SCIENCE CHINA
Information Sciences

• **Supplementary File** •

An Approximation Algorithm for k-Median with Priorities

Zhen ZHANG[1], Qilong FENG[1*], Jinhui XU[2] & Jianxin WANG[1]

[1]*School of Computer Science and Engineering, Central South University, Changsha* 410083*, P.R. China;*
[2]*Department of Computer Science and Engineering, State University of New York at Buffalo, Buffalo, USA*

Appendix A Proof of Lemma 1

Lemma 1. Given an instance $\mathcal{I} = (\mathcal{D}, \mathcal{F}, \mathcal{P}, k, g, f)$ of k-MP and a real number $\epsilon > 0$, we can find in polynomial time either a 3-approximation solution to $\mathcal{I}$, or two integer solutions $\mathcal{G}_1 = (x^1, y^1)$ and $\mathcal{G}_2 = (x^2, y^2)$ to LP2$^\gamma$ such that $\sum_{i\in\mathcal{F},p\in\mathcal{P}} y^1_{ip} = k_1 < k$, $\sum_{i\in\mathcal{F},p\in\mathcal{P}} y^2_{ip} = k_2 > k$, and $aV(\mathcal{G}_1) + bV(\mathcal{G}_2) \leqslant \big(3 + O(\epsilon)\big)\mathrm{OPT}$, where $a = \frac{k_2-k}{k_2-k_1}$ and $b = 1 - a$.

Proof. Denote by $c_{\max} = \max_{i,j\in\mathcal{D}\cup\mathcal{F}} c(i,j)$ and $c_{\min} = \min_{i,j\in\mathcal{D}\cup\mathcal{F}} c(i,j)$ the maximum and minimum distances between any $i, j \in \mathcal{D}\cup\mathcal{F}$, respectively. Similarly, let $f_{\max} = \max_{p\in\mathcal{P}} f(p)$ and $f_{\min} = \min_{p\in\mathcal{P}} f(p)$. Let OPT($\gamma$) denote the cost of an optimal solution to LP2$^\gamma$. The dual of LP2$^\gamma$ is

$$\max \quad \sum_{j\in\mathcal{D}} \alpha_j - \gamma k \qquad \mathrm{DUAL}^\gamma$$

$$\text{s.t.} \quad \alpha_j \leqslant c(j,i) + \beta_{ij} \qquad \forall i \in \mathcal{F}, j \in \mathcal{D} \tag{A1}$$

$$\sum_{j\in\mathcal{D}\,\cap\, g(j)\leqslant p} \beta_{ij} \leqslant f(p) + \gamma \qquad \forall i \in \mathcal{F}, p \in \mathcal{P} \tag{A2}$$

$$\alpha_j, \beta_{ij} \geqslant 0 \qquad \forall i \in \mathcal{F}, j \in \mathcal{D} \tag{A3}$$

Observe that LP2$^\gamma$ is an LP relaxation of a facility location with priorities problem, except that its objective function contains an additional constant term of $-\gamma k$. Using the primal-dual algorithm given in [1], we can get integer solutions to LP2$^\gamma$ that have the following guarantee.

Claim 1 ([1], Theorem 3.12)**.** Given an instance $\mathcal{I} = (\mathcal{D}, \mathcal{F}, \mathcal{P}, k, g, f)$ of k-MP and a real number $\gamma \geqslant 0$, there is a polynomial-time algorithm that yields an integer solution $\mathcal{G} = (x, y)$ to LP2$^\gamma$ and its corresponding dual solution (α, β), such that $3\gamma\sum_{i\in\mathcal{F},p\in\mathcal{P}} y_{ip} + V(\mathcal{G}) \leqslant 3\sum_{j\in\mathcal{D}} \alpha_j$.

