

Data Mining Tutorial

Clusteranalyse – Teil III

Erich Schubert, Arthur Zimek

Ludwig-Maximilians-Universität München

2015-05-27 — KDD Übung

Wiederholung:

Shared-Nearest-Neighbors-Ähnlichkeit:

$$SNN_5(o, p) := |NN(o, 5) \cap NN(p, 5)|$$

Keine Distanz, sondern eine Ähnlichkeitsfunktion!

Aufgabe 6-1

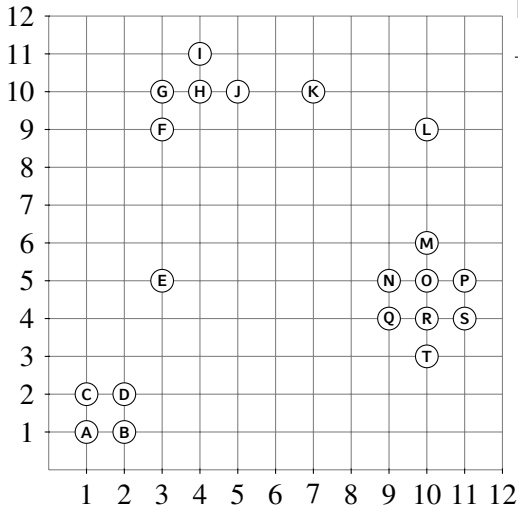
Aufgabe 6-2

Aufgabe 6-3

single-link

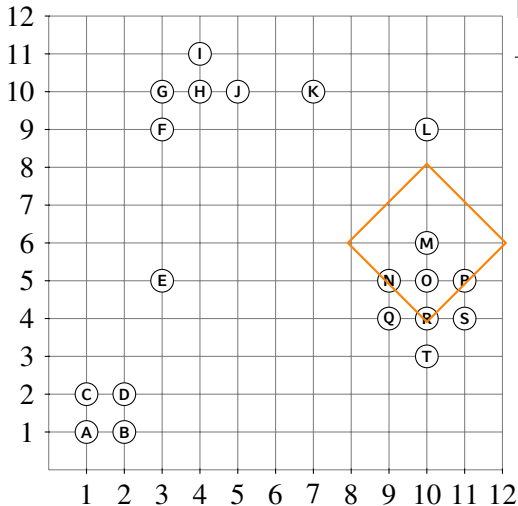
average-link

complete-link



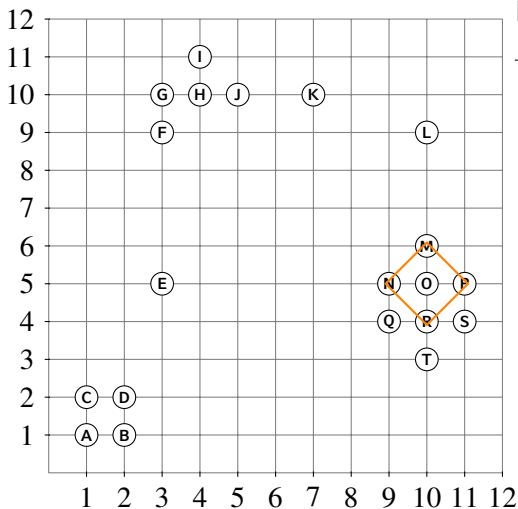
Nachbarschaften:

p	$NN(p, 5)$
M	
O	
R	
T	



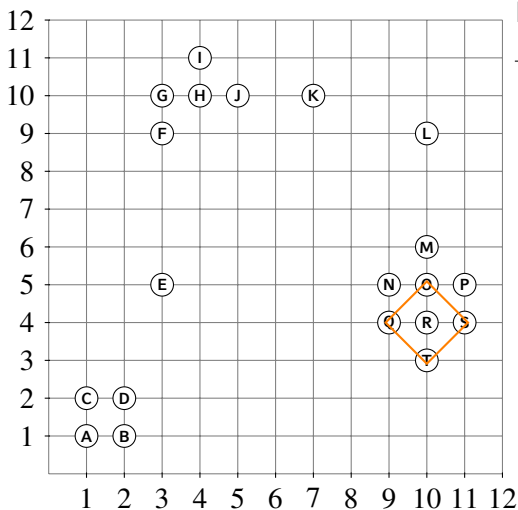
Nachbarschaften:

p	$NN(p, 5)$					
M	M	N	O	P	R	
O						
R						
T						



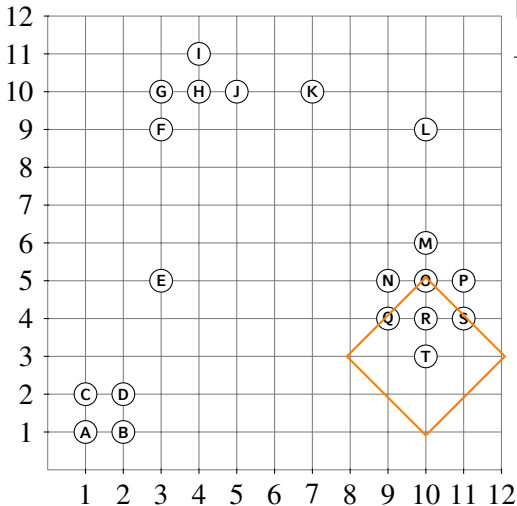
Nachbarschaften:

p	$NN(p, 5)$
M	M N O P R
O	M N O P R
R	
T	



Nachbarschaften:

p	$NN(p, 5)$
M	M N O P R
O	M N O P R
R	O Q R S T
T	



Nachbarschaften:

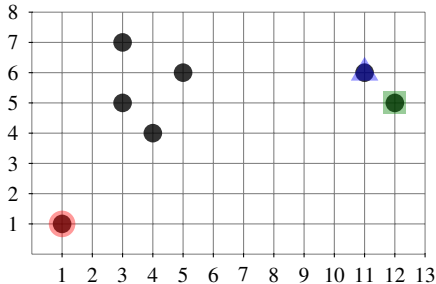
p	$NN(p, 5)$
M	M N O P R
O	M N O P R
R	O Q R S T
T	O Q R S T

Nachbarschaften:

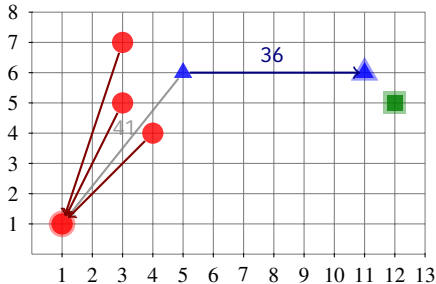
p	$NN(p, 5)$
M	M N O P R
O	M N O P R
R	O Q R S T
T	O Q R S T

Ähnlichkeiten dadurch:

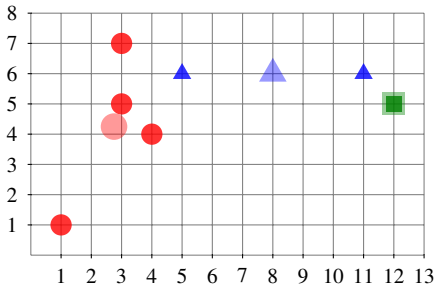
SNN_5	M	O	R	T
M	5	5	2	2
O	5	5	2	2
R	2	2	5	5
T	2	2	5	5



Punkte neu zuordnen



Punkte neu zuordnen

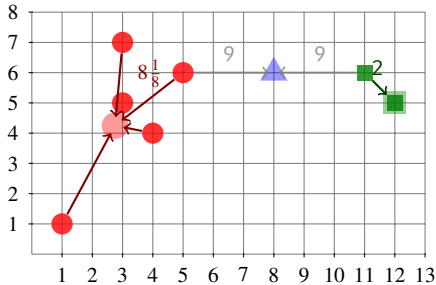


Zentroide neu berechnen:

$$\mu = (2.75, 4.25)$$

$$\mu = (8.0, 6.0)$$

$$\mu = (12.0, 5.0)$$



Punkte neu zuordnen

k-Means Clustering – Lloyd/Forgey Algorithmus

Data Mining
Tutorial

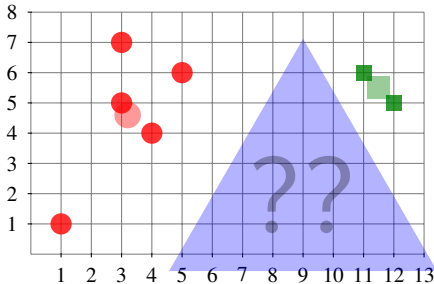
E. Schubert,
A. Zimek

Aufgabe 6-1

Aufgabe 6-2

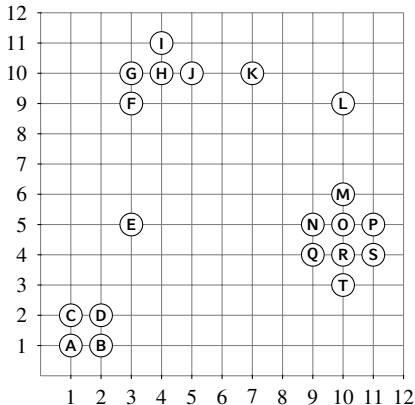
Aufgabe 6-3

single-link
average-link
complete-link



Zentroide neu berechnen:
Oops.
Mittelwert der leeren
Menge?

