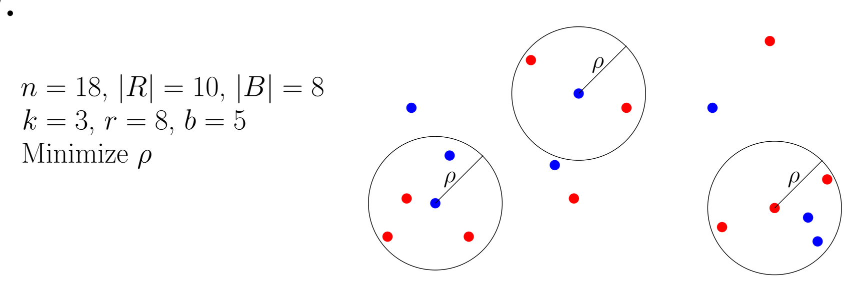
Fair Colorful k-Center Clustering

Xinrui Jia, Kshiteej Sheth, Ola Svensson

1) Problem Statement

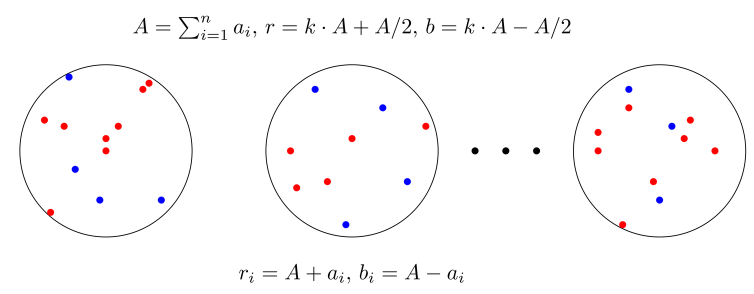
- k-Center: Given n points, choose k of them to minimize largest distance from a point to a center.
- Coverage requirement p: can choose $n - p$ points to omit.
- With colors: each color has a coverage requirement.
- Example: Red points and blue points with coverage requirements r and b.



- Our algorithm easily generalizes to more color classes

2) Motivation

- Fairness
  - ◇In k-center without colors, every member of a certain group may be treated as an outlier
- Algorithmic challenges
  - ◇Subset-sum problem
  - ◇Clustering



3) Previous Results

	Best approx	Tight?	Authors
k-center	2	Yes	T. F. Gonzalez
k-center with outliers	2	Yes	D. Chakrabarty, P. Goyal R. Krishnaswamy

- Colorful k-center: S. Bandyapadhyay, T. Inamdar, S. Pai, K. Varadarajan:

Natural LP Relaxation + Greedy Clustering = 2-Approx $k+1$ centers

4) 3-Approximation Outline

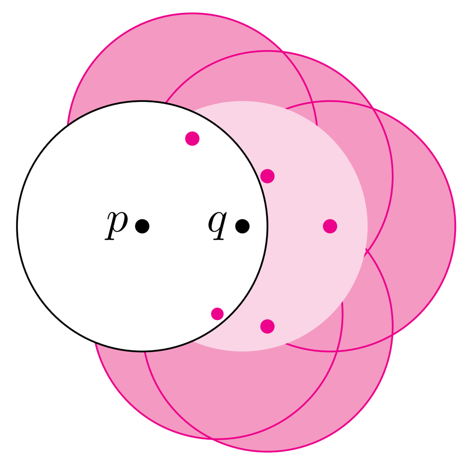
Close one of the extra centers

- Phase I:
  - ◇Guess some centers of optimal solution to maximize number of red points in special *Gain* region
- Phase II:
  - ◇Remove *Dense* sets and solve subset-sum to add centers to solution
  - ◇Run 2-Approx (k+1) centers and close one center

Assume no ball of radius 3·(OPT radius) covers 2 optimal clusters

5) Phase I

- Optimal radius from feasibility LP relaxation.
- *Gain*(p, q): *flower*$(q) \setminus \mathcal{B}(p)$ that maximizes number of red points