Consider an integer solution $\mathcal{G} = (x, y)$ to LP2$^\gamma$ and the corresponding dual solution (α, β) given by Claim 1. We have

$$V(\mathcal{G}) \leqslant 3(\sum_{j\in\mathcal{D}} \alpha_j - \gamma k + \gamma k - \gamma \sum_{i\in\mathcal{F},p\in\mathcal{P}} y_{ip}) \leqslant 3\big(\mathrm{OPT}(\gamma) + \gamma(k - \sum_{i\in\mathcal{F},p\in\mathcal{P}} y_{ip})\big) \leqslant 3\big(\mathrm{OPT} + \gamma(k - \sum_{i\in\mathcal{F},p\in\mathcal{P}} y_{ip})\big), \tag{A4}$$

where the first step follows from Claim 1, the second step is due to the fact that DUAL$^\gamma$ is the dual of LP2$^\gamma$ and $\sum_{j\in\mathcal{D}} \alpha_j - \gamma k$ is the dual objective, and the last step follows from the fact that LP2$^\gamma$ is a relaxation of LP1.

Denote by $\mathcal{G}^0 = (x^0, y^0)$ the solution to LP2^0 returned by Claim 1. If $\mathcal{G}^0$ satisfies constraint (3) (*i.e.*, $\sum_{i\in\mathcal{F},p\in\mathcal{P}} y^0_{ip} \leqslant k$), then $\mathcal{G}^0$ is a feasible solution to k-MP. Using inequality (A4), we have $V(\mathcal{G}^0) \leqslant 3\mathrm{OPT}$, which implies that a 3-approximation for the problem is obtained.

We now consider the case where the solution to LP2^0 given by Claim 1 violates constraint (3). By inequality (A4), we know that for any $\gamma \geqslant 0$, the solution (x, y) to LP2$^\gamma$ given by Claim 1 satisfies $\mathrm{OPT} + \gamma(k - \sum_{i\in\mathcal{F},p\in\mathcal{P}} y_{ip}) \geqslant 0$. Consequently, the solution satisfies $k - \sum_{i\in\mathcal{F},p\in\mathcal{P}} y_{ip} > -1$ when $\gamma > \mathrm{OPT}$. By the fact that (x, y) is an integer solution, we have $k - \sum_{i\in\mathcal{F},p\in\mathcal{P}} y_{ip} \geqslant 0$. This implies that the solution to LP2$^\gamma$ given by Claim 1 satisfies constraint (3) if $\gamma > \mathrm{OPT}$. It can be seen that $\mathrm{OPT} < 2|\mathcal{D}|(c_{\max} + f_{\max})$. We perform a binary search on the interval $[0, 2|\mathcal{D}|(c_{\max} + f_{\max})]$ to find two real numbers γ_1 and γ_2 with $0 < \gamma_1 - \gamma_2 \leqslant \epsilon(c_{\min} + f_{\min})/|\mathcal{F}|$, such that the solution to LP2$^{\gamma_1}$ given by Claim 1 satisfies (but does not tighten) constraint (3), and the solution to LP2$^{\gamma_2}$ given by the claim violates constraint (3) (if we find a real number γ' such that Claim 1 yields a solution to LP2$^{\gamma'}$ tightening constraint (3), then inequality (A4) directly implies a 3-approximation for k-MP). The binary search invokes the algorithm in Claim 1 for $O(\log \frac{|\mathcal{D}||\mathcal{F}|(c_{\max}+f_{\max})}{\epsilon(c_{\min}+f_{\min})})$ times, which can be polynomially bounded based on standard discretization methods [2].

Let $\mathcal{G}_1 = (x^1, y^1)$ and $\mathcal{G}_2 = (x^2, y^2)$ denote the solutions to LP2$^{\gamma_1}$ and LP2$^{\gamma_2}$ obtained using Claim 1 respectively, and let (α^1, β^1) and (α^2, β^2) denote the corresponding dual solutions. Define $\sum_{i\in\mathcal{F},p\in\mathcal{P}} y^1_{ip} = k_1$ and $\sum_{i\in\mathcal{F},p\in\mathcal{P}} y^2_{ip} = k_2$. We have $k_1 < k$ and $k_2 > k$. Observe that

$$V(\mathcal{G}_2) \leqslant 3(\sum_{j\in\mathcal{D}} \alpha^2_j - \gamma_2 k_2) = 3(\sum_{j\in\mathcal{D}} \alpha^2_j - \gamma_1 k_2) + 3(\gamma_1 - \gamma_2)k_2 \leqslant 3(\sum_{j\in\mathcal{D}} \alpha^2_j - \gamma_1 k_2) + \epsilon\frac{3k_2}{|\mathcal{F}|}(c_{\min} + f_{\min})$$