Single-Link Clustering



A B C D E F G H J I K L M O N Q P S R T

Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

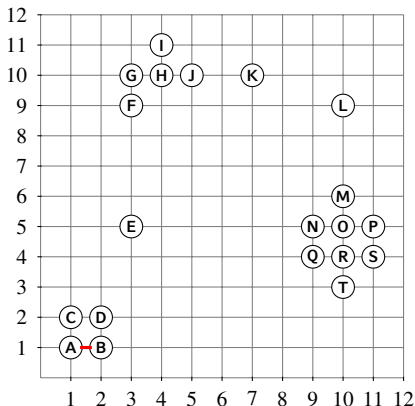
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

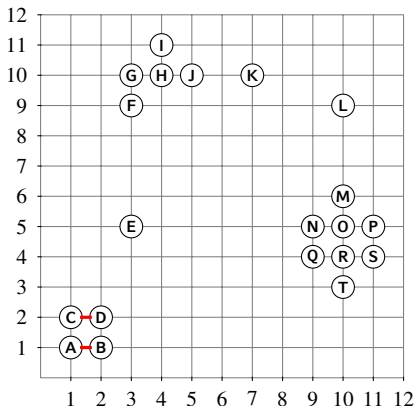
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

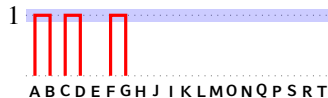
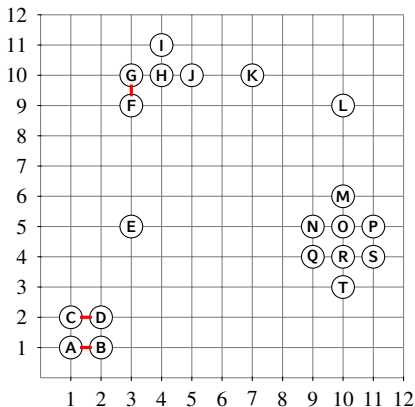
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

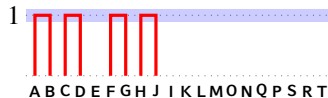
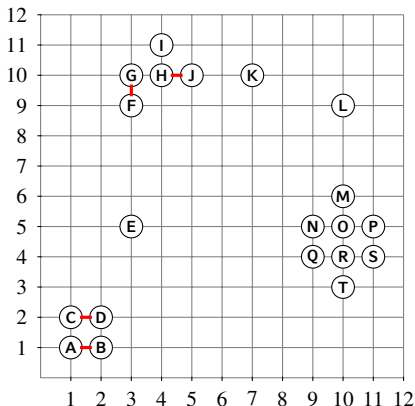
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

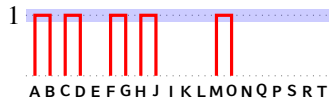
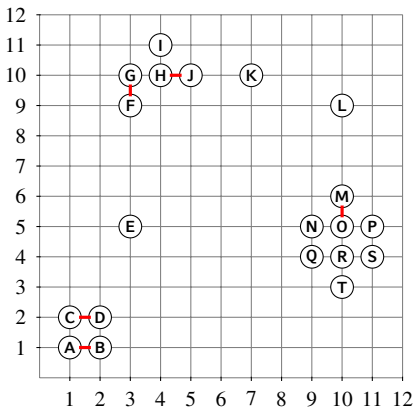
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

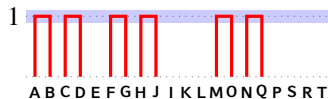
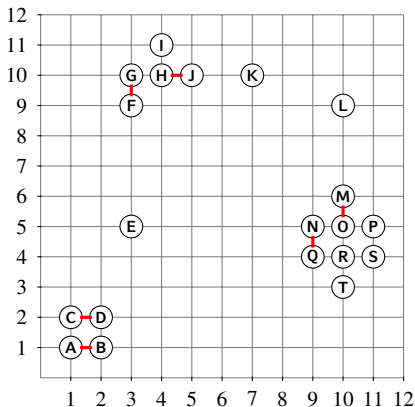
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

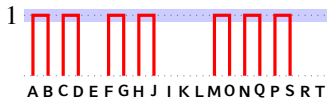
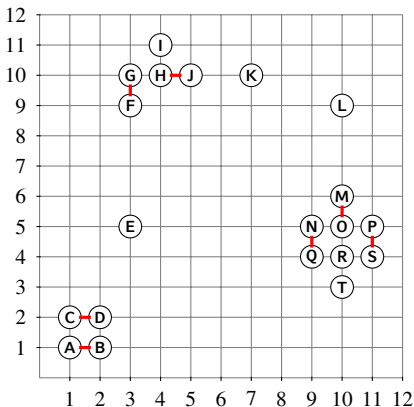
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

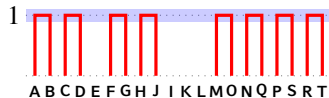
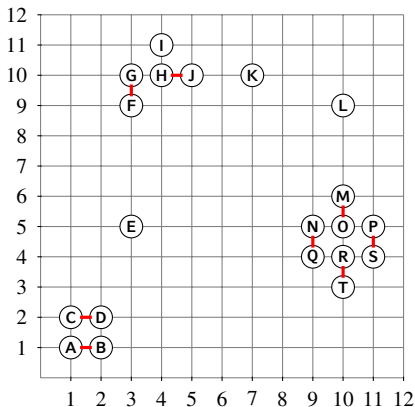
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

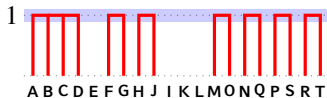
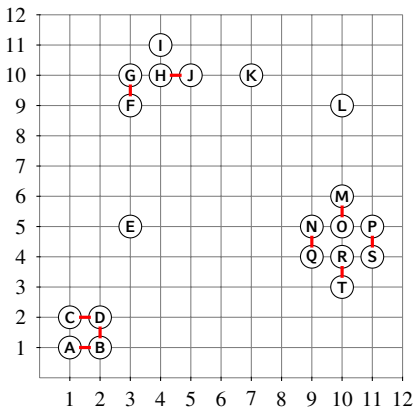
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

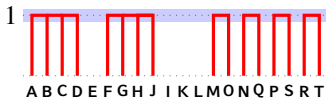
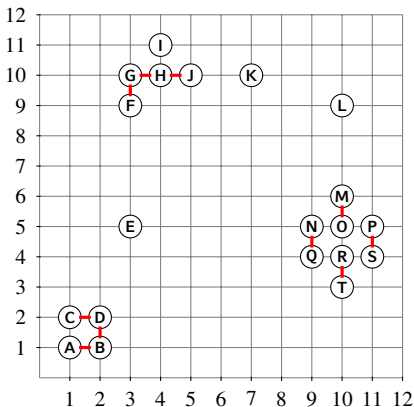
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

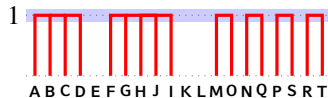
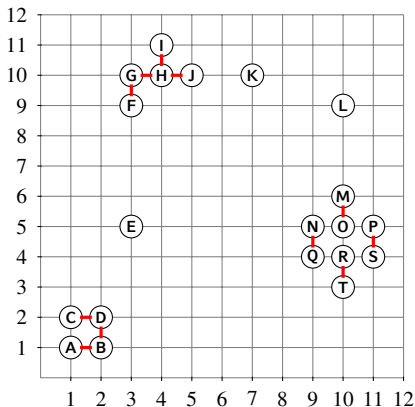
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

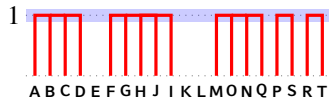
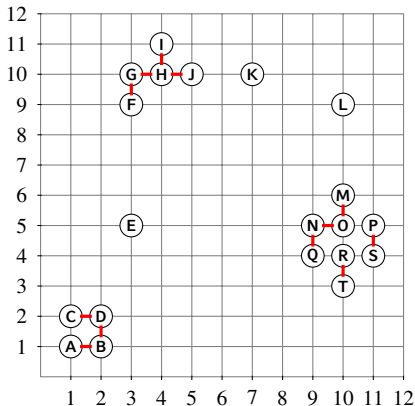
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

