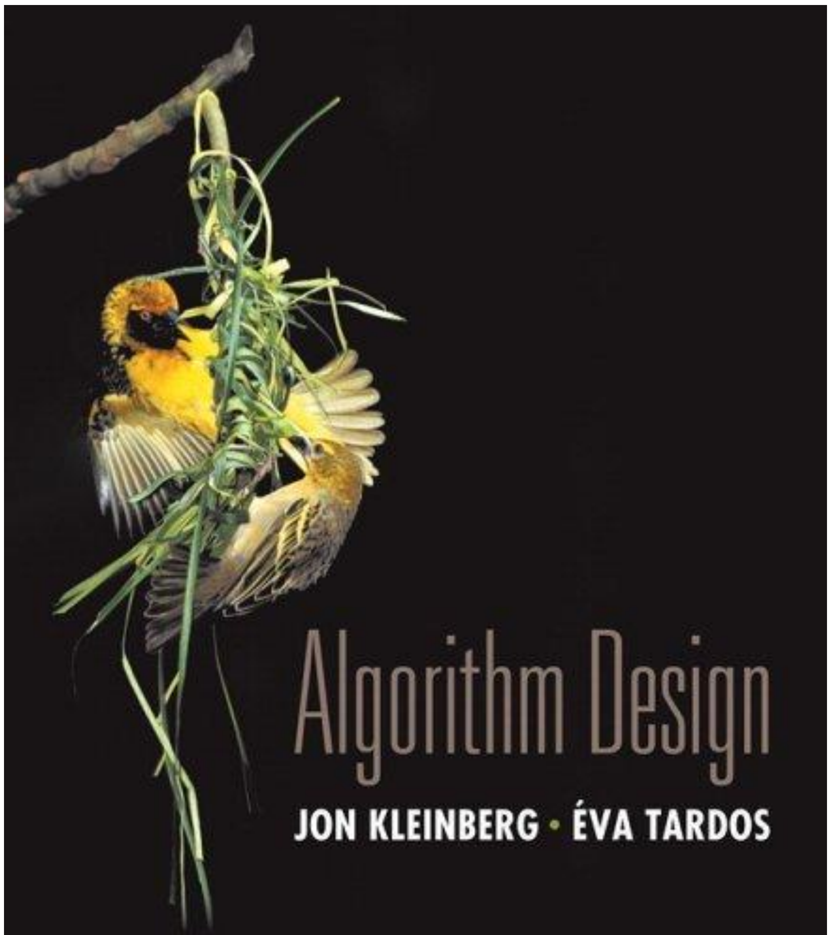
CSE 6140/ CX 4140
Computational Science and Engineering
ALGORITHMS

Coping with NP-completeness - 6
Approximation Algorithms

Instructor: Xiuwei Zhang
Assistant Professor
School of Computational Science and Engineering

