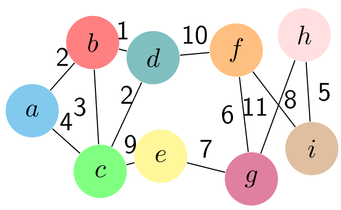
6CCS3AIN, Tutorial 06 (Version 1.0)

1. Perform k-Median on the following input: $\mathbf{x}_1 = \begin{pmatrix}3\\4\end{pmatrix}$ $\mathbf{x}_2 = \begin{pmatrix}4\\2\end{pmatrix}$ $\mathbf{x}_3 = \begin{pmatrix}1\\2\end{pmatrix}$ $\mathbf{x}_4 = \begin{pmatrix}1\\1\end{pmatrix}$ $\mathbf{x}_5 = \begin{pmatrix}1\\3\end{pmatrix}$.
   Assume that k=2 and the initial clusters location of the clusters are
   $$\boldsymbol{\mu}_1 = \begin{pmatrix}1\\1\end{pmatrix} \text{ and } \boldsymbol{\mu}_2 = \begin{pmatrix}2\\4\end{pmatrix}.$$
   Recall, that the distance between two points is the sum of the absolute values. E.g., the distance between $\mathbf{x}_1$ and $\mathbf{x}_2$ is $|3-4|+|4-2| = 3$. In case of a tie, assign the data point to the cluster with the smaller ID.

2. How would the first assignment and update step look if we used k-Means instead?

3. Perform complete linkage on the following similarity graph. Note that the similarity of all edges that aren't present is 0! Also note that $sim(f,g) = 6$, $sim(g,h) = 11$ and $sim(f,i) = 8$. Recall that in case of a tie always pick the cluster with the smallest ID and then merge it with the cluster that has the smallest ID among the ties. The smallest ID here is given by the alphabetical order. E.g., say you have $S_1 = \{a,c\}$, $S_2 = \{b,f\}$, $S_3 = \{d,h\}$ with $sim(S_1,S_2) = 4$, $sim(S_1,S_3) = 5$ and $sim(S_2,S_3) = 5$. So we have a tie here: do we merge S_1 with S_3 or S_2 and S_3? The ID of S_1 is a, the ID of S_2 is b and the ID of S_3 is d. So we pick S_1 and then we merge it with the set having the smallest ID among all ties. In this case there is only S_3. So we merge S_1 and S_3.

   Here is the graph, good luck!



6CCS3AIN, Tutorial 06 Answers (Version 1.0)

1. Perform k-Median on the following input: $\mathbf{x}_1 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$ $\mathbf{x}_2 = \begin{pmatrix} 4 \\ 2 \end{pmatrix}$ $\mathbf{x}_3 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ $\mathbf{x}_4 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ $\mathbf{x}_5 = \begin{pmatrix} 1 \\ 3 \end{pmatrix}$.
   Assume that k=2 and the initial clusters location of the clusters are

   $$\boldsymbol{\mu}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \text{ and } \boldsymbol{\mu}_2 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}.$$

   Recall, that the distance between two points is the sum of the absolute values. E.g., the distance between $\mathbf{x}_1$ and $\mathbf{x}_2$ is $|3-4|+|4-2|=3$. In case of a tie, assign the data point to the cluster with the smaller ID.
   Solution:

   (a) Assignment 1: $S_1 = \{\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$ and $S_2 = \{\mathbf{x}_1\}$.

   (b) Update 1: $\boldsymbol{\mu}_1 = \begin{pmatrix} 7/4 \\ 2 \end{pmatrix}$ and $\boldsymbol{\mu}_2 = \begin{pmatrix} 3 \\ 4 \end{pmatrix}$.

   (c) Assignment 2: $S_1 = \{\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$ and $S_2 = \{\mathbf{x}_1\}$.

   (d) Nothing changes, so the algorithm terminates

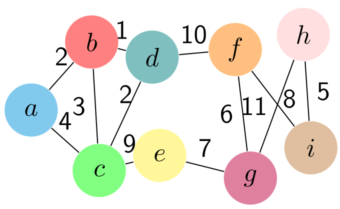
2. How would the first assignment and update step look if we used k-Means instead?
   Solution:

   (a) Assignment 1: $S_1 = \{\mathbf{x}_3, \mathbf{x}_4\}$ and $S_2 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5\}$.