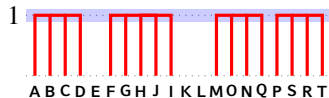
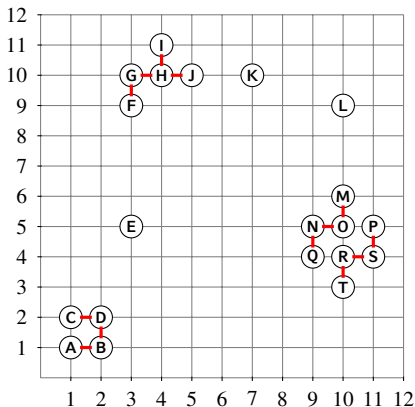
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

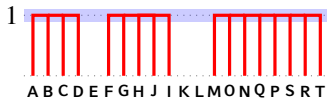
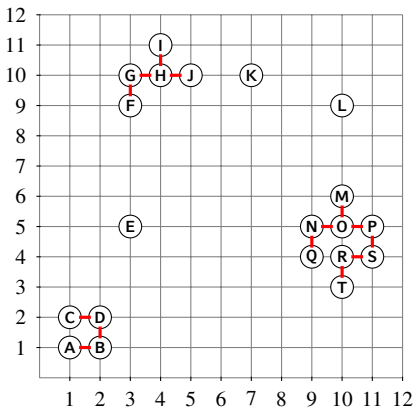
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

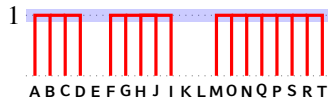
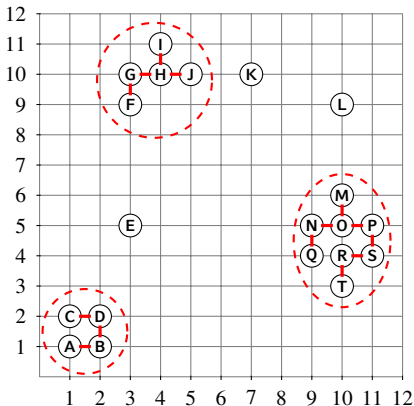
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

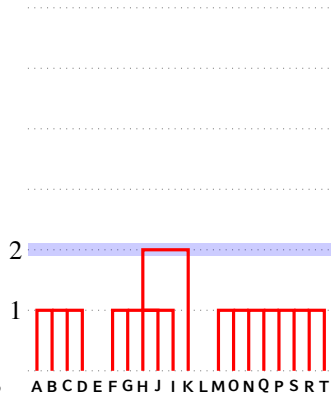
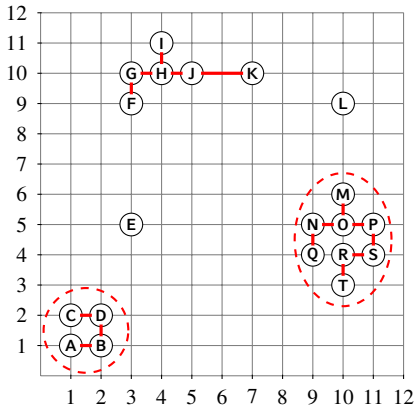
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

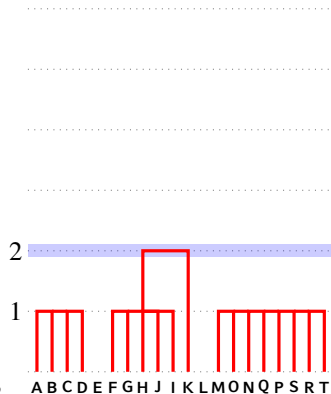
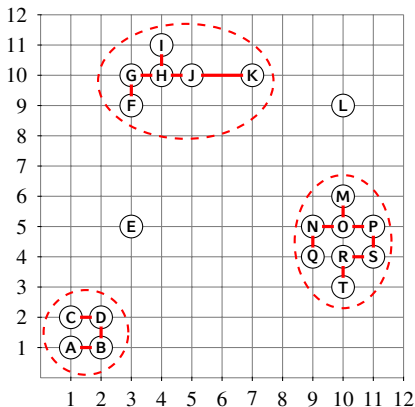
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

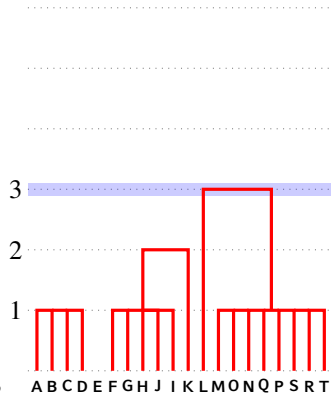
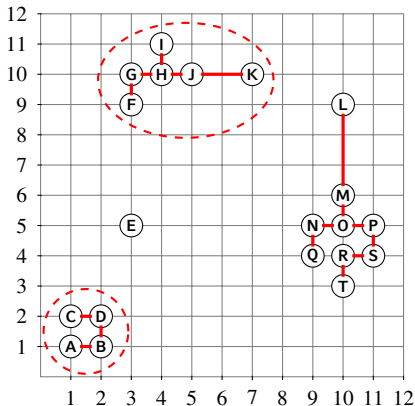
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

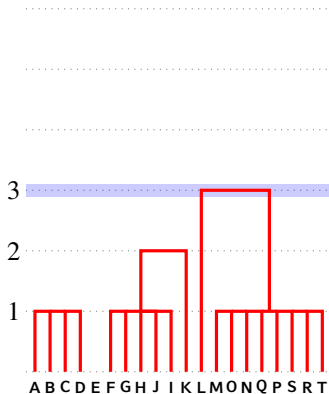
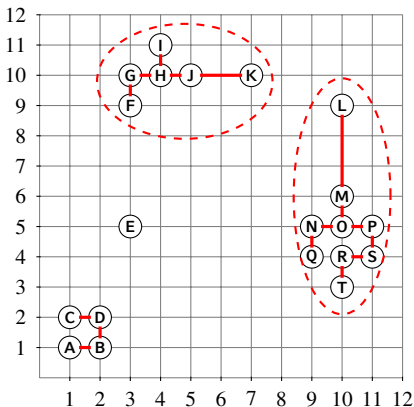
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

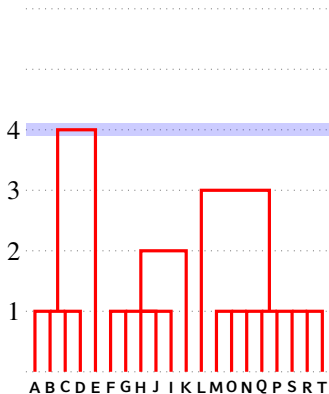
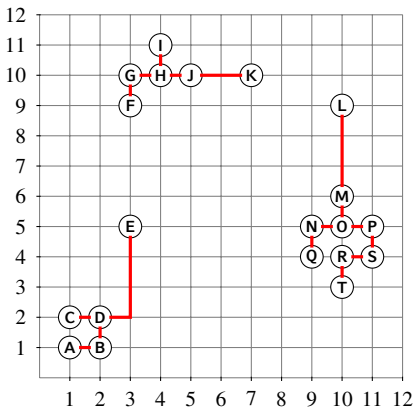
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

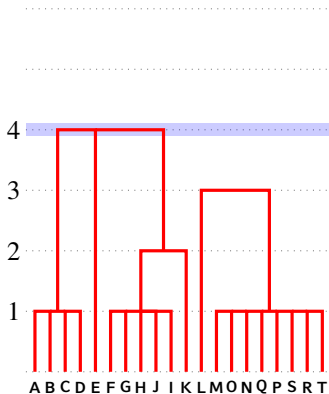
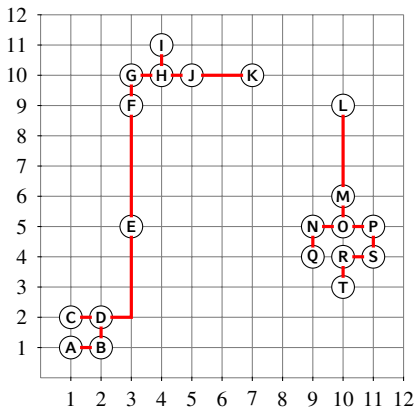
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Single-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

