Approximation Algorithms for the Lower Bounded Correlation Clustering Problem

Sai Ji[1], Yinhong Dong[2], Donglei Du[3], and Dachuan Xu[4(✉)]

[1] Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190, People's Republic of China
jisai@amss.ac.cn

[2] School of Management, South-Central University for Nationalities, Wuhan 430074, Hubei, People's Republic of China
3098610@mail.scuec.edu.cn

[3] Faculty of Business Administration, University of New Brunswick, Fredericton, NB E3B 5A3, Canada
ddu@unb.ca

[4] Beijing Institute for Scientific and Engineering Computing, Beijing University of Technology, Beijing 100124, People's Republic of China
xudc@bjut.edu.cn

Abstract. Lower bounded correlation clustering problem is a generalization of the classical correlation clustering problem, which has many applications in protein interaction networks, cross-lingual link detection, and communication networks, etc. In the lower bounded correlation clustering problem, we are given a complete graph and an integer L. Each edge is labelled either by a positive sign $+$ or a negative sign $-$ whenever the two endpoints of the edge are similar or dissimilar respectively. The goal of this problem is to partition the vertex set into several clusters, subject to an lower bound L on the sizes of clusters so as to minimize the number of disagreements, which is the total number of the edges with positive labels between clusters and the edges with negative labels within clusters. In this paper, we propose the lower bounded correlation clustering problem and formulate the problem as an integer program. Furthermore, we provide two polynomial time algorithms with constant approximate ratios for the lower bounded correlation clustering problem on some special graphs.

Keywords: Lower bounded · Correlation clustering · Approximation algorithm · Polynomial time

1 Introduction

Correlation clustering problem has numerous applications in the areas of machine learning, computer vision, data mining, social networks and data compression. It has been widely studied in the literature [1,11,13,18–20].

The correlation clustering problem was first introduced by Bansal et al. [4]. In this problem, we are given a complete graph $G = (V, E)$, where each edge

D. Mohaisen and R. Jin (Eds.): CSoNet 2021, LNCS 13116, pp. 39–49, 2021.
https://doi.org/10.1007/978-3-030-91434-9_4

$(u, v) \in E$ is labelled by $+$ or $-$ based on the similarity of vertices u and v. The goal is to find a clustering of vertices V so as to make the edges within clusters are mostly positive and the edges between different clusters are mostly negative. Given a clustering, let each positive edge whose two endpoints lie in different clusters and each negative edge whose endpoints lie in the same cluster be a disagreement. Moreover, let each remaining edge be an agreement.

Based on the goal of the correlation clustering problem, there are two versions of the correlation clustering problem: minimizing disagreements and maximizing agreements. The goal of the former problem is to find a clustering so as to minimize the number of disagreements. The goal of the latter problem is to find a clustering so as to maximize the number of agreements. Given any instance of the correlation clustering problem, the two different versions share the same optimal solution. But the two versions of the problem are essentially different from the point of view of approximation algorithm. In the rest of the paper, we only consider the minimizing disagreements version of correlation clustering problem.

The correlation clustering problem is NP-hard. Bansal et al. [4] give a 17433-approximation algorithm, which is the first constant approximation algorithm for the correlation clustering problem. Charikard et al. [7] first provide a very natural linear programming formulation of the problem and prove that the integrality gap of the linear program is 2. Secondly, they propose a 4-approximation algorithm by using the method of region growth, which significantly improves the approximation ratio of the algorithm provided by Bansal et al. [4]. Finally, for the correlation clustering problem on general graphs, they provided an $O(\log n)$-approximation algorithm. The current best approximation algorithm is provided by Chawla et al. [8], which achieves an approximate ratio of 2.06.

Because of the complexity of the practical applications, the correlation clustering problem has some limitations in modelling real-life situations. In order to adapt to the development of society and guide practice more effectively and realistically, various generalizations of the correlation clustering problem have been proposed and widely studied. Such as the min-max correlation clustering [2], the chromatic correlation clustering [5], the overlapping correlation clustering problem [6], the higher-order correlation clustering [9], the capacitated correlation clustering problem [16], and the correlation clustering with noisy input [14,15], among others.

