Introduction to
Approximation Algorithms

Mong-Jen Kao (高孟駿)

Friday 13:20 – 15:10

Outline

- The k-Center Problem
  - 2-approximation by the Parametric Search technique
    - Inherent reduction to dominating set problem
    - Lower-bounding the size of dominating sets
  - 2-approximation by simple Iterative Refining
  - Inapproximability of $2 - \epsilon$
- The weighted k-Center Problem and a 3-approximation

The k-Center Problem

The k-Dominating Set Problem

- The k-dominating set problem is the decision version of the unweighted dominating set problem in graphs.
  - Given a graph $G = (V, E)$ and $k \in \mathbb{N}$, determine if there exists a vertex subset of size k that dominates (covers) all the vertices in V.

Decision Problem
(Yes / No)

- The vertices can also be weighted, and the goal is then to decide the existence of a dominating set with weight at most W.

The k-Center Problem

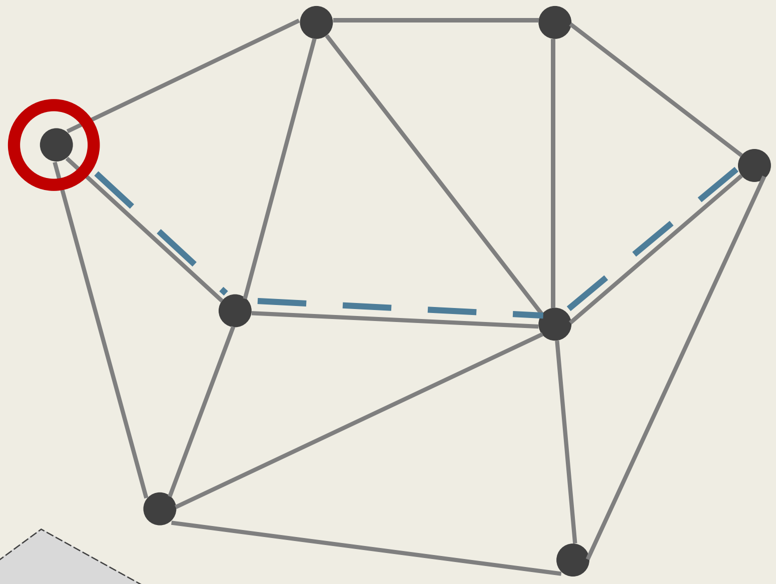
- The k-Center problem is a relaxation of the k-dominating set problem on the dominating (covering) distance.

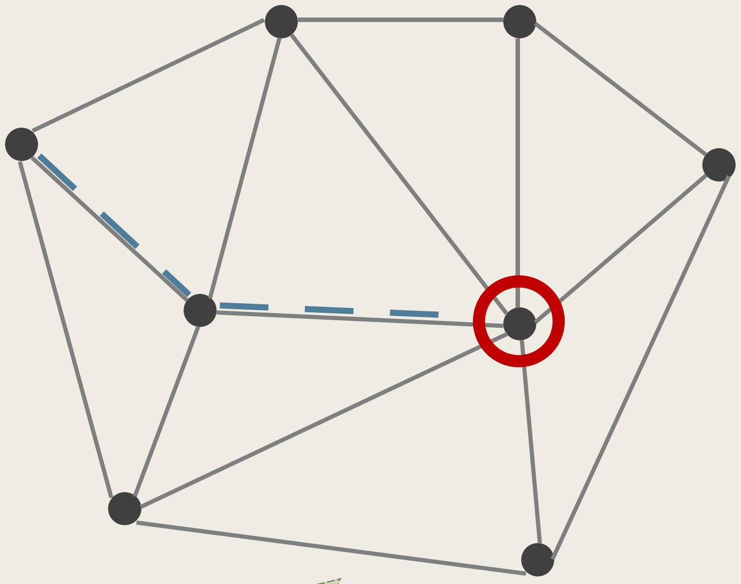
  It asks:

  - What is the *minimum covering radius* it requires, if we want to cover the entire graph with only k vertices?

The k-Center Problem

- Consider the following graph. If we are to select 1 vertex, …

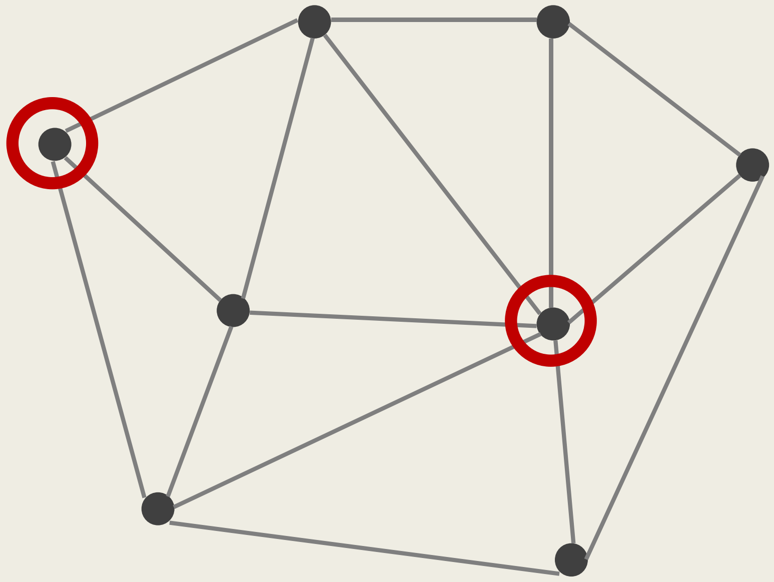




If we select a vertex here,
it covers the entire graph with a distance of 3

If we select a vertex here,
it covers the entire graph with a distance of 2.