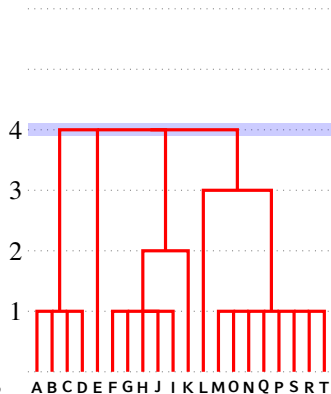
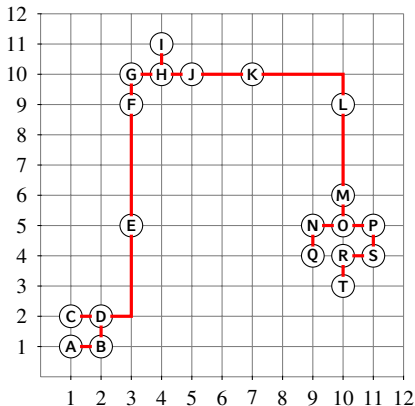
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

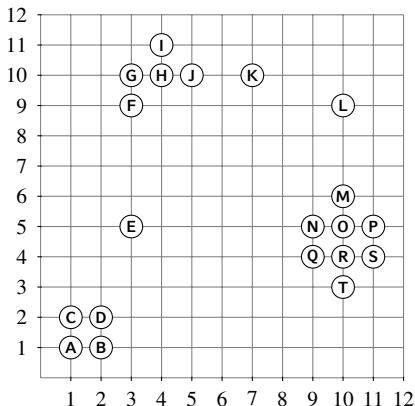
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



A B C D E F G H J I K L M N O P Q R S T

Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

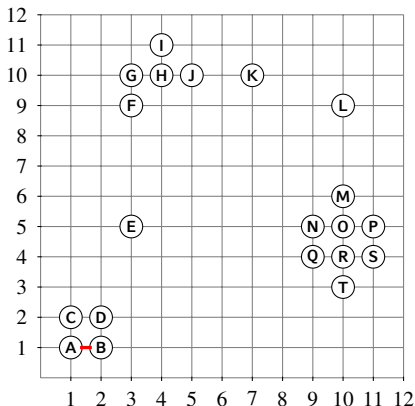
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

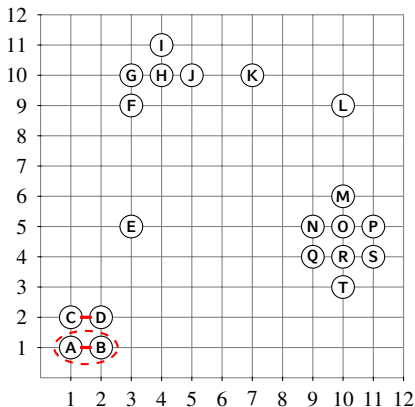
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

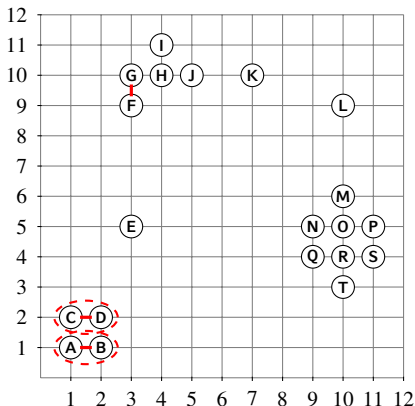
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

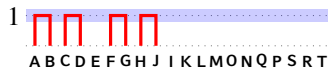
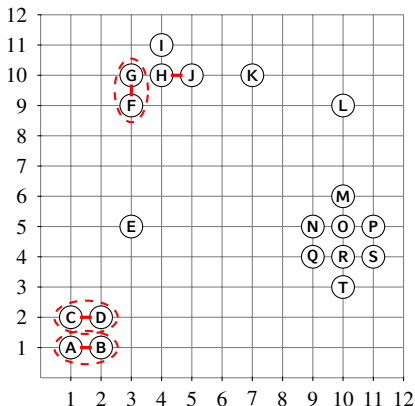
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

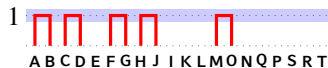
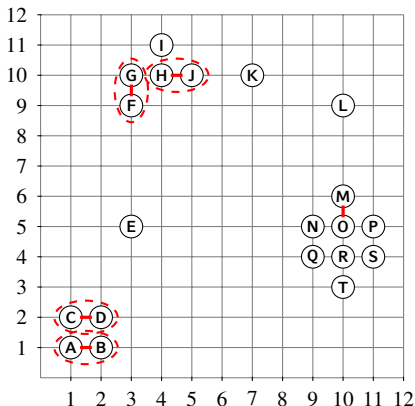
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

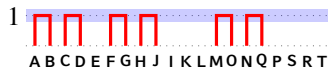
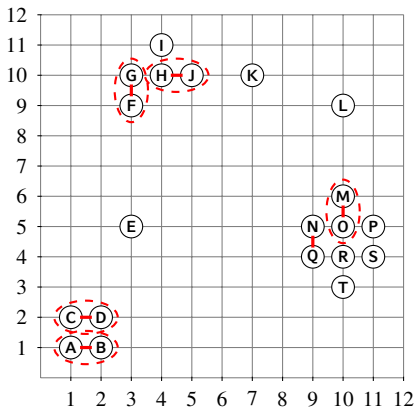
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

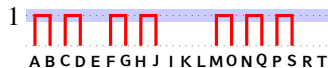
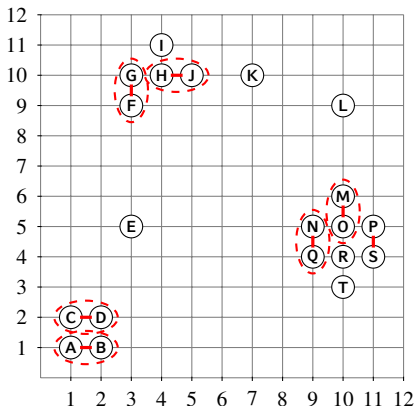
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

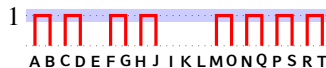
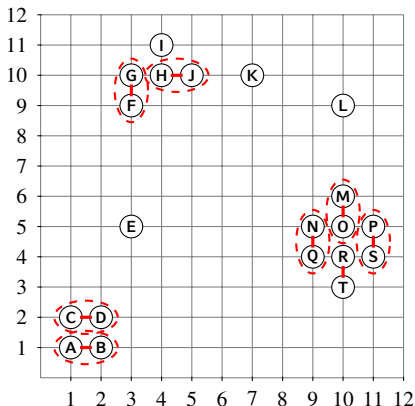
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

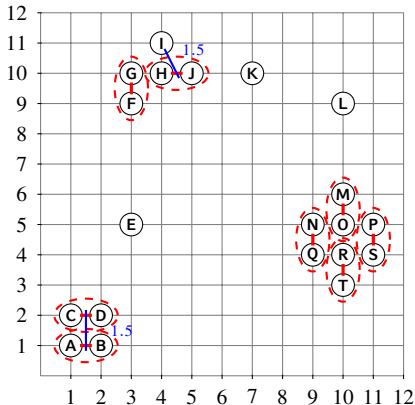
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

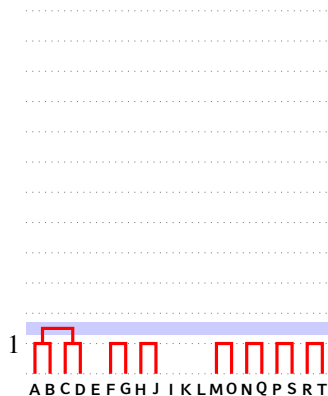
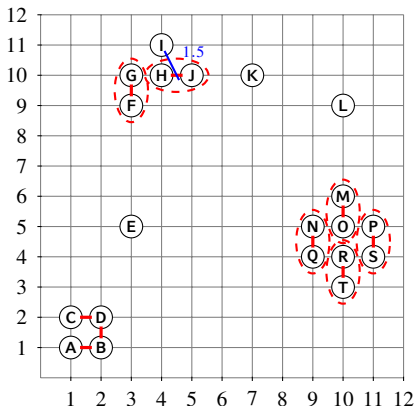
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

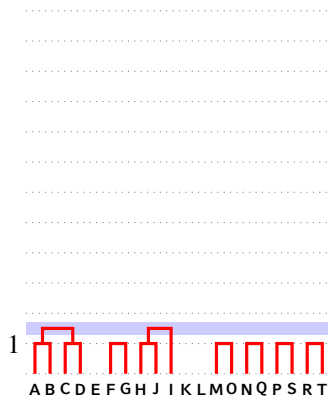
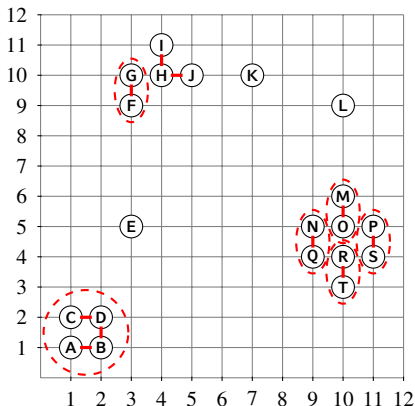
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

