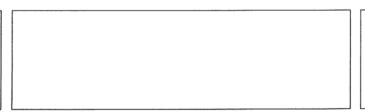Prof. Dr. Christian L. Müller
Ludwig-Maximilians-Universität München
Institut für Statistik
Ludwigstr. 33
80539 München
christian.mueller@stat.uni-muenchen.de

M.Sc. Thesis Proposal: *Analyzing covid-19 data as three-dimensional tensor*

Objective: The purpose of this M.Sc. thesis proposal is to retrospectively analyze the covid-19 incidence data from the German Robert-Koch-Intitut (RKI) in three dimensions: age, region and time. This makes it possible to show which age groups were infected at which point in time in which parts of Germany. To uncover age specific spatio-temporal patterns from the data, we consider the incidence data as a three-dimensional tensor and apply non-negative tensor factorization (also called non-negative PARAFAC decomposition). A german report of the analysis can be found here: *Ein Blick zurück auf die zweite Welle - wer war wann wo infiziert?* https://www.covid19.statistik.uni-muenchen.de/pdfs/codag_bericht_12.pdf

Plan and deliverables: A successful completion of the M.Sc. thesis requires the following computational and scientific advances. The first step is to compare existing implementations in R and Python (potentially Matlab) of non-negative tensor factorization and to extend or reimplement them if necessary. In a further step, model-based approaches -such as Bayesian poisson tensor factorization (https://arxiv.org/abs/1506.03493)- should be taken into account. This allows to include the population size as a model offset, which is neglected in classical tensor factorization of incidence values. Existing models might have to be adapted to the specific problem. In this project, a concise and insightful presentation of the results is of high importance. For this purpose, measures and restrictions (such as school closures) in the individual districts can also be taken into account. A write-up in thesis form and commented code on GitHub are mandatory deliverables at the end of the thesis.