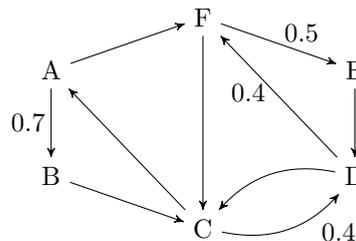


Benutzen Sie – falls nötig – R oder ein ähnliches Programm, um die Lösungen dieses Übungsblattes zu berechnen.

## Aufgabe 1: Markov-Ketten

(1 P.)

Gegeben folgender Zustandsübergangsgraph:



- Fügen Sie die dazu passende Matrix  $P$  in R ein und ergänzen die fehlenden Wahrscheinlichkeiten. Geben Sie eine initiale Verteilung  $\pi^{(0)}$  an und berechnen Sie damit die stationäre Verteilung  $\pi$  mit Hilfe der “Power Iteration”-Methode.
- Ändern Sie die Zustandsübergangswahrscheinlichkeit von  $A \rightarrow F$  auf 0.01 und lassen Sie alle anderen Wahrscheinlichkeiten wie sie sind. Betrachten Sie einige Schritte der “Power Iteration”-Methode. Was passiert mit dem Verteilungsvektor? Konvergiert diese Methode?

## Aufgabe 2: PageRank

(1 P.)

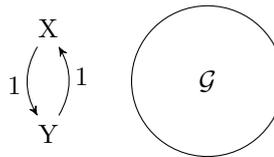
Ein Webcrawler sammelt folgende Informationen über Webseiten:

Seite	Enthält Links zu
A	E
B	C
C	D
D	B, F
E	A
F	C, D

- Geben Sie die passende Zustandsübergangsmatrix  $P$  für die Berechnung von PageRank an und berechnen Sie den PageRank-Vektor mit der “Power Iteration”-Methode für  $\varepsilon = 0.1$ . Wählen Sie ein geeignetes Terminierungskriterium. Wie wirkt sich die Änderung von  $\varepsilon$  auf die Verteilung des PageRanks aus?
- Neben PageRank gibt es auch den HITS Algorithmus von J. Kleinberg zur Linkanalyse. HITS unterscheidet zwischen Hubs und Authorities. Gute Hubs sind Seiten, die auf viele gute Authorities verweisen, und gute Authorities sind Seiten, auf die von guten Hubs verwiesen wird. Für eine Adjazenzmatrix  $A$  ist der Vektor mit den Hub-Scores gegeben durch den (dominanten) Rechtseigenvektor von  $AA^T$ . Der Vektor mit den Authority-Scores ist der (dominante) Rechtseigenvektor von  $A^T A$ .

Berechnen Sie dementsprechend für die Adjazenzmatrix  $A$  (basierend auf den Daten der obigen Tabelle) den Hub- und Authority-Vektor direkt durch Aufruf der Funktion `eigen(A %% t(A))$vectors[,1]` für den Fall des Hub-Vektors (analog für den Authority-Vektor) in R, bzw. in einem Programm Ihrer Wahl. Interpretieren Sie die absoluten Werte anhand des Webgraphen und vergleichen Sie insbesondere den Authority-Vektor mit dem PageRank-Vektor aus a). Berechnen Sie ferner die Singulärwertzerlegung für  $A$  und vergleichen Sie die erste Spalte von  $U$  und  $V$  mit dem Hub- bzw. Authority-Vektor; was fällt dabei auf?

c) Gegeben der folgende Linkgraph:



Hierbei steht  $\mathcal{G}$  für einen vollständigen gerichteten Graphen mit  $|n| = 48$  Knoten,  $X$  und  $Y$  sind nur gegenseitig verbunden. Welchen Pagerank haben  $X$  und  $Y$  für  $\varepsilon = 0.2$ ?

### Aufgabe 3: Warenkorbanalyse

(1 P.)

Der Apriori-Algorithmus benutzt eine Erzeuge-und-zähle-Strategie, um häufige Itemsets zu erzeugen. Kandidaten werden durch die " $F_{k-1} \times F_{k-1}$ "-Strategie erzeugt. Wenden Sie diesen Algorithmus auf die folgende Tabelle mit minimalem Support = 3 an.

TID	Itemset
1	{a, c}
2	{b, c}
3	{a, b, c}
4	{a, b, d}
5	{a, c, d}

a) Zeichnen Sie den Itemsetverband, welcher die Daten in der gegebenen Tabelle repräsentiert. Beschriften Sie jeden Knoten des Verbandes mit den folgenden Buchstaben:

- **N**: Falls das Itemset nicht vom Apriori-Algorithmus als Kandidat vorgeschlagen wird; entweder, weil es garnicht erst erzeugt wurde, oder, weil es später wieder entfernt wurde.
- **F**: Falls das Itemset als *häufig* klassifiziert wurde.
- **I**: Falls das Itemset als *nicht-häufig* klassifiziert wurde.

b) Berechnen Sie alle Assoziationsregeln für  $\text{minconf} = 0.5$  mit dem Algorithmus aus der Vorlesung.