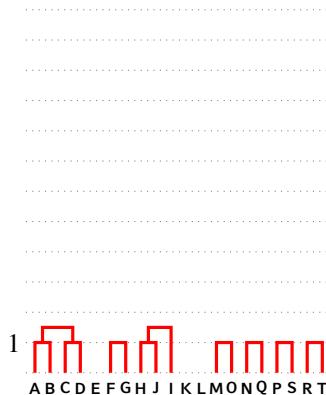
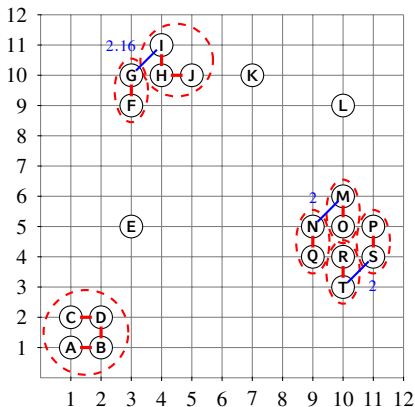
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

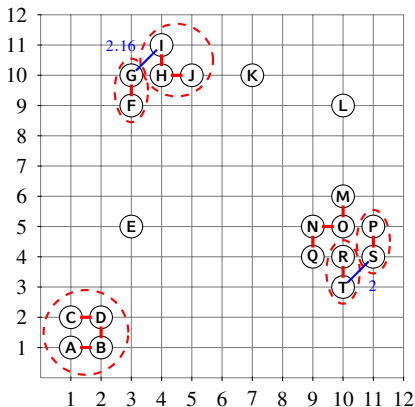
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

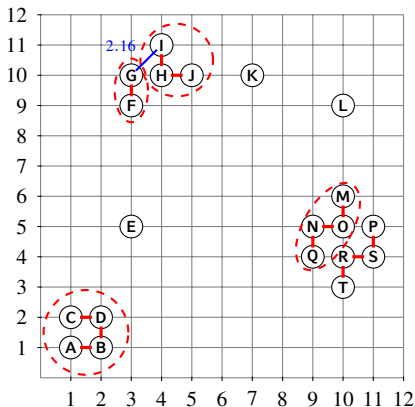
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

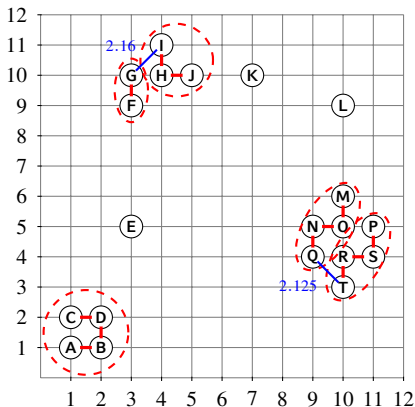
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

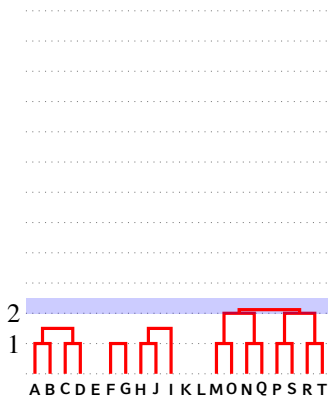
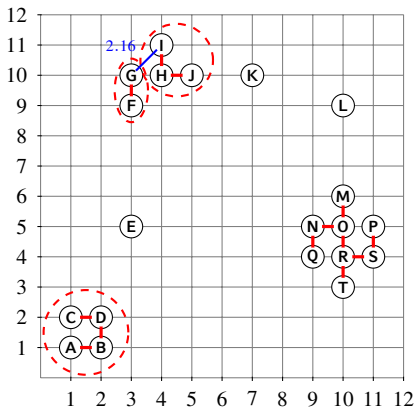
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

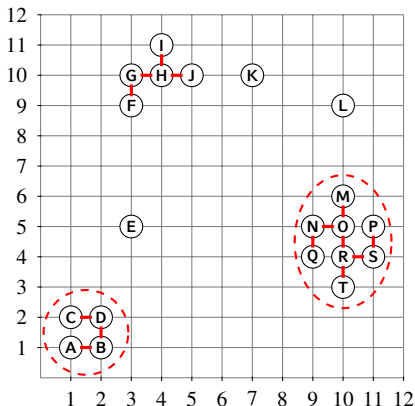
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

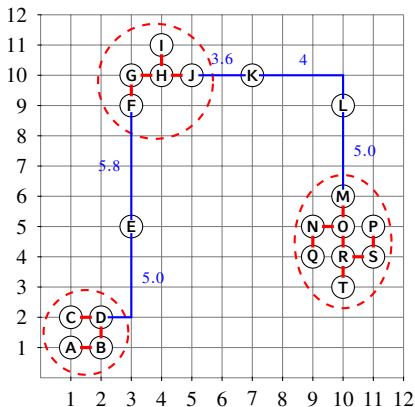
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

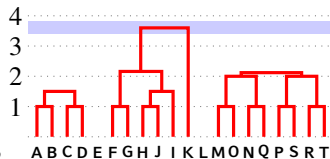
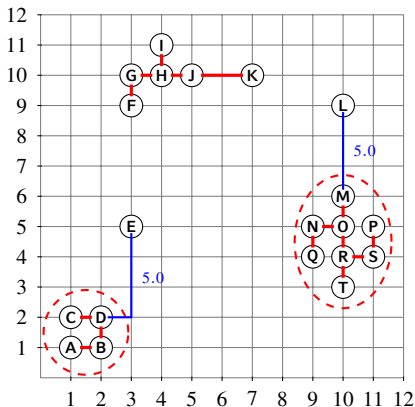
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

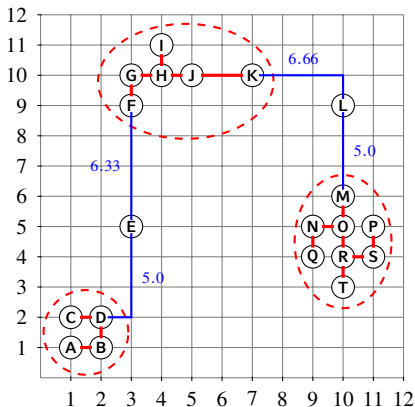
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

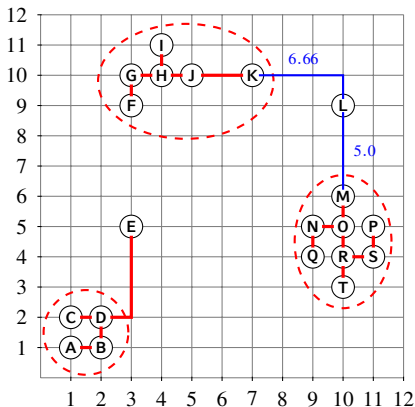
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

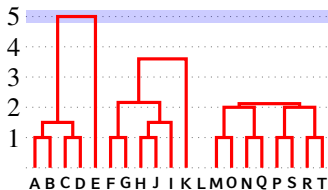
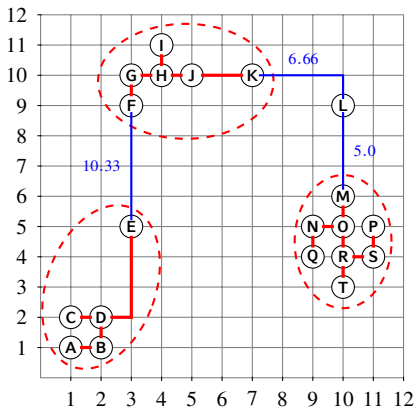
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

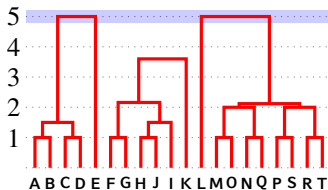
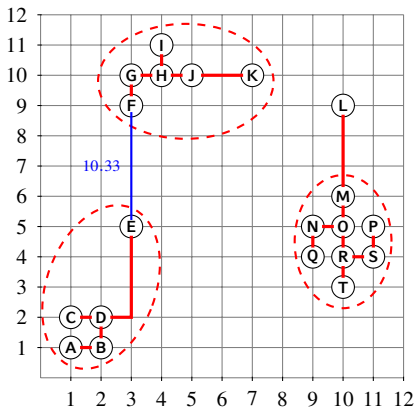
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