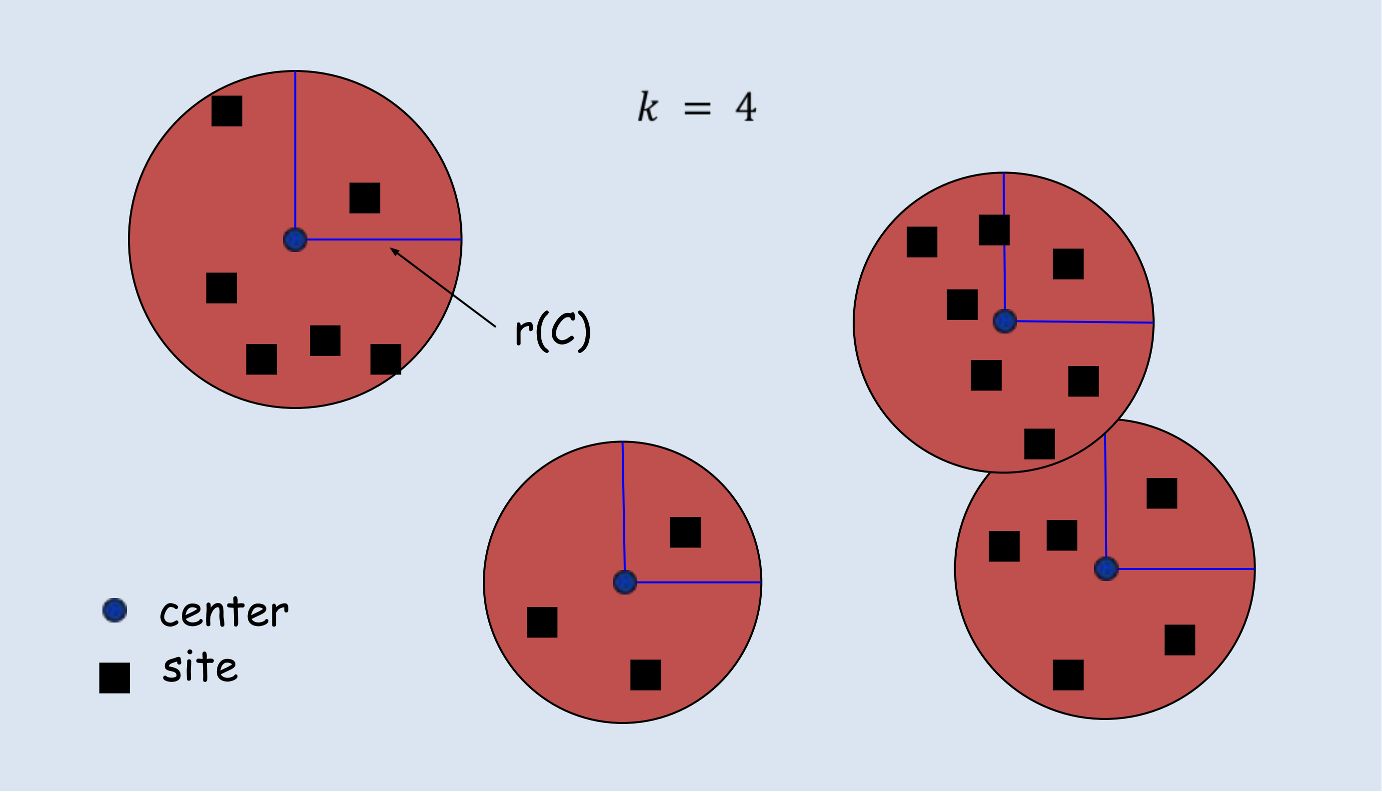
Based on slides by Prof. Ümit V. Çatalyürek and Bistra Dilkina

KT 11.2 Clustering



Center Selection Problem

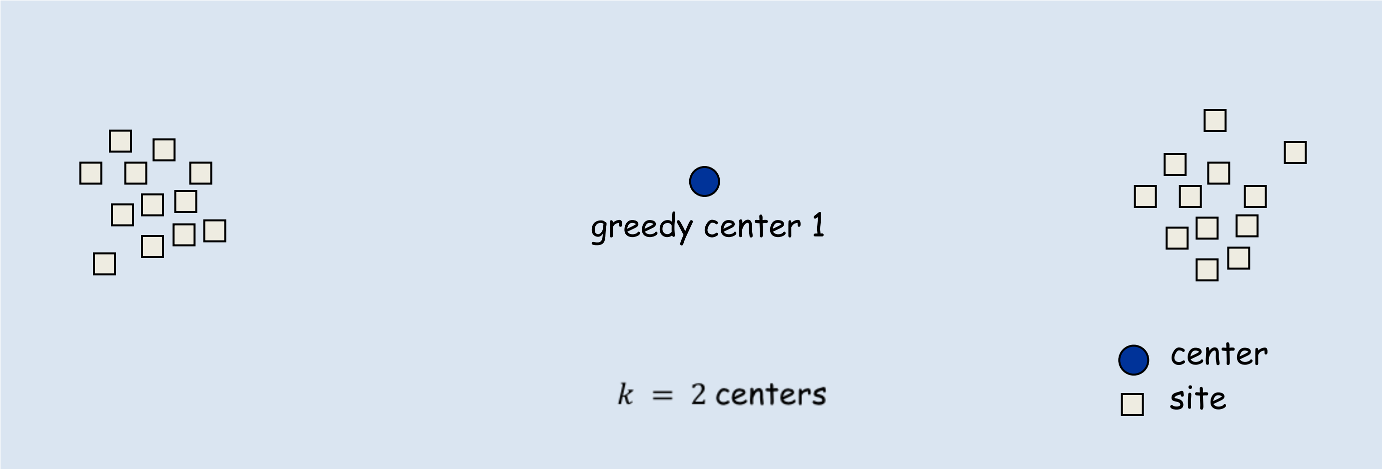
- Input. Set of n sites $s_1, \ldots, s_n$ and integer $k > 0$.
- Center selection problem. Select k centers C so that maximum distance from a site to nearest center is minimized.
- Application: where to put the branch offices w.r.t. clients?



