Clustering to Minimize Cluster-Aware Norm Objectives

Tarek Sallam

Carleton University

$(f, g)-$Clustering

Input:

- X: A set of data points
- Y: A set of potential cluster centers
- $k > 0$: The number of clusters
- d: A metric of X and Y
- f: A monotone symmetric norm (inner-norm)
- g: A monotone symmetric norm (outer-norm)

Goal: Find

- $C \subseteq Y$: A set of cluster centers with $|C| \leq k$
- $\sigma : X \to C$: Assigns points to centers

That minimizes $\Phi = g(f(\delta(c))_{c \in C})$

$(f, g)-$**Clustering**

Well-known problems

- $k-$Median: $f = \mathcal{L}_1,\ g = \mathcal{L}_1$
- $k-$Centers: $f = \mathcal{L}_\infty,\ g = \mathcal{L}_\infty$:
- $k-$Means: $f = \mathcal{L}_2,\ g = \mathcal{L}_2$
- Min-Sum of Radii: $f = \mathcal{L}_\infty,\ g = \mathcal{L}_1$

$(\mathbf{Top}_\ell, \mathcal{L}_1)$-Clustering

$(\textbf{Top}_\ell, \mathcal{L}_1)$-Clustering

Case of (f, g)-clustering where:

- $f = \text{Top}_\ell$: Sum of the largest ℓ distances in the cluster
- $g = \mathcal{L}_1$: Sum over all clusters

Goal becomes: Minimize the sum of the largest ℓ distances in each cluster.

Note: This problem generalizes k-Median ($\ell = |X|$) and Min-Sum of Radii ($\ell = 1$).

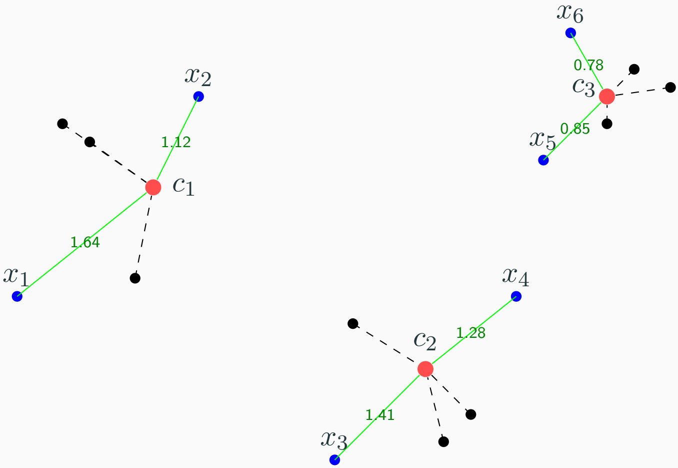
$(\mathbf{Top}_\ell, \mathcal{L}_1)$-Clustering

$\ell = 2$

$k = 3$

$|X| = 15$

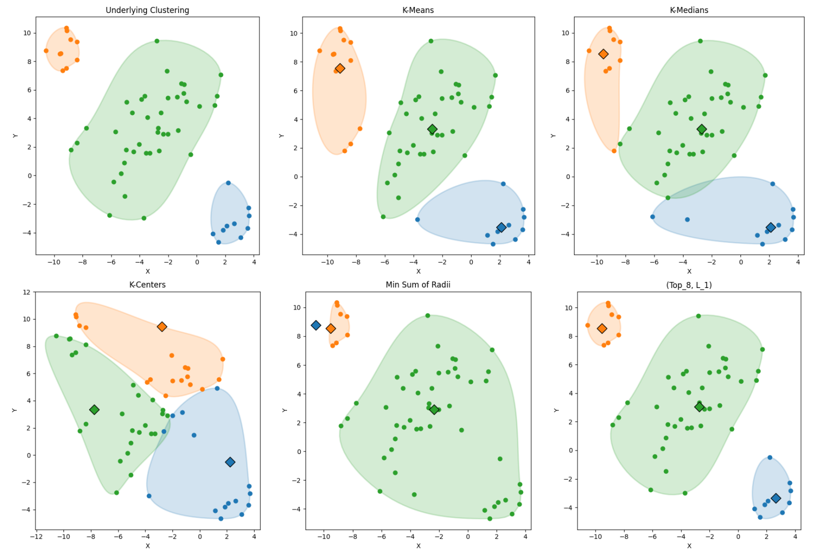
$|Y| = 3$



$$\Phi = 1.64 + 1.12 + 1.41 + 1.28 + 0.85 + 0.78 = 7.08$$

$(\textbf{Top}_\ell, \mathcal{L}_1)$-Clustering

Why $(\mathrm{Top}_\ell, \mathcal{L}_1)$ objective?



$(\mathbf{Top}_\ell, \mathcal{L}_1)$-Clustering

We can map this problem to a problem introduced as Ball-k-Median.

Ball-k-Median

Ball-k-Median

Input:

- X, Y, k, d: As before
- $\rho \in \mathbb{R}^{>0}$: A penalty parameter

Goal: Find

- $C \subseteq Y$: A set of cluster centers
- $r : C \to \mathbb{R}$: Radius value for each center

Define:

- $d^r(x, c) = \max\{d(x, c) - r(c), 0\}$
  - If point is within ball with center x, radius r, distance is 0,
  - If point is outside ball, distance is the distance to the ball

Note: $d^r(x, c)$ does *not* satisfy triangle inequality.

Ball-k-Median

Goal: Find

- $C \subseteq Y$: A set of cluster centers
- $r : C \to \mathbb{R}$: Radius value for each center

That minimizes:

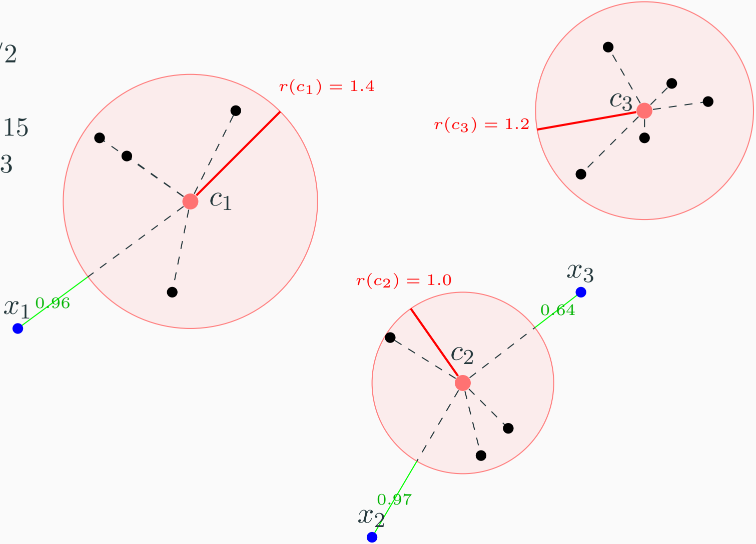
$$\Phi = \sum_{x \in X} \min_{c \in C} d^r(x, c) + \rho \sum_{c \in C} r(c)$$

Let (C^*, r^*) be the optimal solution, with potential Φ^*.

Intuition: We pay for the distance of points outside the balls, and pay a penalty for the size of the balls.

Ball-k-Median

$\rho = 3/2$
$k = 3$
$|X| = 15$
$|Y| = 3$



$$\Phi = (0.96 + 0.97 + 0.64) + 1.5(1.4 + 1.2 + 1) = 7.97$$

Ball-k-Median

Lemma 1

$(\mathrm{Top}_\ell, \mathcal{L}_1)$-*Clustering can be mapped to Ball-*k*-Median by setting* $\rho = \ell$*.*

Ball-k-Median

Note: An important property of this problem, is that any optimal solution (C^*, r^*) will have an equivalently optimal solution where the radii $r(c^*)$ are equal to some point-to-center distance $d(x, c^*)$. for all $c^* \in C^*$.

