - 2\delta.
\end{aligned}$$

Moreover, for each negative edge $(p,q), p \in A_j(q)$, from Property 1-(4) we achieve that

$$\begin{aligned}
Z_{pq}^* \geqslant\ & 1 - X_{pq}^* - 2\delta \geqslant \\
& 1 - \left(X_{pq_j^*}^* + X_{wq_j^*}^* + X_{qw}^*\right) - 2\delta \geqslant \\
& 1 - \left(\frac{1}{7} + \frac{3}{7} + \frac{1}{7}\right) - 2\delta = \frac{2}{7} - 2\delta,
\end{aligned}$$

where w is any vertex in set $B_j(q)$.

From Line 8 of Algorithm 1, we have that

$$|B_j(q)| \leqslant |C_q^*(j)| \leqslant |A_j(q)|.$$

Then, the number of positive cut edge $(p,q), p \in A_j(q) \cup B_j(q)$ equals $|\{(p,q) \in E^+ : A_j(q) \cup B_j(q)\}|$ and it is no more than

$$\begin{aligned}
& |\{(p,q) \in E : p \in A_j(q)\}| + \\
& \quad \{(p,q) \in E^+ : p \in A_j(q)\}| = \\
& 2|\{(p,q) \in E^+ : p \in A_j(q)\}| + \\
& \quad |\{(p,q) \in E^- : p \in A_j(q)\}|.
\end{aligned}$$

Above all, we get

$$\begin{aligned}
& |\{(p,q) \in E^+ : p \in A_j(q) \cup B_j(q)\}| \leqslant \\
& 2|\{(p,q) \in E^+ : p \in A_j(q)\}| + \\
& \quad |\{(p,q) \in E^- : p \in A_j(q)\}| \leqslant \\
& 2 \sum_{\substack{(p,q) \in E^+ \\ p \in A_j(q)}} \frac{7}{2-14\delta} Z_{pq}^* +
\end{aligned}$$

$$\sum_{\substack{(p,q)\in E^-\\ p\in A_j(q)}} \frac{7}{2-14\delta} Z^*_{pq} \leqslant$$

$$\sum_{\substack{(p,q)\in E^+\\ p\in A_j(q)}} \frac{7}{1-7\delta} Z^*_{pq} +$$

$$\sum_{\substack{(p,q)\in E^-\\ p\in A_j(q)}} \frac{7}{2-14\delta} Z^*_{pq}.$$

Therefore, we have

$$|\{(p,q)\in E^+ : p\in A_j(q)\cup B_j(q), j\in[i-1]\}| \leqslant$$

$$\sum_{\substack{(p,q)\in E^+\\ p\in A_j(q)\\ j\in[i-1]}} \frac{7}{1-7\delta} Z^*_{pq} +$$

$$\sum_{\substack{(p,q)\in E^-\\ p\in A_j(q)\\ j\in[i-1]}} \frac{7}{2-14\delta} Z^*_{pq}.$$

The lemma is concluded. ■

Lemma 4 For each vertex $q\in C_{q_i^*}, i\in[m]$, the number of positive cut edges $(p,q), p\in F_i(q)$, is no more than

$$\sum_{\substack{(p,q)\in E^+\\ p\in A_i(q)}} \frac{7}{1-14\delta} Z^*_{pq} + \sum_{\substack{(p,q)\in E^-\\ p\in A_i(q)}} \frac{7}{3-14\delta} Z^*_{pq}.$$

Proof From Line 8 of Algorithm 1, we receive that

$$|F_i(q)| \leqslant |C^*_q(i)| \leqslant |A_i(q)|.$$

Therefore, we have

$$|\{(p,q)\in E^+ : p\in F_i(q)\}| \leqslant$$
$$|\{(p,q)\in E^+ : p\in A_i(q)\}| +$$
$$|\{(p,q)\in E^- : p\in A_i(q)\}|.$$

Similar to Lemma 3, we only need to analyze the lower bound of Z^*_{pq} for each edge $(p,q), p\in A_i(q)$.

