STRUCTURES
CLUSTER OF
EXCELLENCE

UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Exploring data with TDA

3rd Session of the

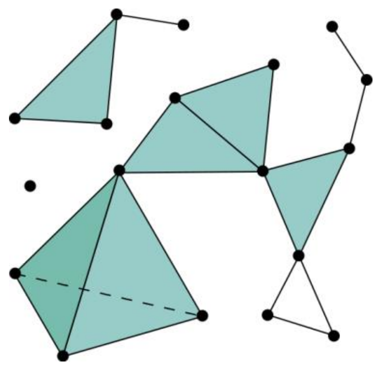
Python Course on Topological Methods in Data Analysis

Daniel Spitz, ITP Heidelberg

Heidelberg, October 28, 2020

The procedure

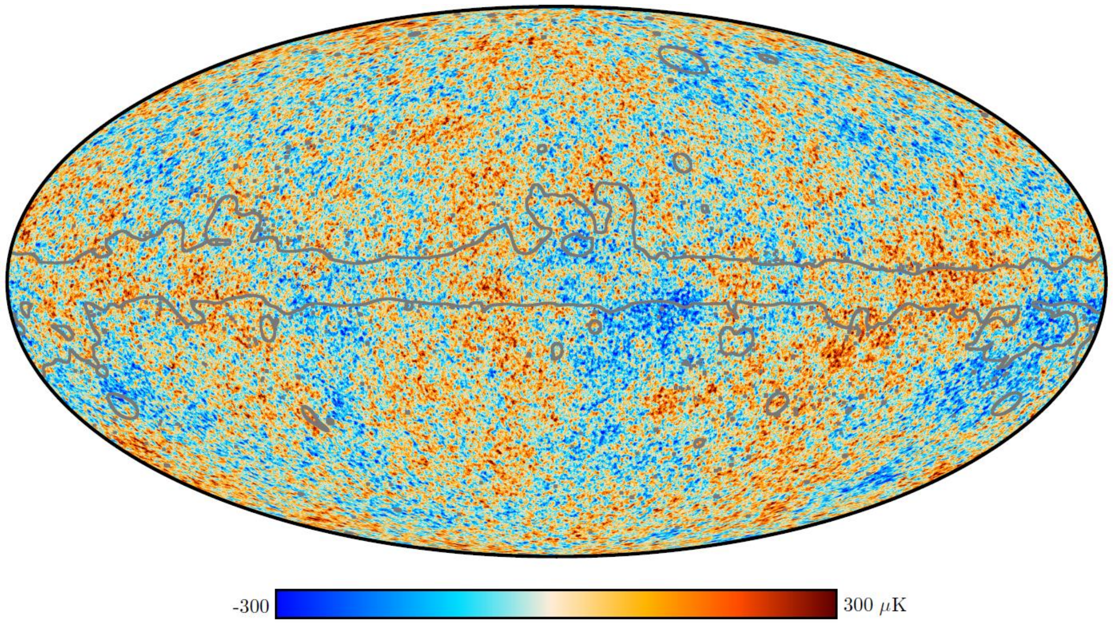
- Brief introduction to the available data
- Individual explorations in subgroups
  until 5:15pm
- Brief presentations of the findings in each group
- Further material
- Final remarks