* Corresponding author (email: csufeng@csu.edu.cn)

$$\leqslant 3(\sum_{j\in\mathcal{D}}\alpha_j^2-\gamma_1k_2)+\epsilon\frac{3k_2}{|\mathcal{F}|}\mathrm{OPT}=3(\sum_{j\in\mathcal{D}}\alpha_j^2-\gamma_1k_2)+O(\epsilon)\mathrm{OPT}, \tag{A5}$$

where the first step follows from Claim 1, and the third step is derived from the fact that $\gamma_1-\gamma_2\leqslant\epsilon(c_{\min}+f_{\min})/|\mathcal{F}|$.

Let $a=\frac{k_2-k}{k_2-k_1}$ and $b=1-a$. We construct a dual solution $(\tilde{\alpha},\tilde{\beta})$ by letting $\tilde{\alpha}_j=a\alpha_j^1+b\alpha_j^2$ and $\tilde{\beta}_j=a\beta_j^1+b\beta_j^2$ for each $j\in\mathcal{D}$. It is the case that

$$\begin{aligned}aV(\mathcal{G}_1)+bV(\mathcal{G}_2)&\leqslant 3a(\sum_{j\in\mathcal{D}}\alpha_j^1-\gamma_1k_1)+3b(\sum_{j\in\mathcal{D}}\alpha_j^2-\gamma_1k_2)+O(\epsilon)\mathrm{OPT}=3[\sum_{j\in\mathcal{D}}(a\alpha_j^1+b\alpha_j^2)-\gamma_1(ak_1+bk_2)]+O(\epsilon)\mathrm{OPT}\\&=3(\sum_{j\in\mathcal{D}}\tilde{\alpha}_j-\gamma_1k)+O(\epsilon)\mathrm{OPT},\end{aligned} \tag{A6}$$

where the first step follows from Claim 1 and inequality (A5).

By the fact that $\gamma_1-\gamma_2>0$, we know that (α^2,β^2) is a feasible solution to DUAL^{γ_1}. Consequently, $(\tilde{\alpha},\tilde{\beta})$ is feasible for DUAL^{γ_1} since it is a convex combination of two feasible solutions, and its cost for DUAL^{γ_1} is $\sum_{j\in\mathcal{D}}\tilde{\alpha}_j-\gamma_1k$. We have $\sum_{j\in\mathcal{D}}\tilde{\alpha}_j-\gamma_1k\leqslant\mathrm{OPT}$ due to the fact that DUAL^{γ_1} is the dual of a relaxation of LP1. Thus, using inequality (A6), we get

$$aV(\mathcal{G}_1)+bV(\mathcal{G}_2)\leqslant 3(\sum_{j\in\mathcal{D}}\tilde{\alpha}_j-\gamma_1k)+O(\epsilon)\mathrm{OPT}\leqslant\big(3+O(\epsilon)\big)\mathrm{OPT}.$$

This completes the proof of Lemma 1. □

Appendix B Proof of Proposition 1

Proposition 1. For any $0<\tau<1$, if $V(\mathcal{G}_1)\geqslant\frac{1}{\tau}V(\mathcal{G}_f)$, then we have $a<\tau$ and $V(\mathcal{G}_2)<V(\mathcal{G}_f)$.

Proof. The assumption that $V(\mathcal{G}_1)\geqslant\frac{1}{\tau}V(\mathcal{G}_f)$ implies that

$$V(\mathcal{G}_1)\geqslant\frac{1}{\tau}\big(aV(\mathcal{G}_1)+bV(\mathcal{G}_2)\big)>\frac{1}{\tau}aV(\mathcal{G}_1),$$

which implies that $a<\tau$. Also by the assumption that $V(\mathcal{G}_1)\geqslant\frac{1}{\tau}V(\mathcal{G}_f)$, we have

$$V(\mathcal{G}_f)=aV(\mathcal{G}_1)+bV(\mathcal{G}_2)\geqslant\frac{a}{\tau}V(\mathcal{G}_f)+bV(\mathcal{G}_2)=\frac{a}{\tau}V(\mathcal{G}_f)+(1-a)V(\mathcal{G}_2).$$

Thus, we get

$$V(\mathcal{G}_2)\leqslant\frac{1-a/\tau}{1-a}V(\mathcal{G}_f)<V(\mathcal{G}_f),$$

where the second step is due to the fact that $a<\tau$ and $0<\tau<1$. □

Appendix C Proof of Proposition 2

Proposition 2. We can find in polynomial time an optimal solution to LP3 that has at most one fractional variable, which is associated with a facility $i\in\mathcal{H}^1$ such that $|\mathcal{L}_i|>1$.