Approximating Ball-k-Median

Approximating Ball-k-Median

Guessing Large Radii

1. Pick a constant $\varepsilon > 0$, $t = \lceil \frac{3}{\varepsilon} \rceil$
2. For each subset $T \subseteq Y$ of size $|T| = t$
   2.1 For each $c \in C$ pick a radius $r(c)$ as a distance to some point $d(x, c)$
   2.2 Output all possible combinations of (T, r)

Approximating Ball-k-Median

Lemma 2 (Guessing Large Radii)

- *The algorithm runs in* $O(|Y|^{1/\varepsilon} + |X|^{1/\varepsilon})$ *time.*
- *One choice of* (T, r) *will have:*
  1. $T \subseteq C^*$ *and for all* $c \in T$, $r(c) = r^*(c)$
  2. *For all* $c^* \in C^* \setminus T$
     - $r^*(c^*) \leq \min_{c \in T} r(c)$
  3. $\min_{c \in T} r(c) \leq \varepsilon/(3\rho) \cdot \Phi^*$

Approximating Ball-k-Median

Relaxing k

We now relax k by allowing more centers, however we introduce $\lambda \in \mathbb{R}^{>0}$ to penalize each center used. The objective becomes:

$$\Phi = \sum_{x \in X} \min_{c \in C} d^r(x, c) + \rho \sum_{c \in C} r(c) + \lambda |C|$$

$$= \sum_{x \in X} \min_{c \in C} d^r(x, c) + \sum_{c \in C} (\rho r(c) + \lambda)$$

Approximating Ball-k-Median

Theorem 3

There exists an algorithm that returns a solution (C, r) *with:*

$$\Phi + 3\lambda|C| \leq 3(\Phi^* + \lambda k)$$

Approximating Ball-k-Median

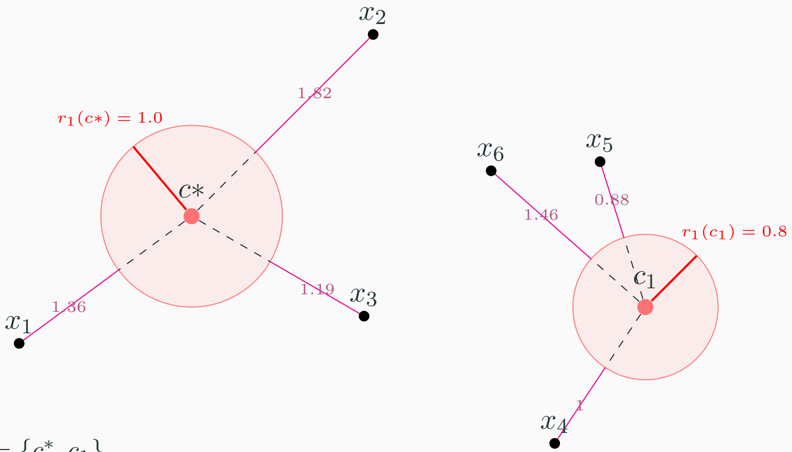
Theorem 4

We can obtain a bi-point solution $(C_1, r_1), (C_2, r_2)$ *using binary search on* λ, *where* $|C_1| \leq k \leq |C_2|$ *with:*

1. $T \subseteq C_1$ *and* $T \subseteq C_2$
2. $c_1 \in C_1 \setminus T \Rightarrow r_1(c_1) \leq \varepsilon/(3\rho) \cdot \Phi^*$
3. $c_2 \in C_2 \setminus T \Rightarrow r_2(c_2) \leq \varepsilon/(3\rho) \cdot \Phi^*$
4. $a\Phi_1 + b\Phi_2 \leq (3+\varepsilon)\Phi^*$, *where* $a = \frac{|C_2|-k}{|C_2|-|C_1|}$ *and* $b = \frac{k-|C_1|}{|C_2|-|C_1|}$

Bi-Point Solution

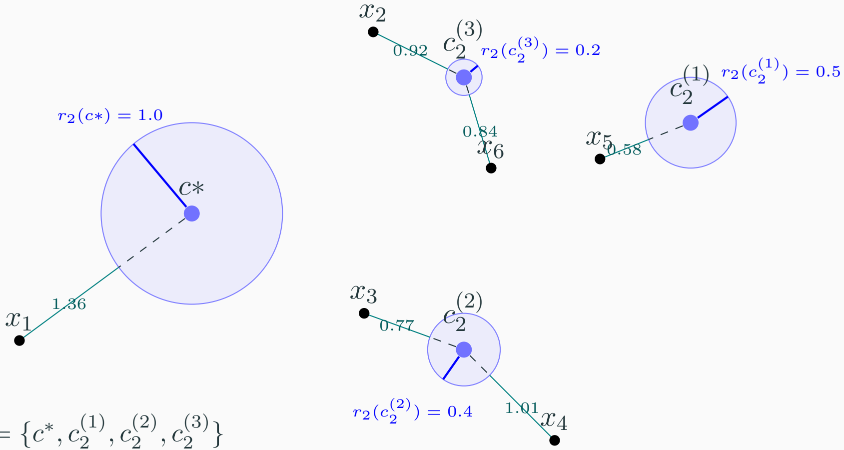
Solution (C_1, r_1), for $k = 3,\ \rho = 3/2$



$C_1 = \{c^*, c_1\}$
$T = \{c^*\}$
$\Phi_1 = (1.36 + 1.82 + 1.19 + 1 + 0.88 + 1.46) + 3/2(1 + 0.8) = 10.41$

Bi-Point Solution

Solution (C_2, r_2), for $k = 3$, $\rho = 3/2$



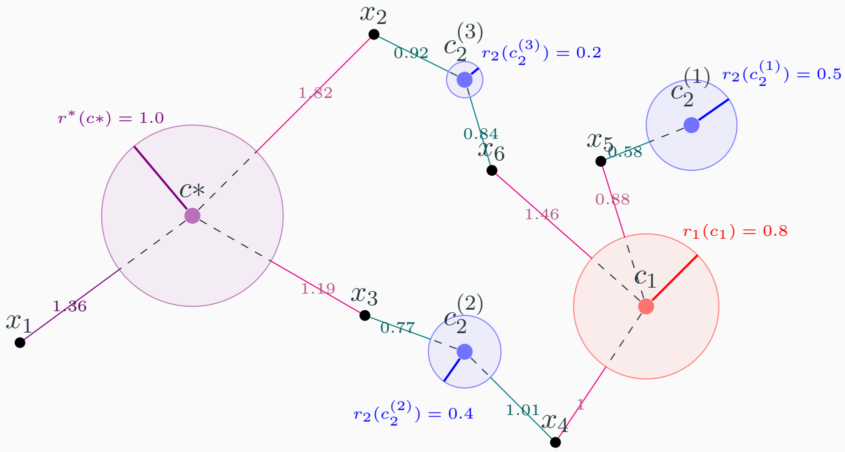
$C_2 = \{c^*, c_2^{(1)}, c_2^{(2)}, c_2^{(3)}\}$

$T = \{c^*\}$

$\Phi_1 = (1.36 + 0.92 + 0.77 + 1.01 + 0.58 + 0.84)$

$+3/2(1 + 0.5 + 0.4 + 0.2) = 8.63$

Both solutions together. $T \subseteq C_1$ and $T \subseteq C_2$



Approximating Ball-k-Median

Bi-Point Rounding

- If $a > 1/4$

$$a\Phi_1 + b\Phi_2 \leq (3+\varepsilon)\Phi*$$

$$\Rightarrow \Phi_1 \leq 4(3+\varepsilon)\Phi*$$

  Then we have a $4(3+\varepsilon)-$approximation using C_1

- If $\Phi_1 \leq \Phi_2$

$$\Phi_1 = (a+b)\Phi_1 \leq a\Phi_1 + b\Phi_2 \leq (3+\varepsilon)\Phi*$$

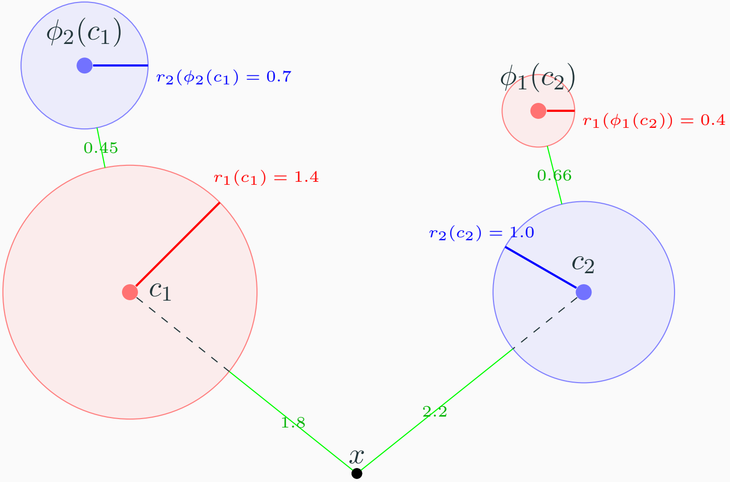
  Then we have a $(3+\varepsilon)-$approximation using C_1