Lower bound constraint is a natural constraint in combinatorial optimization problems and it has been extensively studied [3,10,12,17]. However, there has been no relevant research on the lower bounded correlation clustering problem. Therefore, this work considers the lower bounded correlation clustering problem, which is a new and natural variant of the correlation clustering problem. In this problem, we are given a labeled complete graph $G = (V, E)$ as well as an integer L. The goal of this problem is to partition set V into several clusters with each size at least L so as to minimize the total number of disagreements.

Note that the lower bounded correlation clustering problem includes the classical correlation clustering problem as a special case by letting $L = 1$. Note the case where $L > |V|/2$ is trivial because only one feasible solution exists for the lower bounded correlation clustering problem. Therefore, we assume that $L \leq |V|/2$ for the rest of the discussion.

There are three main contributions in this paper.

(1) We first propose the lower bounded correlation clustering problem and give an integer programming formulation for the problem.
(2) We provide an algorithm which returns V as the cluster. We show that the algorithm always outputs an optimal solution for the lower bounded correlation clustering problem on $(2|V|/L-1)$-positive edge dominant graphs (Theorem 1). Moreover, we prove that the same algorithm is a 20-approximation algorithm for the lower bounded correlation clustering problem on 4-positive edge dominant graphs (Theorem 2).
(3) We present another algorithm which may return multiple clusters and prove the algorithm is a 20-approximation algorithm for the lower bounded correlation clustering problem on $(5|V|/L-1)$-positive edge dominant graphs (Theorem 3).

The rest of our paper is structured as follows. Section 2 presents some definitions as well as the formulation of the lower bounded correlation clustering problem. The two algorithms are presented in Sects. 3 and 4, respectively. Some discussions are provided in Sect. 5.

2 Lower Bounded Correlation Clustering Problem

In this section, we give some definitions used in this paper as well as the formulation of the lower bounded correlation clustering problem. Given a complete graph $G = (V, E)$, let E^+ and E^- be the sets of all positive edges and all negative edges, respectively, For each positive integer k, denote set $[k] = \{1, 2, \ldots, k\}$. The lower bounded correlation clustering is defined in Definition 1.

Definition 1 (Lower bounded correlation clustering problem). *Given a labelled complete graph $G = (V, E)$ as well as an integer L, the goal is to find a partition $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$ of V which satisfies $|C_i| \geq L, i \in [k]$ such that*

$$\sum_{i\in[k]} |(u,v) \in E^- : u, v \in C_i| + \sum_{i,j\in[k]} |(u,v) \in E^+ : u \in C_i, v \in C_j, i \neq j|$$

is minimized.

Definition 2 (M-positive edge dominant graph). *Graph $G = (V, E)$ is an M-positive edge dominant graph if*

$$\inf_{v\in V} \frac{|E_v^+|}{|E_v^-|} \geq M,$$

where $E_v^+ := \{(u,v) \in E^+, u \in V\}$ and $E_v^- := \{(u,v) \in E^-, u \in V\}$.

For each edge (u,v), we introduce a 0-1 variable x_{uv} such that $x_{uv}=1$ when the two vertices of edge (u,v) lie in different clusters and $x_{uv}=0$ when the two vertices of edge (u,v) lie in the same cluster. Based on the above variables and Definition 1, we formulate the lower bounded correlation clustering problem as the following integer program (1).

$$\begin{aligned}
\min \quad & \sum_{(u,v)\in E^+} x_{uv} + \sum_{(u,v)\in E^-} (1-x_{uv}) \\
\text{s. t.} \quad & x_{uv} + x_{vw} \geq x_{uw}, && \forall u,v,w \in V, \\
& \sum_{v\in V}(1-x_{uv}) \geq L, && \forall u \in V, \\
& x_{uu} = 0, && \forall u \in V, \\
& x_{uv} \in \{0,1\}, && \forall u,v \in V.
\end{aligned} \tag{1}$$

