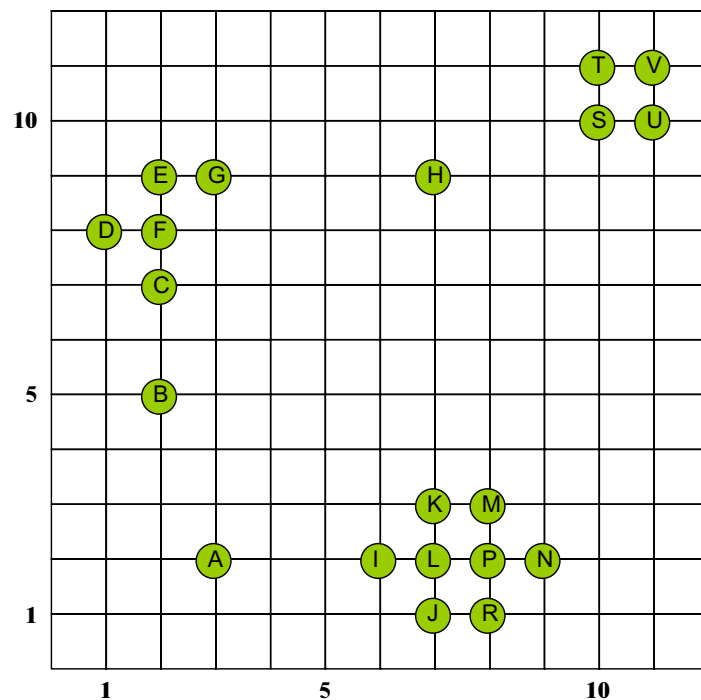


**Knowledge Discovery in Databases**  
 WS 2008/09  
**Übungsblatt 9**

**Aufgabe 9-1** OPTICS

Gegeben sei der folgende 2-dimensionale Datensatz:



Verwenden Sie als Distanzfunktion zwischen den Punkten wieder die Manhattan-Distanz ( $L_1$ -Norm)

Erzeugen Sie mit OPTICS (Pseudocode am Ende des Übungsblattes) jeweils ein Erreichbarkeitsdiagramm für die folgenden Parameter:

- (a)  $\epsilon = 5$  und  $MinPts = 2$
- (b)  $\epsilon = 5$  und  $MinPts = 4$
- (c)  $\epsilon = 2$  und  $MinPts = 4$
- (d)  $\epsilon = \infty$  und  $MinPts = 4$
- (e) Diskutieren Sie, welche Auswirkungen die Parameter  $MinPts$  und  $\epsilon$  haben.

**Aufgabe 9-2** Outlier Detection

Gegeben der Datensatz und die Distanzfunktion aus Aufgabe 1. Berechnen Sie für die Punkte H und L den LOF-Wert für  $MinPts = 3$ .

## Pseudocode OPTICS

```
seedlist =  $\emptyset$  // implemented as a heap
for i = 0 to n-1 do
    if(seedlist =  $\emptyset$ ) then seedlist = {(random_not_handled_point,  $\infty$ )}
    (x, x.reach) = get_and_remove_point_with_min_reach(seedlist)
    x.pos = i
    x.handled = TRUE
    neighbors = rangeQuery(x,  $\epsilon$ )
    x.core = nnDist(x, neighbors, MinPts)
    if(x.core <  $\infty$ )
        for each y  $\in$  neighbors with not(y.handled)
            if (y  $\notin$  seedlist) seedlist = seedlist  $\cup$  {(y, reach-dist(y,x))}
            else
                curr_reach = lookup(seedlist, y)
                update(y, min(curr_reach, reach-dist(y,x)))
        endfor
    endfor
endfor
```