Fair Colorful k-Center Clustering

Xinrui Jia, Kshiteej Sheth,
Ola Svensson

EPFL

FNSNF

A Technique for Obtaining True Approximations for k-Center with Covering Constraints

Georg Anegg, Haris Angelidakis,
Adam Kurpisz, Rico Zenklusen

ETH zürich

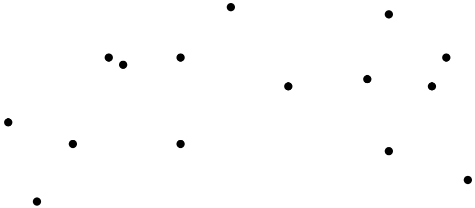
European Research Council
Established by the European Commission

FNSNF

IPCO XXI
2020-06-09

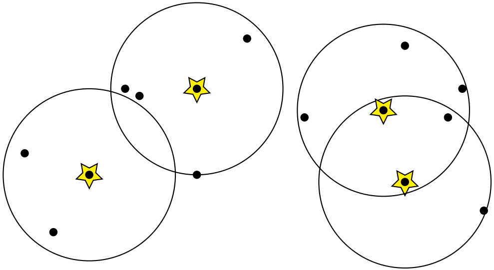
Introduction

k-Center Problem:



Introduction

k-Center Problem:

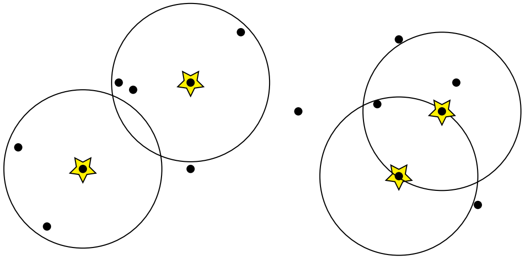


A finite metric space P and integer k, find $C \subseteq P$, $|C| = k$, minimize the maximum distance of $p \in P$ to C.

Introduction

k-Center Problem:

$m = 11$

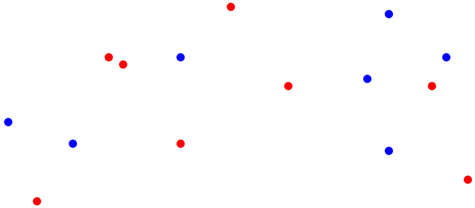


A finite metric space P and integer k, find $C \subseteq P$, $|C| = k$, minimize the maximum distance of $p \in P$ to C.

- Version with outliers: only m of the points need to be covered

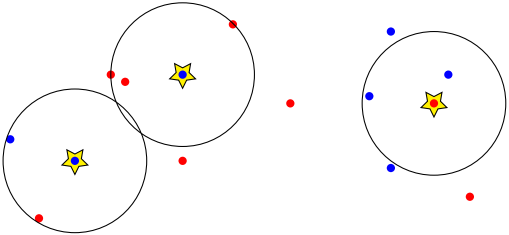
Introduction

Colorful k-Center Problem:



Colorful k-Center Problem:

$$r = 5$$
$$b = 5$$
$$k = 3$$

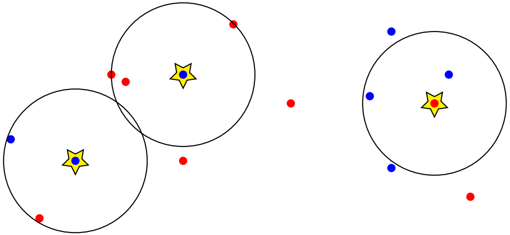


- Points are colored red, R, or blue, B
- Coverage requirements r and b for each color

Introduction

Colorful k-Center Problem:

$r = 5$
$b = 5$
$k = 3$



- Points are colored red, R, or blue, B
- Coverage requirements r and b for each color
- Generalizes to more colors
- Better than 2-approx is NP-hard

Motivation

1) **Fairness**

- Each type of element has some service guarantee

Motivation

1) Fairness

- Each type of element has some service guarantee

2) Clustering

- Generalizes k-center with outliers problem

Motivation

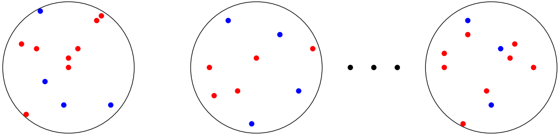
1) Fairness

- Each type of element has some service guarantee

2) Clustering

- Generalizes *k*-center with outliers problem
- A novel algorithmic consideration: involves *k*-subset sum problem

$$A = \Sigma_{i=1}^{n} a_i,\ r = k \cdot A + A/2,\ b = k \cdot A - A/2$$



$$r_i = A + a_i,\ b_i = A - a_i$$

Past Work and Results

Problem introduced in [1]