Greedy Algorithm: A False Start

- Greedy algorithm. Put the first center at the best possible location for a single center, and then keep adding centers so as to reduce the covering radius each time by as much as possible.

- Remark: arbitrarily bad!



Center Selection Problem

- Input. Set of n sites $s_1, \ldots, s_n$ and integer $k > 0$.
- Center selection problem. Select k centers C so that maximum distance from a site to nearest center is minimized.

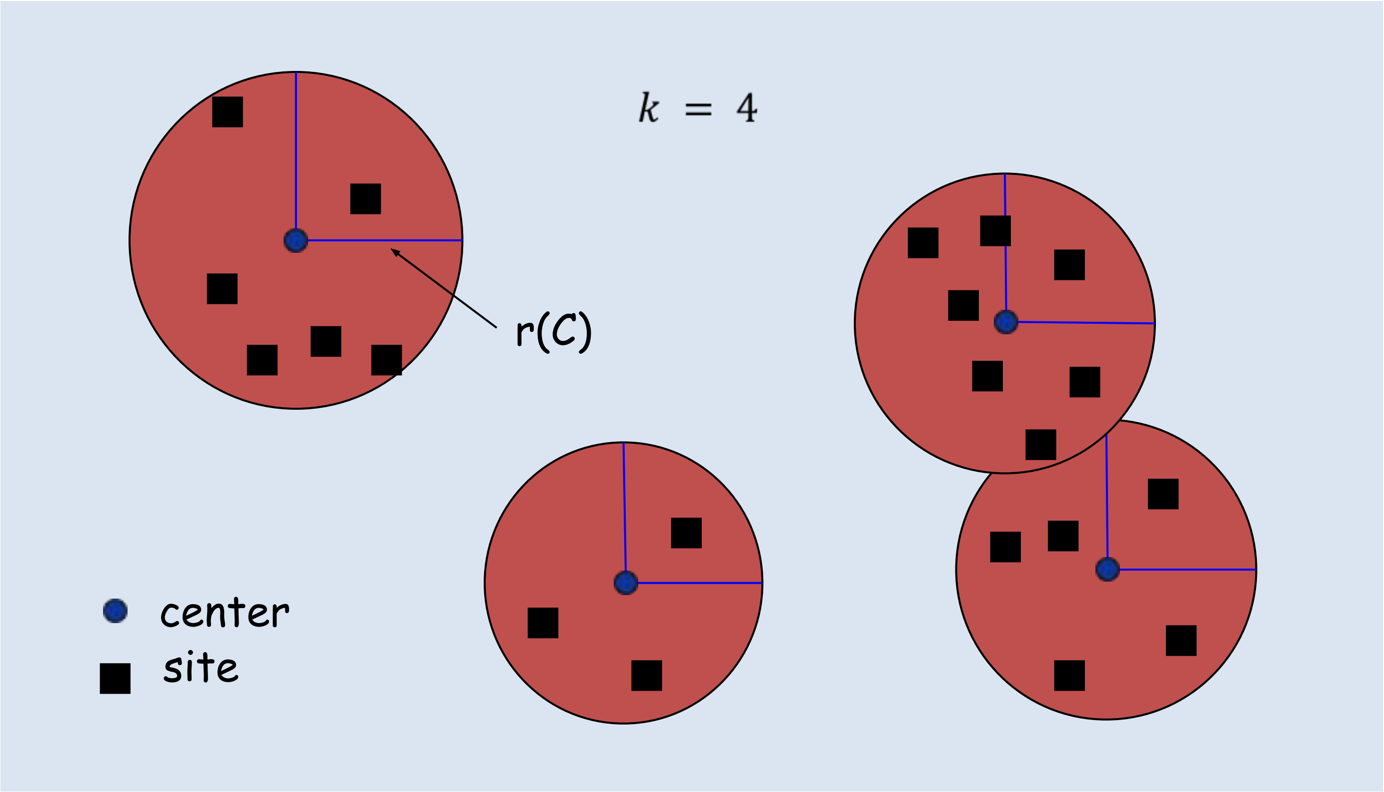
- Notation.
  - dist(x, y) = distance between x and y.
  - $\text{dist}(s_i, C) = \min_{c \in C} \text{dist}(s_i, c)$ = distance from s_i to closest center.
  - $r(C) = \max_i \text{dist}(s_i, C)$ = smallest covering radius.
- Goal. Find set of centers C that minimizes r(C), subject to |C| = k.
- Distance function properties.
  - dist(x, x) = 0 (identity)
  - dist(x, y) = dist(y, x) (symmetry)
  - dist(x, y) ≤ dist(x, z) + dist(z, y) (triangle inequality)