Approximating Ball-k-Median

Bi-Point Rounding

- Now we assume $a \leq 1/4$ and $\Phi_2 \leq \Phi_1$
- Let $\delta(c_1, c_2) = \max\{d(c_1, c_2) - r_1(c_1) - r_2(c_2), 0\}$ be the distance between balls $c_1 \in C_1$ and $c_2 \in C_2$.
- For center $c_2 \in C_2$, let $\phi_1(c_2) = \arg\min_{c_1 \in C_1} \delta(c_1, c_2)$
- Similarly, for center $c_1 \in C_1$, let $\phi_2(c_1) = \arg\min_{c_2 \in C_2} \delta(c_1, c_2)$
- For a point $x \in X$ let $\phi_1(x) = \arg\min_{c_1 \in C_1} d^{r_1}(x, c_1)$
- Similarly, for point $x \in X$, let $\phi_2(x) = \arg\min_{c_2 \in C_2} d^{r_2}(x, c_2)$

$(\mathbf{Top}_\ell, \mathcal{L}_1)$-Clustering



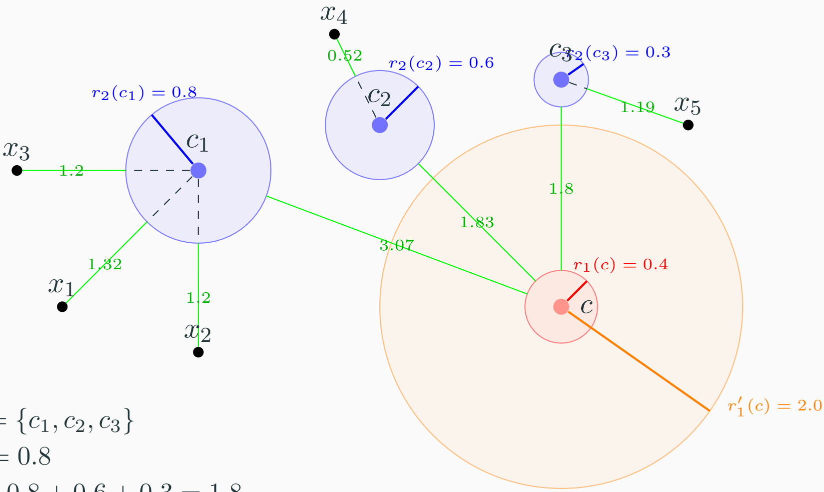
$$c_1 = \phi_1(x), \;\; c_2 = \phi_2(x)$$

Approximating Ball-k-Median

Bi-Point Rounding

- We match each center $c_2 \in C_2$ with $\phi_1(c_2)$. Let $c_1 \in C_1$, denote the subset of centers of C_2 that are closest to c_1: $G_{c_1} = \{c_2 \in C_2 \mid \phi_1(c_2) = c_1\}$
- The largest radius that is closest to c_1 is: $m_{c_1} = \max\limits_{c_2 \in G_{c_1}} r_2(c_2)$
- The sum of all radii closest to c_1 is: $s_{c_1} = \sum_{c_2 \in G_{c_1}} r_2(c_2)$
- Denote the set of points $x \in X$ that are closest to any set of centers $C_2' \subseteq C_2$ as: $\mathcal{X}(C_2') = \{x \in X \mid \phi_2(x) \in C_2'\}$
- Let $r_1'(c_1) = r_1(c_1) + 2m_{c_1}$ be the inflated radius of c_1

$(\mathbf{Top}_\ell, \mathcal{L}_1)$-Clustering



$G_c = \{c_1, c_2, c_3\}$

$m_c = 0.8$

$s_c = 0.8 + 0.6 + 0.3 = 1.8$

$\mathcal{X}(G_c) = \{x_1, x_2, x_3, x_4, x_5\}$

Lemma 5 (Generalized triangle inequality)

Given a point $x \in X$*, with* $c_1 = \phi_1(x),\ c_2 = \phi_2(x),\ c_1' = \phi_1(c_2)$*,*

$$d^{r_1'}(x, c_1') \leq 2d^{r_2}(x, c_2) + d^{r_1}(x, c_1)$$

Approximating Ball-k-Median

Bi-Point Rounding

- For all c_1 in C_1 we have two choices, we either put c_1 into our solution or we put all of G_{c_1} into our solution

1. If we put G_{c_1} into our solution, we contribute a cost of all the $c_2 \in G_{c_1}$ with their respective radii r_{c_2}
2. If we put c_1 into our solution, we enlarge the radii to $r_1(c_1) + m_{c_1}$.

Approximating Ball-k-Median

Linear Program

(relaxed to allow fractional variables):

$$\max \sum_{c_1 \in C_1} u_{c_1}\Big(r_1(c_1)\rho + s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} \big[d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x))\big]\Big)$$

$$\text{s.t.} \sum_{c_1 \in C_1} u_{c_1}(|G_{c_1}| - 1) \leq k - |C_1|$$

$$u_{c_1} \in [0, 1],\ \forall c_1 \in C_1$$

Approximating Ball-k-Median

Fractional Knapsack LP

- Observe that this is the *Fractional Knapsack* problem.
- If we put an item in our bag we add G_{c_1} to our solution. If we don't we add c_1 to our solution.
- The weight of an item u_{c_1} is $(|G_{c_1} - 1)$
- The value of an item is the cost difference between adding G_{c_1} and adding c_1
- Our bag can hold weight at most $k - |C_1|$.
- *Fractional Knapsack*: We want to maximize sum of value while keeping the weight of our bag under the maximum (we can take fractional items)

Approximating Ball-k-Median

Linear Program

- There always exists an optimal solution to *Fractional Knapsack* where at most one variable is fractional, and the rest are integral
- We can find this solution in polynomial time
- If we solve this LP and all variables are integral, we are done
- Suppose there exists one variable: $u_{\tilde{c}_1}$ that is fractional.
- We include the corresponding $\tilde{c}_1$ in our solution with radius $r'_1(\tilde{c}_1)$ and some centers from $G_{\tilde{c}_1} \setminus (T \cap \{\tilde{c}_1\})$

Bi-Point Solution

- For all $c_1 \in C_1 \setminus \{\tilde{c_1}\}$:
  1. If $u_{c_1} = 1$ we include all centers $c_2 \in G_{c_1}$ with corresponding radius $r(c_2) \leftarrow r_2(c_2)$
  2. If $u_{c_1} = 0$ we include center c_1 with radius $r(c_1) \leftarrow r_1'(c_1) = r_1(c_1) + 2m_{c_1}$
- Add $\tilde{c}_1$ with radius $r(\tilde{c}_1) \leftarrow r_1'(\tilde{c}_1) = r_1(\tilde{c}_1) + 2m_{\tilde{c}_1}$, with $m_{\tilde{c}_1} = \max_{c_2 \in G_{\tilde{c}_1} \setminus (T \cap \{\tilde{c}_1\})} r_2(c_2)$
- Add $(\lceil u_{\tilde{c}_1} |G_{\tilde{c}_1}| \rceil - 2)$ centers from $G_{\tilde{c}_1} \setminus (T \cap \{\tilde{c}_1\})$ at random with radius $r(c_2) \leftarrow r_2(c_2)$ for chosen c_2
- Let this constructed solution be (C, r)

Approximating Ball-k-Median

Theorem 6 (Main Claim)

$\Phi \leq (13.5 + 7.5\varepsilon)\Phi^*$

References

[1] Martin G. Herold, Evangelos Kipouridis, and Joachim Spoerhase. *Clustering to Minimize Cluster-Aware Norm Objectives*. 2024. arXiv: 2410.24104 [cs.DS]. URL: https://arxiv.org/abs/2410.24104.

[2] M Charikar et al. "A constant-factor approximation algorithm for the k-median problem". In: *Proceedings of 31st Annual ACM Symposium on Theory of Computing* (Jan. 1999), pp. 1–10.

[3] Moses Charikar and Rina Panigrahy. "Clustering to minimize the sum of cluster diameters". In: *Symposium on the Theory of Computing*. 2001. URL: https://api.semanticscholar.org/CorpusID:10974136.

[4] Shi Li and Ola Svensson. "Approximating k-Median via Pseudo-Approximation". In: *CoRR* abs/1211.0243 (2012). arXiv: 1211.0243. URL: http://arxiv.org/abs/1211.0243.

[5] Sara Ahmadian and Chaitanya Swamy. "Approximation Algorithms for Clustering Problems with Lower Bounds and Outliers". In: *CoRR* abs/1608.01700 (2016). arXiv: 1608.01700. URL: http://arxiv.org/abs/1608.01700.