For each positive edge $(p,q)\in E^+$ with $q\in A_i(q)$, from Property 1-(3) and the first constraint of Formula (2), we can get

$$\begin{aligned} Z^*_{pq} &\geqslant X^*_{pq} - 2\delta \geqslant \\ & X^*_{q_i^* w} - X^*_{pq_i^*} - X^*_{qw} - 2\delta \geqslant \\ & \frac{3}{7} - \frac{1}{7} - \frac{1}{7} - 2\delta = \frac{1}{7} - 2\delta, \end{aligned}$$

where w is any vertex in set $F_i(q)$.

Moreover, from Property 1-(4) and the first constraint of Formula (2), for each $(p,q)\in E^-$ with $p\in A_i(q)$, we achieve

$$\begin{aligned} Z^*_{pq} &\geqslant 1 - X^*_{pq} - 2\delta \geqslant \\ & 1 - \left(X^*_{qq_i^*} + X^*_{pq_i^*}\right) - 2\delta \geqslant \\ & 1 - \left(\frac{3}{7} + \frac{1}{7}\right) - 2\delta = \frac{3}{7} - 2\delta \end{aligned} \tag{3}$$

Above all, we obtain that

$$|\{(p,q)\in E^+ : p\in F_i(q)\}| \leqslant$$
$$|\{(p,q)\in E^+ : p\in A_i(q)\}| +$$
$$|\{(p,q)\in E^- : p\in A_i(q)\}| \leqslant$$

$$\sum_{\substack{(p,q)\in E^+\\ p\in A_i(q)}} \frac{7}{1-14\delta} Z^*_{pq} +$$

$$\sum_{\substack{(p,q)\in E^-\\ p\in A_i(q)}} \frac{7}{3-14\delta} Z^*_{pq}.$$

The lemma is concluded. ■

3.3.2 Negative non-cut edges

In this subsection, we analyze the number of negative non-cut edges for each vertex $q\in C_{q_i^*}, i\in[m]$.

Lemma 5 For each vertex $q\in C_{q_i^*}, i\in[m]$, the number of negative edges $(p,q), p\in C_{q_i^*}$, is no more than

$$\sum_{\substack{(p,q)\in E^-\\ p\in D_i(q)}} \frac{7}{1-14\delta} Z^*_{pq} +$$

$$\sum_{\substack{(p,q)\in E^-\\ p\in A_i(q)}} \frac{7}{3-14\delta} Z^*_{pq}.$$

Proof From Property 1-(4) and the first constraint of Formula (2), for each $(p,q)\in E^-, p\in D_i(q)$, we have

$$\begin{aligned} Z^*_{pq} &\geqslant 1 - X^*_{pq} - 2\delta \geqslant \\ & 1 - X^*_{qq_i^*} - X^*_{pq_i^*} - 2\delta \geqslant \\ & \frac{1}{7} - 2\delta \end{aligned} \tag{4}$$

Combining Formulas (3) and (4), we have

$$\mathrm{dis}^-_q(\mathcal{C}, V\backslash O) \leqslant$$

$$\sum_{\substack{(p,q)\in E^-\\ p\in D_i(q)}} \frac{7}{1-14\delta} Z^*_{pq} +$$

$$\sum_{\substack{(p,q)\in E^-\\ p\in A_i(q)}} \frac{7}{3-14\delta} Z^*_{pq}.$$

The lemma is concluded. ■

From Lemmas 2–5, we can obtain the upper bound of disagreements generated by each vertex.

Lemma 6 For each vertex $q \in C_{q_i^*}, i \in [m]$, we have

$$\text{dis}_q(\mathcal{C}, V \backslash O) \leqslant \frac{7}{1-14\delta} \sum_{(p,q)\in E} Z^*_{pq}.$$