The objective function is the total number of disagreements. The quantity $\sum_{(u,v)\in E^+} x_{uv}$ is the number of disagreements generated by the positive edges, while the quantity $\sum_{(u,v)\in E^-}(1-x_{uv})$ is the number of disagreements generated by the negative edges. There are three types of constraints in (1). The first one is the triangle inequality, which insures the program returns a feasible cluster of the correlation clustering problem. The second one is a lower bound constraint, which guarantees that there are at least L vertices in each cluster. The third one is the natural binary constraint. By relaxing above 0-1 variables, we obtain the LP relaxation of (1).

$$\begin{aligned}
\min \quad & \sum_{(u,v)\in E^+} x_{uv} + \sum_{(u,v)\in E^-} (1-x_{uv}) \\
\text{s. t.} \quad & x_{uv} + x_{vw} \geq x_{uw}, && \forall u,v,w \in V, \\
& \sum_{v\in V}(1-x_{uv}) \geq L, && \forall u \in V, \\
& x_{uu} = 0, && \forall u \in V, \\
& 0 \leq x_{uv} \leq 1, && \forall u,v \in V.
\end{aligned} \tag{2}$$

3 A Simple Effecient Algorithm

In this section, we provide Algorithm 1, which returns set V as the only cluster. Furthermore, we prove that Algorithm 1 achieves a constant approximation ration for the lower bounded correlation clustering problem when restricted to two special graphs.

Theorem 1 below can be shown based on the structure of the feasible solutions of the lower bounded correlation clustering problem.

Theorem 1. *Algorithm 1 is an optimal algorithm for the lower bounded correlation clustering problem on* $(2|V|/L-1)$*-positive edge dominant graphs.*

Algorithm 1

Input: Integer L, a labelled complete graph $G = (V, E)$.
Output: A partition of vertices.
1: Let V be a cluster.
2: **return** cluster V.

Proof. For each $(2|V|/L - 1)$-positive edge dominant graph $G = (V, E)$, we have $|E_v^-| \leq L/2, \forall v \in V$. Moreover, for each feasible solution $\mathcal{C} = \{C_1, \ldots, C_m\}$ of the instance $I = (G, L)$ which contains more than one cluster. There are at least L cut edges generated by each vertex v. Moreover, there are at least $L/2$ disagreements generated by the positive edges among these cut edges since $|E_v^-| \leq L/2$. Therefore, Algorithm 1 returns an optimal solution since the disagreements generated by each vertex $v \in V$ is no more than $|E_v^-|$ and it can be bounded by $L/2$. □

For any instance $I = (G, L)$, where $G = (V, E)$ is a complete 4-positive edge dominant graph satisfying $|E_v^-| \leq (|V| - 1)/5, v \in V$, we solve (2) to obtain an optimal fractional solution x^* of I. For each vertex $v \in V$, compute

$$\text{Avg}_v(V) := \frac{\sum_{t \in V} x^*_{vt}}{|V|}.$$

Let

$$cen(V) := \arg\min_{v \in V} \text{Avg}_v(V).$$

be the center vertex of set V. Then we can analyze the upper bound on the number of disagreements based on the value of $\text{Avg}_{cen(V)}(V)$. Specifically, we consider the following two cases:

(1) $\text{Avg}_{cen(V)}(V) \leq 17/80$;
(2) $\text{Avg}_{cen(V)}(V) > 17/80$.

3.1 $\text{Avg}_{cen(V)}(V) \leq 17/80$

Lemma 1. *For each negative edge (u, w) with $x^*_{ucen(V)}, x^*_{wcen(V)} \leq 19/40$. The number of disagreement generated by the negative edge (u, w) is bounded by*

$$20\left(1 - x^*_{uw}\right).$$

Proof. From the first constraint of (2) and the inequalities $x^*_{ucen(V)}, x^*_{wcen(V)} \leq 19/40$, we have

$$1 - x^*_{uw} \geq 1 - x^*_{ucen(V)} - x^*_{wcen(V)} \geq 1 - \frac{38}{40} = \frac{1}{20}.$$

The number of disagreement generated by edge $(u, w) = 1 \leq 20(1 - x^*_{uw})$.