Extra Proofs and Algorithms

Extra Proofs and algorithms used

Proof of Lemma 2

1. We look through all subsets of size t with all possible radius combinations.
2. Same reasoning as 1.
3.

$$\rho \cdot \min_{c \in T} r(c) \leq \rho \cdot \frac{\sum_{c \in T} r(c)}{|T|} \leq \frac{\rho \cdot \sum_{c^* \in C^*} r^*(c^*)}{\lceil 3/\varepsilon \rceil} \leq \varepsilon/3 \cdot \Phi^*$$

The algorithm for relaxed ball-k-median

LP Relaxation

Let $R_c = \{d(r,c)\}$: The set of possible radii for center c.
Our LP relaxation becomes:

$$\begin{aligned}
\min \quad & \sum_{c \in Y} \sum_{x \in X} \sum_{r \in R_c} d^r(x,c) v_{c,r,x} + \sum_{c \in Y \setminus T} \sum_{r \in R_c} (\rho r + \lambda) u_{c,r} \\
\text{s.t.} \quad & \sum_{c \in Y} \sum_{r \in R_c} v_{c,r,x} \geq 1 & \forall x \in X \\
& v_{c,r,x} \leq u_{c,r} & \forall x \in X, c \in Y, r \in R_c \\
& u_{c,r}, v_{c,r,x} \geq 0 & \forall x \in X, c \in Y, r \in R_c
\end{aligned}$$

$v_{c,r,x}$: Fraction of point x assigned to center c with radius r.
$u_{c,r}$: Fraction of center c opened with radius r

Dual LP

Let $R_c = \{d(r,c)\}$: The set of possible radii for center c.
Our LP relaxation becomes:

$$\begin{aligned}
\max \quad & \sum_{x \in X} \alpha_x & \\
\text{s.t.} \quad & \alpha_x - \beta_{c,r,x} \leq d^r(x,c) & \forall x \in X, c \in Y, r \in R_c \\
& \sum_{x \in X} \beta_{c,r,x} \leq \rho r + \lambda & \forall c \in Y \setminus T, r \in R_c \\
& \sum_{x \in X} \beta_{c,r,x} \leq 0 & \forall c \in T, r \in R_c
\end{aligned}$$

Algorithm Initialization

1. Let T, r' be the guessed centers and radii
2. $\hat{C} \leftarrow \emptyset$
3. $\hat{r}(c) \leftarrow 0$ for all $c \in Y \setminus T$
4. $\alpha_x \leftarrow 0$ for all $x \in X$
5. $\beta_{c,r,x} \leftarrow 0$ for all $x \in X, c \in Y, r \in R_c$

Algorithm Ascent Phase

1. Increase α_x uniformly for all $x \in X$
2. While some α_x is still increasing:

2.1 If $\alpha_x - \beta_{c,r,x} = d^r(x, c)$ for some $c \in Y$, $x \in X, r \in R_c$:

- If $c \in T$: Stop increasing α_x
- If $c \in Y \setminus T$: Increase $\beta_{c,r,c}$ at the same rate as α_x

2.2 If $\sum_{x \in X} \beta_{c,r,x} = \rho r + \lambda$ for some $c \in Y \setminus T, r \in R_c$:

- $\hat{r}(c) \leftarrow \max\{\hat{r}(c), r\}$
- $\hat{C} \leftarrow \hat{C} \cup \{c\}$
- Stop increasing α_x for all x where $\beta_{c,r,x} > 0$
- Stop increasing $\beta_{c,r,x}$ for all $x \in X$

Algorithm Pruning Phase

1. $r(c) \leftarrow r'(c)$ for all $c \in T$
2. $C \leftarrow T$
3. While $\hat{C} \neq \emptyset$:

3.1 Pick $c = \arg\max_{\hat{c} \in \hat{C}} \hat{r}(\hat{c})$

3.2 $C \leftarrow C \cup \{\hat{c}\}$

3.3 $r(c) \leftarrow 3\hat{r}(c)$

3.4 $\hat{C} \leftarrow \hat{C} \setminus \{\hat{c} \in \hat{C} \mid \beta_{c,\hat{r}(c),x}, \beta_{\hat{c},\hat{r}(\hat{c}),x} \geq 0$ for some $x \in X\}$

4. return (C, r)

The analysis for relaxed ball-k-median algorithm

Contributing vs Non-Contributing points

- A point $x \in X$ is **contributing** if for some $c \in C$:
  $\beta_{c,\hat{r}(c),x} > 0$
- Otherwise, x is **non-contributing**.
- We denote X_c as the set of points contributing to center c.
  $X_c = \{x \in X \mid \beta_{c,\hat{r}(c),x} > 0\}$

Lemma 7 (Contributing Points)

$$\sum_{c\in C}\sum_{x\in X_c}\alpha_x = \sum_{c\in C\setminus T}\sum_{x\in X_c} d^{\hat{r}}(x,c) + \rho \sum_{c\in C\setminus T}\hat{r}(c) + \lambda|C \setminus T|$$

Proof of Lemma 7

- For any c added to $\hat{C}$ in the ascent phase:

$$\sum_{x \in X_c} \beta_{c,\hat{r}(c),x} = \rho\hat{r}(c) + \lambda$$

- For any contributing point $x \in X_c$:

$$\alpha_x - \beta_{c,\hat{r}(c),x} = d^{\hat{r}}(p, c)$$

- $X_{c'} = \emptyset$, for all $c' \in T$

Proof of Lemma 7

$$\begin{aligned}\sum_{c\in C}\sum_{x\in X_c}\alpha_x &= \sum_{c\in C\setminus T}\left(\sum_{x\in X_c}(\alpha_x-\beta_{c,\hat{r}(c),x}) + \sum_{x\in X_c}\beta_{c,\hat{r}(c),x}\right)\\ &= \sum_{c\in C\setminus T}\left(\sum_{x\in X_c} d^{\hat{r}}(x,c) + \rho\hat{r}(c) + \lambda\right)\\ &= \sum_{c\in C\setminus T}\sum_{x\in X_c} d^{\hat{r}}(x,c) + \rho\sum_{c\in C\setminus T}\hat{r}(c) + \lambda|C\setminus T|\end{aligned}$$

Lemma 8 (Non-Contributing Points)

For all non-contributing points $x \in X \setminus (\bigcup_{c \in C} X_c)$*,*
there exists a center $c \in C$ *such that:*

$$d^r(x, c) \leq 3\alpha_x$$

Proof of Lemma 8

- Let $\tilde{c}$ and $\tilde{r}$ be the center and radius that caused α_x to stop increasing in the ascent phase.
- If $\tilde{c} \in C$, then:

$$d^r(x,\tilde{c}) \leq d^{\hat{r}}(x,\tilde{c}) \leq d^{\tilde{r}}(x,\tilde{c}) = \alpha_x - \beta_{\tilde{c},\tilde{r},x} \leq \alpha_x \leq 3\alpha_x$$

Proof of Lemma 8

- Suppose $\tilde{c} \notin C$
- There exists some $c \in C$ and $x' \in X$ such that x' contributed to both $\tilde{c}$ and c in the ascent phase
  $\implies \hat{r}(c) \geq \hat{r}(\tilde{c})$
- Since x' contributes to $\tilde{c}$, the $\alpha_{x'}$ stopped increasing no later then when $\tilde{c}$ was added to $\hat{C}$.
- Since x does not contribute to $\tilde{c}$, α_x could have continued to increase after $\tilde{c}$ was added to $\hat{C}$
- So, $\alpha_x \geq \alpha_{x'}$

Proof of Lemma 8

$$
\begin{aligned}
d^r(x,c) &= d^{3\hat{r}}(x,c) \\
&= \max\{d(x,c) - 3\hat{r}(c), 0\} \\
&\leq \max\Big\{d(c,x') + d(x',\tilde{c}) + d(\tilde{c},x) - 3\hat{r}(c), 0\Big\} \\
&\leq \max\{d(c,x') - \hat{r}(c), 0\} + \max\{d(x',\tilde{c}) - \hat{r}(c), 0\} \\
&\quad + \max\{d(\tilde{c},x) - \hat{r}(c), 0\}. \\
&\leq \max\{d(c,x') - \hat{r}(c), 0\} + \max\{d(x',\tilde{c}) - \hat{r}(\tilde{c}), 0\} \\
&\quad + \max\{d(\tilde{c},x) - \hat{r}(\tilde{c}), 0\} \\
&= d^{\hat{r}}(x',c) + d^{\hat{r}}(x',\tilde{c}) + d^{\hat{r}}(x,\tilde{c}) \\
&\leq \alpha_x + 2\alpha_{x'} \\
&\leq 3\alpha_x
\end{aligned}
$$