Proof For each vertex $q \in C_{q_i^*}, i \in [m]$, from Lemmas 2–4, we have

$$\begin{aligned}
&\text{dis}_q^+(\mathcal{C}, V \backslash O) = \\
&|\{(p,q) \in E^+ : p \in M_i(q), j \in [i-1]\}| + \\
&|\{(p,q) \in E^+ : p \in D_j(q), j \in [i-1]\}| + \\
&|\{(p,q) \in E^+ : p \in A_j(q), j \in [i-1]\}| + \\
&|\{(p,q) \in E^+ : p \in B_j(q), j \in [i-1]\}| + \\
&|\{(p,q) \in E^+ : p \in F_i(q)\}| \leqslant \\
&\frac{7}{1-14\delta} \sum_{\substack{(p,q)\in E^+ \\ p \in M_i(q)\cup D_j(q) \\ j\in[i-1]}} Z^*_{pq} + \\
&\frac{7}{1-7\delta} \sum_{\substack{(p,q)\in E^+ \\ p \in A_j(q) \\ j\in[i-1]}} Z^*_{pq} + \\
&\frac{7}{2-14\delta} \sum_{\substack{(p,q)\in E^- \\ p \in A_j(q) \\ j\in[i-1]}} Z^*_{pq} + \\
&\frac{7}{1-14\delta} \sum_{\substack{(p,q)\in E^+ \\ p \in A_i(q)}} Z^*_{pq} + \\
&\frac{7}{3-14\delta} \sum_{\substack{(p,q)\in E^- \\ p \in A_i(q)}} Z^*_{pq} \leqslant \\
&\frac{7}{1-14\delta} \sum_{(p,q)\in E^+} Z^*_{pq} + \\
&\frac{7}{2-14\delta} \sum_{\substack{(p,q)\in E^- \\ p \in A_j(q) \\ j\in[i-1]}} Z^*_{pq} + \\
&\frac{7}{3-14\delta} \sum_{\substack{(p,q)\in E^- \\ p \in A_i(q)}} Z^*_{pq}
\end{aligned} \tag{5}$$

Combining Lemma 5 and Formula (5), we can get

$$\begin{aligned}
&\text{dis}_q(\mathcal{C}, V \backslash O) = \\
&\text{dis}_q^+(\mathcal{C}, V \backslash O) + \text{dis}_q^+(\mathcal{C}, V \backslash O) \leqslant \\
&\frac{7}{1-14\delta} \sum_{(p,q)\in E^+} Z^*_{pq} + \\
&\frac{7}{2-14\delta} \sum_{\substack{(p,q)\in E^- \\ p \in A_j(q) \\ j\in[i-1]}} Z^*_{pq} + \\
&\frac{14}{3-14\delta} \sum_{\substack{(p,q)\in E^- \\ p \in A_i(q)}} Z^*_{pq} + \\
&\frac{7}{1-14\delta} \sum_{\substack{(p,q)\in E^- \\ p \in D_i(q)}} Z^*_{pq} \leqslant \\
&\frac{7}{1-14\delta} \sum_{(p,q)\in E} Z^*_{pq}.
\end{aligned}$$

The lemma is concluded. ■

From Lemmas 1 and 6, we can obtain Theorem 1.

Theorem 1 For each instance $\mathcal{I} = \{G = (V, E), r\}$ of the min-max disagreements with outliers, Algorithm 1 can return a set O with $|O| \leqslant r/\delta$ as well as a clustering of $V \backslash O$, such that

$$\max_{q\in V\backslash O} \text{dis}_q(\mathcal{C}, V\backslash O) \leqslant \frac{7}{1-14\delta} \max_{q\in V\backslash O^*} \text{dis}_q(\mathcal{C}^*, V\backslash O^*),$$

where O^* and $\mathcal{C}^*$ are the set of outliers and the clustering of non-outliers returned by the optimal algorithm, respectively.

4 Min-Max Disagreements with Outliers on One-Sided Complete Bipartite Graphs

In this section, we study the min-max disagreements with outliers on one-sided complete bipartite graphs.

Definition 4 (Min-max disagreements with outliers on one-sided complete bipartite graphs) Given an instance $\mathcal{I} = \{G = (V, E), r\}$, where G is a marked complete bipartite graphs with N vertices and r is the upper bound on the number of outliers. Let V_1 and V_2 be the partite sets of G. The purpose of this problem is to find a set O with r vertices as well as a clustering $\mathcal{C} = \{C_1, C_2, \ldots, C_m\}, m \in [N-r]$ of $V\backslash O$, such that

$$\max_{\substack{q\in V_1\cap C_i \\ i\in[m]}} \text{dis}_q(\mathcal{C}, V\backslash O)$$

is minimized.