The k-Center Problem



If we select the 2 vertices,
they cover the graph with a distance 1.

What is the minimum covering distance, if we are to select k vertices?

The k-Center Problem

Satisfies identity of indiscernible, symmetry, and ***the triangle inequality***.

- Let $M = (V, d)$ be a metric space with *distance function* d defined over V.
  - For any **vertex subset** $\mathbf{A} \subseteq \mathbf{V}$ and any $v \in V$, let

$$d(v, A) \coloneqq \min_{u \in A} d(v, u)$$

    denote minimum distance between v and any vertex in the subset A.

  - The ***covering radius*** of A is defined as $\max_{v \in V} d(v, A)$,

    i.e., the maximum distance between any vertex and the set A.

The k-Center Problem

- Let $M = (V, d)$ be a metric space with *distance function* d defined over V and $k \in \mathbb{N}$ be a positive integer.

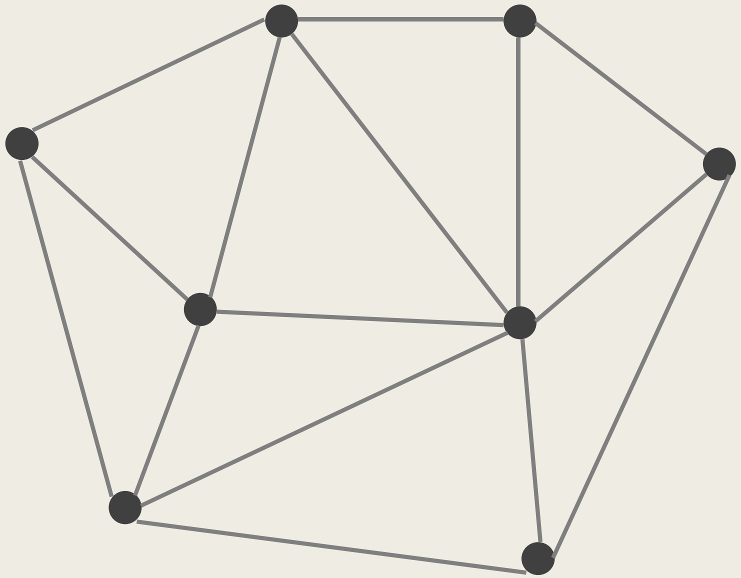
  The metric k-center problem is to compute a subset $A \subseteq V$ with $|A| = k$ such that the **<u>covering radius</u> of A is minimized**.

  - That is, $\max_{v \in V} d(v, A)$, is minimized.

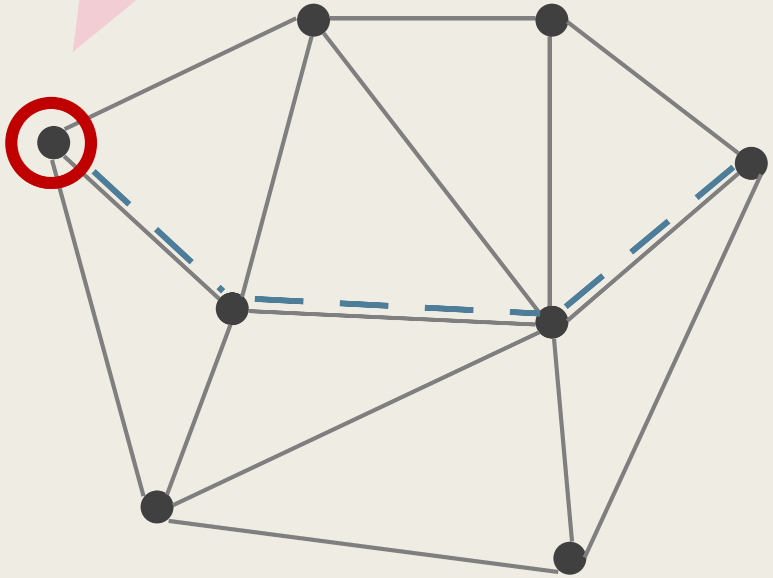
Place the centers so as to minimize the covering radius.

The k-Center Problem

- Consider the following graph.



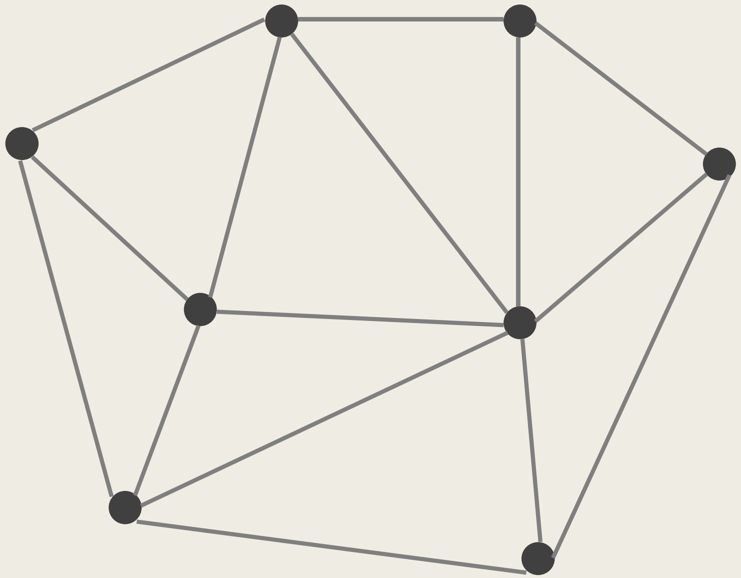
Placing a center here gives a covering radius of 3.

