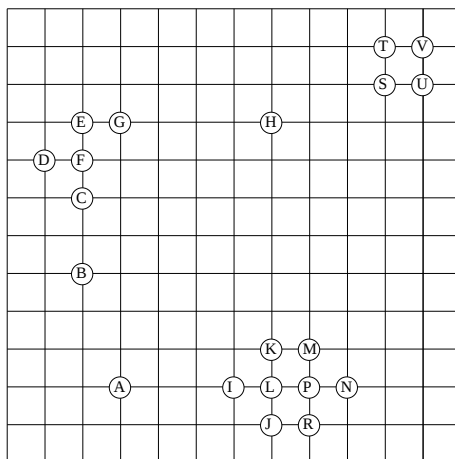


5. Übung „Knowledge Discovery“

Sommersemester 2006

1 Hierarchische Clusterverfahren

1. Entwickeln Sie einen Algorithmus zum Update einer Distanz-Matrix D , die die Abstände zwischen Clustern beim Single-Link-Verfahren speichert.
2. Wie muß die Funktion zum Aktualisieren der zusammengelegten Matrixelemente aussehen, wenn statt des Single-Link-Verfahrens das Average-Link-Verfahren verwendet werden soll?
3. Überlegen Sie sich ein (möglichst kleines) Beispiel, in dem die Single-Link-Methode ein anderes Dendrogramm liefert als die Average-Link-Methode.
4. Berechnen Sie für den folgenden Datensatz jeweils ein Dendrogramm für den Single-Link-Ansatz und für den Average-Link-Ansatz. Nutzen Sie als Distanzfunktion zwischen Punkten die Manhattan-Metrik.



2 Datenkompression zum Vor-Clustering

Erzeugen Sie einen CF-Baum aus den Datenpunkten der folgenden Tabelle. Nutzen Sie die Parameter $B = 2$, $T = 1.5$ und $L = 2$.

1	5	1.5	9	6	6,2	2
---	---	-----	---	---	-----	---