We conclude the lemma. □

Lemma 2. *For each vertex $w \in V$ with $x^*_{wcen(V)} > 19/40$, the number of disagreements generated by the negative edges (u, w) with $x^*_{ucen(V)} \leq x^*_{wcen(V)}$ can be bounded by*

$$20 \left[\sum_{(u,w)\in E^+, x^*_{ucen(V)} \leq x^*_{wcen(V)}} x^*_{uw} + \sum_{(u,w)\in E^-, x^*_{ucen(V)} \leq x^*_{wcen(V)}} (1 - x^*_{uw}) \right].$$

Proof. For each vertex $w \in V$, denote

$$P_w(V) := \left\{ u \in V : (u, w) \in E^+, x^*_{ucen(V)} \leq x^*_{wcen(V)} \right\},$$
$$N_w(V) := \left\{ u \in V : (u, w) \in E^-, x^*_{ucen(V)} \leq x^*_{wcen(V)} \right\}.$$

Recall $\text{Avg}_{cen(V)}(V) \leq 17/80$ and $x^*_{wcen(V)} > 19/40$. We obtain that there are at least $|V|/2$ vertices in $P_w(V) \cup N_w(V)$. Moreover we have $|P_w(V)| \geq |N_w(V)|$ since

$$|N_w(V)| \leq |E^-_w| \leq \frac{|V| - 1}{5}.$$

Then, we get

$$\begin{aligned}
&\sum_{u\in P_w(V)} x^*_{uw} + \sum_{u\in N_w(V)} (1 - x^*_{uw}) \\
\geq &\sum_{u\in P_w(V)} \left(x^*_{wcen(V)} - x^*_{ucen(V)} \right) + \sum_{u\in N_w(V)} \left(1 - x^*_{wcen(V)} - x^*_{ucen(V)} \right) \\
= &\, x^*_{wcen(V)} |P_w(V)| + \left(1 - x^*_{wcen(V)} \right) |N_w(V)| - \sum_{u\in P_w(V)\cup N_w(V)} x^*_{ucen(V)} \\
\geq &\frac{19}{40} |P_w(V)| - \frac{17}{80} (|P_w(V)| + |N_w(V)|) \\
\geq &\frac{1}{20} |P_w(V)|.
\end{aligned}$$

Therefore, the number of disagreements generated by the negative edges (u, w) with $x^*_{ucen(V)} \leq x^*_{wcen(V)}$ is bounded by

$$20 \left[\sum_{u\in P_w(V)} x^*_{uw} + \sum_{u\in N_w(V)} (1 - x^*_{uw}) \right].$$

We conclude the lemma. □

3.2 $\text{Avg}_{cen(V)}(V) > 17/80$

Lemma 3. *For each vertex $w \in V$, the number of disagreements generated by the negative edges (u, w) is bounded by*

$$20 \sum_{(u,w)\in E^+, u\in V} x^*_{uw}.$$

Proof. From the definition of $cen(V)$, we obtain that if $\mathrm{Avg}_{cen(V)}(V) > 17/80$, then for each vertex $w \in V$

$$\frac{\sum_{u \in V} x^*_{uw}}{|V|} > \frac{17}{80} \tag{3}$$

holds. Moreover, for each vertex $w \in V$, we have

$$|E^-_w| \leq \frac{(|V|-1)}{5} \text{ and } |E^+_w| \geq \frac{4(|V|-1)}{5}. \tag{4}$$

Combining (3) and (4), we obtain that for each vertex $w \in V$,

$$\frac{\sum_{u \in E^+_w} x^*_{uw}}{|E^+_w|} \geq \frac{17}{80} - \frac{1}{5} = \frac{1}{80}$$

holds. Therefore, for each vertex $w \in V$, the number of disagreements generated by the negative edges (u, w) is no more than $|E^+_w|/4$ and it is bounded by

$$20 \sum_{u \in E^+_w} x^*_{uw}.$$

$\square$.

Combining Lemma 1–3, we obtain Theorem 2.