Proof of Theorem 3

Define $X_C = \bigcup_{c \in C} X_c$

$$
\begin{aligned}
&\Phi + 3\lambda|C \setminus T| - \sum_{c \in T} \rho r(c) \\
&\quad \leq \sum_{x \in X} \min_{c \in C} d^r(x, c) + \sum_{c \in C \setminus T} \rho r(c) + 3\lambda|C \setminus T| \\
&\quad \leq \sum_{c \in C} \sum_{x \in X_c} d^r(x, c) + \sum_{x \in X \setminus X_C} \min_{c \in C} d^r(x, c) \\
&\qquad + \sum_{c \in C \setminus T} \rho r(c) + 3\lambda|C \setminus T|
\end{aligned}
$$

Proof of Theorem 3

$$
\begin{aligned}
&\leq \sum_{c\in C}\sum_{x\in X_c} d^r(x,c) + \sum_{x\in X\setminus X_C} \min_{c\in C} d^r(x,c) \\
&\quad + \sum_{c\in C\setminus T} 3\hat{r}(c)\rho + 3\lambda|C\setminus T| \\
&\leq \sum_{c\in C}\sum_{x\in X_c} \alpha_x + \sum_{x\in X\setminus X_C} \min_{c\in c} d^r(x,c) \\
&\leq 3\sum_{x\in X}\alpha_x \\
&\leq 3\sum_{x\in X}\min_{c^*\in C*} d^{r^*}(x,c^*) + 3\sum_{c^*\in C^*\setminus T}(\rho r^*(c^*)+\lambda)
\end{aligned}
$$

Proof of Theorem 3

$$
\begin{aligned}
&\Phi + 3\lambda|C \setminus T| - \sum_{c \in T} \rho r(c) + \sum_{c \in T} (\rho r(c) + 3\lambda) \\
&\leq 3 \sum_{x \in X} \min_{c^* \in C*} d^{r^*}(x, c^*) + 3 \sum_{c^* \in C^* \setminus T} (\rho r^*(c^*) + \lambda) + 3 \sum_{c \in T} (\rho r(c) + \lambda) \\
\Rightarrow \ &\Phi + 3\lambda|C| \leq 3\Phi^* + 3\lambda k
\end{aligned}
$$

The binary search algorithm on λ

Binary search on λ

1. $\lambda_1 \leftarrow |X| \cdot d_{max}$
2. Consider the solution that would be returned by algorithm, (C'_1, r'_1)
3. Now by removing all centers $c \in C$ where $c \notin T$, we increase the cost by at most $|X| \cdot d_{max}$
4. We decrease cost by at least $|X| \cdot d_{max}$
5. This gives us a new solution (C_1, r_1) with $|C_1| \leq k$
6. And $\Phi_1 + 3\lambda_1|C_1| \leq \Phi'_1 + 3\lambda_1|C'_1| \leq 3(\Phi^* + \lambda_1 k)$

Extra Proofs and Algorithms

Binary search on λ

1. $\lambda_2 \leftarrow 0$
2. Consider the solution that would be returned by algorithm, (C_2', r_2')
3. We add centers $c \in Y$ where $c \notin C_2'$ with radius $r(c) = 0$ until we have at least k centers
4. This gives us a new solution (C_2, r_2) with $|C_2| \geq k$
5. We cannot increase the cost $\Phi_1 = \Phi_1'$, since $C_2 \supseteq C_2'$ and $r(c) = 0$ for all new centers.
6. So, $\Phi_2 + 3\lambda_2|C_2| \leq \Phi_2' + 3\lambda_2|C_2'| \leq 3(\Phi^* + \lambda_2 k)$

Binary search on λ

- Notice $|C_1| \cap |C_2| \subseteq T$
- If $c_2 \in C_2$ but $c_2 \notin C_1$, then $r_2(c_2) = 0$
  $\Rightarrow \rho r_2(c_2) \leq \varepsilon\Phi^*$
- If $c_1 \in C_1$ but $c_1 \notin C_2$, then $c_1 \in T$
  $\Rightarrow \rho r_1(c_1) \leq \varepsilon\Phi^*$
- Also note that increasing λ decreases $|C|$, since we larger λ penalizes opening centers more.

Binary search on λ

Since we have a monotone relationship between λ and $|C|$, and our two constructed endpoint solutions have $|C_1| \leq k$ and $|C_2| \geq k$, with potential preserved, we can perform binary search to find solutions that are close to k.

1. Perform binary search with $\lambda \in [0, |X| \cdot d_{max}]$.
2. We find the midpoint of the interval $\lambda_m = (\lambda_1 + \lambda_2)/2$
3. With the solution (C'_m, r'_m) returned by the algorithm:
4. If $|C'_m| \geq k$, set $\lambda_2 \leftarrow \lambda_m$ and $(C_2, r_2) \leftarrow (C'_m, r'_m)$
5. If $|C'_m| < k$, set $\lambda_1 \leftarrow \lambda_m$ and $(C_1, r_1) \leftarrow (C'_m, r'_m)$
6. Repeat until $|\lambda_1 - \lambda_2| \leq (\varepsilon d_{min})/3|Y|$

Proof of Theorem 4

1. We fix T for any solution in the interval $(0, |X| \cdot d_{max})$ and we have shown that T is a subset for both endpoint solutions.
2. For any solution in the interval $(0, |X| \cdot d_{max})$,
   $3r(c) = \hat{r}(c) \leq \min_{c' \in T} r'(c') \leq \varepsilon/(3\rho)\Phi^*$
   $\Rightarrow r(c) \leq \varepsilon/\rho\Phi^*$
3. We have also shown this for both endpoint solutions already

Proof of Theorem 4.4

Observe $a + b = 1$ and $a|C_1| + b|C_2| = k$.

$$
\begin{aligned}
a\Phi_1 + b\Phi_2 &= a(\Phi_1 + 3\lambda_1|C_1|) + b(\Phi_2 + 3\lambda_2|C_2|) - 3(a\lambda_1|C_1| + b\lambda_2|C_2|) \\
&\leq 3a(\Phi^* + \lambda_1 k) + 3b(\Phi^* + \lambda_2 k) - 3(a\lambda_1|C_1| + b\lambda_2|C_2|) \\
&= 3(a+b)\Phi^* + 3a\lambda_1 k + 3b\lambda_2 k - 3(a\lambda_1|C_1| + b\lambda_2|C_2|) \\
&= 3\Phi^* + 3(a+b)\lambda_1 k + 3b(\lambda_2 - \lambda_1)k - 3(a\lambda_1|C_1| + b\lambda_2|C_2|) \\
&= 3\Phi^* + 3\lambda_1 k + 3b(\lambda_2 - \lambda_1)k - 3(a\lambda_1|C_1| + b\lambda_2|C_2|) \\
&= 3\Phi^* + 3\lambda_1(a|C_1| + b|C_2|) + 3b(\lambda_2 - \lambda_1)k - 3(a\lambda_1|C_1| + b\lambda_2|C_2|) \\
&= 3\Phi^* + 3b\lambda_1|C_2| + 3b(\lambda_2 - \lambda_1)k - 3b\lambda_2|C_2| \\
&= 3\Phi^* + 3b(\lambda_2 - \lambda_1)k - 3b|C_2|(\lambda_2 - \lambda_1) \\
&= 3\Phi^* + (\lambda_2 - \lambda_1)(3bk - 3b|C_2|) \\
&\leq 3\Phi^* + |\lambda_1 - \lambda_2||3bk - 3b|C_2| \\
&\leq 3\Phi^* + |\lambda_1 - \lambda_2|3b|C_2| \text{ because } 0 \leq 3bk \leq 3b|C_2| \\
&\leq 3\Phi^* + \varepsilon d_{min} b|C_2|/|Y| \\
&\leq 3\Phi^* + \varepsilon r(c) b|C_2|/|Y| \text{ for some } c \in T \\
&\leq 3\Phi^* + \varepsilon r(c) \text{ since } b|C_2| \leq k \leq |Y| \\
&\leq (3 + \varepsilon)\Phi^*
\end{aligned}
$$

