Distributed Information Systems: Spring Semester 2017 - Quiz 2

1. Given the 2-itemsets {1, 2}, {1, 3}, {2, 3}, {2, 5}, {3, 5}, when generating the 3-itemset we will:

   - ☐ a. Have **4** 3-itemsets after the **join** and **4** 3-itemsets after the **prune**
   - ☐ b. Have **4** 3-itemsets after the **join** and **2** 3-itemsets after the **prune**
   - ☐ c. Have **3** 3-itemsets after the **join** and **3** 3-itemsets after the **prune**
   - ☐ **d. Have 2 3-itemsets after the join and 2 3-itemsets after the prune**

2. Given the following transactions {milk, bread}, {eggs, bread}, {milk, eggs, bread}, {eggs}, {milk, eggs}, {milk}

   - ☐ **a. bread⇒milk has support ⅓ and confidence ⅔**
   - ☐ b. eggs⇒milk has support ⅓ and confidence ⅔
   - ☐ c. milk⇒bread has support ⅓ and confidence ⅔
   - ☐ d. milk⇒eggs has support ⅓ and confidence ⅔

3. If we have a graph with nodes {1, 2, 3, 4} and edges {1→2, 1→3, 1→4, 2→3} then the authority values, without normalization, are:

   - ☐ **a. (0, ¼, ½, ¼)**
   - ☐ b. (¾, ¼, 0, 0)
   - ☐ c. (0, ⅓, ⅓, ⅓)
   - ☐ d. (½, ½, 0, 0)

4. If milk⇒{bread, eggs} has confidence c1 and milk⇒bread has confidence c2, then:

   - ☐ **a. c1 <= c2**
   - ☐ b. c2 <= c1
   - ☐ c. c1 < c2 and c2 < c1 are possible
   - ☐ d. c1 = c2

5. Given the following matrix for teleporting in a random walker model:

$$\begin{bmatrix} 0 & 0 & 1 \\ 0 & ½ & 0 \\ 1 & ½ & 0 \end{bmatrix}$$

Which of the following is true (independent of how the link matrix is given):

- ☐ a. A random walker can always reach node 2
- ☐ b. A random walker can always reach any node
- ☐ **c. A random walker can always leave node 2**
- ☐ d. A random walker can never reach node 2

6. Which of the following statements concerning compression of adjacency lists for link indexing is wrong:

   - ☐ a. Compression can exploit the fact that most links of a page point to the page itself
   - ☐ b. Compression can exploit the fact that pages with similar URLs typically have also many outgoing links in common
   - ☐ **c. Exploiting similarity among different adjacency lists will always decrease the cost of encoding of adjacency lists**
   - ☐ d. Compression works well, even if we consider similarity of adjacency lists only for a fraction of neighbouring URLs in the lexicographically order

7. Given the graph 1→2, 1→ 3, 2→3, 3→2, the *PageRank* value of this graph is (without random jumps)

   - ☐ a. (0, 1, 1)
   - ☐ **b. (0, ½, ½)**
   - ☐ c. (⅕, ⅖, ⅖)
   - ☐ d. (1, 0, 0)

8. When computing *PageRank* iteratively the computation ends when

   - ☐ a. The norm of the rank vector exceeds a predefined threshold
   - ☐ b. All nodes of the graph have been visited a predefined number of times
   - ☐ **c. The norm of the difference of rank vectors of two subsequent iterations falls below a predefined threshold**
   - ☐ d. The difference among the Eigenvalues of two subsequent iterations falls below a predefined threshold