

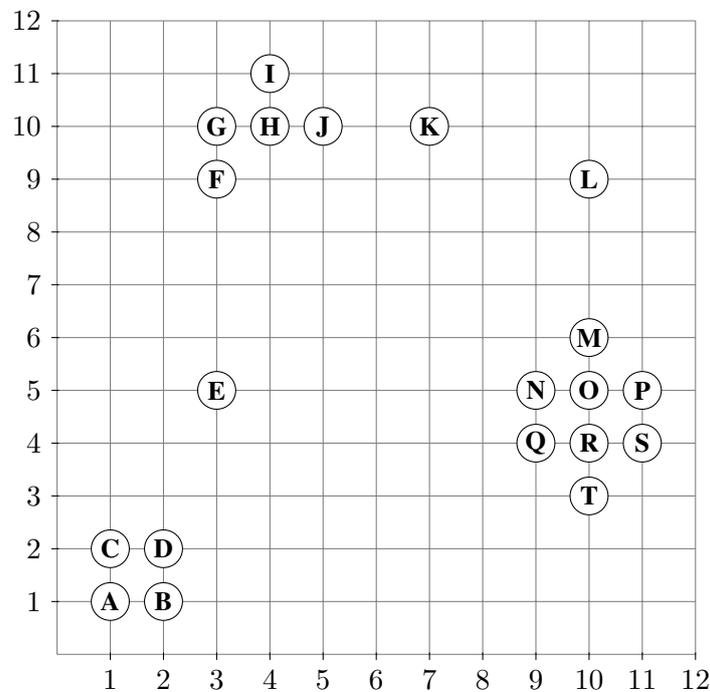
Managing Massive Multiplayer Online Games  
 SS 2012

Übungsblatt 7: Klassifikation und Clustering

Besprechung: 28.06.2012

**Aufgabe 7-1** DBSCAN

Gegeben sei folgender Datensatz:



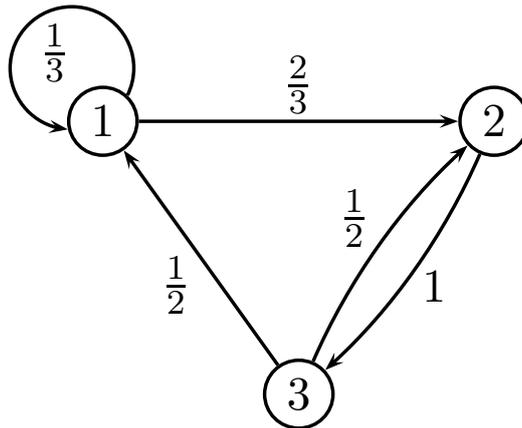
Clustern Sie diesen Datensatz mit Hilfe des DBSCAN algorithmus. Verwenden Sie als Distanzfunktion die Manhattan Distanz

$$L_1(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

und verwenden Sie dabei die Parameter  $\epsilon = 1.1$  und  $minPts = 3$ .

**Aufgabe 7-2** Markov Ketten

Gegeben sei folgende Markov Kette  $M$  in Graph-Darstellung. Dabei stehen Knoten für Zustände, und Kanten für mögliche Übergänge. Kantenlabels entsprechen den Übergangswahrscheinlichkeiten.



- (a) Geben Sie  $M$  in Matrix Schreibweise an. Nehmen Sie dabei an, dass die Startzustände gleichverteilt sind, und Sequenzen nur nach Zustand 3 enden, und zwar mit einer Wahrscheinlichkeit von 50%.
- (b) Wie hoch ist die Wahrscheinlichkeit, die Sequenz  $3 - 1 - 1 - 2 - 3$  zu beobachten?
- (c) Wie hoch ist die Wahrscheinlichkeit, die Sequenz  $2 - 3 - 2 - 1 - 2$  zu beobachten?

**Aufgabe 7-3**    *Levenshtein Distanz*

Bestimmen Sie die Levenshtein Distanz zwischen den Sequenzen *BANANE* und *ANANAS*.

**Aufgabe 7-4**    *OPTICS*

Gegeben sei folgendes OPTICS Erreichbarkeits-Diagramm.



- (a) Enthält der zugrunde liegende Datensatz Cluster? Wenn ja, wie viele?
- (b) Suchen Sie die Cluster. Was ist eine geeignete Schwelle zur Trennung von Clustern?
- (c) Enthält der Datensatz hierarchisch geschachtelte Cluster? Wenn ja, identifizieren Sie sie.
- (d) Welcher Anteil des Datensatzes sind Noise Objekte?