Based on Formula (1), we can formulate the min-max disagreements with outliers on one-sided complete bipartite graphs problem as follows:

$$\begin{aligned}
&\min \quad \max_{q\in V_1}\sum_{p\in V_2} Z_{pq},\\
\text{s.t.,}\quad & X_{pq}+X_{qw}\geqslant X_{pw}, \quad \forall p,q,w\in V,\\
& Y_p+Y_q+Z_{pq}\geqslant 1-X_{pq},\ \forall (p,q)\in E^-,\\
& Y_p+Y_q+Z_{pq}\geqslant X_{pq},\ \forall (p,q)\in E^+,\\
& \sum_{q\in V} Y_q = r,\\
& X_{qq}=0, \quad \forall q\in V,\\
& X_{pq},Z_{pq},Y_p\in\{0,1\}, \quad \forall p,q\in V
\end{aligned}\tag{6}$$

By relaxing each variable to interval [0, 1], we obtain the following LP relaxation of Formula (6).

$$\begin{aligned}
&\min \quad \max_{q\in V_1}\sum_{p\in V_2} Z_{pq},\\
\text{s.t.,}\quad & X_{pq}+X_{qw}\geqslant X_{pw}, \quad \forall p,q,w\in V,\\
& Y_p+Y_q+Z_{pq}\geqslant 1-X_{pq},\ (p,q)\in E^-,\\
& Y_p+Y_q+Z_{pq}\geqslant X_{pq},\ \forall (p,q)\in E^+,\\
& \sum_{q\in V} Y_q = r,\\
& X_{qq}=0, \quad \forall q\in V,\\
& X_{pq},Z_{pq},Y_p\in[0,1], \quad \forall p,q\in V
\end{aligned}\tag{7}$$

We can get Theorem 2 directly from Theorem 1. We omit the proof.

Theorem 2 For each instance $\mathcal{I}=\{G=(V,E),r\}$ of the min-max disagreements with outliers on one-sided complete bipartite graphs, Algorithm 1 can return a set O with $|O|\leqslant r/\delta$ as well as a clustering of $V\backslash O$, such that

$$\max_{q\in V_1\backslash O}\mathrm{dis}_q(\mathcal{C},V\backslash O)\leqslant \frac{7}{1-14\delta}\max_{q\in V_1\backslash O^*}\mathrm{dis}_q(\mathcal{C}^*,V\backslash O^*),$$

where O^* and $\mathcal{C}^*$ are the set of outliers and the clustering of non-outliers returned by the optimal algorithm, respectively.

5 Min-Max Disagreements with Penalties

In this section, we study the min-max disagreements with penalties. First, we give the definition as well as two formulations for problem in Section 5.1. Then, we prove that Algorithm 1 can achieve an approximation ratio of 21 for the min-max disagreements with penalties in Section 5.2.

5.1 Definition and formulation of the min-max disagreements with penalties

Definition 5 (Min-max disagreements with penalties) Let $\mathcal{I}=\{G=(V,E),p_q,q\in V\}$, where G is a marked complete graph with N vertices, and p_q is the penalty cost of vertex q. The target is to find a penalized set P as well as a clustering $\mathcal{C}=\{C_1,C_2,\ldots,C_m\}, m\in[N-r]$ of $V\backslash P$, such that

$$\max\left\{\max_{\substack{q\in C_i\\ i\in[m]}}\mathrm{dis}_q(\mathcal{C},V\backslash P),\max_{q\in P}p_q\right\}$$

is minimized.

