Logic and Theory of Discrete Systems
RWTH AACHEN UNIVERSITY

Prof. Dr. M. Grohe E. Fluck, N. Runde

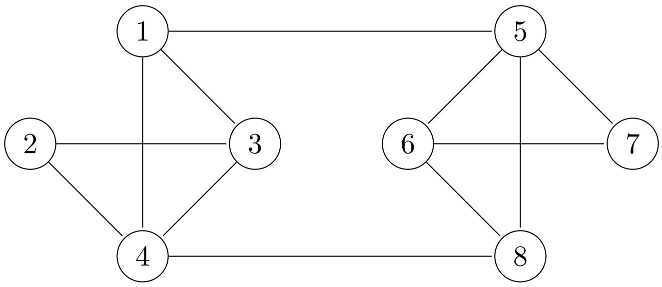
Exercise Sheet 5

Due date: Monday, June 12 until 13:00

- Please upload your solutions to Moodle.
- Hand in your solutions in groups of **two to three students**.
- Hand in the solutions of your group as a single PDF file.
- The solutions for this exercise sheet will be published after the deadline.
- A discussion regarding this exercise sheet will take place on **Friday, June 16 14:30** in room AH II.
- Note that Exercise 1 gives 0 points and will not be corrected.

Exercise 1 (Spectral Clustering) **0 points**

Consider the following graph $G = (V, E)$ with node set $V = \{1, \ldots, 8\}$.



Let $s\colon \{1, \ldots, 8\}^2 \to \mathbb{R}_{\geq 0}$ be the similarity measure that is defined by

$$s(v, w) = \begin{cases} 1 & \text{if } v = w, \\ 1 & \text{if } (v, w) \in E, \\ 0 & \text{otherwise.} \end{cases}$$

We want to cluster the vertices of this graph using spectral clustering methods. Solve the following tasks.

a) Compute the Laplacian L of the similarity matrix S associated with s.

b) Compute the two smallest eigenvalues λ_1 and λ_2 of L along with corresponding eigenvectors $\mathbf{u}_1, \mathbf{u}_2 \in \mathbb{R}^8$.

Hint: Use a computer to solve this task. Your solution should be sufficiently precise but needs not be exact. Round your final values to three decimal places.

c) What is special about the eigenpair belonging to the smallest eigenvalue of L? Justify your answer.

d) Plot the points $\big((\mathbf{u}_1)_i, (\mathbf{u}_2)_i\big)$ for all $i = 1, \ldots, 8$.

e) Using your plot from part d), discuss which clustering of V is returned by the Spectral Clustering algorithm on S with $k = 2$.

Exercise 2 (Naive DNF Counting) **1+4+5=10 points**

We consider the following naive rejection sampling algorithm for counting the number μ of satisfying DNF assignments:

1. For some m, independently sample m assignment $\alpha_1, \ldots, \alpha_m$ for the n variables, uniformly at random from the 2^n possible assignments

2. For each i, let $y_i := \begin{cases} 1 & \text{if satisfing assigment,} \\ 0 & \text{else} \end{cases}$

3. Return $\hat{\mu} = \frac{2^n}{m} \sum_{i=1}^{m} y_i$ as the estimate.

a) Determine the minimum number of samples that is needed to return an estimate that deviates from μ by at most ε with a probability of $1 - \delta$, using Lemma 6.1.

b) Improve the bound computed in (a) to $m \geq \frac{2^{n-1} \ln(2/\delta)}{\varepsilon^2}$. Proof your result.

Hint: You can follow the general idea of the proof of Theorem 6.1 in the lecture.

c) Assume μ can be bounded by some polynomial $\alpha(n)$. Show that even after sampling $2^{n/2}$ assignments the probability of finding even a single satisfying assignment is exponentially small in n.

Hint: The following inequality may be useful for achieving the desired bound (you may use it without showing it to hold):

$$1 + \alpha x \leq (1 + \alpha)^x \qquad \text{for all } x \geq 1 \text{ and } \alpha \geq -1. \tag{1}$$

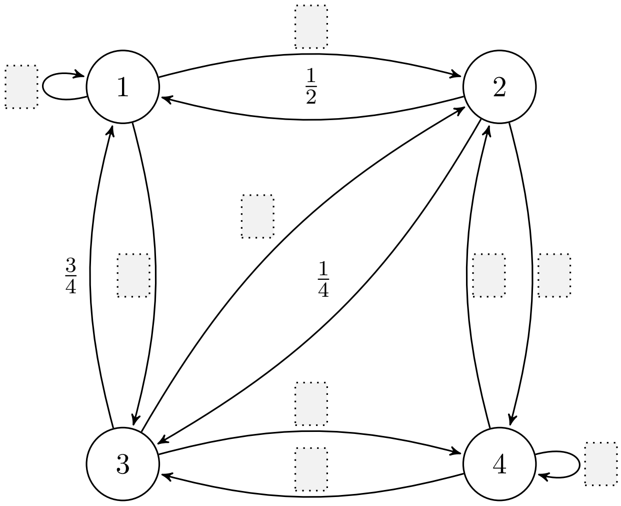
Exercise 3 (Symmetric Markov Chains) **5 points**

A Markov chain $\mathcal{Q}$ is called *symmetric* if its transition matrix Q is symmetric.

Show that there exists a unique probability vector $\boldsymbol{\pi} \in \mathbb{R}^{1 \times n}$ such that $\boldsymbol{\pi}$ is the stationary distribution of $\mathcal{Q}$ for all connected, symmetric Markov chains $\mathcal{Q}$ with state space $[n]$.

Exercise 4 (Completing Markov Chains) 5 points

Consider the following incomplete graphical representation of a Markov chain with 4 states. Therein, all missing edges have probability 0.



Fill in the gaps in the graphical representation such that the stationary distribution of the Markov chain becomes $\boldsymbol{\pi} = (\pi_1, \pi_2, \pi_3, \pi_4) = \left(\frac{1}{2}, \frac{1}{12}, \frac{1}{4}, \frac{1}{6}\right)$ where π_i denotes the probability of being in state i. Justify your solution.

Hint: You may use Lemma 6.11 to solve this exercise.