

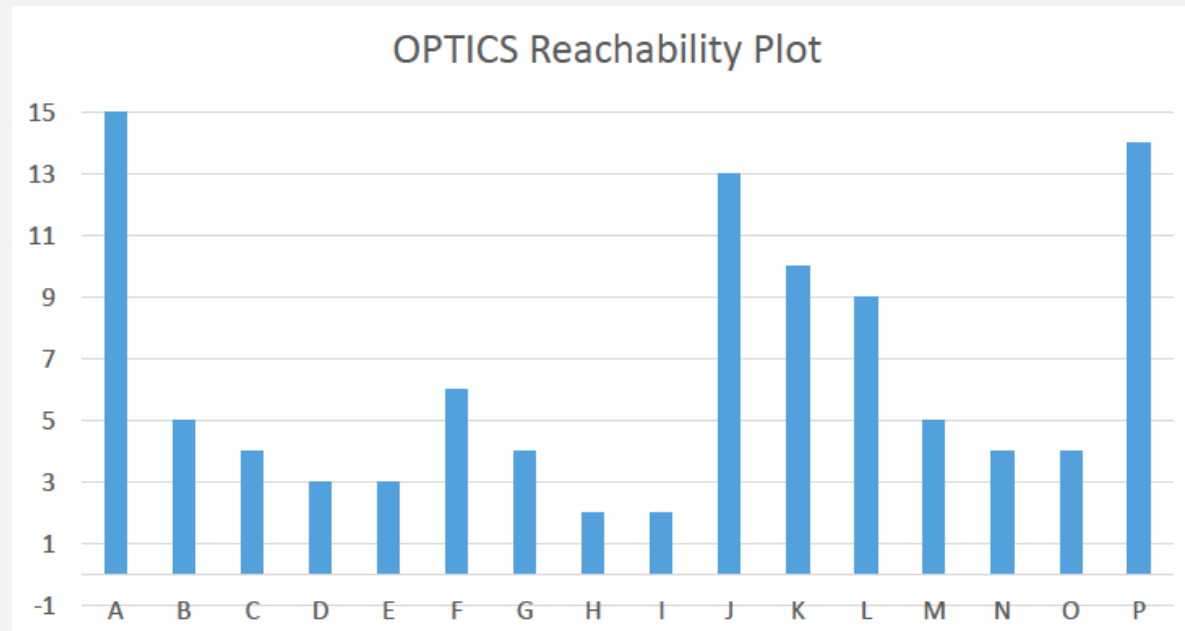
Übung 9: Clustering und Outlier Detection