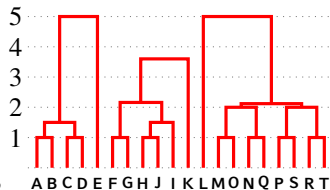
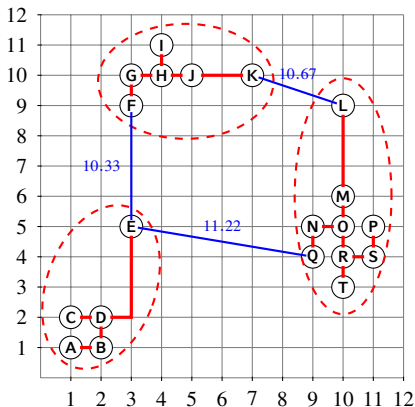
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

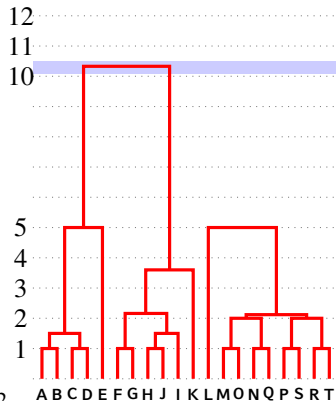
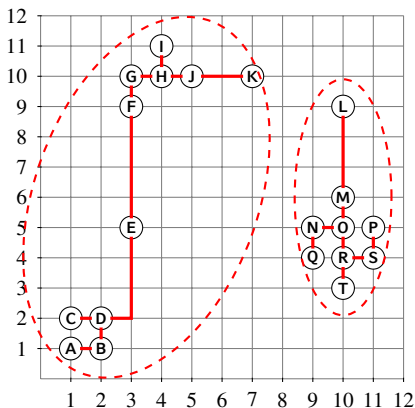
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

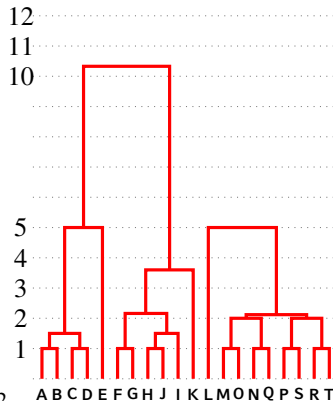
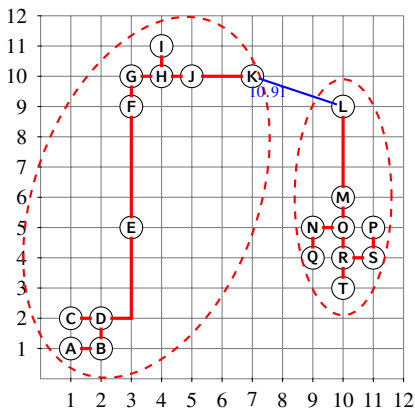
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Group-Average-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

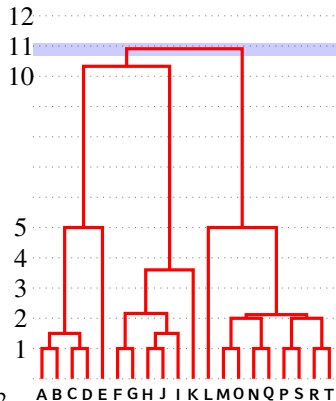
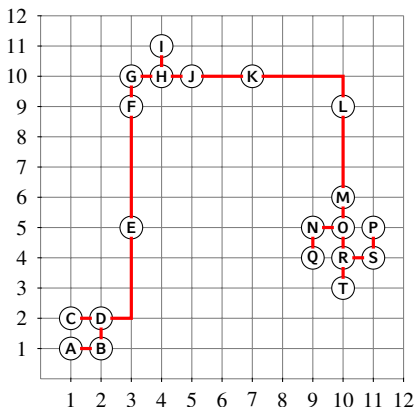
Aufgabe 6-2

Aufgabe 6-3

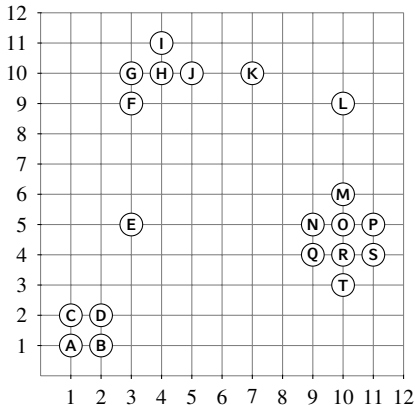
single-link

average-link

complete-link



Complete-Link Clustering



A B C D E F G H J I K L M O N Q P S R T

Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

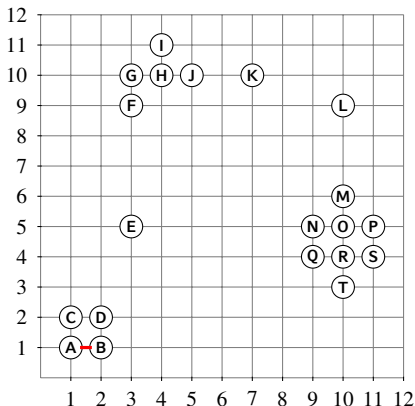
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

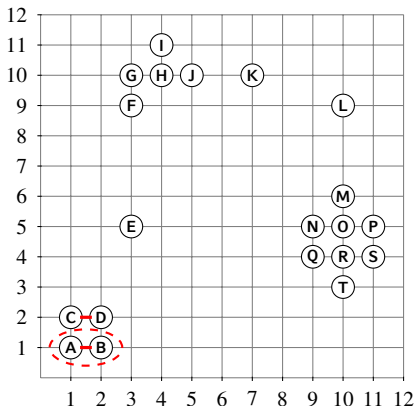
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

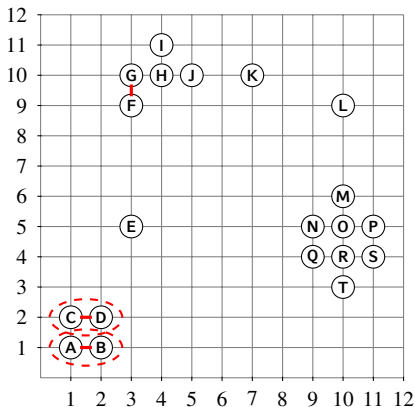
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

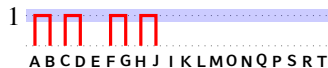
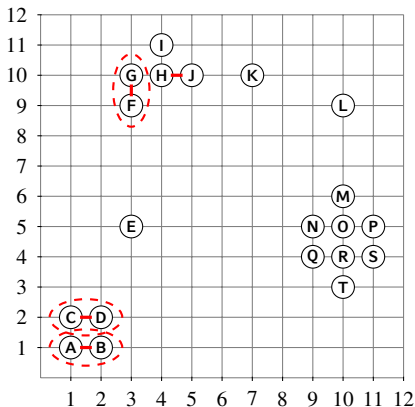
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

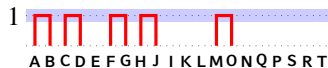
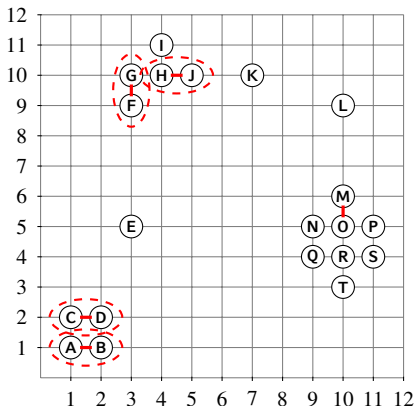
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

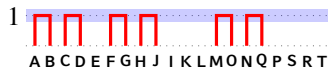
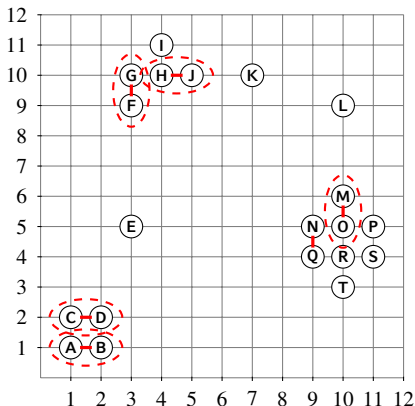
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

