# Tutorial Network Analysis

Freie Universität Berlin, SS 2016/17
Martin Vingron · Alena van Bömmel

**Assignment 3**
**Due date: 14.7.2017 9:00AM before the lecture**

Include all important steps of your calculations/solutions. Give the important parts of your code or send the complete code to: `alena.vanboemmel@molgen.mpg.de`. Form groups of max. 2 students to solve the problems.

Name(s): Matrikelnr.:

**Problem 1** (*30 Points; SVD*)**.** Generate in R matrix $X$ with positive entries from Gaussian distribution, with $m$ rows and 10 columns:

```
X <- matrix(abs(rnorm(m*10)),m,10)
```

Let $m$ vary as follows: $m \in \{5, 10, 15, \ldots, 300\}$. Now calculate the (estimated) covariance matrix $S = X \cdot X^T$.

(A) What is the rank of matrix $S$ for different $m$? Create a scatterplot and justify your result.
*Hint: Use function* `rankMatrix` *from library* `Matrix`*.*

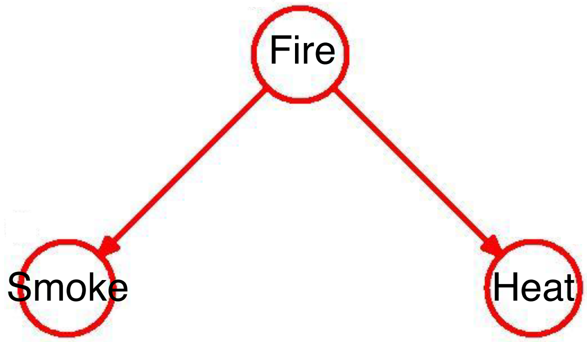
(B) Calculate the Singular Value Decomposition (SVD) of matrix $S$ and compute the pseudoinverse $S^+$.
*Hint: Use the standard function* `svd` *and the formula from the lecture.*

(C) Calculate the product $S \cdot S^+$. By definition, this should be an identity matrix for singular matrices.

(D) Calculate the Hamming distance between $S \cdot S^+$ and identity matrix of rank $m$ $I_m$.
*Hint: Use function* `hamming.distance` *from library* `e1071`*. This function calculates the distance between 2 vectors, not between matrices!*

(E) Plot the Hamming distance versus $m$. Interpret your results.

**Problem 2** (*40 Points; Bayesian Network*)**.** Consider the following Baysian network with 3 variables: {F=Fire,S=Smoke,H=Heat}.




Variable $F$ represents an existence of a fire ($F = 1$), $S = 1$ if we se a smoke; $H = 1$ if we observe heat and vice versa:

$$
\begin{aligned}
F &= \text{Fire} \quad (1 = \text{Fire}, 0 = \text{No fire}) \\
S &= \text{Smoke} \quad (1 = \text{Observed smoke}, 0 = \text{No observed smoke}) \\
H &= \text{Heat} \quad (1 = \text{Observed heat}, 0 = \text{No observed heat})
\end{aligned}
$$

We know the following probabilities:

$$
\begin{aligned}
P(F = 1) &= 0.1 && \text{(suppose I light up a fire in my fireplace every 10 days)} \\
P(S = 1|F = 1) &= 0.9 && \text{(if there is a fire, we will observe very likely the smoke...etc. )} \\
P(S = 0|F = 1) &= 0.1 && \text{(obviously...the same for all other complementary probabilities)} \\
P(S = 1|F = 0) &= 0.001 \\
P(H = 1|F = 1) &= 0.99 \\
P(H = 1|F = 0) &= 0.0001
\end{aligned}
$$

(A) Write down the joint distribution of the Bayesian network $P(F, S, H)$ using the conditionally independent effects. Are Smoke and Heat somehow independent?

(B) Before we observe any data, what is the prior probability that there is no fire?

(C) Now suppose that we observe smoke. Use Bayes' theorem to evaluate the posterior probability that there is no fire observing smoke and compare it to prior probability that there is no fire.

(D) Next suppose that we also check the temperature and find that it is very hot. We have now observed the states of both smoke and heat. Compute the posterior probability that there is fire given the observations of both heat and smoke state and compare it to the posterior probability of $P(F = 0|S = 1)$.

*Hint: Use Bayesian formula and the total probability property:* $P(A, B) = \sum_c P(A, B, C)$

**Problem 3** (*30 Points; Bayesian networks, Parameter estimation*)**.** Consider a Bayesian network with 2 variables and with the following structure: $A- > B$. We observed 4 samples for these variables:

| A | B |
|---|---|
| 1 | 1 |
| 0 | 0 |
| 1 | 1 |
| 0 | 0 |

(A) Calculate the maximum likelihood (ML) estimator of the parameter $\theta = P(B = 1|A)$. Recall $\hat{\theta}_{ML} = \max_\theta P(D|G)$, where $D$ =Data, $G$ =Graph(Model).

(B) Calculate the model evidence $P(D|G)$. The prior distribution of $\theta$ is: $P(\theta) = P(\theta|G) \sim \beta(3, 3)$.

(C) Calculate the posterior distribution $P(\theta|D, G)$.