Also known as Metric Facility Location problem

Center Selection Example

- **Ex**: each site is a point in the plane, a center can be any point in the plane, dist(x, y) = Euclidean distance.

- **Remark**: search can be infinite!



Center Selection: Greedy Algorithm

- Greedy algorithm. Repeatedly choose the next center to be the site farthest from any existing center.

```
Greedy-Center-Selection(k, n, s1,s2,…,sn) {

   C = φ
   repeat k times {
      Select a site si with maximum dist(si, C)
      Add si to C
   }
   return C
}
```

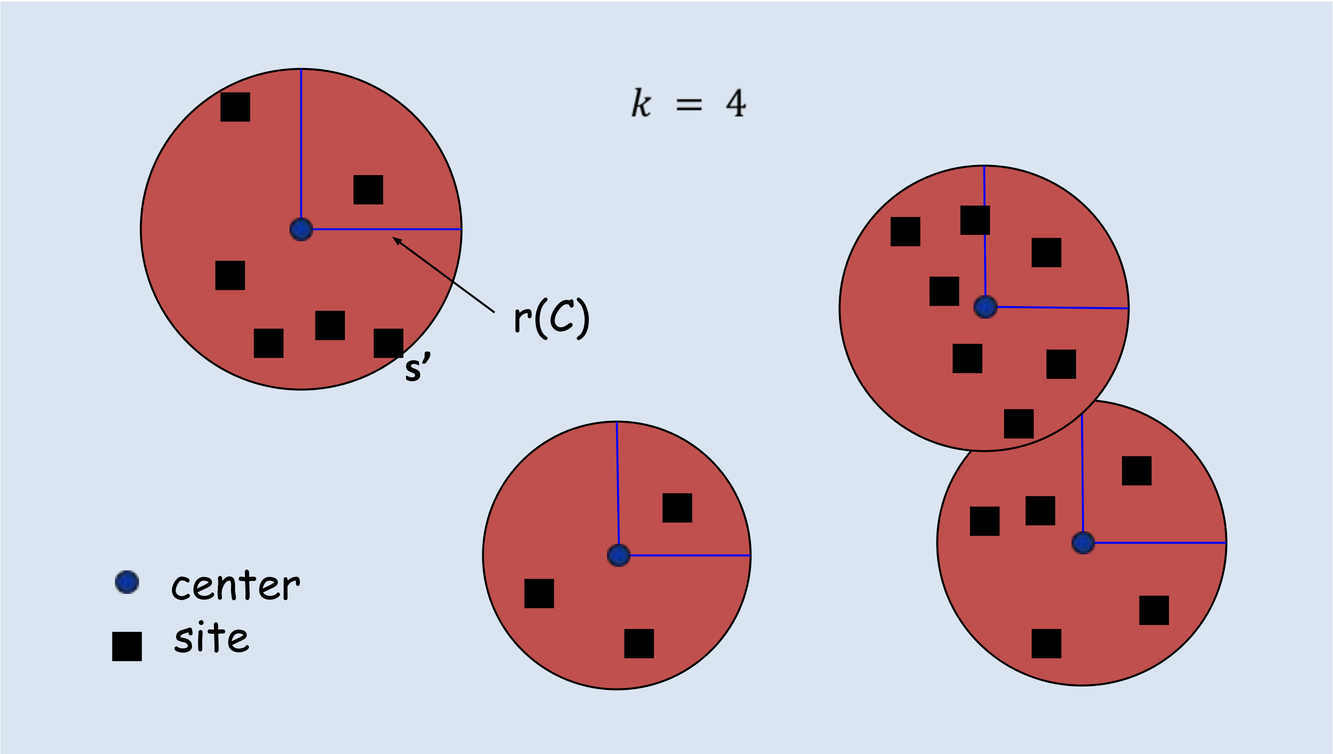
site farthest from any center

Center Selection: Greedy Algorithm

Observation. Upon termination all centers in C are pairwise at least r(C) apart.

Pf.

- Remember that $r(C) = \max_i \text{dist}(s_i, C)$
- Let us call the point that achieves this maximum radius s'
  - clearly s' is not one of the chosen centers, k < n
  - s' is at least r(C) away from any chosen center



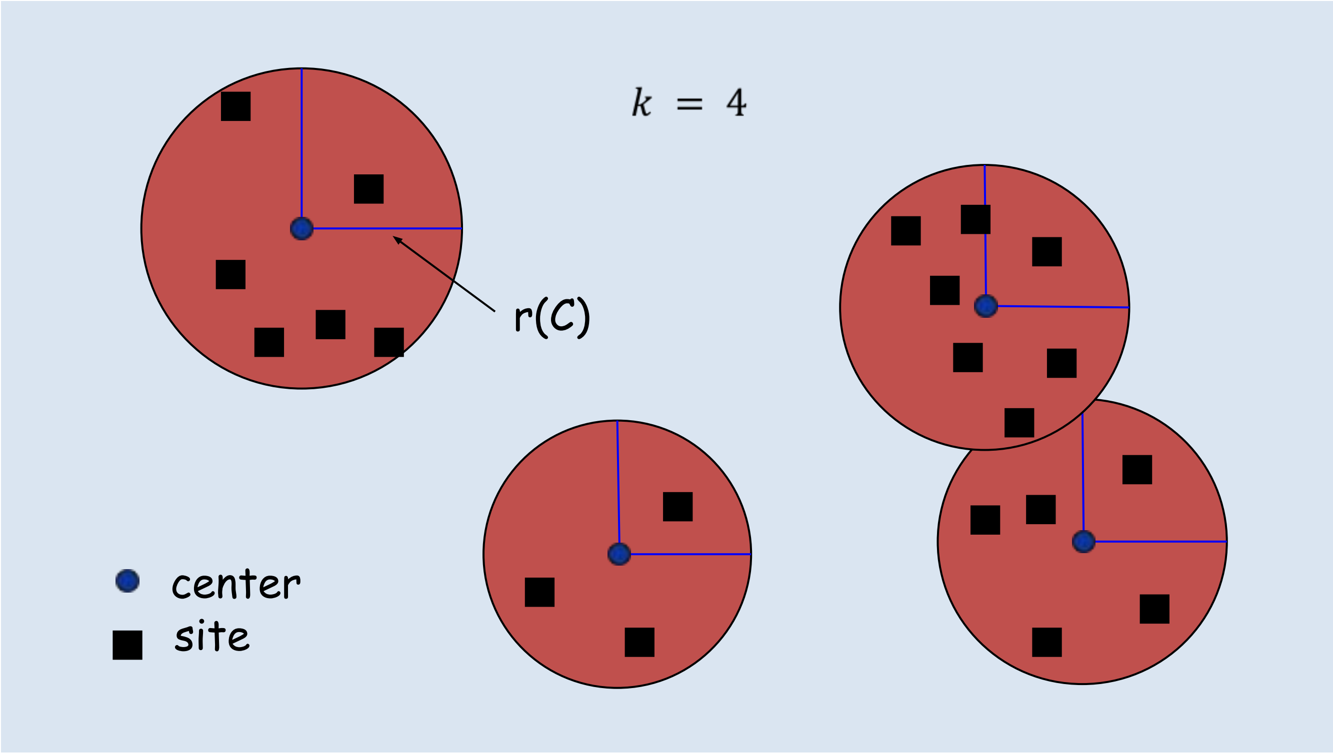
Center Selection: Greedy Algorithm

Observation. Upon termination all centers in C are pairwise at least r(C) apart.