   (b) Update 1: $\boldsymbol{\mu}_1 = \begin{pmatrix} 1 \\ 1.5 \end{pmatrix}$ and $\boldsymbol{\mu}_2 = \begin{pmatrix} 8/3 \\ 3 \end{pmatrix}$.

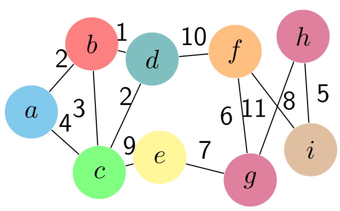
3. Perform complete linkage on the following similarity graph. Note that the similarity of all edges that aren't present is 0! Also note that $sim(f,g) = 6$, $sim(g,h) = 11$ and $sim(f,i) = 8$. Recall that in case of a tie always pick the cluster with the smallest ID and then merge it with the cluster that has the smallest ID among the ties. The smallest ID here is given by the alphabetical order. E.g., say you have $S_1 = \{a, c\}$, $S_2 = \{b, f\}$, $S_3 = \{d, h\}$ with $sim(S_1, S_2) = 4$, $sim(S_1, S_3) = 5$ and $sim(S_2, S_3) = 5$. So we have a tie here: do we merge S_1 with S_3 or S_2 and S_3? The ID of S_1 is a, the ID of S_2 is b and the ID of S_3 is d. So we pick S_1 and then we merge it with the set having the smallest ID among all ties. In this case there is only S_3. So we merge S_1 and S_3.

   Here is the graph, good luck!

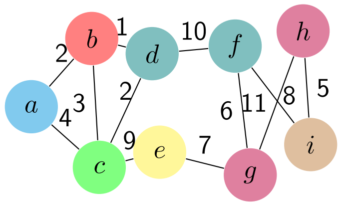


   Solution:

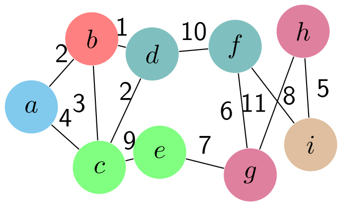
   - Round 1:



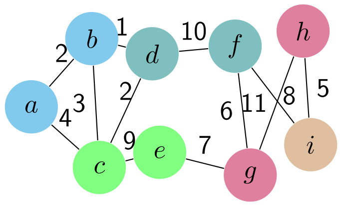
- Round 2:



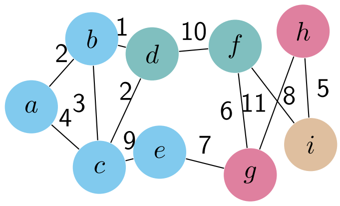
- Round 3:



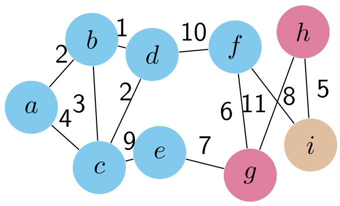
- Round 4:
  Note that an edge that is not present has a value of 0. $sim(\{d,f\},\{i\}) = 0$ for complete linkage. The same holds for $sim(\{c,e\},\{g\})$, $sim(\{d,f\},\{g,h\})$, $sim(\{h,g\},\{i\})$ and for all other edges until we reach the edge between a and b.



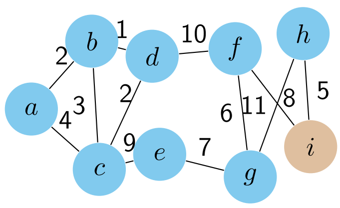
- Round 5:



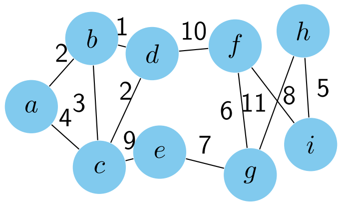
- Round 6:



- Round 7:



- Round 8:



a
b
c
d
e
f
g
h
i
2
1
10
3
2
4
9
7
6
11
8
5