- 2-pseudo approximation opening $(k+1)$ centers
- Constant-factor for 2 colors in $\mathbb{R}^2$

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Past Work and Results

Problem introduced in [1]

- 2-pseudo approximation opening $(k+1)$ centers
- Constant-factor for 2 colors in $\mathbb{R}^2$

Our Results

- 3-approximation for γ number of colors in $|P|^{O(\gamma^2)}$
- Unbounded integrality gap for a linear number of rounds of the Lasserre/Sum-of-Squares hierarchy

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Past Work and Results

Problem introduced in [1]

- 2-pseudo approximation opening $(k+1)$ centers
- Constant-factor for 2 colors in $\mathbb{R}^2$

Our Results

- 3-approximation for γ number of colors in $|P|^{O(\gamma^2)}$
- Unbounded integrality gap for a linear number of rounds of the Lasserre/Sum-of-Squares hierarchy

Second Half

- 4-approximation in time $|P|^{O(\gamma)}$
- Assuming $P \neq NP$, inapproximable for unbounded γ
- Assuming ETH, inapproximable for $\gamma = \omega(\log |P|)$

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Challenges

Standard linear programming relaxation has unbounded integrality gap after a linear number of rounds of the Lasserre/Sum-of-Squares hierarchy [2].

$$\sum_{i=1}^{n} 2y_i = k, \quad k \in \mathbb{Z} \text{ odd}$$

$$y_i \in \{0, 1\}$$

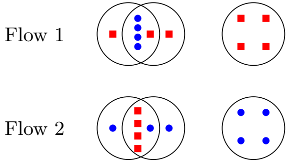
[2] Grigoriev, D.: Complexity of Positivstellensatz Proofs for the Knapsack. Comput. Complex. 10(2), 139–154 (2001).

Challenges

Standard linear programming relaxation has unbounded integrality gap after a linear number of rounds of the Lasserre/Sum-of-Squares hierarchy [2].

$$\sum_{i=1}^{n} 2y_i = k, \quad k \in \mathbb{Z} \text{ odd}$$

$$y_i \in \{0,1\}$$



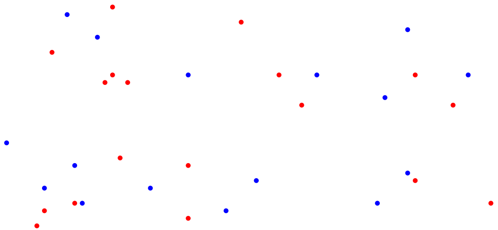
Adding flow-based inequalities to solve a subset-sum problem also does not work.

[2] Grigoriev, D.: Complexity of Positivstellensatz Proofs for the Knapsack. Comput. Complex. 10(2), 139–154 (2001).

Main Ideas

$(k+1)$-pseudo approx. algorithm[1]:

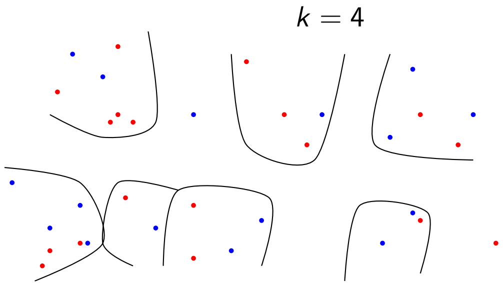
$k = 4$



[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Main Ideas

$(k+1)$-pseudo approx. algorithm[1]:

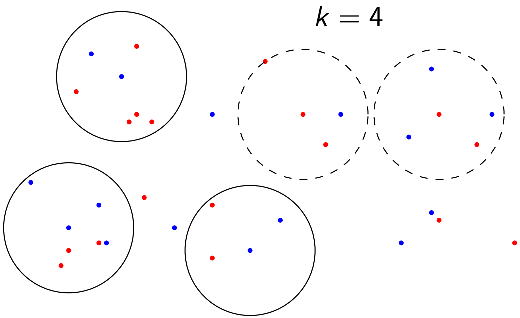


- Use feasible point to natural LP relaxation to get clustering

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Main Ideas

$(k+1)$-pseudo approx. algorithm[1]:

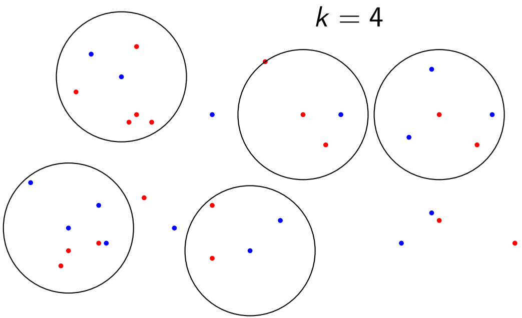


- Use feasible point to natural LP relaxation to get clustering
- Extreme point to simplified LP has at most 2 strictly fractional variables

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Main Ideas

$(k+1)$-pseudo approx. algorithm[1]:

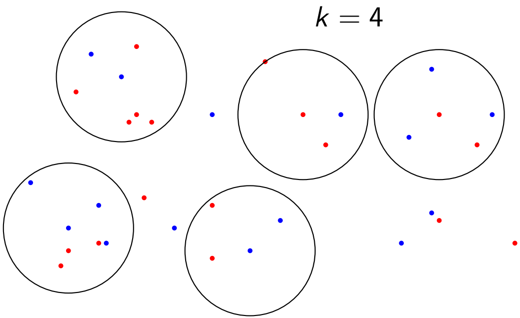


- Use feasible point to natural LP relaxation to get clustering
- Extreme point to simplified LP has at most 2 strictly fractional variables
- Open both to get $k+1$ centers

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Main Ideas

$(k+1)$-pseudo approx. algorithm[1]:

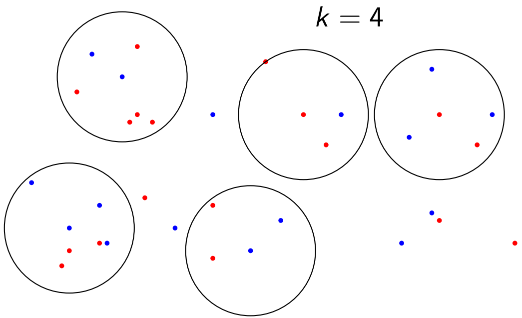


- Use feasible point to natural LP relaxation to get clustering
- Extreme point to simplified LP has at most 2 strictly fractional variables
- Open both to get $k+1$ centers
- Assume OPT radius is 1

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Main Ideas

$(k+1)$-pseudo approx. algorithm[1]:



- Use feasible point to natural LP relaxation to get clustering
- Extreme point to simplified LP has at most 2 strictly fractional variables
- Open both to get $k+1$ centers
- Assume OPT radius is 1

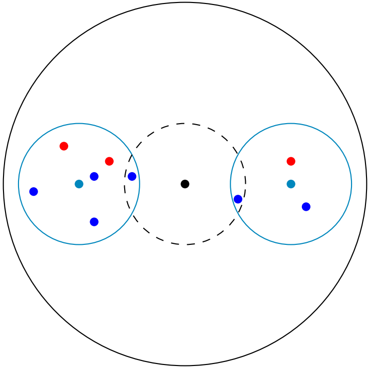
Idea: Expand balls to cover enough red points, to close fractional center with fewer blue points.

[1] S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan.: *A Constant Approximation for Colorful k-Center.* In: ESA, pp. 12:1–12:14 (2019).

Main Ideas

Intuition:

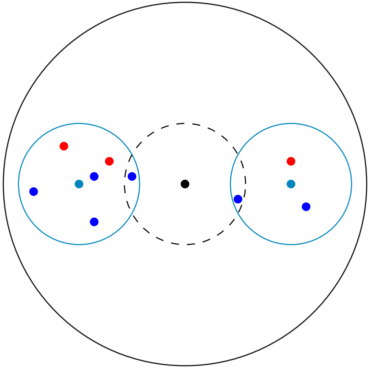
- OPT solution not *well-separated*: easy



Main Ideas

Intuition:

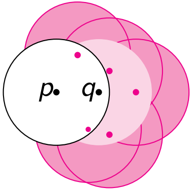
- **OPT** solution not *well-separated*: easy



- **OPT** solution well-separated: subset-sum problem

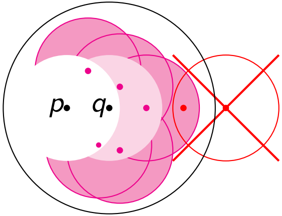
Well-Separated Instances

Phase I: Guess points in an OPT solution, that maximizes the number of red points in a *flower* structure.