In order to give an integer programming for the min-max disagreements with penalties, we first introduce a binary variable W_q for each vertex $q\in V$. If variable $W_q=1$, then vertex q is penalized. If variable $W_q=0$, vertex q is clustered. Next, based on the variables X and Z defined in Section 3.1, we can formulate the min-max disagreements with penalties as follows:

$$\begin{aligned}
&\min \quad \max_{q\in V}\left(\sum_{p\in V} Z_{pq}+p_qW_q\right),\\
\text{s.t.,}\quad & X_{pq}+X_{qw}\geqslant X_{pw}, \quad \forall p,q,w\in V,\\
& W_p+W_q+Z_{pq}\geqslant 1-X_{pq},\ \forall (p,q)\in E^-,\\
& W_p+W_q+Z_{pq}\geqslant X_{pq},\ \forall (p,q)\in E^+,\\
& X_{qq}=0, \quad \forall q\in V,\\
& X_{pq},Z_{pq},W_p\in\{0,1\}, \quad \forall p,q\in V
\end{aligned}\tag{8}$$

From the second and the third constraints, we can obtain that if vertex q is penalized, then for each edge (p,q), variable $Z_{pq}=0$. Therefore, the objective function is the maximum value of the number of disagreements or penalty cost generated by each vertex q. By relaxing each variable to interval [0, 1], we get the LP relaxation of Formula (8):

$$\begin{aligned}
&\min \quad \max_{q\in V}\left(\sum_{p\in V} Z_{pq}+p_qW_q\right),\\
\text{s.t.,}\quad & X_{pq}+X_{qw}\geqslant X_{pw}, \quad \forall p,q,w\in V,\\
& Y_p+Y_q+Z_{pq}\geqslant 1-X_{pq},\ \forall (p,q)\in E^-,\\
& Y_p+Y_q+Z_{pq}\geqslant X_{pq},\ \forall (p,q)\in E^+,\\
& X_{qq}=0, \quad \forall q\in V,\\
& X_{pq},Z_{pq},Y_p\in[0,1], \quad \forall p,q\in V
\end{aligned}\tag{9}$$

5.2 Theoretical analysis of the constant ratio

In this section, we prove that Algorithm 1 is a 21-approximation algorithm for the min-max disagreements problem with penalties. For each instance $\mathcal{I}=\{G=(V,E),p_q,q\in V\}$, we can construct a feasible solution based on Algorithm 1.

(1) Let (X^*,W^*,Z^*) be the optimal fractional solution by solving Formula (9).

(2) Set parameter $\delta = 1/21$, and let $P := \{q : q \in V, W_q^* \geqslant 1/21\}$ be the penalized set.

(3) Let $\mathcal{U} := V \backslash P$ and let $\mathcal{C}$ be the clustering of $V \backslash P$ returned by running Lines 3–11 of Algorithm 1.

From Lemma 6, we achieve the following theorem.

Theorem 3 For each instance $\mathcal{I} = \{G = (V, E), p_q, q \in V\}$ of the min-max disagreements with penalties, P and $\mathcal{C}$ are the penalized set and clustering of $V \backslash P$ based on above construction, respectively. We have

$$\max\left\{\max_{q \in V\backslash P} \text{dis}_q(\mathcal{C}, V\backslash P), \max_{q \in P} p_q\right\} \leqslant$$
$$21 \max\left\{\max_{q \in V\backslash P^*} \text{dis}_q(\mathcal{C}^*, V\backslash P^*), \max_{q \in P^*} p_q\right\},$$

where P^* and $\mathcal{C}^*$ are the set of un-clustered vertices and the clustering of clustered vertices returned by the optimal algorithm, respectively.

Proof For each vertex $q \in P$, from the construction of P, we have

$$p_q \leqslant 21 p_q W_q^*,$$

which indicates that

$$\max_{q \in P} p_q \leqslant 21 \max_{q \in P} p_q W_q^* \leqslant 21 \max_{q \in V} p_q W_q^*.$$

Moreover, for each $q \in V \backslash P$, from Lemma 6 we have

$$\text{dis}_q(\mathcal{C}, V\backslash P) \leqslant 21 \sum_{p \in V} Z_{pq}^*,$$

which indicates that

$$\max_{q \in V\backslash P} \text{dis}_q(\mathcal{C}, V\backslash P) \leqslant 21 \max_{q \in V\backslash P} \sum_{p \in V} Z_{pq}^* \leqslant 21 \max_{q \in V} \sum_{p \in V} Z_{pq}^*.$$