Guess c_1, c_2, c_3 optimal centers and find $q_i \in \mathcal{B}(c_i)$ such that number of red points in $\mathbf{Gain}(c_i, q_i) \cap P_i$ is maximized over all possible c_i, where

$$P_1 = P$$
$$P_i = P_{i-1} \setminus flower(q_i) \quad \text{for } 2 \le i \le 4.$$

- Define $\mathbf{Guess} := \mathcal{B}(c_1) \cup \mathcal{B}(c_2) \cup \mathcal{B}(c_3)$
- Define $\tau = |\mathbf{Gain}(c_3, q_3) \cap P_3|$

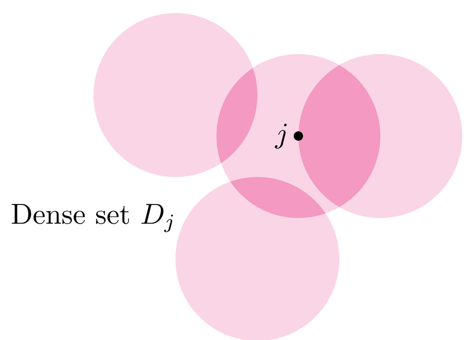
6) Phase II

- Point j is **dense** if $\mathcal{B}(j)$ contains strictly more than $2 \cdot \tau$ red points of P_s
- Define: $I_j \subseteq P_s$ contain those points $i \in P_s$ such that $\mathcal{B}(i) \cap \mathcal{B}(j)$ contains strictly more than τ red points of P_s

Initially, let $I = \emptyset$ and $P_s = P_4$. While there is a dense point $j \in P_s$:

- Add I_j to I, update P_s by removing $D_j = \cup_{i \in I_j} \mathcal{B}(i) \cap P_s$.

- Let $P_d = \cup_j D_j$
- Use flow network/dynamic programming to find dense points that belong in solution set

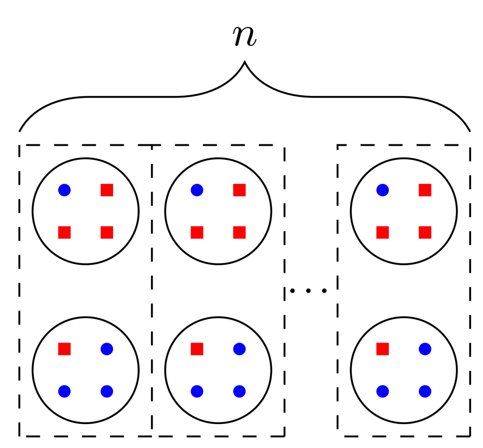


- 2-Approx (k+1) centers on P_s to complete solution
- Can remove a center, since number of red points in all flowers from P_s is bounded by $3 \cdot \tau$, which **Guess** makes up for

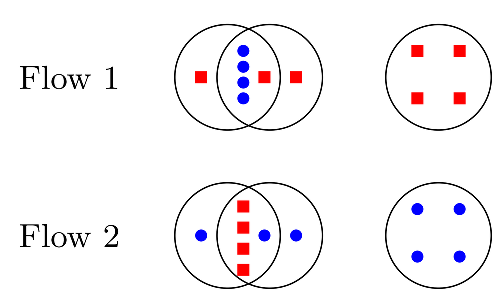
7) LP Integrality Gaps

Sum-of-Squares

- Linear rounds of SoS are required to close integrality gap for following example, $k = n$, $r = b = 2n$, n odd
- Need radius to cover 2 balls



Knapsack Constraints



- Let $k = 3$, $r = b = 8$ and add flow constraints to LP to model knapsack problem
- Fractional assignment of $1/2$ to each ball satisfies flow constraints

8) Open Questions

- Tight 2-approximation?
- Integrality gap example for SoS doesn't fool knapsack, and vice versa
  - ◇SoS hierarchy on LP with added flow constraints?