



# Mathematisches Kolloquium

## *Density-Based Cluster Analysis*

**Sprecher: Prof. Dr. Ingo Steinwart | Universität Stuttgart**

Eingeladen von Prof. Dr. Anna Dall'Acqua

**06.05.2022 | 14:30 Uhr | Hörsaal H8**

A central, initial task in data science is cluster analysis, where the goal is to find clusters in unlabeled data. One widely accepted definition of clusters has its roots in a paper by Carmichael et al., where clusters are described to be densely populated areas in the input space that are separated by less populated areas. The mathematical translation of this idea usually assumes that the data is generated by some unknown probability measure that has a density with respect to the Lebesgue measure. Given a threshold level, the clusters are then defined to be the connected components of the density level set. However, choosing this threshold and possible width parameters of a density estimator, which is left to the user, is a notoriously difficult problem, typically only addressed by heuristics.

In this talk, I will describe a simple algorithm based on a rather generic density estimator that can find the smallest level for which there are more than one connected component in the level set. Unlike other cluster algorithms this approach is fully adaptive in the sense that it does not require the user to guess crucial hyper-parameters.

I will discuss some practical aspects of the algorithm and finally, I present some numerical illustrations.

**Der Vortrag ist für ein breites Publikum geeignet**