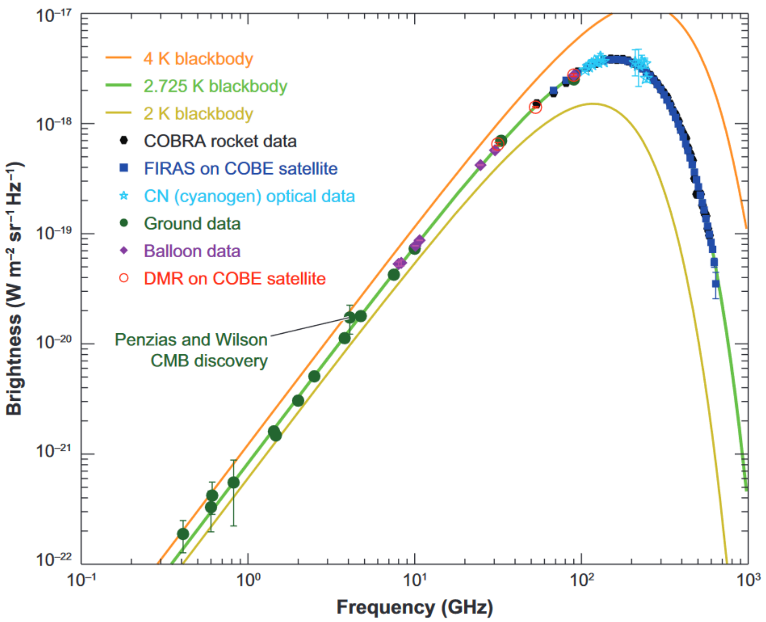


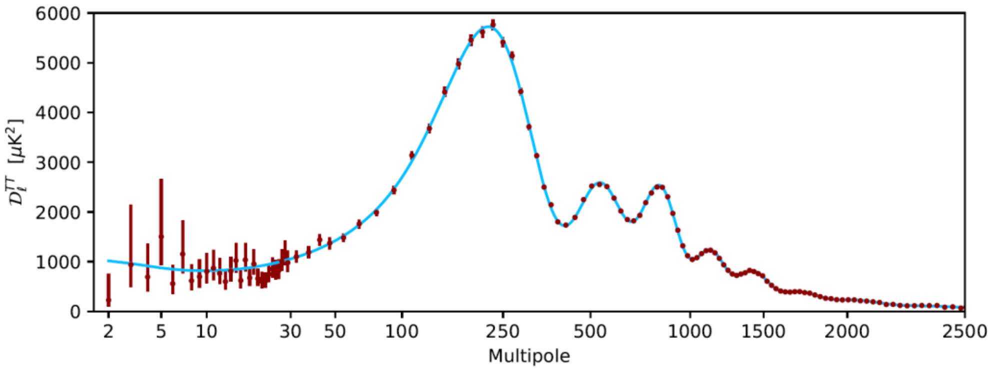
Available data

1. Cosmic Microwave Background (CMB) temperature maps
2. scRNAseq: finding sparse representations of gene expression matrices via Mapper
3. EEG recordings of a human adult performing a visuospatial working memory task
4. Action and outcome activity state patterns of the anterior cingulate cortex of rats
5. EEG measurement data of different empathy tasks

1: CMB temperature fluctuations



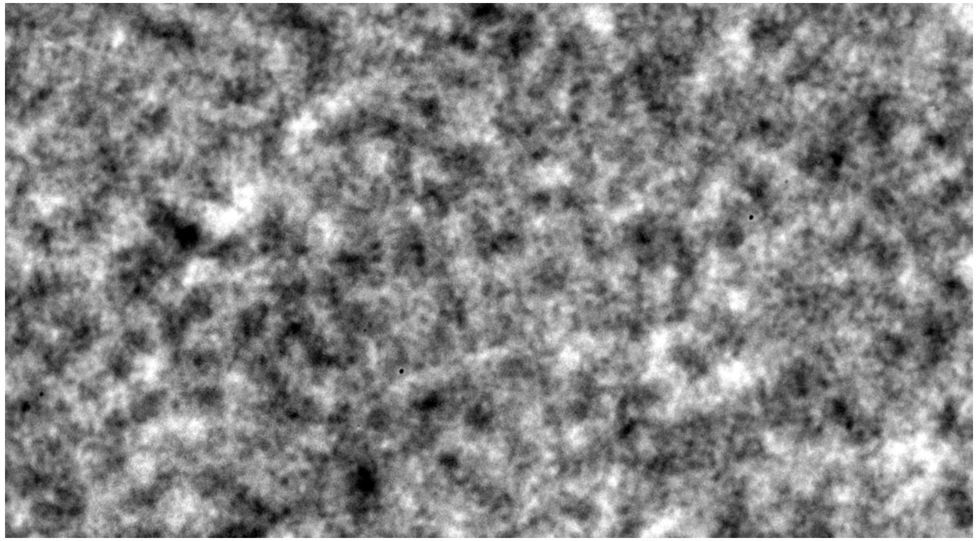




Top left and bottom right: ESA and the Planck Collaboration, 2018 Results
Top right: Samtleben, Staggs, Winstein, Annu. Rev. Nucl. Part. Sci. 57, 2007

1: CMB temperature fluctuations

- Use excerpts of CMB temperature fluctuation maps from the Planck Collaboration
- Compare CMB temperature fluctuation maps with random field models in spirit of a comparison with simple estimators of inflationary models
- Further information on Planck website
- Inspired by Pranav *et al.*, Astronomy & Astrophysics 627, 2019, arXiv:1812.07678