Pf. (continued...)

- **Assume** there are two centers c_i and c_j at distance < r(C) (let i < j)
  - when we chose j, its distance to the current C was < r(C) due to c_i
  - s' was an option to choose as center and it was at least r(C) away from all centers in current C ⇒ s' is further than j from the current centers

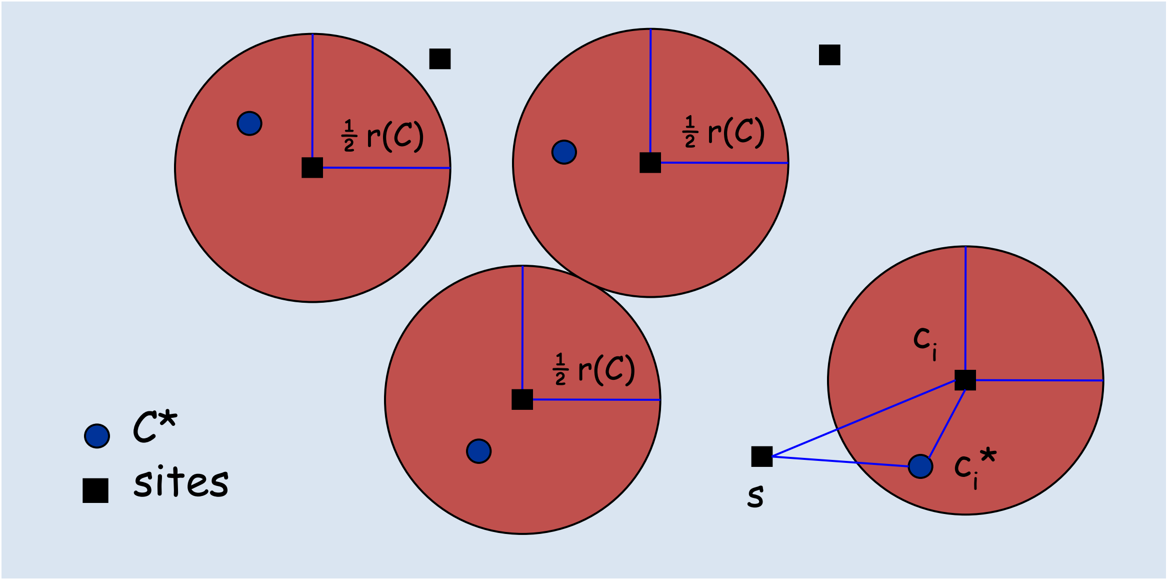
- By construction of algorithm, we always choose the furthest point from the current C
  ⇒ Contradict with that s' is not a center