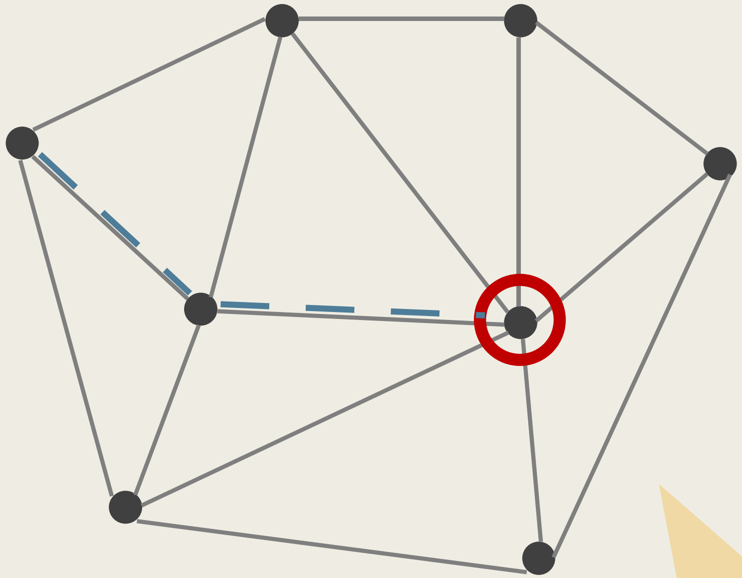


The covering radius is the maximum distance from the vertices to the center set, i.e., $\max_{v \in V} \min_{u \in A} d(v, u)$.

The k-Center Problem

- Consider the following graph.

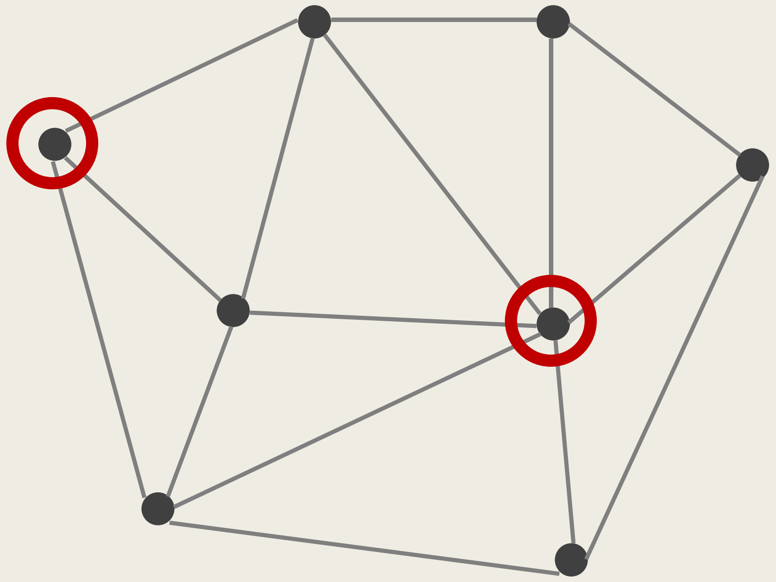




Placing a center here gives a covering radius of 2.

The covering radius is the maximum distance from the vertices to the center set, i.e., $\max_{v \in V} \min_{u \in A} d(v, u)$.