Proof. Observe that LP3 is an LP relaxation of a 0-1 Knapsack problem, and we can obtain the desired solution using a simple greedy method. Define $\mathcal{H}^\dagger=\{i\in\mathcal{H}^1:|\mathcal{L}_i|>1\}$. We sort each $i\in\mathcal{H}^\dagger$ by increasing value of $\Psi_1(i)/(|\mathcal{L}_i|-1)$, and let i^q denote the q-th facility in this order for each $1\leqslant q\leqslant|\mathcal{H}^\dagger|$. Define $\mathcal{H}_q^\dagger=\{i^{\tilde{q}}:1\leqslant\tilde{q}\leqslant q\}$ for each $1\leqslant q\leqslant|\mathcal{H}^\dagger|$. Let $\mathcal{H}_0^\dagger=\emptyset$ and $\sum_{i\in\mathcal{H}_0^\dagger}(|\mathcal{L}_i|-1)=0$. The definition of $\mathcal{H}^\dagger$ implies that $\sum_{i\in\mathcal{H}^\dagger}(|\mathcal{L}_i|-1)\geqslant\sum_{i\in\mathcal{H}^1}(|\mathcal{L}_i|-1)=|\mathcal{H}^2|-|\mathcal{H}^1|>|\mathcal{H}^2|-k$. Consequently, we can find an integer $1\leqslant q\leqslant|\mathcal{H}^\dagger|$ that satisfies

$$\sum_{i\in\mathcal{H}_{q-1}^\dagger}(|\mathcal{L}_i|-1)\leqslant|\mathcal{H}^2|-k<\sum_{i\in\mathcal{H}_q^\dagger}(|\mathcal{L}_i|-1). \tag{C1}$$

We construct a solution z^* to LP3 as follows. Let $z_i^*=1$ for each $i\in\mathcal{H}_{q-1}^\dagger$ and $z_i^*=0$ for each $i\in\mathcal{H}^1\backslash\mathcal{H}_q^\dagger$. Let $z_{i^q}^*=\frac{1}{|\mathcal{L}_{i^q}|-1}\big(|\mathcal{H}^2|-k-\sum_{i\in\mathcal{H}_{q-1}^\dagger}(|\mathcal{L}_i|-1)\big)$ to satisfy constraint (6). By inequality (C1), we have $0\leqslant z_{i^q}^*<1$. Since we sort each $i\in\mathcal{H}^\dagger$ by increasing value of $\Psi_1(i)/(|\mathcal{L}_i|-1)$, it is easy to show that z^* is an optimal solution to LP3. The solution has at most one fractional variable $z_{i^q}^*$, which is associated with a facility from $\mathcal{H}^\dagger$. Thus, z^* is the desired solution to LP3. □

Appendix D Proof of Lemma 2

Lemma 2. $|\mathcal{H}'|\leqslant k$.

Proof. If none of the facilities from $\mathcal{H}^1$ is associated with a fractional variable, then we have $|\mathcal{H}'|=\sum_{i\in\mathcal{O}_0}|\mathcal{L}_i|+|\mathcal{O}_1|$. Using constraint (6), we get $\sum_{i\in\mathcal{O}_1}|\mathcal{L}_i|-|\mathcal{O}_1|=|\mathcal{H}^2|-k$, which implies that $k=|\mathcal{H}^2|-\sum_{i\in\mathcal{O}_1}|\mathcal{L}_i|+|\mathcal{O}_1|=\sum_{i\in\mathcal{O}_0}|\mathcal{L}_i|+|\mathcal{O}_1|$. Thus, we have $|\mathcal{H}'|=k$, as desired.