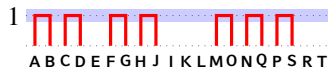
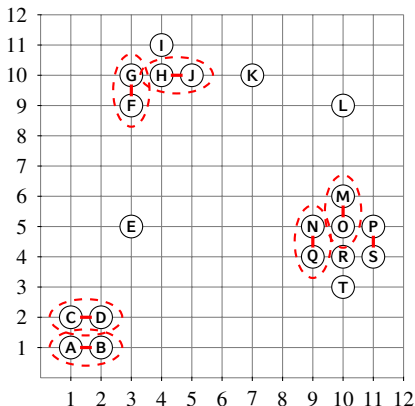
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

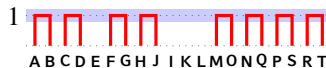
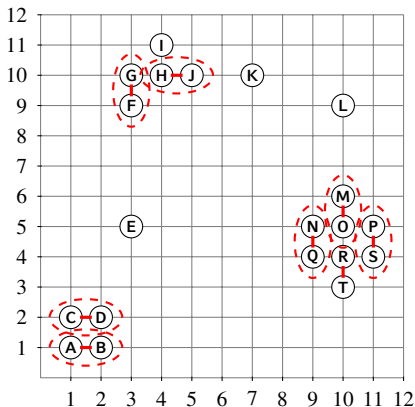
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

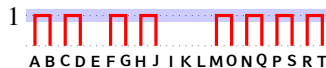
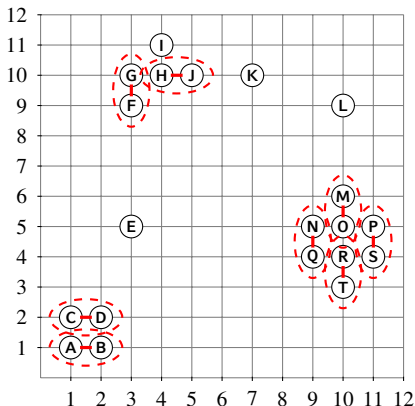
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

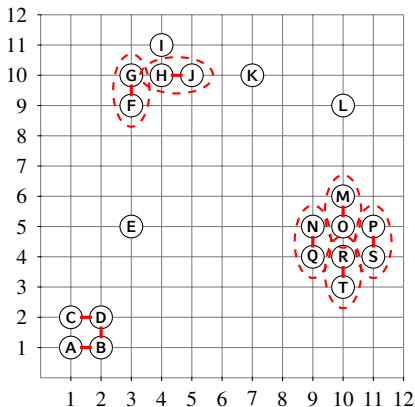
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

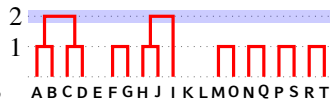
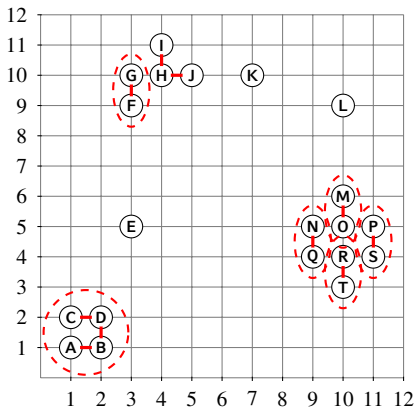
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

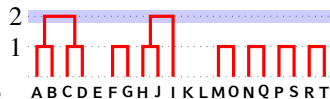
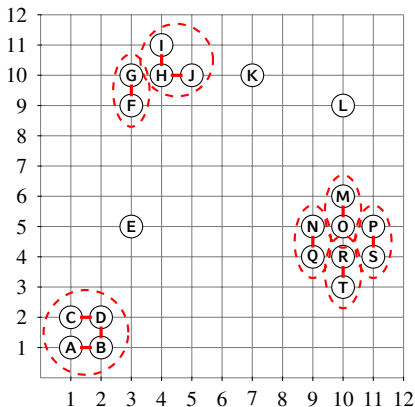
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

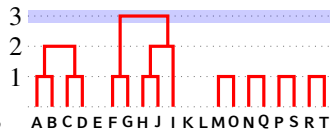
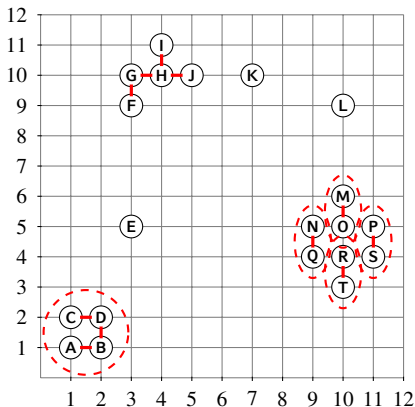
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

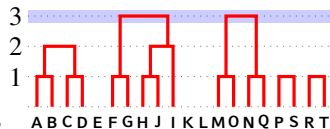
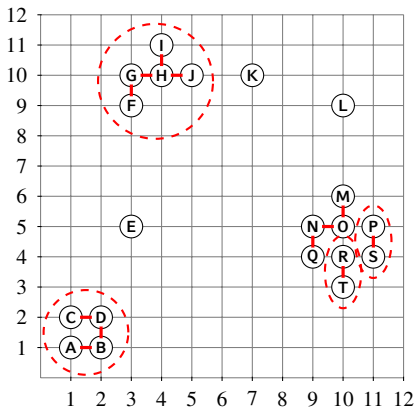
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

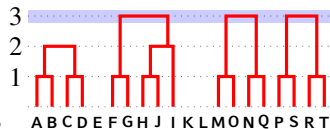
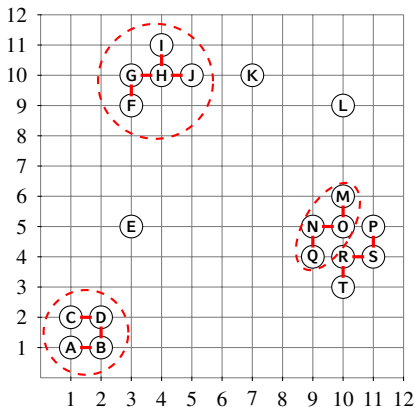
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

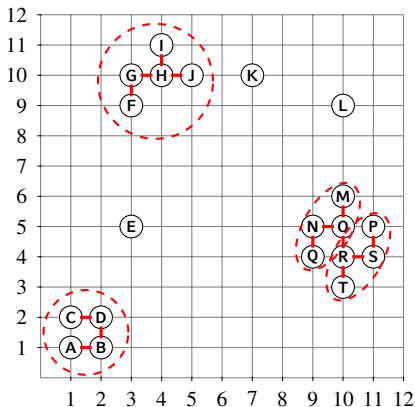
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

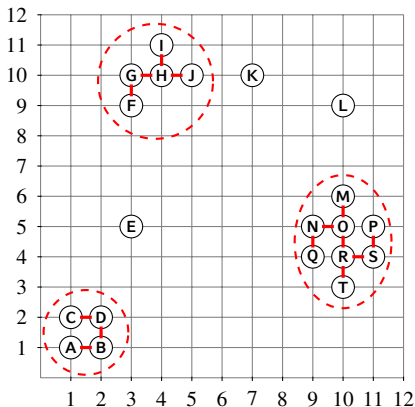
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

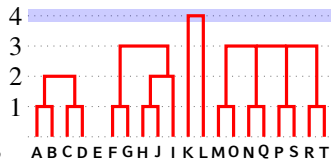
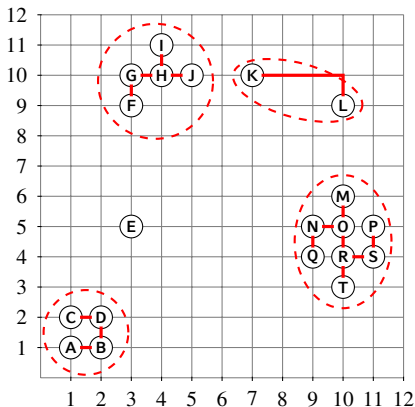
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

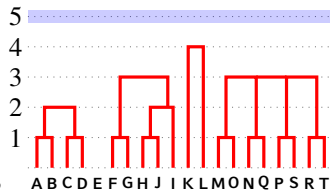
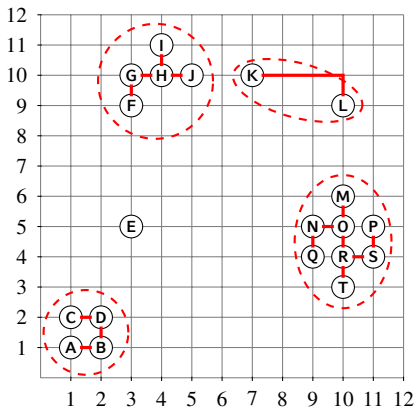
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