Theorem 2. *Algorithm 1 is a* 20*-approximation algorithm for the lower bounded correlation clustering problem on* 4*-positive edge dominant graphs.*

4 A Complex Algorithm May Outputs Multiple Clusters

In Sect. 3, we give a algorithm which only return one cluster. However, in some applications, we may need to output more than one clusters. Therefore, we provide Algorithm 2 for some special graphs in this section which may output multiple clusters. The detailed algorithm is shown in Algorithm 2.

We assume without loss of generality that the solution returned by Algorithm 2 contains exactly k clusters. The center set of vertices of the solution is $C := \{v_1, v_2, \ldots, v_k\}$. The corresponding clusters are $C_{v_1}, C_{v_2}, \ldots, C_{v_k}$. The number of the disagreements generated by partition $\mathcal{C}$ is

$$\sum_{i \in [k-1]} |(u, v) \in E^+ : u \in C_{v_i}, v \in \cup_{t \in [k] \setminus [i]} C_{v_t}| + \sum_{i \in [k]} |(u, v) \in E^- : u, v \in C_{v_i}|.$$

The first part is the number of disagreements generated by the positive edges and the upper bound on the number of these disagreements is analyzed in Subsect. 4.1. The second part is the number of disagreements generated by the negative edges and the upper bound on the number of these disagreements is analyzed in Subsect. 4.2.

Algorithm 2

Input: Integer L, and a labelled complete graph $G=(V,E)$ with $|E_v^-|\leq L/5, v\in V$.
Output: A partition of vertices.

```
Solve (2) to obtain an optimal fractional solution x*.
Initialize the un-cluster set S := V, and the center set of vertices C := ∅.
while S ≠ ∅ do
    for each vertex v ∈ S do
        Order the vertices in S in nondecreasing value of x* from v. Let set T_v^1 be the
        first L vertices in S according to above order and T_v^2 := {t ∈ S : x*_vt ≤ 1/2}.
        Set T_v = T_v^1 ∪ T_v^2 and compute
            Avg_v(T_v) := (Σ_{t∈T_v} x*_vt) / |T_v|.
    end for
    Choose vertex v with minimum Avg_v(T_v).
    if |S\T_v| ≥ L and Avg_v(T_v) ≤ 17/80 then
        Let C_v := T_v be a cluster.
        Update S := S\C_v and C := C ∪ {v}.
    else
        Select
            v := arg min_{s∈S} (Σ_{t∈S} x*_st) / |S|.
        Let C_v := S be a cluster.
        Update S := ∅ and C := C ∪ {v}.
    end if
end while
return Set C and 𝒞 = {C_v : v ∈ C}.
```

4.1 Disagreements Generated by Positive Edges

Recall Algorithm 2. For each $i\in[k-1]$, we have $\mathrm{Avg}_{v_i}(C_{v_i})\leq 17/80$. We analyze the upper bound on the number of disagreements generated by the positive edges in the following lemma.

Lemma 4. *For each $i\in[k-1]$ and vertex $v\in V\setminus\cup_{t\in[k]\setminus[i]}C_{v_t}$, the number of disagreements generated by the positive edges $(q,v), q\in C_{v_i}$ is bounded by*

$$\frac{64}{17}\left[\sum_{(q,v)\in E^+,q\in C_{v_i}} x^*_{qv}+\sum_{(q,v)\in E^-,q\in C_{v_i}}\left(1-x^*_{qv}\right)\right].$$

Proof. For each $v\in V\setminus\cup_{t\in[k]\setminus[i]}C_{v_t}$, denote

$$E_v^+(C_{v_i}):=\left\{(q,v)\in E^+ : q\in C_{v_i}\right\},$$
$$E_v^-(C_{v_i}):=\left\{(q,v)\in E^- : q\in C_{v_i}\right\}.$$