Knowledge Discovery in Databases I
WS 2017/18

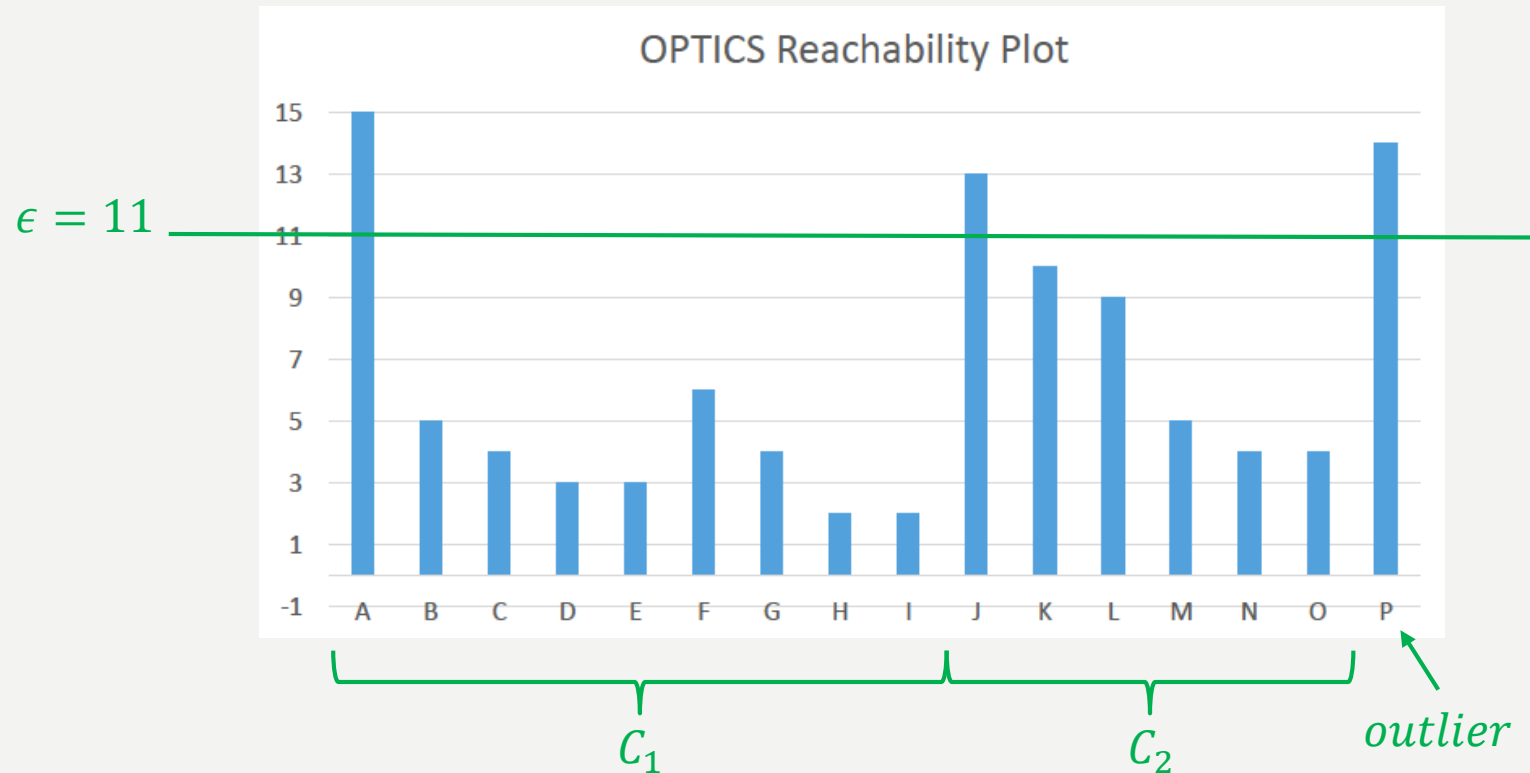




Gegeben ist der folgende OPTICS Erreichbarkeitsgraph:

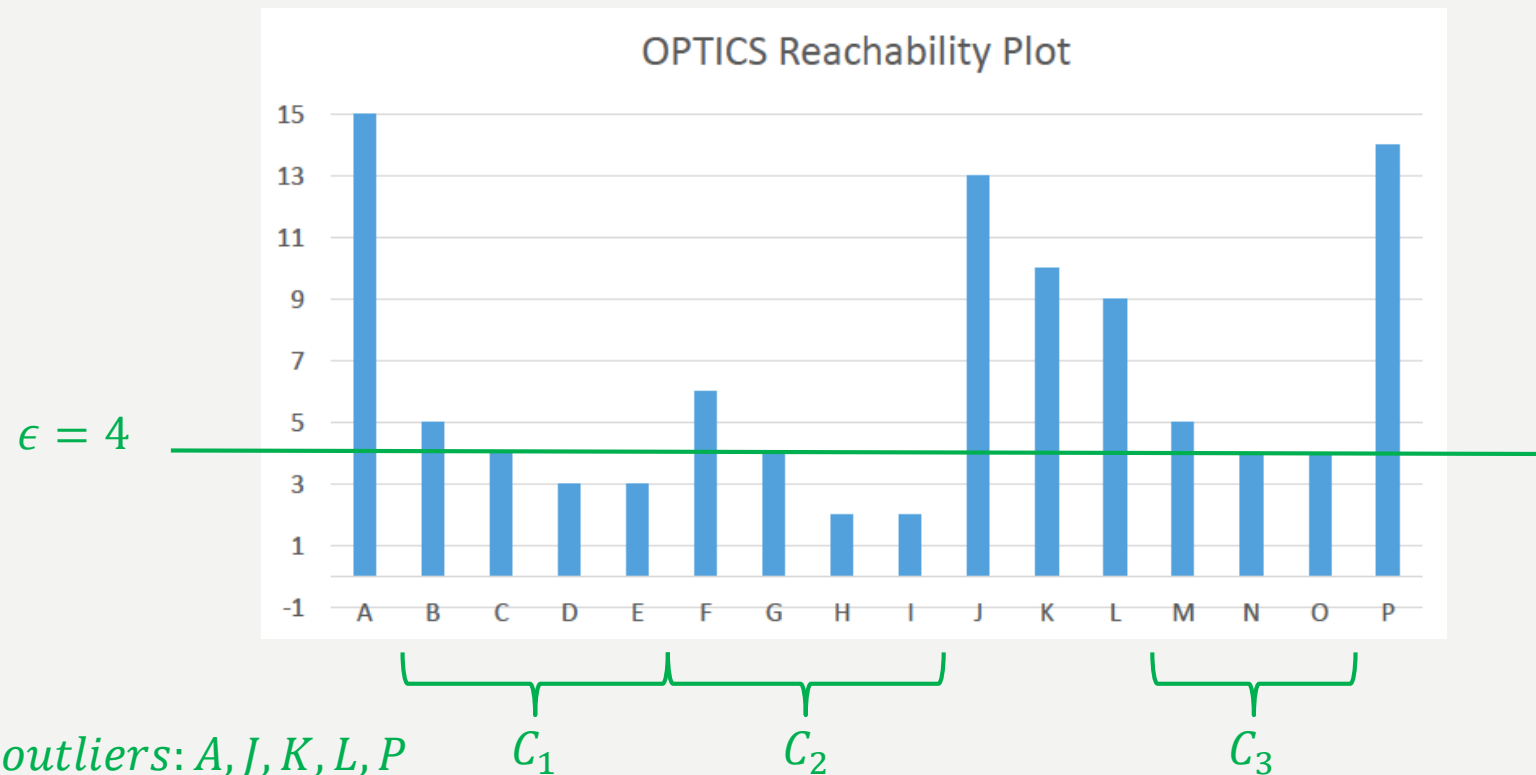


- a) Welches Clustering erhält man für $\epsilon = 11$?
- b) Welche Objekte sind Outlier?





- c) Welches Clustering erhält man für $\epsilon = 4$?
- d) Welche Objekte sind Outlier?





d) Für welches ϵ ist P kein Outlier??

- Für $\epsilon \geq r - \text{dist}(P) = 14$

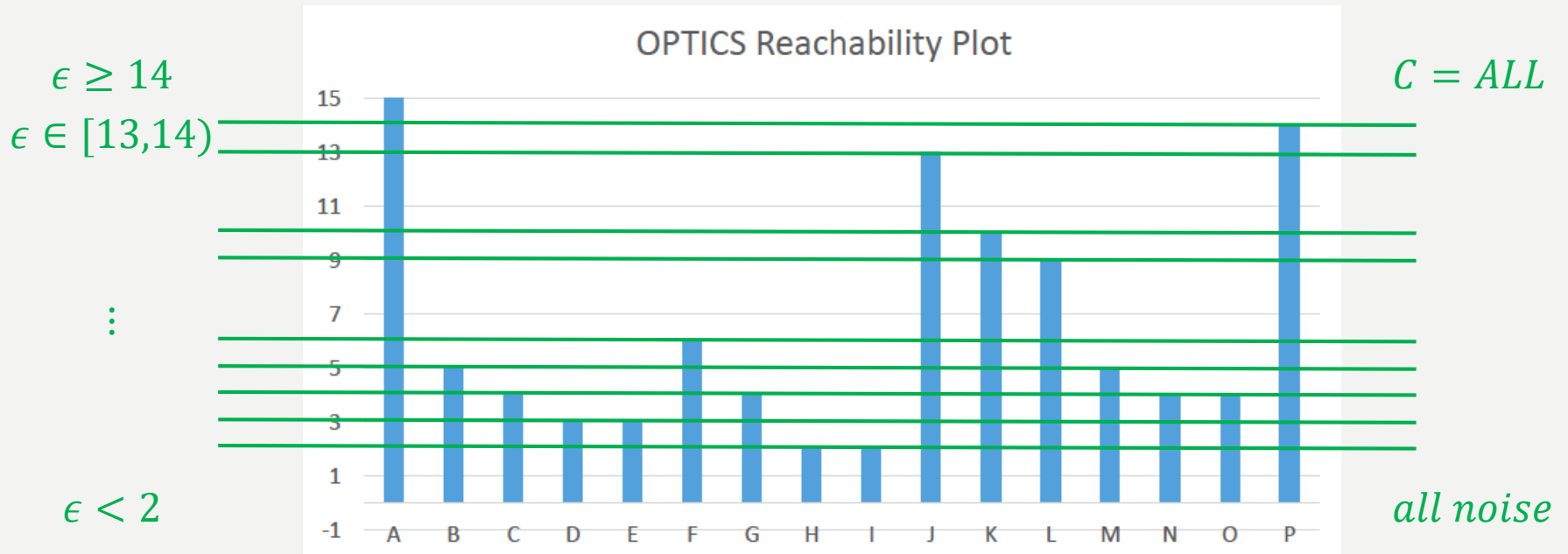
e) Für welches ϵ erhält man das Cluster $\{G, H, I\}$?

- $\epsilon < r - \text{dist}(G) = 4$ (sonst wäre F Teil des Clusters)
- $\epsilon \geq r - \text{dist}(H) = r - \text{dist}(I) = 2$ (sonst wären H und I nicht Teil des Clusters)

f) Ist es möglich, ϵ so zu wählen, dass man ein einelementiges Cluster erhält?

- Ein einelementiges Cluster ist ein Rauschobjekt.
- Wir haben Rauschobjekte z.B. in (a) und (c) gesehen.

g) Wie viele verschiedene Clusterings können in diesem Dataset durch unterschiedliche ϵ gefunden werden?



- Insgesamt können 10 verschiedene Clusterings gefunden werden.