Bi-Point Rounding

Proof of Lemma 5

$$
\begin{aligned}
d^{r_1'}(x, c_1') &= \max\{d(c_1', x) - r_1'(c_1'), 0\} \\
&= \max\{d(c_1', x) - r_1(c_1') - 2m_{c_1'}, 0\} \\
&\leq \max\{d(c_1', x) - r_1(c_1') - 2r_2(c_2), 0\} \\
&\leq \max\{d(x, c_2) + d(c_2, c_1') - r_1(c_1') - 2r_2(c_2), 0\} \\
&\leq \max\{d(x, c_2) - r_2(c_2), 0\} + \max\{d(c_2, c_1') - r_1(c_1') - r_2(c_2), 0\} \\
&= d^{r_2}(x, c_2) + \delta(c_1', c_2) \\
&\leq d^{r_2}(x, c_2) + \delta(c_1, c_2) \\
&= d^{r_2}(x, c_2) + \max\{d(c_1, c_2) - r_1(c_1) - r_2(c_2), 0\} \\
&\leq d^{r_2}(x, c_2) + \max\{d(c_1, x) + d(x, c_2) - r_1(c_1) - r_2(c_2), 0\} \\
&\leq d^{r_2}(x, c_2) + \max\{d(x, c_2) - r_2(x_2), 0\} + \max\{d(c_1, x) - r_1(x_1), 0\} \\
&= 2d^{r_2}(x, c_2) + d^{r_1}(x, c_1)
\end{aligned}
$$

Approximating Ball-k-Median

Bi-Point Rounding

- For all c_1 in C_1 we have two choices, we either put c_1 into our solution or we put all of G_{c_1} into our solution. Then our

1. If we put G_{c_1} into our solution, we contribute a cost of all the $c_2 \in G_{c_1}$ with their respective radii r_{c_2}
2. If we put c_1 into our solution, we enlarge the radii to $r_1(c_1) + 2m_{c_1}$.

- If we choose all centers in G_{c_1}, $x \in \mathcal{X}(G_{c_1})$ is either closest to some $c_2 \in G_{c_1}$ or some other $c_1' \in C_1$
- So the cost of choosing this is at most: $\sum_{x \in \mathcal{X}(G_{c_1})} \min_{c_2 \in G_{c_1}} d^{r_2}(x, c_2) + \sum_{c_2 \in G_{c_1}} \rho r_2(c_2)$

Bi-Point Rounding

- If we choose c_1 with radius $r_1'(c_1) = r_1(c_1) + 2m_{c_1}$
- Our cost becomes: $\sum_{x \in X | \phi_1(x) = c_1} d^{r'_1}(x, c_1) + \rho(r_1(c_1) + 2m_{c_1})$

Bi-Point Rounding

- Notice that $(r_1(c_1) + 2m_{c_1})\rho \leq (r_1(c_1) + 2s_{c_1})\rho$
- So we save at most $(r_1(c_1) + s_{c_1})\rho$ from the radii cost, by choosing G_{c_1} for our solution
- Notice that $d^{r_1'}(x, c) \leq d^{r_1'}(x, c_1)$ with $c = \arg\min_{c \in C_1} d^{r_1'}(x, c)$ and by Lemma 6, $d^{r_1'}(x, c_1) \leq 2d^{r_2}(x, c_2) + d^{r_1}(x, c)$, where $c_2 \in G_{c_1}$, for any $x \in \mathcal{X}(G_{c_1})$
- So we save at most $d^{r_2}(x, c_2) + d^{r_1}(x, \phi(x))$ for each point $x \in \mathcal{X}(G_{c_1})$ from the point cost, by choosing G_{c_1} for our solution.

Extra Proofs and Algorithms

Bi-Point Rounding

- Naturally we want to save as much cost as possible.
- Let $u_{c_1} = 1$ if we add G_{c_1} to the solution and $u_{c_1} = 0$ if we add c_1 to the soluton.
- Then we save at most

$$\sum_{c_1 \in C_1} u_{c_1} \Big(\big(r_1(c_1) + s_{c_1} \big) \rho + \sum_{x \in \mathcal{X}(G_{c_1})} \left[d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x)) \right] \Big)$$

- However, we can't select every G_{c_1} because we must keep the size of our center set $\leq k$
- Some u_x will be 0. So we have an LP

Knapsack LP

- We know the objective, but what are the constraints?
- The number of centers we have must be $\leq k$ so:

$$\sum_{c_1 \in C_1} u_{c_1} \cdot |G_{c_1}| + (1 - u_{c_1}) \cdot 1 \leq k$$

$$\Rightarrow \sum_{c_1 \in C_1} (u_{c_1}|G_{c_1}| - u_{c_1}) + \sum_{c_1 \in C_1} 1 \leq k$$

$$\Rightarrow \sum_{c_1 \in C_1} u_{c_1} (|G_{c_1}| - 1) \leq k - |C_1|$$

Approximating Ball-k-Median

Lemma 9 (Center feasibility)

$|C| \leq k$

Proof of Lemma 9

- We add at most $u_{\tilde{c}_1}|G_{\tilde{c}_1}| - 1$ centers from $G_{\tilde{c}_1}$ and we add $\tilde{c}_1$ itself, which totals to $u_{\tilde{c}_1}|G_{\tilde{c}_1}|$ centers
- For $c_1 \neq \tilde{c}_1$ we add 1 center if $u_{c_1} = 0$ and we add $|G_{c_1}|$ if $u_{c_1} = 1$, which totals to

$$\sum_{c_1 \in C_1 \setminus \{\tilde{c}_1\}} u_{c_1}(|G_{c_1}|) + (1 - u_{c_1}) \cdot 1$$

$$= |C_1| - 1 + \sum_{c_1 \in C_1 \setminus \{\tilde{c}_1\}} u_{c_1}(|G_{c_1}| - 1)$$

Proof of Lemma 9

By our LP constraint:

$$\sum_{c_1 \in C_1} u_{c_1}(|G_{c_1}| - 1) + |C_1| \leq k$$

$$\Rightarrow u_{\tilde{c}_1}(|G_{\tilde{c}_1}| - 1) + \sum_{c_1 \in C_1 \setminus \{\tilde{c}_1\}} u_{c_1}(|G_{c_1}| - 1) + |C_1| \leq k$$

$$\Rightarrow u_{\tilde{c}_1}|G_{\tilde{c}_1}| + \sum_{c_1 \in C_1 \setminus \{\tilde{c}_1\}} u_{c_1}(|G_{c_1}| - 1) + |C_1| - 1 \leq k - 1 + u_{\tilde{c}_1}$$

$$\Rightarrow |C| \leq k - 1 + u_{\tilde{c}_1} \leq k$$

Approximating Ball-k-Median

Lemma 10 (Portion of centers added)

Let $G'_{\tilde{c}_1} = G_{\tilde{c}_1} \setminus (T \cap \{\tilde{c}_1\})$ *and let the ratio of centers added to our solution from* $G_{\tilde{c}_1}$ *be* $p_{\tilde{c}_1}$*. That is* $p_{\tilde{c}_1} = \frac{\lceil u_{\tilde{c}_1} |G'_{\tilde{c}_1}| \rceil - 2}{|G'_{\tilde{c}_1}|}$ *Then*

$$\frac{(1 - p_{\tilde{c}_1})}{(1 - u_{\tilde{c}_1})} \leq 3$$

Proof of Lemma 10

- If $u_{\tilde{c}_1} \leq 2/3$ then $1 - u_{\tilde{c}_1} \geq 1/3$, so we have:

$$\frac{(1 - p_{\tilde{c}_1})}{(1 - u_{\tilde{c}_1})} \leq (1 - p_{\tilde{c}_1}) \cdot 3 \leq 3$$

- We now assume $u_{\tilde{c}_1} > 2/3$

Proof of Lemma 10

- Let ψ be the integer such that $\frac{\psi}{\psi+1} < u_{\tilde{c}_1} \leq \frac{\psi+1}{\psi+2}$
- Such a ψ exists because

1. $u_{\tilde{c}_1} \in (0, 1)$
2. $0 \leq \frac{\psi}{\psi+1}$ for $\psi \in \mathbb{Z}^{\geqslant 0}$
3. $\lim_{\psi \to +\infty} \frac{\psi}{\psi+1} = 1$

- Then we have: $\frac{1}{\psi+1} > 1 - u_{\tilde{c}_1} \geq \frac{1}{\psi+2}$

Proof of Lemma 10

- The capacity constraint in an optimal solution of *Fractional Knapsack* is binding
- So we have that $\sum_{c_1 \in C_1} u_{c_1}(|G_{c_1}| - 1) = k - |X_1|$
- $k - |X_1|$ is an integer
- $u_{c_1}(|G_{c_1}| - 1) = (|G_{c_1}| - 1)$ or $u_{c_1}(|G_{c_1}| - 1) = 0$, for all $c_1 \in C_1 \setminus \{\tilde{c}_1\}$. In either case $u_{c_1}(|G_{c_1}| - 1)$ is an integer.
- Since the sum all integers and $u_{\tilde{c_1}}(|G_{\tilde{c}_1}| - 1)$ equals an integer, $u_{\tilde{c_1}}(|G_{\tilde{c}_1}| - 1)$ must be an integer.