Center Selection: Analysis of Greedy Algorithm

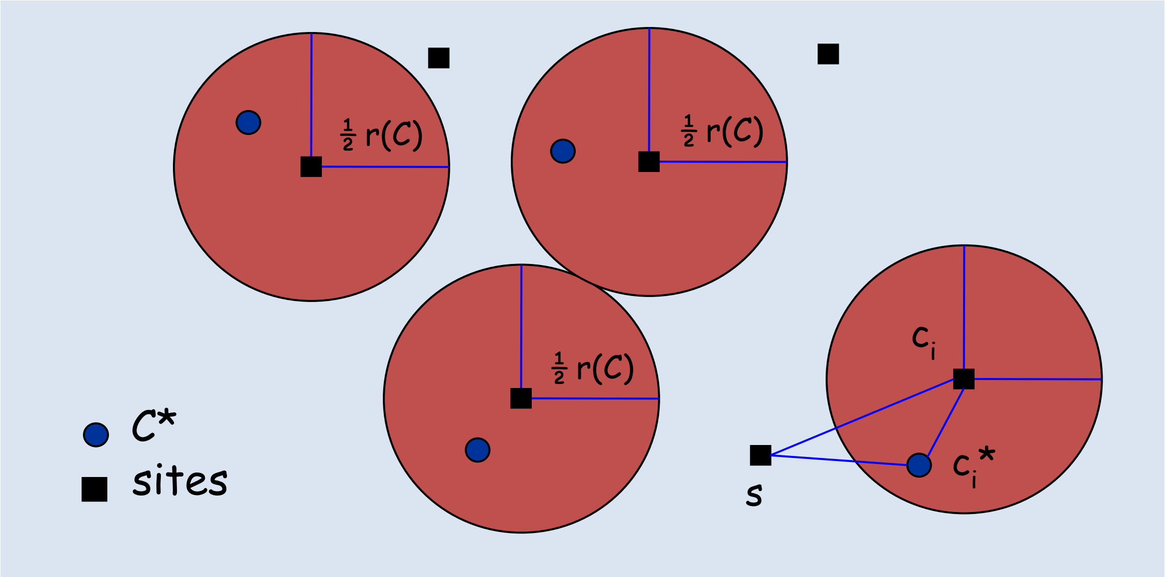
- **Theorem.** Let C* be an optimal set of centers. Then r(C) ≤ 2r(C*).
  greedy optimum
- **Pf.** (by contradiction)
- Assume r(C) > 2r(C*), i.e., r(C*) < ½ r(C).
  - For each center c_i in C, consider ball of radius ½ r(C) around it. (no 2 balls overlaps) (by observation: every two centers are at least r(C) apart)
  - dist(c_i, C*) ≤ r(C*) < ½ r(C) , **so <u>at least</u> one c_i* in each ball in C**

Greedy center is site By our assumption



Center Selection: Analysis of Greedy Algorithm

- **Theorem**. Let C^* be an optimal set of centers. Then $r(C) \le 2r(C^*)$.
- **Pf**. (by contradiction) Assume $r(C) > 2r(C^*)$, i.e. $r(C^*) < ½\, r(C)$.
  - at least one c_i^* in each ball in C
  - Every pair of c_i's in C are at least r(C) apart (by alg.), so each ball around a c_i in C does not intersect with any other ball
  - Each ball has at least one c_i^* and $|C|=|C^*|=k$, so at most one c_i^* in each ball
  - Therefore exactly one c_i^* in each ball



Center Selection

Greedy algorithm:
repeatedly choose the next center to be the site farthest from any existing center.

Notations:

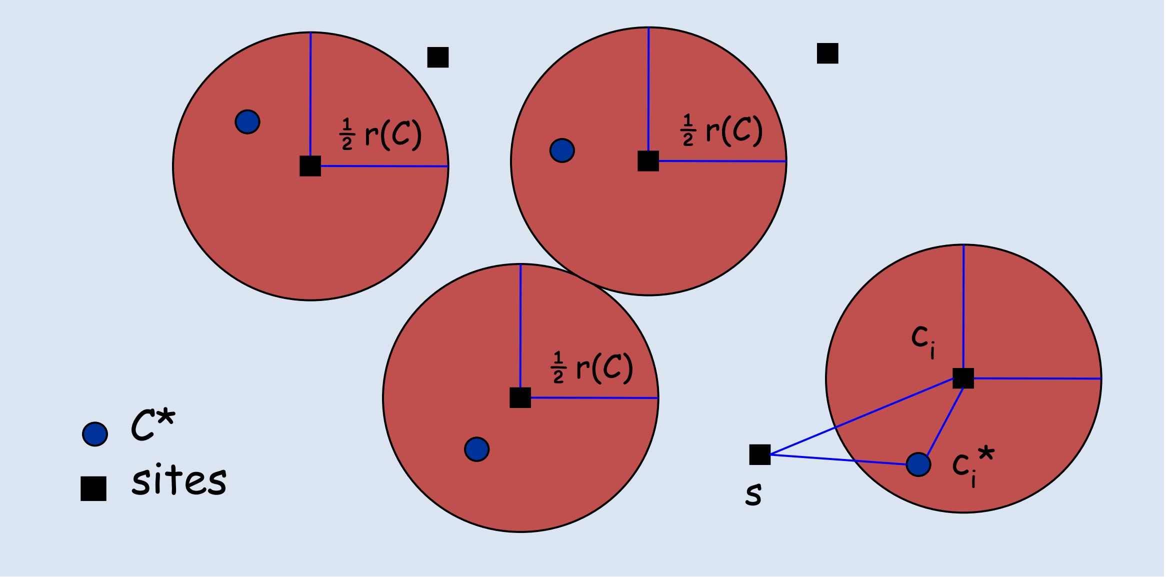
C: the centers selected by Greedy

C*: the centers selected by Optimal

$r(C) = \max_i \text{dist}(s_i, C)$ (we want to minimize this value)

