

November 30, 2018

## Mathematics and Information, Exercise sheet 12

### Problem 1: (10 points)

Some collection contains six documents, in which the following words occur with frequencies given in parenthesis:

D<sub>1</sub>: *Shannon* (7), *Entropy* (1), *Information* (2) *Cryptanalysis* (1)

D<sub>2</sub>: *Shannon* (1), *Boltzmann* (1), *Entropy* (8), *Information* (5)

D<sub>3</sub>: *Boltzmann* (2), *Entropy* (4), *Clausius* (1), *Information* (3)

D<sub>4</sub>: *Boltzmann* (5), *Entropy* (1), *Clausius* (1), *Heat* (5)

D<sub>5</sub>: *Information* (4), *Shannon* (2), *Code* (6), *Cryptanalysis* (3), *Key* (4)

D<sub>6</sub>: *Kelly* (5), *Shannon* (1), *Bet* (6), *Portfolio* (8), *Information* (3)

D<sub>7</sub>: *Las Vegas* (5), *Shannon* (3), *Kelly* (1)

D<sub>8</sub>: *Shannon* (20), *Juggling* (2), *Roboter* (4), *Unicycle* (1)

Construct two term-document-matrix for this collection in which all column vectors have length one, one with entry one if the term occurs and zero otherwise, the other one with entry equal to the frequency of the term! Then rank the eight documents by their relevance for the following queries according to either of these matrices:

- Shannon*
- Information Shannon*
- Information Entropy*

### Problem 2: (10 points)

In the course, we computed the singular value decomposition of the matrix

$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 \end{pmatrix}$$

and used it to replace A by a nearby matrix of rank two.

- Do the same using the QR decomposition!
- How does this matrix rank the relevance of the five web pages for the query *ranking of web pages*?