Proof of Lemma 10

- $1 - u_{\tilde{c}_1} = \frac{|G_{\tilde{c}_1}|-1-u_{\tilde{c}_1}(|G_{\tilde{c}_1}|-1)}{|G_{\tilde{c}_1}|-1} < \frac{1}{\psi+1}$
- So: $1 \leq |G_{\tilde{c}_1}| - 1 - u_{\tilde{c}_1}(|G_{\tilde{c}_1}| - 1)$ is an integer
- And: $|G_{\tilde{c}_1}| - 1$ is an integer
- So $|G_{\tilde{c}_1}| - 1 > \psi + 1 \Rightarrow |G'_{\tilde{c}_1}| \geq \psi + 2$

Proof of Lemma 10

$$\begin{aligned}
\frac{|G'_{\tilde{c}_1}| - (\lceil u_{\tilde{c}_1} |G'_{\tilde{c}_1}| \rceil - 2)}{|G'_{\tilde{c}_1}|(1 - u_{\tilde{c}_1})} &\leq \frac{1 - u_{\tilde{c}_1} + 2/|G'_{\tilde{c}_1}|}{1 - u_{\tilde{c}_1}} \\
&= 1 + \frac{2}{|G'_{\tilde{c}_1}|(1 - u_{\tilde{c}_1})} \\
&\leq 1 + \frac{2(\psi + 2)}{|G'_{\tilde{c}_1}|} \\
&\leq 1 + \frac{2(\psi + 2)}{\psi + 2} \\
&\leq 3
\end{aligned}$$

Lemma 11 (LP Feasible Solution Bound)

The solution $u_{c_1} = b$ for all $c_1 \in C_1$ is a feasible solution to the LP and has value equal to:

$$b \sum_{c_1 \in C_1} \Big(r_1(c_1)\rho + s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} \big[d^{r_1}(x, \phi_1(x)) + d^{r_2}(x, \phi_2(x)) \big] \Big)$$

Proof of Lemma 11

- The value of the solution follows directly from the LP
- It is feasible since

1. $b \in [0,1]$, since $b = \frac{k-|C_1|}{|C_2|-|C_1|}$ and $|C_2| \geq k$
2. $$\sum_{c_1 \in C_1} u_{c_1}(|G_{c_1}| - 1) = b \sum_{c_1 \in C_1} (|G_{c_1} - 1) = b(\sum_{c_1 \in C_1} |G_{c_1}| - \sum_{c_1 \in C_1} 1)$$

$$= b(|C_2|-|C_1|) = b|C_2|-(1-a)|C_1| = a|C_1|+b|C_2|-|C_1| = k-|C_1| \leq k$$

Lemma 12 (Bounding the cost to the LP)

Let U be the value of the optimal solution to the LP and let Φ be the cost of our solution (C, r) constructed from (C_1, r_1) and (C_2, r_2). Then we have

$$\Phi \leq 3 \sum_{c \in C_1} \Big(r_1(c_1)\rho + 2 s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} \left[d^{r_1}(x, \phi_1(x)) + 2 d^{r_2}(x, \phi_2(x)) \right] \Big)$$

$$-3U + 3\varepsilon\Phi^*$$

Proof of Lemma 12

By Lemma 6, for some $c_1 \in C_1$ where $u_{c_1} = 0$ we have the cost associated to c_1 as:

$$
\begin{aligned}
& r'_1(c_1)\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d^{r'_1}(x, c_1) \\
\leq & r_1(c_1)\rho + 2m_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d^{r'_1}(x, \phi_1(\phi_2(x))) \\
\leq & r_1(c_1)\rho + 2s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} 2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x)) \\
= & r_1(c_1)\rho + 2s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} 2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x)) \\
& - .u_{c_1}\Big(r_1(c_1)\rho + s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x))\Big)
\end{aligned}
$$

Proof of Lemma 12

For some $c_1 \in C_1$ where $u_{c_1} = 1$ we have the cost associated to c_1 as:

$$\begin{aligned}
&\sum_{c_2 \in G_{c_1}} r_2(c_2)\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d_2^r(x, \phi(c_2)) \\
=&s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d_2^r(x, \phi(c_2)) \\
=&r_1(c_1)\rho + 2s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} 2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x)) \\
&- u_{c_1}\Big(r_1(c_1)\rho + s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x))\Big)
\end{aligned}$$

Proof of Lemma 12

$$\sum_{c \in C_1; u_{c_1}=0} \Big(r_1(x)\rho + 2m_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d^{r_1'}(x, c_1) \Big)$$

$$+ \sum_{c_1 \in C_1; u_{c_1}=1} \Big(s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} d^{r_2}(x, \phi_2(x)))$$

$$\leq \sum_{c_1 \in C_1; u_{c_1} \in \{0,1\}} \Big(r_1(c_1)\rho + 2s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} \big[d^{r_1}(x, \phi_1(x)) + 2d^{r_2}(x, \phi_2(x)) \big] \Big)$$

$$- \sum_{c_1 \in C_1; u_{c_1} \in \{0,1\}} u_{c_1} \Big(r_1(c_1)\rho + s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} \big[d^{r_1}(x, \phi_1(x)) + d^{r_2}(x, \phi_2(x)) \big] \Big)$$

Proof of Lemma 12

- Now we have to bound $\tilde{c_1}$ cost.
- For any $c_2 \in G'_{\tilde{c_1}}$ we have $r_2(c_2)\rho \leq \varepsilon\Phi^*.$ This inequality holds for all $c \notin T$, so it is only not true if $c_2 \in T$, and by definition of $G'_{\tilde{c}_1}$, $c_2 \neq \tilde{c}_1.$ However if $c_2 \in T$ then $c_2 \in C_1$ since $T \subseteq C_1.$ Then $c_2 \in G_{c_2}$ rather than $G'_{\tilde{c}_1}$ which is a contradiction.
- The radii cost of using center $\tilde{c}_1$ is $r_1(\tilde{c}_1)\rho + 2m_{\tilde{c}_1}\rho \leq r_1(\tilde{c}_1)\rho + 2\varepsilon\Phi^*$

Proof of Lemma 12

- The expected radii cost of using center $c_2 \in G'_{\tilde{c}_1}$ is:

$$p_{\tilde{c}_1} r_2(c_2) \leq u_{\tilde{c}_1} r_2(c_2) \rho$$

Proof of Lemma 12

- If we connect some point $x \in \mathcal{X}(G_{\tilde{c}_1})$ to $\tilde{c}_1$ we have a cost of $d^{r'_1}(x, \tilde{c}_1) \leq 2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x))$
- If we connect some point $x \in \mathcal{X}(G_{\tilde{c}_1})$ to some $c_2 \in G_{\tilde{c}_1}$ we have a cost of $d^{r_2}(x, \phi_2(x))$
- So the expected point cost for some point $x \in \mathcal{X}(G_{\tilde{c}_1})$ is:

$$(1 - p_{\tilde{c}_1})(2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x))) + p_{\tilde{c}_1} d^{r_2}(x, \phi_2(x))$$

Proof of Lemma 12

Using Lemma 8 we have:

$$\sum_{x\in\mathcal{X}(G_{\tilde{c}_1})} (1-p_{\tilde{c}_1})(2d^{r_2}(x,\phi_2(x)) + d^{r_1}(x,\phi_1(x))) + p_{\tilde{c}_1} d^{r_2}(x,\phi_2(x))$$

$$= \sum_{x\in\mathcal{X}(G_{\tilde{c}_1})} (1-p_{\tilde{c}_1})(d^{r_2}(x,\phi_2(x))) + (1-p_{\tilde{c}_1})d^{r_1}(x,\phi_1(x)) + d^{r_2}(x,\phi_2(x))$$

$$\leq 3 \sum_{x\in\mathcal{X}(G_{\tilde{c}_1})} (1-u_{\tilde{c}_1})(d^{r_2}(x,\phi_2(x))) + (1-u_{\tilde{c}_1})d^{r_1}(x,\phi_1(x)) + d^{r_2}(x,\phi_2(x))$$

$$= 3 \sum_{x\in\mathcal{X}(G_{\tilde{c}_1})} (2-u_{\tilde{c}_1})(d^{r_2}(x,\phi_2(x))) + (1-u_{\tilde{c}_1})d^{r_1}(x,\phi_1(x))$$