Above all, we can obtain

$$\max\left\{\max_{q \in V\backslash P} \text{dis}_q(\mathcal{C}, V\backslash P), \max_{q \in P} p_q\right\} \leqslant$$
$$21 \max\left\{\max_{q \in V} \sum_{p \in V} Z_{pq}^*, \max_{q \in V} p_q W_q^*\right\} \leqslant$$
$$21 \max_{q \in V}\left(\sum_{p \in V} Z_{pq}^* + p_q W_q^*\right).$$

The theorem is concluded. ■

6 Conclusion

In this paper, we study two generalizations of the min-max disagreements: min-max disagreements with outliers and min-max disagreements with penalties. We design an approximation algorithm based on LP-rounding, and then prove that the algorithm is effective to solve the min-max disagreements with outliers, min-max disagreements with outliers on one-sided complete bipartite graphs, and the min-max disagreements with penalties. For the future research work of min-max disagreements, we have the following two directions:

(1) We will continue to study above two problems. We hope to design improved algorithms to reduce the approximation ratio of the existing algorithms for the above two problems.

(2) Capacitated constraint is a common constraint in combinatorial optimization problems and has been widely studied. So the second direction is to study the capacitated min-max disagreements and its generalizations.

Acknowledgment

This work was supported by the Science and Technology Project of Hebei Education Department (No. BJK2023076), the National Natural Science Foundation of China (Nos. 12101594, 12001025, and 12131003), and the Natural Science Foundation of Shandong Province of China (No. ZR2020MA029).

References

[1] J. P. Hou, A. Emad, G. J. Puleo, J. Ma, and O. Milenkovic, A new correlation clustering method for cancer mutation analysis, *Bioinformatics*, vol. 32, no. 24, pp. 3717–3728, 2016.

[2] J. Jafarov, S. Kalhan, K. Makarychev, and Y. Makarychev, Correlation clustering with asymmetric classification errors, in *Proc. 37th Int. Conf. Machine Learning*, Virtual, 2020, pp. 4641–4650.

[3] S. Kim, C. D. Yoo, S. Nowozin, and P. Kohli, Image segmentation using higher-order correlation clustering, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 9, pp. 1761–1774, 2014.

[4] P. Li, G. J. Puleo, and O. Milenkovic, Motif and hypergraph correlation clustering, *IEEE Trans. Inf. Theory*, vol. 66, no. 5, pp. 3065–3078, 2019.

[5] A. Ukkonen, Crowdsourced correlation clustering with relative distance comparisons, in *Proc. 2017 IEEE Int. Conf. Data Mining* (*ICDM*), New Orleans, LA, USA, 2017, pp. 1117–1122.

[6] N. Veldt, D. F. Gleich, and A. Wirth, A correlation clustering framework for community detection, in *Proc. 2018 World Wide Web Conf.*, Lyon, France, 2018, pp. 439–448.

[7] N. Bansal, A. Blum, and S. Chawla, Correlation clustering, *Mach. Learn.*, vol. 56, pp. 89–113, 2004.

[8] N. Ailon, N. Avigdor-Elgrabli, E. Liberty, and A. V. Zuylen, Improved approximation algorithms for bipartite correlation clustering, *SIAM J. Comput.*, vol. 41, no. 5, pp. 1110–1121, 2012.

[9] N. Ailon, M. Charikar, and A. Newman, Aggregating inconsistent information, *J. ACM*, vol. 55, no. 5, pp. 1–27, 2008.

[10] M. Bressan, N. Cesa-Bianchi, A. Paudice, and F. Vitale, Correlation clustering with adaptive similarity queries, in *Proc. 33rd Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2019, pp. 12531–12540.

[11] M. Charikar, V. Guruswami, and A. Wirth, Clustering with qualitative information, *J. Comput. Syst. Sci.*, vol. 71, no. 3, pp. 360–383, 2005.