The k-Center Problem

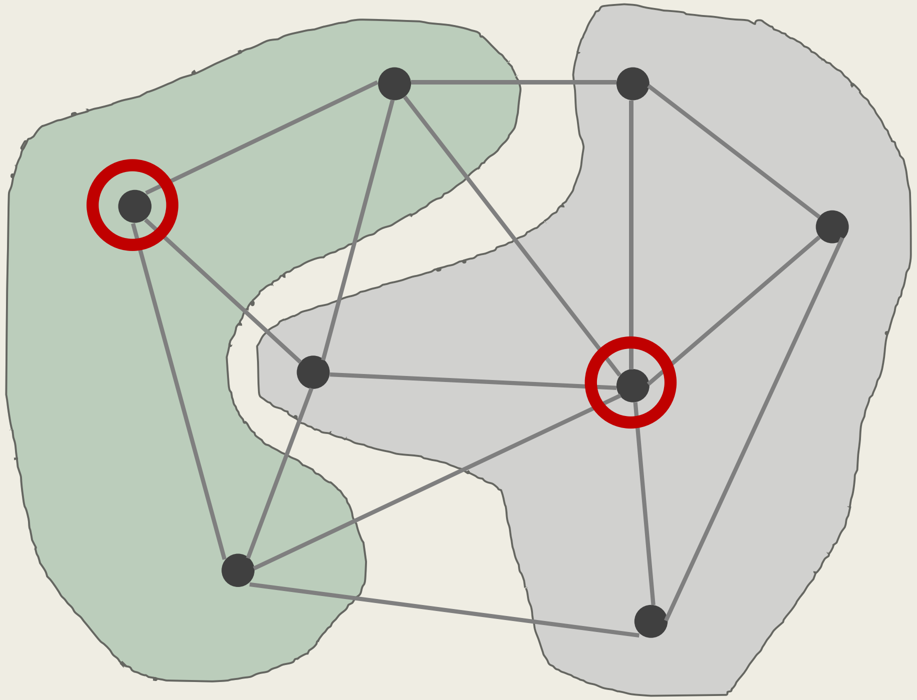


For placing 2 centers,
the optimal covering radius is 1.

The k-center problem is ***to place the centers*** so as ***to minimize the covering radius***.

k-Center as a Clustering Problem

- The k-center problem is a type of clustering problems.
  - Placing the centers to form clusters such that,
    the distance of intra-cluster communications is minimized.



(Brief)

Status of the k-Center Problem

The Status of k-Center

- The k-center problem is NP-hard to solve.
  - It can be approximated to a factor of 2,
    either by *parametric search* or simple *iterative refining*.
  - It cannot be approximated to $2 - \epsilon$ for any $\epsilon > 0$, unless P = NP.
- For the vertex-weighted version,
  parametric search yields a 3-approximation.

Inherent reduction to the Dominating Set Problem

The k-center problem is tightly connected to the existence of dominating sets.

The incidence graph G(t)

- Let $M = (V, d)$ and $k \in \mathbb{N}$ be an instance of k-center, and $t \geq 0$ *be a* ***target radius***.
  - Define the ***incidence graph*** $\boldsymbol{G}(\boldsymbol{t}) = (V, E_t)$ with vertex set V and edge set

$$\boldsymbol{E_t} \coloneqq \{ (\boldsymbol{u}, \boldsymbol{v}) : u, v \in V, \boldsymbol{d}(\boldsymbol{u}, \boldsymbol{v}) \leq \boldsymbol{t} \}.$$

In $G(t)$, we connect vertices that are within distance t.

- Let $M = (V, d)$ and $k \in \mathbb{N}$ be an instance of k-center, and $t \geq 0$ *be a* ***target radius***.
  - Define the ***incidence graph*** $\boldsymbol{G(t)} = (V, E_t)$ with vertex set V and edge set

$$\boldsymbol{E_t} \coloneqq \{ (\boldsymbol{u}, \boldsymbol{v}) : u, v \in V, \boldsymbol{d(u, v)} \leq \boldsymbol{t} \}.$$

In $G(t)$, we connect vertices that are within distance t.

- Let t^* denote the optimal radius that can be achieved.

Lemma 1.

For any $t \geq 0$,

$G(t)$ has a dominating set of size k if and only if $t \geq t^*$.

Lemma 1.

For any $t \geq 0$,

$G(t)$ has a dominating set of size k if and only if $t \geq t^*$.

- If $G(t)$ has a dominating set S with size k,
  then selecting S to be the center set yields a covering radius at most t.

  Since t^* is the optimal radius that can be achieved, $t^* \leq t$.

- Conversely, if $t \geq t^*$, then let A^* be an optimal center set.

  For any $v \in V$, we have $d(v, A^*) \leq t^* \leq t$,
  which means that in $G(t)$, v is dominated by some vertex in A^*.

  Hence A^* is a dominating set for $G(t)$ with size k.

An Inherent Reduction to Dominating Set

Lemma 1.

For any $t \geq 0$,

$G(t)$ has a dominating set of size k if and only if $t \geq t^*$.

- By Lemma 1, the optimal radius is the smallest t such that $G(t)$ has a dominating set of size at most k.

This reduction illustrates the nature of the k-center problem.

Solving the k-dominating set problem, however, is *NP-hard*.

Discretizing Possible Values for t

Discretizing Possible Values for t

Lemma 1.

For any $t \geq 0$,

$G(t)$ has a dominating set of size k if and only if $t \geq t^*$.

- The optimal radius is the smallest t such that $G(t)$ has a dominating set of size at most k.

- Let's, for now, leave aside the solvability of dominating set problem.

Do we really have infinitely many possible $G(t)$ to consider ?