For the case where facility $t\in\mathcal{H}^1$ is associated with a fractional variable z_t^*, we have $\mathcal{H}'=(\bigcup_{i\in\mathcal{O}_0}\mathcal{L}_i)\cup\mathcal{O}_1\cup\{t\}\cup(\mathcal{L}_t\backslash\mathcal{L}^\dagger)$, and thus $|\mathcal{H}'|\leqslant\sum_{i\in\mathcal{O}_0}|\mathcal{L}_i|+|\mathcal{O}_1|+1+(1-z_t^*)|\mathcal{L}_t|$. Constraint (6) implies that

$$k=|\mathcal{H}^2|-\sum_{i\in\mathcal{O}_1}|\mathcal{L}_i|+|\mathcal{O}_1|-z_t^*|\mathcal{L}_t|+z_t^*=\sum_{i\in\mathcal{O}_0}|\mathcal{L}_i|+|\mathcal{O}_1|+(1-z_t^*)|\mathcal{L}_t|+z_t^*.$$

Consequently, we get $|\mathcal{H}'|\leqslant k+1-z_t^*$. By the fact that z_t^* is a fractional variable and $|\mathcal{H}'|$ is an integer, we have $|\mathcal{H}'|\leqslant k$. This completes the proof of Lemma 2. □

Appendix E Proof of Lemma 3

Lemma 3. For any $0 < \tau < 1$, if $V(\mathcal{G}_1) \geqslant \frac{1}{\tau} V(\mathcal{G}_f)$, then $\Phi < \max\{2, \frac{3+\tau}{2-\tau}\} V(\mathcal{G}_f)$.

Proof. We first show that Φ is near to $V(\mathcal{G}_2)$.

- For each $i \in \mathcal{O}_0$ and $i' \in \mathcal{L}_i$, our solution opens i' at priority $p_2(i')$, and assigns each $j \in \phi(i')$ to i'. The total opening cost of the facilities from $\mathcal{L}_i$ induced by the solution is $\sum_{i' \in \mathcal{L}_i} f(p_2(i'))$, and the assignment cost of the clients from $\phi(\mathcal{L}_i)$ is $\sum_{j \in \phi(\mathcal{L}_i)} c_2(j)$.
- For each $i \in \mathcal{O}_1$, the solution opens i at priority $\max_{i' \in \mathcal{L}_i} p_2(i')$, and assigns each $j \in \phi(\mathcal{L}_i)$ to i. The opening cost of i is $\max_{i' \in \mathcal{L}_i} f(p_2(i')) \leqslant \sum_{i' \in \mathcal{L}_i} f(p_2(i'))$, and the assignment cost of the clients from $\phi(\mathcal{L}_i)$ is no more than $\sum_{j \in \phi(\mathcal{L}_i)} \big(2c_2(j) + c_1(j)\big)$ due to inequality (5).
- If facility $t \in \mathcal{H}^1$ is associated with a fractional variable, then the solution opens t at priority $\max_{i' \in \mathcal{L}^\dagger} p_2(i')$ and assigns each $j \in \phi(\mathcal{L}^\dagger)$ to t. For each $i' \in \mathcal{L}_t \backslash \mathcal{L}^\dagger$, the solution opens i' at priority $p_2(i')$, and assigns each $j \in \phi(i')$ to i'. The cost of opening the facilities from $\{t\} \cup (\mathcal{L}_t \backslash \mathcal{L}^\dagger)$ is $\max_{i' \in \mathcal{L}^\dagger} f(p_2(i')) + \sum_{i' \in \mathcal{L}_t \backslash \mathcal{L}^\dagger} f(p_2(i')) \leqslant \sum_{i' \in \mathcal{L}_t} f(p_2(i'))$, and the assignment cost of the clients from $\phi(\mathcal{L}_t)$ is at most $\sum_{j \in \phi(\mathcal{L}^\dagger)} \big(2c_2(j) + c_1(j)\big) + \sum_{j \in \phi(\mathcal{L}_t \backslash \mathcal{L}^\dagger)} c_2(j)$ due to inequality (5).