Well-Separated Instances

Phase I: Guess points in an OPT solution, that maximizes the number of red points in a *flower* structure.

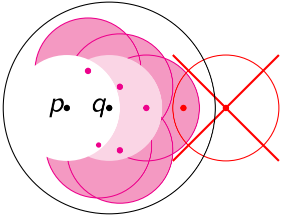


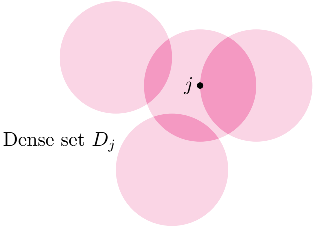
Key Point:

- Separatedness ensures that no ball of OPT intersects expanded region

Well-Separated Instances

Phase I: Guess points in an OPT solution, that maximizes the number of red points in a *flower* structure.





Key Point:

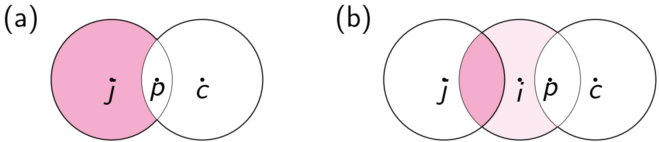
- Separatedness ensures that no ball of OPT intersects expanded region

Phase II:

- Remove subsets of remaining points that are *dense*
- Run pseudo-approx. on remaining *sparse* points

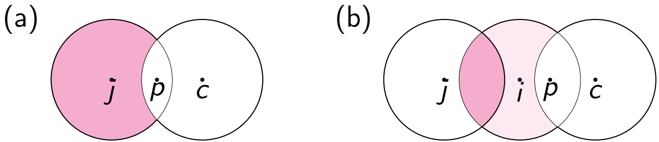
Phase II

Required: Removed sets shouldn't "interact" with balls in OPT



Phase II

Required: Removed sets shouldn't "interact" with balls in OPT



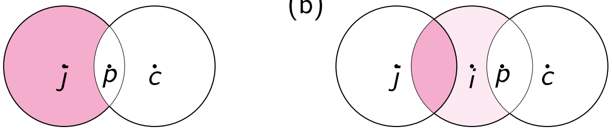
Dynamic program on sets D_j:

- Guess number of red and blue points covered by D_j's, and verify with dynamic program

Phase II

Required: Removed sets shouldn't "interact" with balls in OPT

(a) (b)

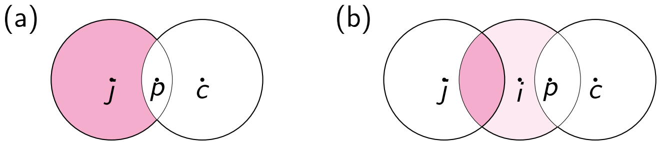


Dynamic program on sets D_j:

- Guess number of red and blue points covered by D_j's, and verify with dynamic program
- Can run pseudo-approx on remaining points, and close center with fewer blue points

Phase II

Required: Removed sets shouldn't "interact" with balls in OPT



Dynamic program on sets D_j:

- Guess number of red and blue points covered by D_j's, and verify with dynamic program
- Can run pseudo-approx on remaining points, and close center with fewer blue points

$\Rightarrow$ covered all required points to get poly-time 2-approximation for well-separated instances!

Conclusion Part I

Result:

- Easily generalizes to γ color classes

Conclusion Part I

Result:

- Easily generalizes to γ color classes
- 3-approximation with running time having $|P|^{O(\gamma^2)}$ term
- 2-approximation for well-separated instances

Conclusion Part I

Result:

- Easily generalizes to γ color classes
- 3-approximation with running time having $|P|^{O(\gamma^2)}$ term
- 2-approximation for well-separated instances

Open Questions:

- 2-approximation for all instances?

Conclusion Part I

Result:

- Easily generalizes to γ color classes
- 3-approximation with running time having $|P|^{O(\gamma^2)}$ term
- 2-approximation for well-separated instances

Open Questions:

- 2-approximation for all instances?
- Lasserre hierarchy with flow-based inequalities?

Conclusion Part I

Result:

- Easily generalizes to γ color classes
- 3-approximation with running time having $|P|^{O(\gamma^2)}$ term
- 2-approximation for well-separated instances

Open Questions:

- 2-approximation for all instances?
- Lasserre hierarchy with flow-based inequalities?

And now...