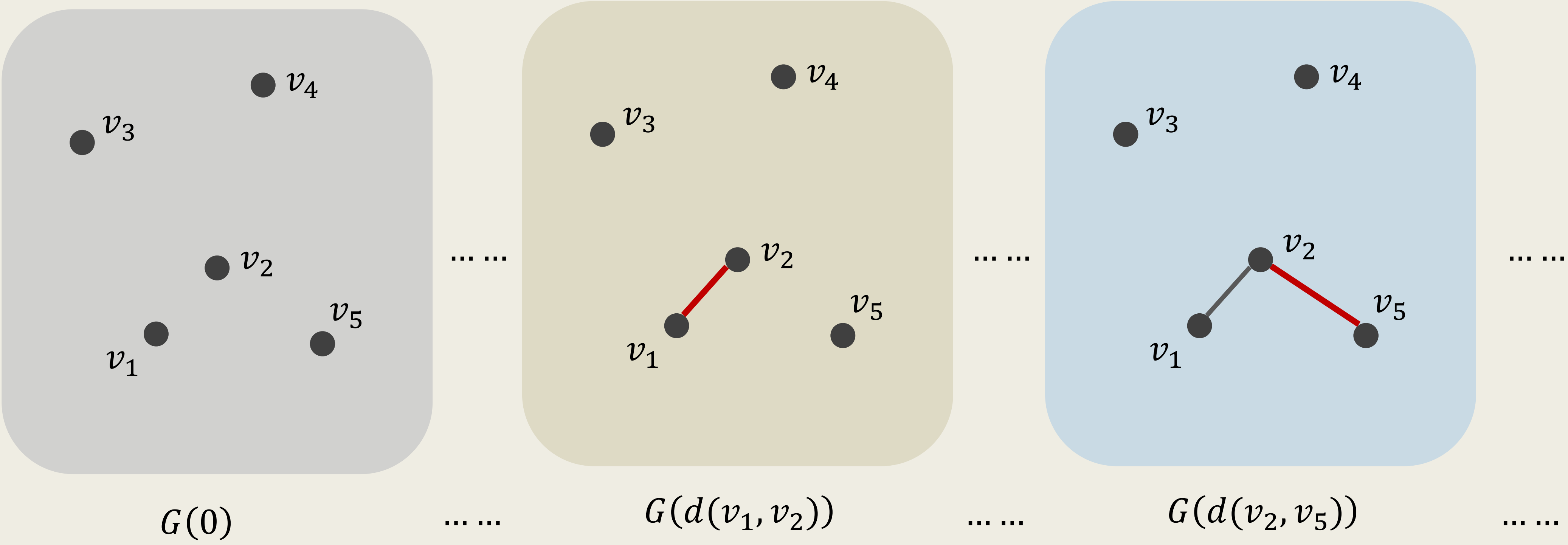
The answer turns out to be no.

Discretizing Possible Values for t

- Consider the following example.

  When t goes from zero to infinity, we have……

New edges pop up in $G(t)$ only when t passes the distance between a pair.



Discretizing Possible Values for t

- When t goes from zero to infinity, we know that……
  - $G(t)$ changes only when the value of t reaches the distance between any pair of vertices.

    In that case, new edges will pop up in $G(t)$.

- Let $d_1, d_2, \dots, d_m$ denote the distances between all pair of vertices, sorted in ascending order.
  - Then, $G(t)$, where $t \in \{d_1, d_2, \dots, d_m\}$, are exactly the set of graphs that will appear when t goes from zero to infinity.

Lower-Bounding

the Size of any Dominating Set

Lower-bounding the size of dominating sets

Lemma 1.

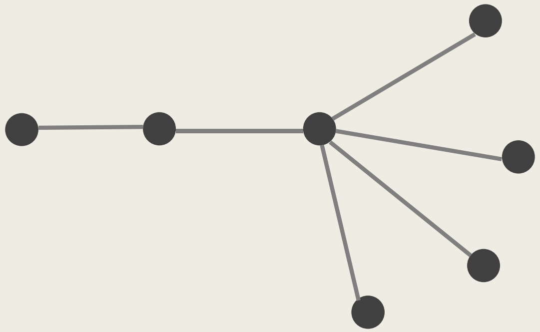
For any $t \geq 0$, **$\boldsymbol{G(t)}$ has a dominating set of size $\boldsymbol{k}$** if and only if $t \geq t^*$.

- In the following, we derive a *beautiful lower-bound* on the size of any dominating set in a graph.

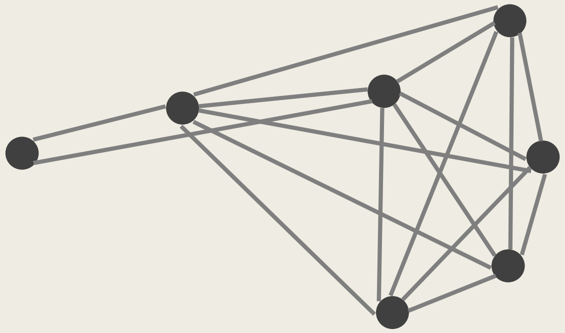
Some Notations – Graph Closure

- Let $G = (V, E)$ be a graph.
  - For any positive constant c, define the graph $G^c = (V, E^c)$ with

$$E^c \coloneqq \{ (u, v) : d_G(u, v) \leq c \}.$$



G



G^2

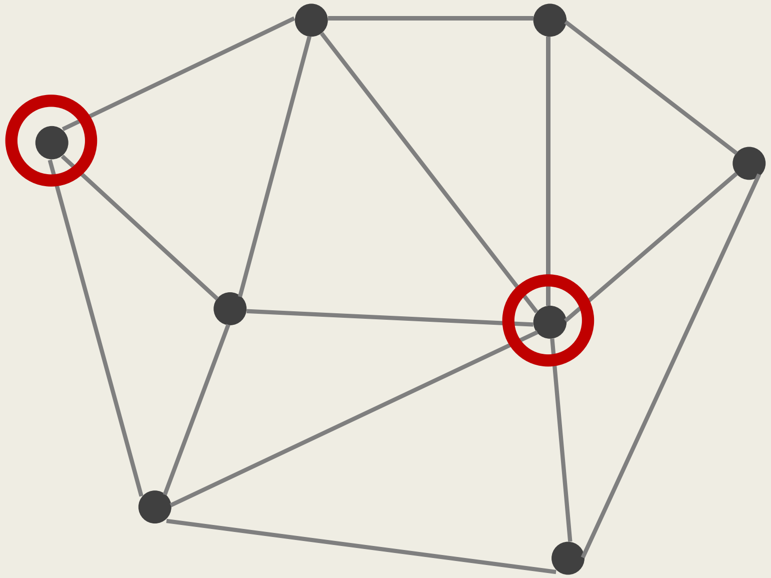
Every pair of vertices
that has distance at most 2 in G
is connected in G^2.