From $|E_v^-|\leq L/5$ and $|E_v^-(C_{v_i})|+|E_v^+(C_{v_i})|\geq L$, we have

$$|E_v^-(C_{v_i})|\leq|E_v^+(C_{v_i})|/4. \tag{5}$$

Combining (5) and Step 5 of Algorithm 2, we have

$$\begin{aligned}
&\sum_{q \in E_v^+(C_{v_i})} x_{qv}^* + \sum_{q \in E_v^-(C_{v_i})} \left(1 - x_{qv}^*\right) \\
&\geq \sum_{q \in E_v^+(C_{v_i})} \left(x_{vv_i}^* - x_{qv_i}^*\right) + \sum_{q \in E_v^-(C_{v_i})} \left(1 - x_{vv_i}^* - x_{qv_i}^*\right) \\
&= x_{vv_i}^* |E_v^+(C_{v_i})| + \left(1 - x_{vv_i}^*\right) |E_v^-(C_{v_i})| - \sum_{q \in C_{v_i}} x_{qv_i}^* \\
&\geq \frac{1}{2}|E_v^+(C_{v_i})| - \frac{17}{80}(|E_v^+(C_{v_i})| + |E_v^-(C_{v_i})|) \\
&\geq \frac{17}{64}|E_v^+(C_{v_i})|.
\end{aligned}$$

Therefore, the number of disagreements generated by the positive edges $(q, v), q \in C_{v_i}$ equals $|E_v^+(C_{v_i})|$ and it is bounded by

$$\frac{64}{17}\left[\sum_{(q,v) \in E^+, q \in C_{v_i}} x_{qv}^* + \sum_{(q,v) \in E^-, q \in C_{v_i}} \left(1 - x_{qv}^*\right)\right].$$

We conclude the lemma. □

4.2 Disagreements Generated by Negative Edges

In this subsection, we consider the disagreements generated by the negative edges. Similar to Lemmas 2 and 3, we obtain the following two lemmas.

Lemma 5. *For each cluster C_{v_i} with $Avg_{v_i}(C_{v_i}) \leq 17/80$ and vertex $w \in C_{v_i}$, if $x_{wv_i}^* > 19/40$, then the number of disagreements generated by the negative edges (u, w) with $x_{uv_i}^* \leq x_{wv_i}^*$ is bounded by*

$$20\left[\sum_{u \in C_{v_i}:(u,w) \in E^+, x_{uv_i}^* \leq x_{wv_i}^*} x_{uw}^* + \sum_{u \in C_{v_i}:(u,w) \in E^+, x_{uv_i}^* \leq x_{wv_i}^*} (1 - x_{uw}^*)\right].$$

Lemma 6. *If $Avg_{v_k}(C_{v_k}) > 17/80$, then for each vertex $w \in C_{v_k}$, the number of disagreements generated by the negative edges $(u, w), u \in C_{v_k}$ is bounded by*

$$20 \sum_{(u,w) \in E^+, u \in C_{v_k}} x_{uw}^*.$$

Combining Lemmas 1, 4–Lemma 6, we obtain Theorem 3.

Theorem 3. *Algorithm 2 is a 20-approximation algorithm for the lower bounded correlation clustering problem on $(5|V|/L - 1)$-positive edge dominant graphs.*

5 Discussions

In this paper, we first study the lower bounded correlation clustering problem and give an integer programming formulation for the problem. We provide two polynomial time approximation algorithms and prove that the algorithms in this paper achieve constant approximate ratios for the lower bounded correlation clustering problem on some special graphs. About the lower bounded correlation clustering problem, we propose the following future research questions:

– In this paper, we prove that our algorithms achieve constant ratio for the correlation clustering problem on some special graphs. It will be interesting to design a polynomial time constant approximation algorithm for the lower bounded correlation clustering problem on general complete graphs.
– In this paper, we study the minimizing disagreements version of the lower bounded correlation clustering problem. There is no relevant research on the maximizing agreements version of the lower bounded correlation clustering problem. Therefore, another interesting future work is to study the maximizing agreements version of the lower bounded correlation clustering problem.

Acknowledgement. The first author is supported by National Natural Science Foundation of China (No. 12101594) and the Project funded by China Postdoctoral Science Foundation (No. 2021M693337). The second author is supported by National Social Science Foundation (NO. 20BGL259). The third author is supported by the NSERC (06446) and National Natural Science Foundation of China (Nos. 11771386, 11971349, 71771117). The fourth author is supported by National Natural Science Foundation of China (No. 12131003) and Beijing Natural Science Foundation Project No. Z200002.