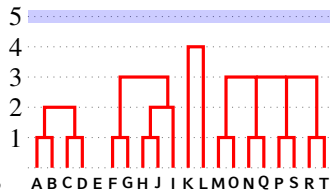
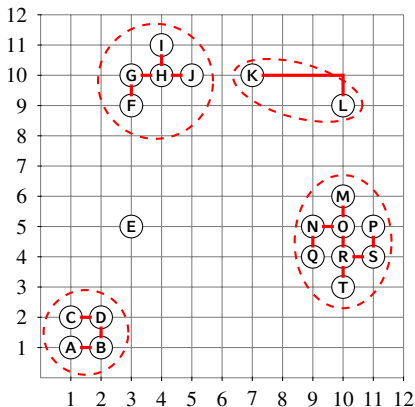
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

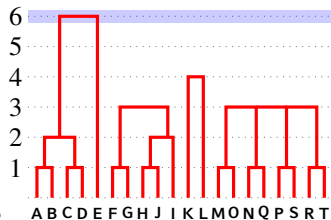
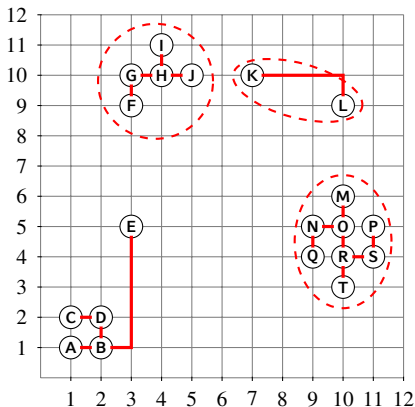
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

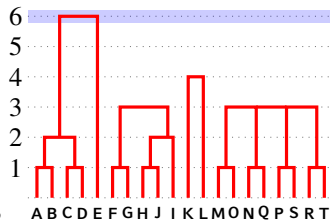
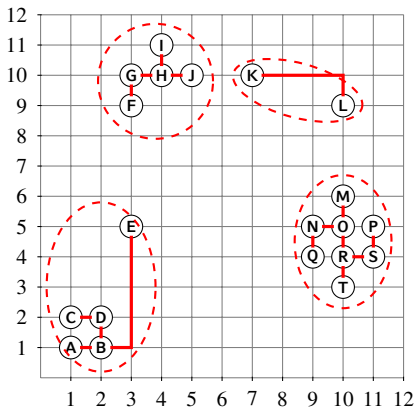
E. Schubert,
A. Zimek

Aufgabe 6-1

Aufgabe 6-2

Aufgabe 6-3

single-link
average-link
complete-link



Complete-Link Clustering

Data Mining
Tutorial

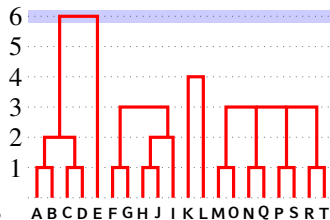
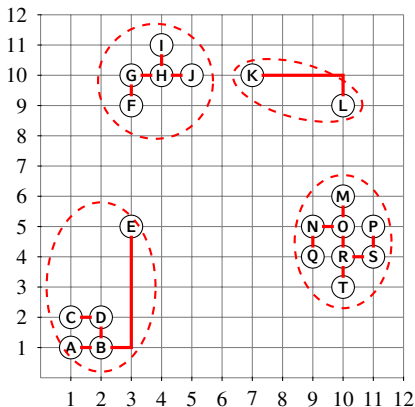
E. Schubert,
A. Zimek

Aufgabe 6-1

Aufgabe 6-2

Aufgabe 6-3

single-link
average-link
complete-link



Complete-Link Clustering

Data Mining
Tutorial

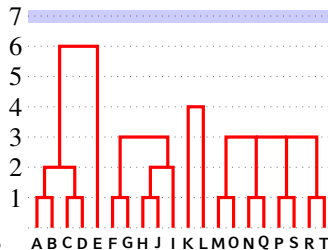
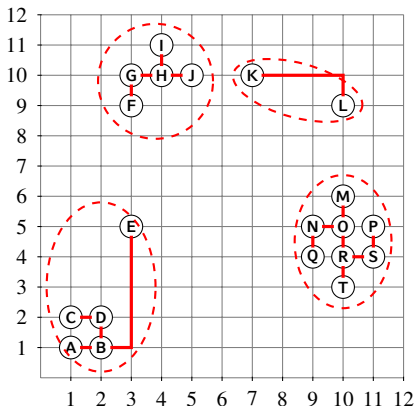
E. Schubert,
A. Zimek

Aufgabe 6-1

Aufgabe 6-2

Aufgabe 6-3

single-link
average-link
complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

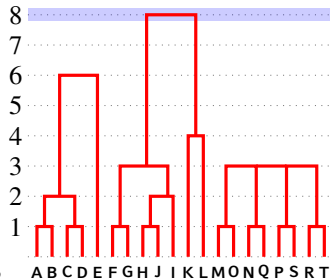
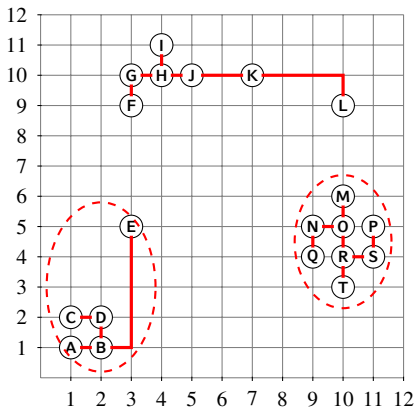
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

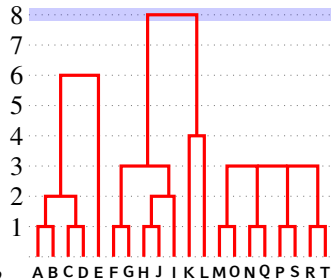
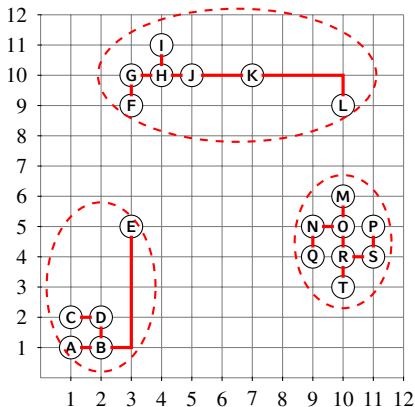
E. Schubert,
A. Zimek

Aufgabe 6-1

Aufgabe 6-2

Aufgabe 6-3

single-link
average-link
complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

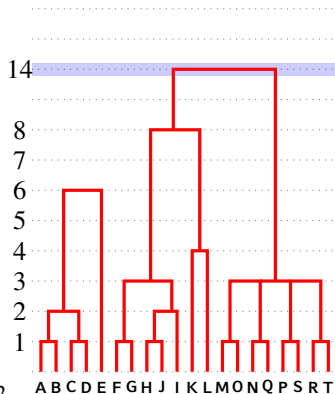
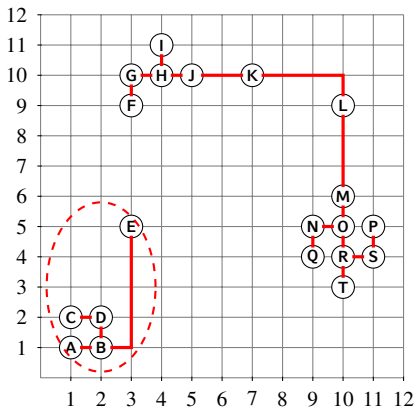
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

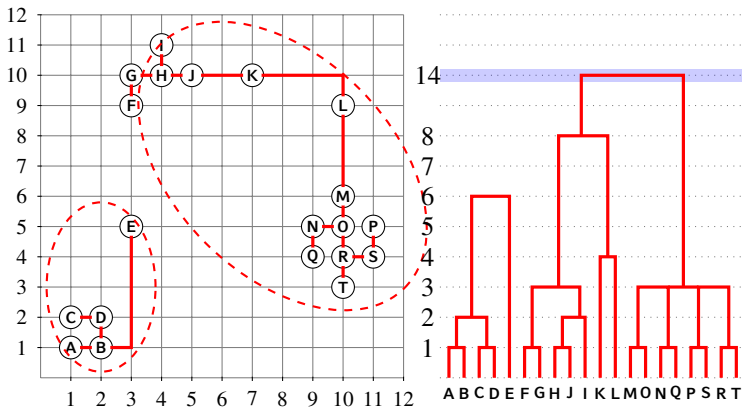
Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link



Complete-Link Clustering

Data Mining
Tutorial

E. Schubert,
A. Zimek

Aufgabe 6-1

Aufgabe 6-2

Aufgabe 6-3

single-link

average-link

complete-link