Some Notations – Maximal Independent Set

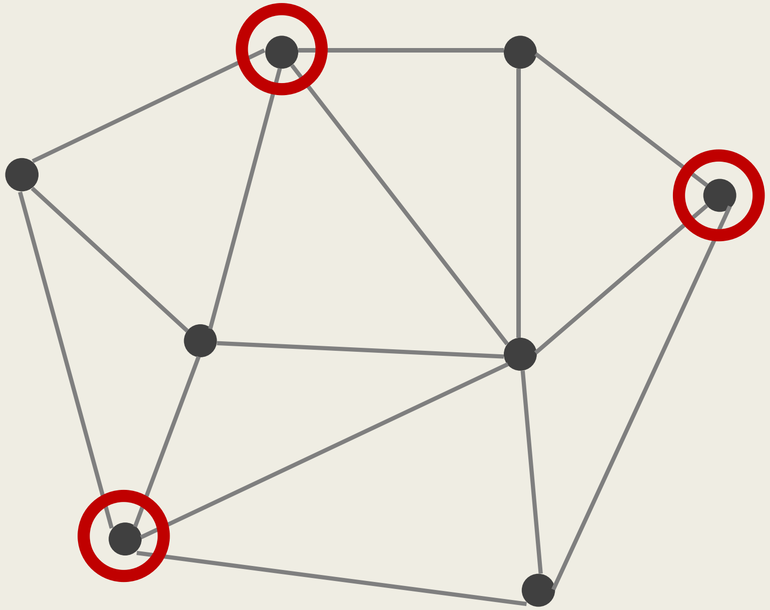
- Let $G = (V, E)$ be a graph.
  - We say that a vertex subset $I \subseteq V$ is an ***independent set*** for G if ***none of vertex pairs*** $u, v \in I$ ***is connected by an edge*** in G, i.e., the induced subgraph of I has no edges at all.
  - We say that *an independent set* I is ***maximal*** if it is not contained in any other independent set as a subset.

Intuitively, the size of a maximal independent set ***cannot be extended by adding any new vertex***.

Maximal Independent Sets



I_1



I_2

Two maximal independents I_1, I_2 for the graph.

No more vertex can be added to the two sets.

The Maximal Independent Sets for G

- Let $G = (V, E)$ be a graph.

Lemma 2.

Any maximal independent set for G is also a dominating set for G.

- Let I be an MIS for G.
  - If I is not dominating in G,
    then there exist a $v \in V$ such that, $v \notin I$ and $(v, u) \notin E$ for all $u \in I$.
  - Hence, $I \cup \{v\}$ is an independent set, a contradiction.

The Maximal Independent Sets in G^2

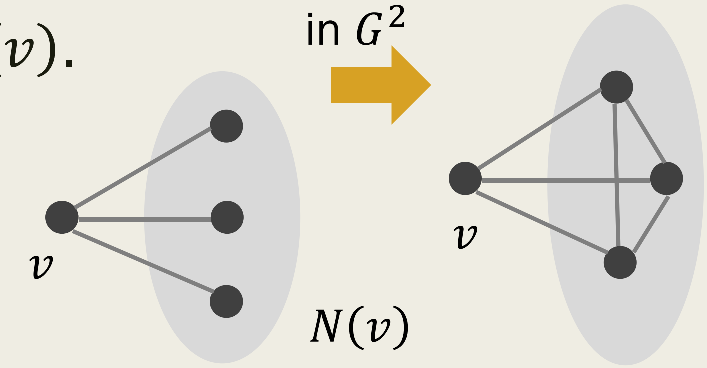
- Let $G = (V, E)$ be a graph.

Lemma 3.

For any feasible dominating set D for G and any independent set I for G^2, we have $|I| \leq |D|$.

- Consider any $v \in D$ and the neighbors $N(v)$ of v.
  - The vertices $\{v\} \cup N(v)$ form a clique in G^2.
  - Hence, I contains at most one vertex from $\{v\} \cup N(v)$.
- This holds for all $v \in D$.

  Hence, we have $|I| \leq |D|$.



The Maximal Independent Sets in G^2

- Let $G = (V, E)$ be a graph.

Lemma 2.

Any maximal independent set for G is also a dominating set for G.

Lemma 3.

For any feasible dominating set D for G and any independent set I for G^2, we have $|I| \leq |D|$.