References

1. Ailon, N., Avigdor-Elgrabli, N., Liberty, E., Zuylen, A.V.: Improved approximation algorithms for bipartite correlation clustering. SIAM J. Comput. **41**(5), 1110–1121 (2012)
2. Ahmadi, S., Khuller, S., Saha, B.: Min-max correlation clustering via MultiCut. In: Lodi, A., Nagarajan, V. (eds.) IPCO 2019. LNCS, vol. 11480, pp. 13–26. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-17953-3_2
3. Ahmadian, S., Swamy, C.: Improved approximation guarantees for lower-bounded facility location. In: Erlebach, T., Persiano, G. (eds.) WAOA 2012. LNCS, vol. 7846, pp. 257–271. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38016-7_21
4. Bansal, N., Blum, A., Chawla, S.: Correlation clustering. Mach. Learn. **56**(1–3), 89–113 (2004)
5. Bonchi, F., Gionis, A., Gullo, F., Tsourakakis, C.E., Ukkonen, A.: Chromatic correlation clustering. ACM Trans. Knowl. Discov. Data **9**(4), 1–24 (2015)
6. Bonchi, F., Gionis, A., Ukkonen, A.: Overlapping correlation clustering. Knowl. Inf. Syst. **35**(1), 1–32 (2013)
7. Charikar, M., Guruswami, V., Wirth, A.: Clustering with qualitative information. J. Comput. Syst. Sci. **71**(3), 360–383 (2005)

8. Chawla, S., Makarychev, K., Schramm, T., Yaroslavtsev, G.: Near optimal LP rounding algorithm for correlation clustering on complete and complete k-partite graphs. In: Proceedings of the 47th ACM Symposium on Theory of Computing, pp. 219–228 (2015)
9. Fukunaga, T.: LP-based pivoting algorithm for higher-order correlation clustering. J. Comb. Optim. **37**(4), 1312–1326 (2018). https://doi.org/10.1007/s10878-018-0354-y
10. Han, L., Hao, C., Wu, C., Zhang, Z.: Approximation algorithms for the lower-bounded k-median and its generalizations. In: Kim, D., Uma, R.N., Cai, Z., Lee, D.H. (eds.) COCOON 2020. LNCS, vol. 12273, pp. 627–639. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-58150-3_51
11. Jafarov, J., Kalhan, S., Makarychev, K., Makarychev, Y.: Correlation clustering with asymmetric classification errors. In: Proceedings of the 37th International Conference on Machine Learning, pp. 4641–4650 (2020)
12. Li, S.: On facility location with general lower bounds. In: Proceedings of the 30th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 2279–2290 (2019)
13. Li, P., Puleo, G.J., Milenkovic, O.: Motif and hypergraph correlation clustering. IEEE Trans. Inf. Theory **66**(5), 3065–3078 (2019)
14. Makarychev, K., Makarychev, Y., Vijayaraghavan, A.: Correlation clustering with noisy partial information. In: Proceedings of the 28th Annual Conference Computational Learning Theory, pp. 1321–1342 (2015)
15. Mathieu, C., Schudy, W.: Correlation clustering with noisy input. In: Proceedings of the 21th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 712–728 (2010)
16. Puleo, G.J., Milenkovic, O.: Correlation clustering with constrained cluster sizes and extended weights bounds. SIAM J. Optim. **25**(3), 1857–1872 (2015)
17. Svitkina, Z.: Lower-bounded facility location. ACM Trans. Algorithms **6**(4), 1–16 (2010)
18. Swamy, C.: Correlation clustering: maximizing agreements via semidefinite programming. In: Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 526–527 (2004)
19. Saha, B., Subramanian, S.: Correlation clustering with same-cluster queries bounded by optimal cost. In: Proceedings of the 27th Annual European Symposium on Algorithms, pp. 81:1–81:17 (2019)
20. Veldt, N., Gleich, D.F., Wirth, A.: A correlation clustering framework for community detection. In: Proceedings of the 27th World Wide Web Conference, pp. 439–448 (2018)