By the argument above, the assignment cost induced by the solution can be upper-bounded by $\sum_{j \in \phi(\bigcup_{i \in \mathcal{O}_1} \mathcal{L}_i) \cup \phi(\mathcal{L}^\dagger)} \big(2c_2(j) + c_1(j)\big) + \sum_{j \in \phi(\bigcup_{i \in \mathcal{O}_0} \mathcal{L}_i) \cup \phi(\mathcal{L}_t \backslash \mathcal{L}^\dagger)} c_2(j)$, which is at most $C(\mathcal{G}_2) + \sum_{i \in \mathcal{O}_1} \Psi_1(i) + \sum_{i \in \mathcal{L}^\dagger} \Psi_2(i)$ due to the definitions of $\Psi_1(i)$ and $\Psi_2(i)$, and the cost of opening facilities induced by the solution is no more than $\sum_{i \in \mathcal{H}^1} \sum_{i' \in \mathcal{L}_i} f(p_2(i')) = \sum_{i' \in \mathcal{H}^2} f(p_2(i')) \leqslant F(\mathcal{G}_2)$. Compared with $\mathcal{G}_2$, this solution increases the cost for the problem by no more than $\sum_{i \in \mathcal{O}_1} \Psi_1(i) + \sum_{i \in \mathcal{L}^\dagger} \Psi_2(i)$. Let $R = \sum_{i \in \mathcal{O}_1} \Psi_1(i) + \sum_{i \in \mathcal{L}^\dagger} \Psi_2(i)$ denote this upper bound on the increased cost. We now show that R is quite small compared with $V(\mathcal{G}_1)$ and $V(\mathcal{G}_2)$. Denote by $opt = \sum_{i \in \mathcal{H}^1} z_i^* \Psi_1(i)$ the value of z^* for LP3. Observe that

$$\frac{|\mathcal{H}^2| - k}{|\mathcal{H}^2| - |\mathcal{H}^1|} \sum_{i \in \mathcal{H}^1} (|\mathcal{L}_i| - 1) = \frac{|\mathcal{H}^2| - k}{|\mathcal{H}^2| - |\mathcal{H}^1|} (|\mathcal{H}^2| - |\mathcal{H}^1|) = |\mathcal{H}^2| - k.$$

This implies that the solution taking $z_i = \frac{|\mathcal{H}^2| - k}{|\mathcal{H}^2| - |\mathcal{H}^1|}$ for each $i \in \mathcal{H}^1$ is feasible for LP3, whose value is

$$\frac{|\mathcal{H}^2| - k}{|\mathcal{H}^2| - |\mathcal{H}^1|} \sum_{i \in \mathcal{H}^1} \Psi_1(i) \leqslant \frac{k_2 - k}{k_2 - k_1} \sum_{i \in \mathcal{H}^1} \Psi_1(i) = a \sum_{i \in \mathcal{H}^1} \Psi_1(i) < a\big(V(\mathcal{G}_1) + V(\mathcal{G}_2)\big),$$

where the first step follows from the fact that $|\mathcal{H}^1| \leqslant k_1$ and $|\mathcal{H}^2| \leqslant k_2$, and the last step is due to the definition of $\Psi_1(i)$. The fact that z^* is an optimal solution to LP3 implies that *opt* is no more than the value of the solution considered above, and thus

$$opt < a\big(V(\mathcal{G}_1) + V(\mathcal{G}_2)\big). \tag{E1}$$

If we can show that

$$R < opt + \frac{a}{1+a}\big(V(\mathcal{G}_1) + V(\mathcal{G}_2)\big), \tag{E2}$$

then using inequality (E1), we have

$$R < a\frac{2+a}{1+a}\big(V(\mathcal{G}_1) + V(\mathcal{G}_2)\big). \tag{E3}$$

It remains to show inequality (E2). If there is not a fractional variable in z^*, then we have $\mathcal{L}^\dagger = \emptyset$ and $R = \sum_{i \in \mathcal{O}_1} \Psi_1(i) + \sum_{i \in \mathcal{L}^\dagger} \Psi_2(i) = \sum_{i \in \mathcal{O}_1} \Psi_1(i)$. By the definition of $\mathcal{O}_1$, we get $R = \sum_{i \in \mathcal{O}_1} \Psi_1(i) = opt$, as desired. We now consider the case where facility $t \in \mathcal{H}^1$ is associated with a fractional variable z_t^*. For this case, we have