- By Lemma 2 and 3, any maximal independent set for G^2
  - *Lower-bounds the size of any dominating set of G, and*
  - *Dominates the vertices in G within a distance of at most 2.*

The Parametric Search Technique

&

2-Approximation for k-Center

MIS as a Tool for “Approximate-or-Refute”

- Consider the k-Center problem.
- Let $t > 0$ be a target parameter to be tested, and
  let $I(t)$ be a maximal independent set for $G^2(t)$.
  - If $|I(t)| > k$,
    then by Lemma 3, $G(t)$ has no dominating set of size k, and $t < t^*$.
  - If $|I(t)| \leq k$,
    then by Lemma 2, $I(t)$ has a covering radius of $2t$.
  - The smallest t with $|I(t)| \leq k$ must satisfy $t \leq t^*$ and
    will be a 2-approximation.

The “Approximate-or-Refute” Search Process

- The algorithm goes as follows.

  1. Let $d_1, d_2, \ldots, d_m$ be the all-pair distances between the vertices, sorted in ascending order.
  2. Greedily compute a maximal independent set I_i for $G^2(d_i)$. Let i' be the smallest index such that $\left|I_{i'}\right| \leq k$.
  3. Output $I_{i'}$ as the approximate solution for the metric k-center problem.

Step 2 can either be done by sequential search or binary search.

2-Approximation

by Simple Iterative Refining

Simple Iterative Refinement

- We can also obtain a 2-approximation by ***simple iterative refinement***.
  - The idea is to greedily insert new centers so as to minimize the current assignment radius.
  - The algorithm goes as follows.
    1. Let $\mathcal{C} \leftarrow \emptyset$ be the current of centers.
    2. For $i = 1, 2, \ldots, k$ do
       - Pick $u \in V$ that maximize $d(u, \mathcal{C})$, i.e., $u = \operatorname{argmax}_{v \in V} d(v, \mathcal{C})$.
       - $\mathcal{C} \leftarrow \mathcal{C} \cup \{u\}$.

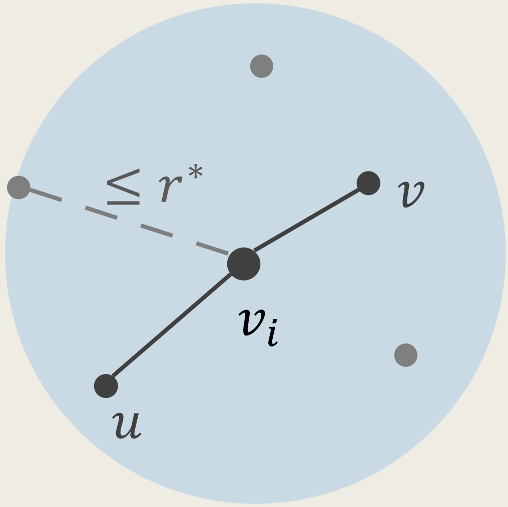
Pick a vertex farthest from $\mathcal{C}$ and add it to $\mathcal{C}$.

The Approximation Guarantee

- To see that the set $\mathcal{C}$ computed by the algorithm is a 2-approximation, consider any optimal solution $\mathcal{S}^* = \{v_1, v_2, \ldots, v_k\}$ with radius r^*.
  - For any $1 \leq i \leq k$, and any $u, v \in N(v_i) \cup \{v_i\}$, we have

$$d(u, v) \leq 2 \cdot r^*$$

by the triangle inequality.



The reason is that,

$$d(u, v) \leq d(u, v_i) + d(v_i, v)$$
$$\leq r^* + r^*$$
$$\leq 2 \cdot r^*$$

by triangle inequality.

- For any $1 \leq i \leq k$, and any $u, v \in N(v_i) \cup \{v_i\}$, we have

$$\boldsymbol{d(u,v) \leq 2 \cdot r^*}$$

Inequality (*)

by the triangle inequality.

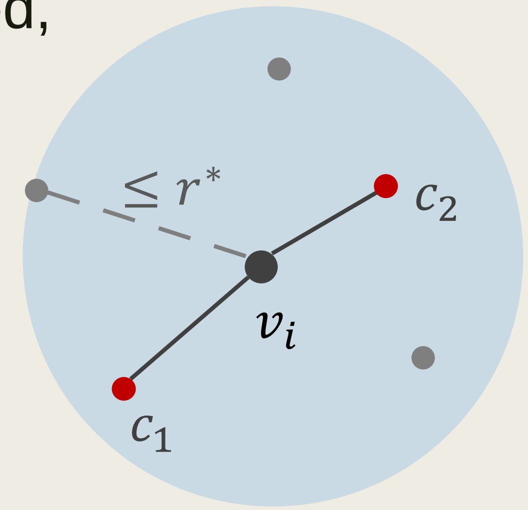
- Hence,
  - If $\mathcal{C}$ includes one vertex from $N(v_i) \cup \{v_i\}$ for each $1 \leq i \leq k$, then by (*) we know that, $d(v, \mathcal{C}) \leq 2 \cdot r^*$ holds for all $v \in V$.
  - If $\mathcal{C}$ includes more than one vertex from $N(v_i) \cup \{v_i\}$ for some i, then at the moment when the second center is placed,

    for any $v \in V$, we have

$$d(v, \mathcal{C}) \leq d(c_1, c_2) \leq 2 \cdot r^*$$

as well.

By the design of the greedy algorithm.