Shown earlier:

➔ There is exactly one c_i^* in each ball



Center Selection: Analysis of Greedy Algorithm

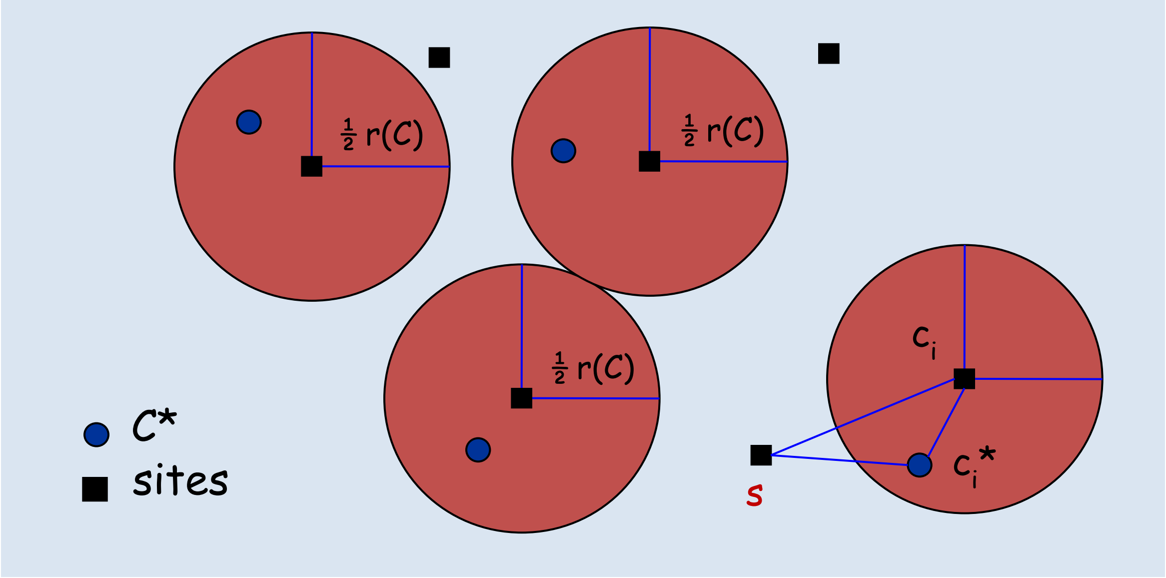
- **Theorem**. Let C^* be an optimal set of centers. Then $r(C) \le 2r(C^*)$.
- **Pf.** (by contradiction) Assume $r(C) > 2r(C^*)$, i.e. $r(C^*) < ½\, r(C)$.
  - exactly one c_k^* in each ball in C ; let c_k be the site paired with c_k^*
  - Consider **any** site s and its closest center c_i^* in C^*:
  - $\mathrm{dist}(s, C) \le \mathrm{dist}(s, c_i) \le \mathrm{dist}(s, c_i^*) + \mathrm{dist}(c_i^*, c_i) \le 2r(C^*)$.

min across all c_i

Δ-inequality

$\le r(C^*)$ since c_i^* is closest center to both s and c_i

  - true for any site s including the one that has dist(s,C)=r(C)
  - Thus $r(C) \le 2r(C^*)$, this is a contradiction with our assumption ▪
  - The assumption is wrong
  - $r(C) \le 2r(C^*)$ ▪



Center Selection

- **Theorem**. Let C* be an optimal set of centers. Then $r(C) \leq 2r(C^*)$.

- **Theorem**. Greedy algorithm is a 2-approximation for center selection problem.

- Remark. Greedy algorithm always places centers at sites, but is still within a factor of 2 of best solution that is allowed to place centers anywhere.

e.g., points in the plane