1. Na danih podatkih (X) izvedi hierarhično gručenje in nariši dendrograme za naslednje vrste povezanosti (linkage):
    a. Enojna (single)
    b. Popolna (complete)
    c. Povprečna (average)

X = [1, 2, 4, 8, 9, 14]

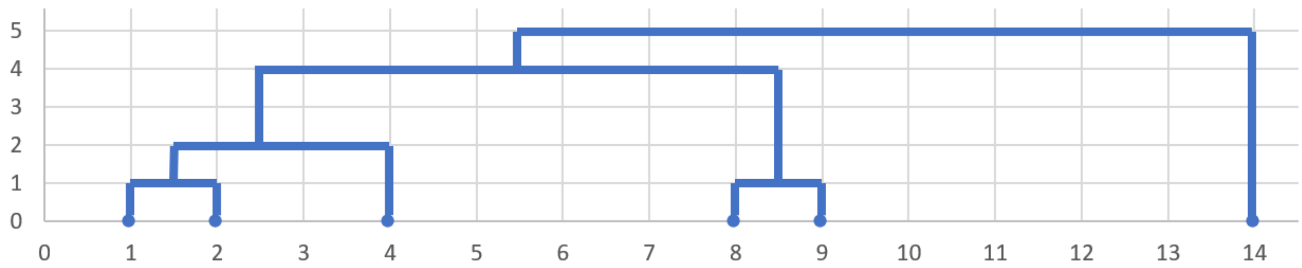
REŠITEV:

a) Izračunamo razdalje med vsemi pari točk in izberemo najmanjšo:
d(1,2)=1, d(1,4)=3, d(1,8)=7, ... d(1,14)=13
d(2,4)=2, ... d(2,14)=12
d(4,8)=4, ... d(4,14)=10
d(8,9)=1, ... d(8,14)=6
d(9,14)=5
V dendrogramu narišemo povezavi na višini 1.

V našem primeru sta dva para z isto minimalno razdaljo – vzamemo oba in ju združimo v gruči G1 in G2:
Izračunamo razdalje med novima gručama in preostalimi točkami:
d(G1,4)=min(d(1,4),d(2,4))=2, (V dendrogramu narišemo povezavo na višini 2)
d(G1,G2)=min(d(1,8),d(2,8),d(1,9),d(2,9))=6,
d(G1,14)=min(d(1,14),d(2,14))=12
d(G2,14)=5

Gručo G1 povežemo s točko 4 v novo gručo G3 in izračunamo nove razdalje:
d(G3,G2)=min(...)=d(4,8)=4 (kar je še vedno manjše od najmanjše naslednje razdalje, d(G2,14)=5)
Zato v dendrogramu povežemo gruči G2 in G3 v novo gručo G4 na višini 4.

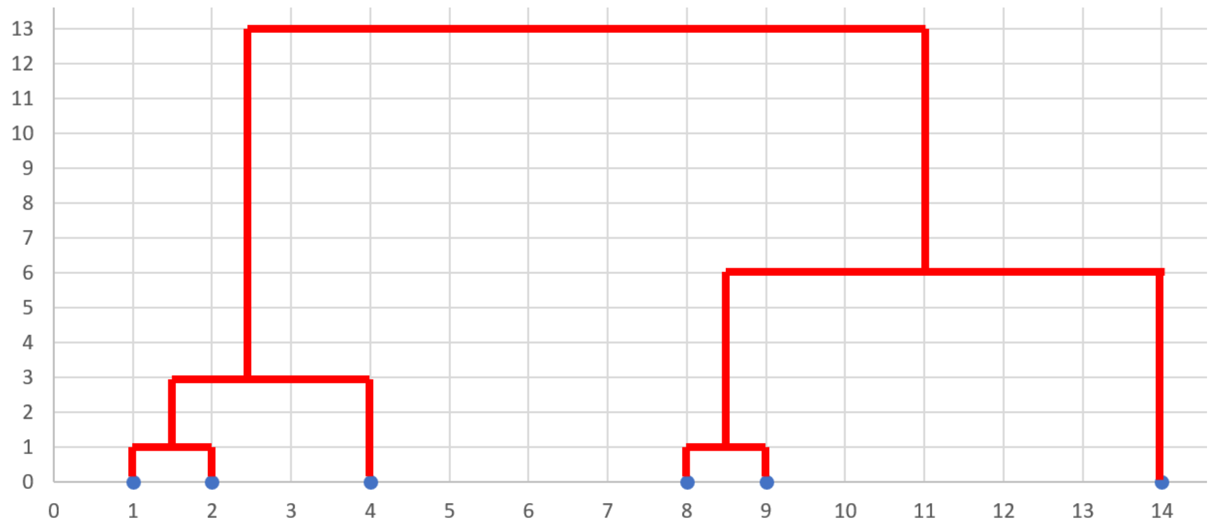
Preostane nam le še to, da povežemo gručo G4 s točko 14 – razdalja med njima je 5.
Dendrogram za enojno povezanost je torej:



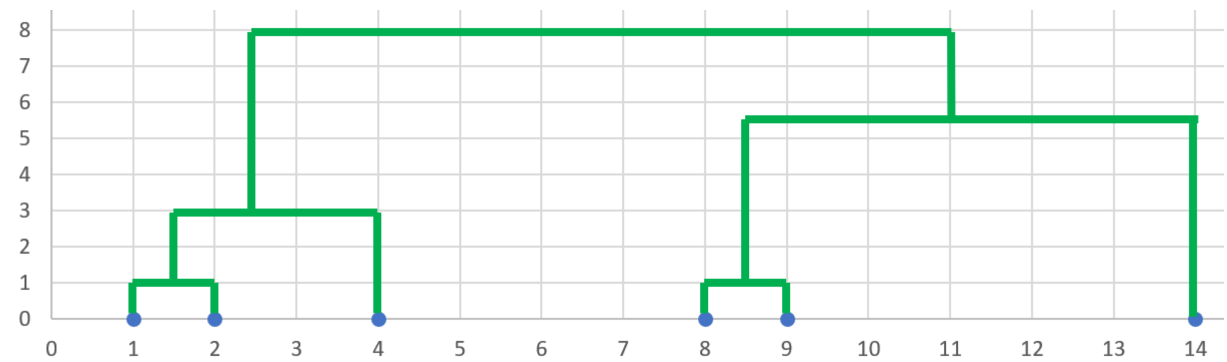
b) Popolna povezanost:
začetek je isti kot pri enojni povezanosti, do razlik pride šele v drugem koraku, ko računamo razdalje med gručo in točko ali med dvema gručama:

d(G1,4)=**max**(d(1,4),d(2,4))=3, (V dendrogramu narišemo povezavo na višini 3)
d(G2,14)=max(d(8,14),d(9,14))=6

Naredimo torej novo gručo G3, ki povezuje G1 in 4. Izračunamo nove razdalje:
d(G3,G2)=...najdaljša razdalja med točkami v G1 in gruči G2...=d(1,9)=8
d(G2,14)=max(d(8,14),d(9,14))=6
Na naslednjem koraku povežemo G2 in 14 v gručo G4. V zadnjem koraku nam preostane le še povezava med G3 in G4, na višini 13 (ker je max razdalja med dvema točkama iz teh grup enaka d(1,14)=13).



c) Pri povprečni povezanosti je prvi korak spet isti kot prej, nato pa namesto min oz. max računamo povprečje vseh razdalj med dvema gručama oz. gručo in točko.
d(G1,4)=2.5, ... naredimo G3
d(G2,14)=5.5 ...
Dendrogram za povprečno povezanost:



2. Na danih podatkih simuliraj metodo k-voditeljev (k-Means) za k=3. Začetni centroidi naj bodo
   C1 = 2, C2 = 5, C3 = 11
   X = [1, 2, 4, 5, 8, 9, 11, 12, 15, 16, 17, 18]

REŠITEV

V tabeli so za vsako točko izračunane razdalje do vseh treh centroidov. Z barvo so označene minimalne razdalje v vsaki vrstici, ki kažejo tudi na pripadnost točke najbližjemu centroidu.

x	C1	C2	C3
	2	5	11
1	**1**	4	10
2	**0**	3	9
4	2	**1**	7
5	3	**0**	6
8	6	3	**3**
9	7	4	**2**
11	9	6	**0**
12	10	7	**1**
15	13	10	**4**
16	14	11	**5**
17	15	12	**6**
18	16	13	**7**

Na novo izračunamo centroide:

C1 = (1+2)/2 = 1,5

C2 = (4+5+8)/3 = 4,5

C3 = (8+9+11+12+15+16+17+18)/7 = 13,25

... in ponavljamo postopek, dokler se pripadnost gručam ne ustali:

x	C1	C2	C3		x	C1	C2	C3		x	C1	C2	C3
	1,5	4,5	13,25			1,5	5,67	14			1,5	6,5	14,8333
1	**0,5**	3,5	12,25		1	**0,5**	4,67	13		1	**0,5**	5,5	13,8333
2	**0,5**	2,5	11,25		2	**0,5**	3,67	12		2	**0,5**	4,5	12,8333
4	2,5	**0,5**	9,25		4	2,5	**1,67**	10		4	2,5	**2,5**	10,8333
5	3,5	**0,5**	8,25		5	3,5	**0,67**	9		5	3,5	**1,5**	9,83333
8	6,5	**3,5**	5,25		8	6,5	**2,33**	6		8	6,5	**1,5**	6,83333
9	7,5	4,5	**4,25**		9	7,5	**3,33**	5		9	7,5	**2,5**	5,83333
11	9,5	6,5	**2,25**		11	9,5	5,33	**3**		11	9,5	4,5	**3,83333**
12	11	7,5	**1,25**		12	10,5	6,33	**2**		12	10,5	5,5	**2,83333**
15	14	10,5	**1,75**		15	13,5	9,33	**1**		15	13,5	8,5	**0,16667**
16	15	11,5	**2,75**		16	14,5	10,3	**2**		16	14,5	9,5	**1,16667**
17	16	12,5	**3,75**		17	15,5	11,3	**3**		17	15,5	10,5	**2,16667**
18	17	13,5	**4,75**		18	16,5	12,3	**4**		18	16,5	11,5	**3,16667**

Preverimo rezultat še s programom:

```
from sklearn.cluster import KMeans
import numpy as np
X = np.array([[1], [2], [4], [5], [8], [9], [11], [12], [15], [16], [17], [18]])
C = np.array([[2], [5], [11]], np.float64)
kmeans = KMeans(n_clusters=3, init=C).fit(X)
kmeans.cluster_centers_
```

Izpis centroidov se ujema z našim izračunom:

```
array([[ 1.5       ],
       [ 6.5       ],
       [14.83333333]])
```