ETH zürich
Departement Informatik

Approximations- und Online-Algorithmen

Dr. Hans-Joachim Böckenhauer
Prof. Dr. Dennis Komm
Dr. Manuel Wettstein

courses.csedu.inf.ethz.ch/approx_online_alg_24

Lösungsvorschläge – Blatt 2

Zürich, 7. März 2024

Lösung zu Aufgabe 2

(a) Für eine Zentrenmenge Z ist der kürzeste Weg von einem Knoten v zu einem Zentrum $d(v, Z) := \min_{z \in Z} d(\{v, z\})$ und die Kosten sind $\rho := \max_{v \in V} d(v, Z)$.

Die Menge der Knoten, die tatsächlich diese Maximaldistanz zu einem Zentrum haben und die Kosten somit hoch halten, können wir als $\operatorname{argmax}_{v \in V} d(v, Z) := \{v \in V \mid d(v, Z) = \rho\}$ bezeichnen. Den beschriebenen Greedy-Algorithmus können wir also folgendermassen umsetzen:

Algorithmus GreedyCenter

Eingabe: $k \geq 1$ und eine Instanz von k-Center.

Arbeitsweise:
Wähle einen beliebigen Knoten $z_1 \in V$ und setze $Z := \{z_1\}$.
for i **from** 2 **to** k:
Wähle einen Knoten $z_i \in \operatorname{argmax}_{v \in V} d(v, Z)$ und setze $Z := Z \cup \{z_i\}$.

Ausgabe: Die Zentrenmenge Z mit $|Z| = k$.

(b) Wir bezeichnen mit $B_r(z) := \{v \in V \mid d(\{v, z\}) \leq r\}$ den r-Ball um das Zentrum z, also die Menge der Knoten, die höchstens eine Distanz von r zu z haben.

Wir beobachten, dass ein r-Ball, wenn er einen Knoten z enthält, vollständig in dem $2r$-Ball um z liegt, also

$$\forall z \in B_r(z^*)\colon B_r(z^*) \subseteq B_{2r}(z) \ .$$

Dies liegt daran, dass alle Punkte in dem r-Ball höchstens einen Abstand von r zum Mittelpunkt z^* haben, aufgrund der Dreiecksungleichung metrischer Graphen also höchstens einen Abstand von $2r$ zueinander.

Sei nun eine beliebige Instanz von k-Center gegeben und sei $Z^* = (z_1^*, \ldots, z_k^*)$ eine optimale Lösung dieser Instanz mit Kosten ρ^*.

Die ρ^*-Bälle um die k optimalen Zentren decken somit alle Knoten des Graphen ab, es gilt

$$\bigcup_{z^* \in Z^*} B_{\rho^*}(z^*) = V \; . \tag{1}$$

Wir sollen zeigen, dass auch die $2\rho^*$-Bälle um die k von GreedyCenter gewählten Zentren $Z = (z_1, \ldots, z_k)$ alle Knoten des Graphen abdecken, in Gleichungsform also

$$\bigcup_{z \in Z} B_{2\rho^*}(z) = V \; . \tag{2}$$

Dazu beobachten wir zunächst den Wert von $\max_{v \in V} d(v, Z)$ während des Ablaufs des Algorithmus. Mit jedem neu dazukommenden Zentrum in Z wird dieser Wert kleiner oder er verändert sich nicht.

Fall 1: Die Maximaldistanz zu einem Zentrum wird noch vor der Wahl des letzten Zentrums z_k kleiner als $2\rho^*$, es gilt also

$$\max_{v \in V} d(v, Z) \leq 2\rho^* \; .$$

Ab diesem Moment befinden sich also bereits alle Knoten innerhalb eines $2\rho^*$-Balles um eines der Zentren z_i und wir haben eine Approximationsgüte von schlimmstenfalls 2.

Fall 2: Die Maximaldistanz zu einem Zentrum bleibt bis zur Wahl des letzten Zentrums grösser als $2\rho^*$, es gilt also

$$\max_{v \in V} d(v, Z) > 2\rho^* \; .$$

Dies bedeutet, dass für jedes neu gewählte Zentrum z_i die Distanz zu den bisherigen Zentren grösser als $2\rho^*$ ist; die k Bälle $B_{\rho^*}(z_i)$ sind also paarweise disjunkt.

