Visualization
Prof. Bernhard Schmitzer, Uni Göttingen, summer term 2023

Problem sheet 5

- *Submission by 2023-07-11 18:00 via StudIP* ***as a single PDF****. Please combine all results into one PDF. If you work in another format (markdown, jupyter notebooks), convert them to PDF before submission.*
- *Use Python 3 for the programming tasks as shown in the lecture. If you cannot install Python on your system, the GWDG jupyter server at* `https://jupyter-cloud.gwdg.de/` *might help. Your submission should contain the final images as well as the code that was used to generate them.*
- *Work in groups of two to three. Clearly indicate names and matrikelnr of all group members at the beginning of the submission.*

Exercise 5.1: professors, students and papers.
In our dataset there are 7 persons, indexed by numbers $\{0, \ldots, 6\}$. Persons 0 and 1 are professors, persons 2 to 6 are doctoral students. Students $\{2, 4, 5\}$ are advised by professor 0; students $\{3, 6\}$ are advised by professor 1. Together they have published four papers, indexed by numbers $\{0, \ldots, 3\}$. The amount of contribution to each paper is captured by the following list where each triple $[a, b, c]$ indicates that person b contributed to paper a a fraction of c, normalized such that the total contribution to each paper sums to 1:

```
[[0, 0, 0.4], [0, 2, 0.6], [1, 0, 0.3], [1, 4, 0.3], [1, 5, 0.2], [1, 3, 0.2],
 [2, 0, 0.2], [2, 1, 0.4], [2, 3, 0.4], [3, 1, 0.3], [3, 3, 0.2], [3, 6, 0.5]]
```

Visualize this dataset. Your visualization should contain the following information:

- role of each person, student-advisor relation,
- contribution (and strength thereof) to papers, total contribution of each person

You can complete this exercise with any software you want, simply include the resulting figure to your submitted pdf.

Exercise 5.2: animating the law of large numbers.
Consider a univariate standard normal distribution (i.e. zero mean and unit standard deviation) and let $(x_i)_{i=1}^N$ be independent samples from it (which can numerically easily be approximated with `numpy.random.normal`). As shown earlier in the lecture, for given $N > 0$, one can use the points, e.g. in a histogram or via a kernel density estimation, to estimate the density of the underlying normal distribution. For fixed parameters of the density estimation (width of bins or bandwidth of kernel), this estimate should stabilize as $N \to \infty$. This is an instance of the law of large numbers. Create an animation that shows this effect, i.e. the evolution of the estimated density with increasing N, in comparison with the true underlying density.
Hints: Feel free to use code for creating animations that was presented in the lecture. Your submission should contain a screenshot of your animation as well as the generating code, in a way that can easily be extracted for reproduction.