Proof of Lemma 12

If $\tilde{c}_1 \in T$ it is in $G_{\tilde{c}_1}$ but not $G'_{\tilde{c}_1}$ Then from summing the terms found above we have:

$$r_1(\tilde{c}_2)\rho + 2\varepsilon\Phi^* + \sum_{c_2 \in G'_{\tilde{c}_1}} u_{\tilde{c}_1} r_2(c_2)\rho$$

$$+3 \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2 - u_{\tilde{c}_1})(d^{r_2}(x, \phi_2(x))) + (1 - u_{\tilde{c}_1})d^{r_1}(x, \phi_1(x))$$

$$= (1 - u_{\tilde{c}_1})r_1(\tilde{c}_1) + u_{\tilde{c}_1} r_1(\tilde{c}_1)\rho + 2\varepsilon\Phi^* + \sum_{c_2 \in G'_{\tilde{c}_1}} u_{\tilde{c}_1} r_2(c_2)\rho$$

$$+3 \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2 - u_{\tilde{c}_1})(d^{r_2}(x, \phi_2(x))) + (1 - u_{\tilde{c}_1})d^{r_1}(x, \phi_1(x))$$

Proof of Lemma 12

$$= (1 - u_{\tilde{c}_1}) r_1(\tilde{c}_1)\rho + 2\varepsilon\Phi^* + \sum_{c_2 \in G_{\tilde{c}_1}} u_{\tilde{c}_1} r_2(c_2)\rho$$

$$+3 \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2 - u_{\tilde{c}_1})(d^{r_2}(x, \phi_2(x))) + (1 - u_{\tilde{c}_1}) d^{r_1}(x, \phi_1(x))$$

$$= (1 - u_{\tilde{c}_1}) r_1(\tilde{c}_1)\rho + 2\varepsilon\Phi^* + u_{\tilde{c}_1} s_{\tilde{c}_1}\rho$$

$$+3 \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2 - u_{\tilde{c}_1})(d^{r_2}(x, \phi_2(x))) + (1 - u_{\tilde{c}_1}) d^{r_1}(x, \phi_1(x))$$

Proof of Lemma 12

$$\leq (1 - u_{\tilde{c}_1}) r_1(\tilde{c}_1)\rho + 2\varepsilon\Phi^* + (2 - u_{\tilde{c}_1}) s_{\tilde{c}_1}\rho$$

$$+3 \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2 - u_{\tilde{c}_1})(d^{r_2}(x, \phi_2(x))) + (1 - u_{\tilde{c}_1}) d^{r_1}(x, \phi_1(x))$$

$$\leq 3\Big(r_1(\tilde{c}_2)\rho + 2s_{\tilde{c}_1}\rho + \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x)))\Big)$$

$$-3u_{\tilde{c}_1}\Big(r_1(\tilde{c}_1)\rho + s_{\tilde{c}_1}\rho + \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (d^{r_2}(x, \phi_2(x)) d^{r_1}(x, \phi_1(x)))\Big) + 2\varepsilon\Phi^*$$

Proof of Lemma 12

If $\tilde{c}_1 \notin T$ then $r_1(\tilde{c_1}\rho) \leq \varepsilon\Phi^*$ so we have:

$$r_1(\tilde{c}_2)\rho + 2\varepsilon\Phi^* + \sum_{c_2 \in G'_{\tilde{c}_1}} u_{\tilde{c_1}} r_2(c_2)\rho$$

$$+3 \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2 - u_{\tilde{c}_1})(d^{r_2}(x, \phi_2(x))) + (1 - u_{\tilde{c}_1})d^{r_1}(x, \phi_1(x))$$

$$\leq 3\varepsilon\Phi^* + \sum_{c_2 \in G_{\tilde{c}_1}} u_{\tilde{c_1}} r_2(c_2)\rho$$

$$+3 \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2 - u_{\tilde{c}_1})(d^{r_2}(x, \phi_2(x))) + (1 - u_{\tilde{c}_1})d^{r_1}(x, \phi_1(x))$$

Proof of Lemma 12

$$= 3\varepsilon\Phi^* + u\tilde{c}_1 s_{\tilde{c}_1}\rho + 3\sum_{x\in\mathcal{X}(G_{\tilde{c}_1})}(2-u_{\tilde{c}_1})(d^{r_2}(x,\phi_2(x))) + (1-u_{\tilde{c}_1})d^{r_1}(x,\phi_1(x))$$

$$\leq 3\varepsilon\Phi^* + (2-u\tilde{c}_1)s_{\tilde{c}_1}\rho$$

$$+3\sum_{x\in\mathcal{X}(G_{\tilde{c}_1})}(2-u_{\tilde{c}_1})(d^{r_2}(x,\phi_2(x))) + (1-u_{\tilde{c}_1})d^{r_1}(x,\phi_1(x))$$

Proof of Lemma 12

$$\leq 3\Big(r_1(\tilde{c}_2)\rho + 2s_{\tilde{c}_1}\rho + \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x)))\Big)$$

$$-3u_{\tilde{c}_1}\Big(r_1(\tilde{c}_1)\rho + s_{\tilde{c}_1}\rho + \sum_{x \in \mathcal{X}(G_{\tilde{c}_1})} (d^{r_2}(x, \phi_2(x))d^{r_1}(x, \phi_1(x)))\Big) + 3\varepsilon\Phi^*$$

Proof of Lemma 12

By summing over each $c_1 \in C_1; u_{\tilde{c_1} \in \{0,1\}}$ and adding the cost associated with $\tilde{c}_1$ we obtain:

$$\Phi \leq 3 \sum_{c \in C_1} \Big(r_1(c_1)\rho + 2 s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} \left[d^{r_1}(x, \phi_1(x)) + 2 d^{r_2}(x, \phi_2(x)) \right] \Big)$$

$$-3U + 3\varepsilon\Phi^*$$

Proof of Main Claim

- By Lemma 10 we have:

$$\Phi \leq 3 \sum_{c \in C_1} (r_1(c_1)\rho + 2s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} (2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x))))$$

$$-3U + 3\varepsilon\Phi^*$$

- Then substituting in Lemma 9 we have:

$$3 \sum_{c \in C_1} (r_1(c_1)\rho + 2s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} (2d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x))))$$

$$-3b \sum_{c \in C_1} (r_1(c_1)\rho + s_{c_1}\rho + \sum_{x \in \mathcal{X}(G_{c_1})} (d^{r_2}(x, \phi_2(x)) + d^{r_1}(x, \phi_1(x)))) + 3\varepsilon\Phi^*$$

Proof of Main Claim

- We know that $a + b = 1$:

$$\Phi \leq 3 \sum_{c_1 \in C_1} \Big(a r_1(c_1)\rho + (1+a) s_{c_1} \rho$$

$$+ \sum_{x \in \mathcal{X}(G_{c_1})} (a d^{r_1}(x, \phi_1(x)) + (1+a) d^{r_2}(x, \phi_2(x))) \Big) + 3\varepsilon\Phi^*$$

$$= 3a\Phi_1 + 3(1+a)\Phi_2 + 3\varepsilon\Phi^*$$

Proof of Main Claim

- Since $a \leq 1/4$, $a \leq b$
- And $a\Phi_1 + b\Phi_2 \leq (3+\varepsilon)\Phi^*$
- So we have:

$$\Phi \leq 3(a\Phi_1 + (1+a)\Phi_2) + 3\varepsilon\Phi^*$$

$$= 3(a\Phi_1 + b\Phi_2 + (1+a-b)\Phi_2) + 3\varepsilon\Phi^*$$

$$\leq 3(3+\varepsilon)\Phi^* + 3(1+a-b)\Phi_2 + 3\varepsilon\Phi^*$$

$$= (9+6\varepsilon)\Phi^* + 6a\Phi_2$$

Proof of Main Claim

- $\Phi_2 \leq (3+\varepsilon)\Phi^* \leq \Phi_1$, since $a+b=1$, $a\Phi_1 + b\Phi_2 \leq (3+\varepsilon)\Phi^*$, and $\Phi_1 \geq \Phi_2$. So

$$\Phi \leq (9+6\varepsilon)\Phi^* + 6a\Phi_2$$

$$\leq (9+6\varepsilon)\Phi^* + 6/4\Phi_2$$

$$\leq (9+6\varepsilon)\Phi^* + 6/4(3+\varepsilon)\Phi^*$$

$$\leq (13.5+7.5\varepsilon)\Phi^*$$