Inapproximability of $2 - \epsilon$

Creating the Gap for k-Center

- As hinted in Lemma 1, the metric k-center problem is closely related to the k-dominating set problem.
  - Given an instance $G = (V, E)$ of k-dominating set problem, we create an instance (V, d) of metric k-center problem such that,
    - If the answer for G is “yes”, then ***there exists a feasible solution*** for (V, d) ***with radius 1***.
    - If the answer for G is “no”, then ***any feasible solution*** for (V, d) ***has radius at least 2***.

Optimal radius = 1

Optimal radius ≥ 2

The ratio of the gap corresponds to the hardness of approximation.

The Reduction

- Let $G = (V, E)$ be an instance of the k-dominating set problem.

  Define a distance metric as

$$\text{for any } u, v \in V, \qquad d(u,v) := \begin{cases} 1, & \text{if } (u,v) \in E, \\ 2, & \text{otherwise.} \end{cases}$$

- We have the following lemma.

Lemma 3.

G has a dominating set of size k if and only if (V, d) has a k-center set with radius 1.

The Weighted k-Center Problem

&

3-Approximation by Parametric Search

The Weighted k-Center Problem

- In the weighted metric k-center problem,
  the vertices are weighted by a weight function $w : V \rightarrow \mathbb{R}^{+}$, and
  the goal is to compute a subset $A \subseteq V$ such that
  - The total weight of A does not exceed the given budget K, i.e., $w(A) \leq K$,
  - The covering radius A is minimized.

Place the centers ***under the given budgets*** *to* ***minimize the covering radius***.

Parametric Search for the Weighted k-Center

- We will obtain a simple 3-approximation by parametric search technique.
  - Let t^* be the optimal radius.
  - The following lemma reduces this problem to the ***weighted dominating set*** problem.

Lemma 5.

For any $t \geq 0$, the graph $G(t)$ has a dominating set of weight k if and only if $t \geq t^*$.

The proof is the same as Lemma 1.

Parametric Search for the Weighted k-Center

- In order to perform parametric search,
  we need to establish the testing process for the weighted dominating set.

  For any t, the testing process either

  - Computes a solution with radius at most $c \cdot t$ for some constant c, or,
  - Asserts that $t < t^*$ and refutes t.

Then, by Lemma 5,
the smallest t that is not refuted by the process will be a c-approximation.

The Testing Process for Weighted Dominating Set

For any t, the testing process either

- Computes a solution with radius at most $c \cdot t^*$ for some constant t, or
- Asserts that $t < t^*$ and refutes t.

The basic properties
for maximal independent sets still hold.

- To form a valid lower-bound, we can observe that…
  - Any maximal independent set I for G^2 still covers G with a distance at most 2.
  - Any maximal independent set I for G^2 still bounds any dominating set of G ***in size***. (***but not in weight***)

– Any maximal independent set I for G^2 still bounds any dominating set of G ***in size*** (***but not in weight***)

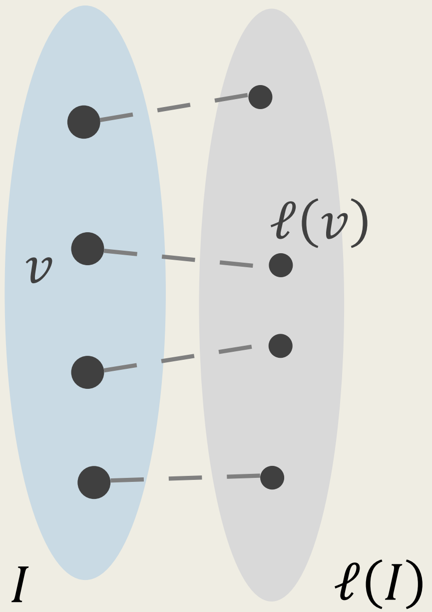
By selecting the lightest neighbor for each $v \in I$,
we can lower-bound the weight of any dominating set D.

For any vertex $v \in V$, let $\ell(v)$ denote the lightest vertex in $N(v) \cup \{v\}$,
i.e., $\ell(v) \coloneqq \operatorname{argmin}_{u \in \{v\} \cup N(v)} w(u)$.

Define

$$\ell(I) \coloneqq \{\, \ell(v) : v \in I \,\}.$$

Then, $w(\ell(I))$ lower-bounds $w(D)$, and
$\ell(I)$ covers G within a distance of 3!



The Maximal Independent Sets in G^2

- Let $G = (V, E)$ be a graph with weight function $w : V \to \mathbb{R}^+$.

Lemma 6.

For any maximal independent set I for G^2,

- $\ell(I)$ dominates the vertices of V with a distance at most 3.
- $w\big(\ell(I)\big) \leq w(D)$, for any feasible dominating set D for G.

- The proof is based on the same idea.

The Parametric Search Process

- The algorithm goes as follows.

1. Let $d_1, d_2, \ldots, d_m$ be the all-pair distances between the vertices, sorted in ascending order.
2. Greedily compute a maximal independent set I_i for $G^2(d_i)$.

   Let i' be the smallest index such that $w\left(\ell\left(I_{i'}\right)\right) \leq k$.
3. Output $\ell\left(I_{i'}\right)$ as the approximate solution for the weighted metric k-center problem.

Step 2 can either be done by sequential search or binary search.

That’s all for k-Center so far.

Let’s proceed to our next problem.