Prof. Dr. Katharina Morik,
Prof. Dr. Claus Weihs,
Dr. Wouter Duivesteijn,
M.Sc. Sarah Schnackenberg,
B.Sc. Melanie Dagge

Übungen zur Vorlesung
Wissensentdeckung in Datenbanken
Sommersemester 2015

Blatt 4

Aufgabe 4.1 (6 Punkte)

Last week, we have used the Apriori algorithm to find frequent sets of films, visited by viewers $z_1, \ldots, z_{10}$, in the transaction database extracted from the following table:

Rep.	Titel	Jahr	z_1	z_2	z_3	z_4	z_5	z_6	z_7	z_8	z_9	z_{10}
a	Star Wars	1977	1	1	0	0	1	0	1	0	1	1
b	E.T. der Ausserirdische	1982	1	1	0	1	1	0	1	0	1	1
c	Indiana Jones	1989	1	1	1	0	0	0	1	0	1	1
d	Otto - der Ausserfriesische	1989	0	0	0	0	0	0	1	0	1	1
e	Wayne's World	1992	1	1	0	1	0	1	0	1	0	1
f	Bang Boom Bang	1999	1	1	0	1	1	0	0	0	1	1
g	Bridget Jones	2001	1	0	0	1	0	0	0	1	0	0
h	Simpsons (Film)	2007	0	0	0	1	1	0	0	0	0	1

1. (1 Punkt) Which set of films is the closure of {Simpsons (Film)}?

2. (3 Punkte) Last week, we have seen that with a minimum support of $\frac{3}{5}$, the frequent itemsets are (using the one-letter representation of the films as provided by the first column): $a, b, c, e, f, \{a, b\}, \{b, f\}$. Which of these itemsets are closed?
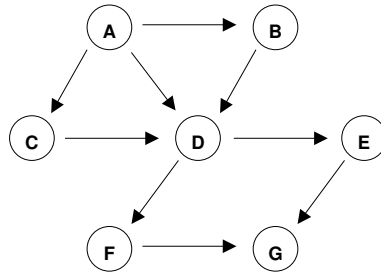
An itemset S is called *free* if S is not included in the closure of any proper subset of S. Formally, S is free if and only if:

$$S' \subset S \Rightarrow S \not\subseteq \text{closure}(S')$$

3. (2 Punkte) Which of the itemsets $\{a, b\}, \{b, c\}, \{b, f\}$ are free?

Aufgabe 4.2 (4 Punkte)
Consider the following web graph:

1. (1 Punkt) What are the Clustering Coefficients C_A and C_D of nodes A and D?

On this web graph, we are going to study part of the HIT procedure (Hyperlink-Induced Topic search). Assume that at the start of the procedure, every node i in the web graph has hubness value $h_i = 1$ and authority value $a_i = 1$.

3. (2 Punkte) Compute for all nodes in the graph the hubness and authority values after one iteration.

4. (1 Punkt) If the initial hubness of node D would have been $h_D = 2$ instead of 1, which nodes would have a higher authority value after one iteration? What would happen to the authority values of the other nodes?