$$\begin{aligned} R - opt &= \sum_{i \in \mathcal{O}_1} \Psi_1(i) + \sum_{i \in \mathcal{L}^\dagger} \Psi_2(i) - opt = \sum_{i \in \mathcal{O}_1} \Psi_1(i) + \sum_{i \in \mathcal{L}^\dagger} \Psi_2(i) - \sum_{i \in \mathcal{O}_1} \Psi_1(i) - z_t^* \Psi_1(t) = \sum_{i \in \mathcal{L}^\dagger} \Psi_2(i) - z_t^* \Psi_1(t) \\ &\leqslant \frac{\lceil z_t^* |\mathcal{L}_t| \rceil}{|\mathcal{L}_t|} \Psi_1(t) - z_t^* \Psi_1(t) < \frac{1}{|\mathcal{L}_t|} \Psi_1(t), \end{aligned} \tag{E4}$$

where the second step follows from the definitions of $\mathcal{O}_1$ and $\mathcal{O}_2$, and the fourth step is derived from the fact that $\mathcal{L}^\dagger = \arg\min_{\mathcal{L} \subseteq \mathcal{L}_t \cap |\mathcal{L}| = \lceil z_t^* |\mathcal{L}_t| \rceil} \sum_{i' \in \mathcal{L}} \Psi_2(i')$ and $\Psi_1(t) = \sum_{i' \in \mathcal{L}_t} \Psi_2(i')$. We break the analysis into the following two subcases: (2.1) $\frac{1}{|\mathcal{L}_t|} < \frac{a}{1+a}$, and (2.2) $\frac{1}{|\mathcal{L}_t|} \geqslant \frac{a}{1+a}$.

We first consider case (2.1). We have

$$\frac{1}{|\mathcal{L}_t|} \Psi_1(t) < \frac{a}{1+a} \Psi_1(t) < \frac{a}{1+a}\big(V(\mathcal{G}_1) + V(\mathcal{G}_2)\big), \tag{E5}$$

where the second step follows from the definition of $\Psi_1(t)$. Using inequalities (E4) and (E5), we get $R < opt + \Psi_1(t)/|\mathcal{L}_t| < opt + \frac{a}{1+a}\big(V(\mathcal{G}_1) + V(\mathcal{G}_2)\big)$, as desired.

We now consider case (2.2). Proposition 2 implies that $|\mathcal{L}_t| > 1$. We have

$$\frac{1}{|\mathcal{L}_t|} \Psi_1(t) = (1 - \frac{1}{|\mathcal{L}_t|}) \frac{1}{|\mathcal{L}_t| - 1} \Psi_1(t) \leqslant \frac{1}{(1+a)(|\mathcal{L}_t| - 1)} \Psi_1(t), \tag{E6}$$

where the second step is due to the condition of case (2.2). By the fact that z_t^* is a fractional variable and $|\mathcal{L}_t| > 1$, we know that $z_t^*(|\mathcal{L}_t| - 1) > 0$. Constraint (6) implies that $z_t^*(|\mathcal{L}_t| - 1) = |\mathcal{H}_2| - k - \sum_{i \in \mathcal{O}_1} (|\mathcal{L}_i| - 1)$, which in turn implies that $z_t^*(|\mathcal{L}_t| - 1)$ is an integer. Thus, we get $z_t^*(|\mathcal{L}_t| - 1) \geqslant 1$. Consequently, we have

$$R - opt < \frac{1}{|\mathcal{L}_t|} \Psi_1(t) \leqslant \frac{1}{(1+a)(|\mathcal{L}_t| - 1)} \Psi_1(t) \leqslant \frac{1}{1+a} z_t^* \Psi_1(t) \leqslant \frac{1}{1+a} opt < \frac{a}{1+a}\big(V(\mathcal{G}_1) + V(\mathcal{G}_2)\big),$$

where the first step is derived from inequality (E4), the second step follows from inequality (E6), the third step follows from the fact that $z_t^*(|\mathcal{L}_t|-1) \geqslant 1$, and the last step is due to inequality (E1). This completes the proof of inequality (E2), and thus inequality (E3) is true.