Alle k Knoten z_i sind (wie auch jeder andere Knoten) in einem ρ^*-Ball um eines der optimalen Zentren z_i^* enthalten. Jeder solche Ball kann maximal ein z_i enthalten, weil er dann gemäss der Beobachtung am Anfang komplett in dem zugehörigen disjunkten $2\rho^*$-Ball liegt. Wir haben also k verschiedene z_i und jeder der k Bälle $B_{\rho^*}(z_i^*)$ kann nur jeweils eines enthalten, daher enthält jeder solche Ball genau ein z_i und ist somit in dem Ball $B_{2\rho^*}(z_i)$ enthalten. Diese decken somit ebenfalls alle Knoten ab, was wir noch in Formeln ausdrücken können:

Durch allfälliges Umbenennen der Knoten können wir voraussetzen, dass z_i in $B_{\rho^*}(z_i^*)$ liegt, also

$$\forall i \in \{1, \ldots, k\} \colon B_{\rho^*}(z_i^*) \subseteq B_{2\rho^*}(z_i) \, .$$

Daraus folgt

$$\bigcup_{z^* \in Z^*} B_{\rho^*}(z^*) \subseteq \bigcup_{z \in Z} B_{2\rho^*}(z)$$

und wegen (1) ist die geforderte Gleichung (2) erfüllt.

Somit ist auch für Fall 2 bewiesen, dass eine Approximationsgüte von 2 oder besser erreicht wird.

(c) Wir können als Instanz eine Menge von Punkten im euklidischen Vektorraum $\mathbb{R}^m$ beliebiger Dimension m angeben; mit der euklidischen Distanz bilden diese immer einen metrischen Graphen. Wir beschränken uns auf $\mathbb{R}$.

Für $k = 1$ genügen die drei Punkte -1, 0 und 1; siehe Abbildung 1. Die optimale Lösung mit den Kosten 1 ist Wahl von 0 als Zentrum. Wenn das von GreedyCenter zufällig gewählte Zentrum hingegen -1 oder 1 ist, so hat der jeweils andere Punkt Abstand 2 zum einzigen Zentrum und die Kosten sind 2.

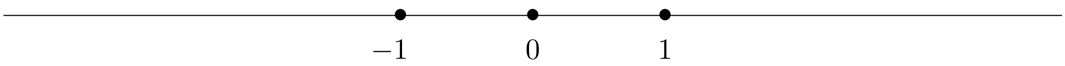

Abbildung 1: Die schwere Eingabe für $k = 1$ aus Teilaufgabe (c).

Für den allgemeinen Fall $k \geq 1$ können wir einfach die obige Dreipunktekonstellation k Mal in ausreichender Entfernung voneinander verwenden. Eine Möglichkeit sind die die Punkte $\bigcup_{i=1}^{k}\{4k-1, 4k, 4k+1\}$; siehe Abbildung 2. Die einzige optimale Lösung mit Kosten 1 besteht darin, den mittleren Punkt jeder Dreiergruppe zu wählen, also im Beispiel die k Punkte $4, 8, \ldots, 4k$. Wir zeigen, dass alle anderen Lösungen mindestens Kosten von 2 verursachen.

Nehmen wir an, dass einer der mittleren Punkte nicht als Zentrum gewählt wird. Dann verursachen die beiden benachbarten Punkte Kosten von mindestens 2, sofern nicht beide als Zentrum gewählt werden. Dadurch würden aber zwei Zentren für diese Dreiergruppe verbraucht, womit in einer anderen der k Dreiergruppe kein Zentrum mehr enthalten sein kann. Dies führt wiederum zu Kosten von mindestens 2.

Wenn das erste, von GreedyCenter zufällig gewählte Zentrum nicht Teil der optimalen Lösung ist, kann diese nicht mehr erreicht werden. Damit ist die Approximationsgüte bestenfalls 2.

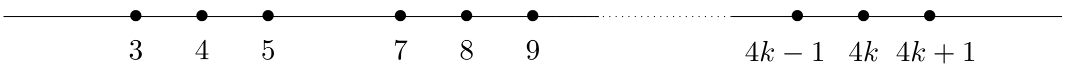

Abbildung 2: Die schwere Eingabe für $k \geq 1$ aus Teilaufgabe (c).