[12] J. H. Lange, A. Karrenbauer, and B. Andres, Partial optimality and fast lower bounds for weighted correlation clustering, in *Proc.* 35th *Int. Conf. Machine Learning* (*ICML*), Stockholm, Sweden, 2018, pp. 2892–2901.

[13] E. Thiel, M. H. Chehreghani, and D. Dubhashi, A non-convex optimization approach to correlation clustering, *Proc. AAAI Conf. Artif. Intell.*, vol. 33, no. 1, pp. 5159–5166, 2019.

[14] S. Chawla, K. Makarychev, T. Schramm, and G. Yaroslavtsev, Near optimal LP rounding algorithm for correlation clustering on complete and complete k-partite graphs, in *Proc. Forty-Seventh Annual ACM Symp. on Theory of Computing*, Portland, OR, USA, 2015, pp. 219–228.

[15] G. J. Puleo and O. Milenkovic, Correlation clustering and biclustering with locally bounded errors, *IEEE Trans. Inf. Theory*, vol. 64, no. 6, pp. 4105–4119, 2018.

[16] M. Charikar, N. Gupta, and R. Schwartz, Local guarantees in graph cuts and clustering, in *Proc. 19th Int. Conf. Integer Programming and Combinatorial Optimization*, Waterloo, Canada, 2017, pp. 136–147.

[17] Devvrit, R. Krishnaswamy, and N. Rajaraman, Robust correlation clustering, in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques* (*APPROX/RANDOM 2019*), D. Achlioptas and L. A. Végh, eds. Dagstuhl, Germany: Schloss Dagstuhl, 2019, pp. 33:1–33:18.

[18] M. H. Chehreghani, Hierarchical correlation clustering and tree preserving embedding, arXiv preprint arXiv: 2002.07756, 2020.

[19] D. Vainstein, V. Chatziafratis, G. Citovsky, A. Rajagopalan, M. Mahdian, and Y. Azar, Hierarchical clustering via sketches and hierarchical correlation clustering, arXiv preprint arXiv: 2101.10639, 2021.

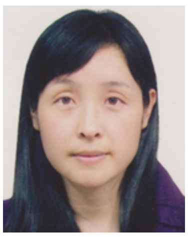
[20] K. Makarychev, Y. Makarychev, and A. Vijayaraghavan, Correlation clustering with noisy partial information, in *Proc.* 28th *Conf. Learn. Theory* (*COLT*), Paris, France, 2015, pp. 1321–1342.

[21] C. Mathieu and W. Schudy, Correlation clustering with noisy input, in *Proc. 2010 Annu. ACM-SIAM Symp. on Discrete Algorithms* (*SODA*), Austin, TX, USA, 2010, pp. 712–728.

[22] I. Giotis and V. Guruswami, Correlation clustering with a fixed number of clusters, in *Proc. Seventeenth Annu. ACM-SIAM Symp. on Discrete Algorithm* (*SODA*), Miami, FL, USA, 2006, pp. 1167–1176.

[23] A. Aboud and Y. Rabani, Correlation clustering with penalties and approximating the reordering buffer management problem, Doctor dissertation, Computer Science Department, Technion, Haifa, Israel, 2008.

Sai Ji received the PhD degree from Beijing University of Technology, China in 2020. Currently she is a lecturer at the Institute of Mathematics, Hebei University of Technology, China. Her research interests mainly include combinatorial optimization, approximation algorithm, and machine learning.



Min Li received the PhD degree from University of Coimbra, Portugal in 2013. Currently she is an associate professor at the School of Mathematics and Statistics, Shandong Normal University, China. Her research interests mainly include combinatorial optimization, approximation algorithm, submodular optimization, and integer programming.

Mei Liang received the BS degree from Guangxi Normal University, China in 2021. Currently she is pursuing the MS degree at the College of Statistics and Data Science, Beijing University of Technology, China. Her research interests mainly include robust optimization, combinatorial optimization, and approximation algorithm.

Zhenning Zhang received the PhD degree from Beijing Institute of Technology, China in 2010. Currently she is an associate professor at Beijing University of Technology, China. Her research interests mainly include submodular maximization, combinatorial optimization, and approximation algorithm.