Using inequality (E3), we have

$$\frac{\Phi}{V(\mathcal{G}_f)} \leqslant \frac{V(\mathcal{G}_2)+R}{V(\mathcal{G}_f)} = \frac{V(\mathcal{G}_2)+R}{aV(\mathcal{G}_1)+(1-a)V(\mathcal{G}_2)} < \frac{1}{1+a} \cdot \frac{a(2+a)V(\mathcal{G}_1)+(a^2+3a+1)V(\mathcal{G}_2)}{aV(\mathcal{G}_1)+(1-a)V(\mathcal{G}_2)}. \tag{E7}$$

If $a < \frac{\sqrt{6}}{2} - 1$, then we have $a^2+3a+1 < (1-a)(2+a)$, and thus

$$\frac{a(2+a)V(\mathcal{G}_1)+(a^2+3a+1)V(\mathcal{G}_2)}{aV(\mathcal{G}_1)+(1-a)V(\mathcal{G}_2)} < \frac{a(2+a)V(\mathcal{G}_1)+(1-a)(2+a)V(\mathcal{G}_2)}{aV(\mathcal{G}_1)+(1-a)V(\mathcal{G}_2)} = 2+a. \tag{E8}$$

Using inequalities (E7) and (E8), we have

$$\frac{\Phi}{V(\mathcal{G}_f)} < \frac{2+a}{1+a} < 2,$$

where the second step is derived from the fact that $a > 0$. This implies that Lemma 3 is true for the case where $a < \frac{\sqrt{6}}{2} - 1$.

We now consider the case where $a \geqslant \frac{\sqrt{6}}{2} - 1$. For this case, it can be seen that the right hand side of inequality (E7) decreases monotonously with increasing value of $V(\mathcal{G}_1)/V(\mathcal{G}_2)$. Consequently, we get

$$\begin{aligned}\frac{\Phi}{V(\mathcal{G}_f)} &< \frac{1}{1+a} \cdot \frac{a(2+a)V(\mathcal{G}_1)+(a^2+3a+1)V(\mathcal{G}_2)}{aV(\mathcal{G}_1)+(1-a)V(\mathcal{G}_2)} < \frac{1}{1+a} \cdot \frac{a(2+a)/\tau+a^2+3a+1}{a/\tau+1-a} \\ &= \frac{(1/\tau+1)a+1}{(1/\tau-1)a+1} + \frac{(1/\tau+1)a}{(1/\tau-1)a^2+a/\tau+1},\end{aligned} \tag{E9}$$

where the first step is due to inequality (E7), and the second step follows from the fact that $V(\mathcal{G}_1)/V(\mathcal{G}_2) > V(\mathcal{G}_1)/V(\mathcal{G}_f) \geqslant \frac{1}{\tau}$, which is derived from Proposition 1. It can be easily verified that the right hand side of inequality (E9) increases monotonously with increasing value of a for $0 < a \leqslant \tau$ and $0 < \tau < 1$. Thus, using inequality (E9), we get

$$\frac{\Phi}{V(\mathcal{G}_f)} < \frac{(1/\tau+1)a+1}{(1/\tau-1)a+1} + \frac{(1/\tau+1)a}{(1/\tau-1)a^2+a/\tau+1} < \frac{(1/\tau+1)\tau+1}{(1/\tau-1)\tau+1} + \frac{(1/\tau+1)\tau}{(1/\tau-1)\tau^2+\tau/\tau+1} = \frac{3+\tau}{2-\tau},$$

where the second step is due to the fact that $a < \tau$, which follows from Proposition 1. This completes the proof of Lemma 3. □

References

1 Mahdian M. Facility location and the analysis of algorithms through factor-revealing problems, Ph.D.'s degree thesis. Massachusetts Institute of Technology, Cambridge, MA, 2004

2 Saulpic D, Cohen-Addad V, Feldmann A E. Near-linear time approximations schemes for clustering in doubling metrics. In: Proceedings of the 60th IEEE Annual Symposium on Foundations of Computer Science, 2019: 540–559