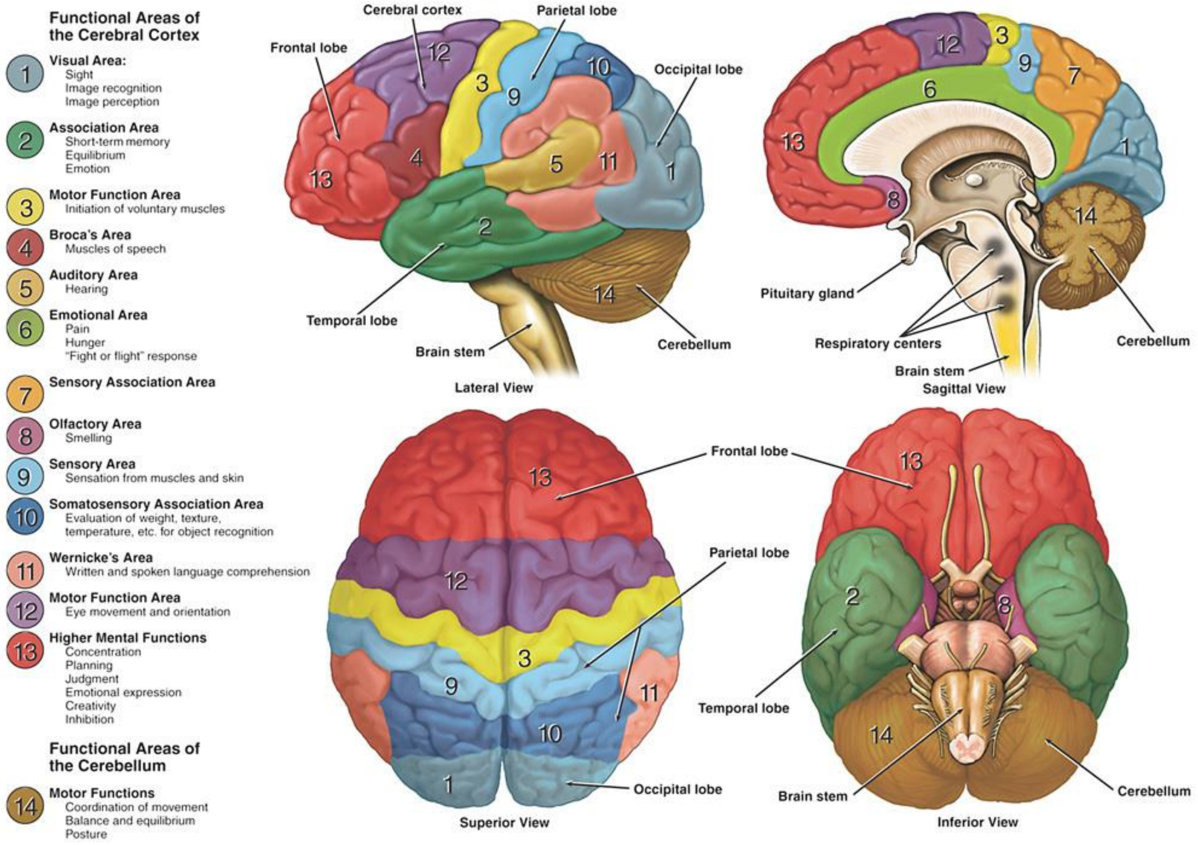
Example data set

2: scRNAseq - finding sparse representations of gene expression matrices via Mapper

- Provided by Sebastian Damrich, IWR Heidelberg

3: EEG recordings of a human adult performing a visuospatial working memory task

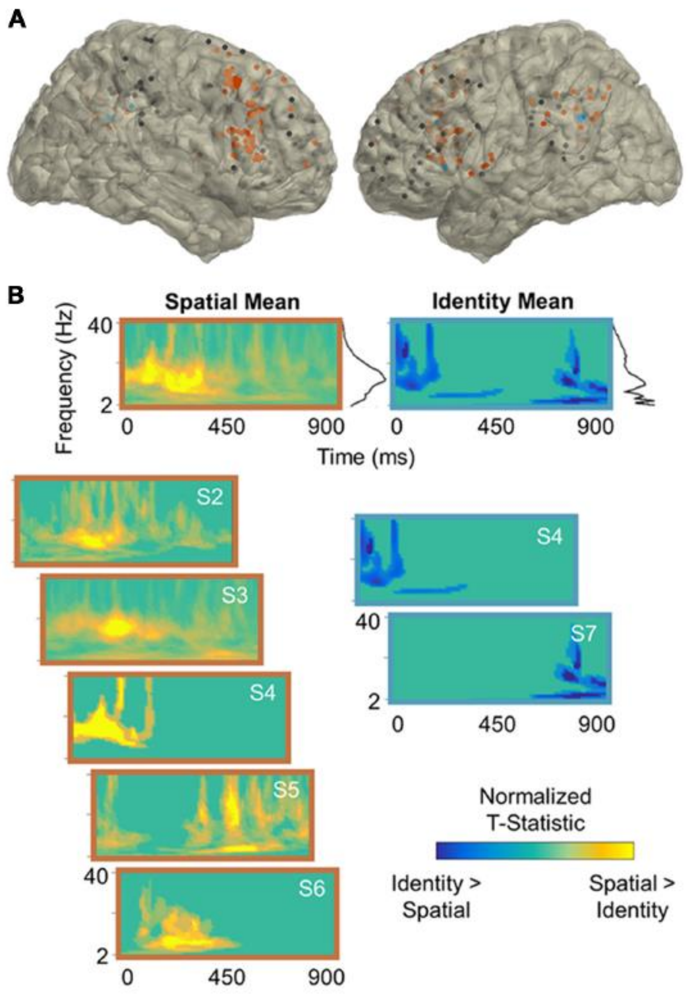
- How does the human brain rapidly process information in working memory?
- Recent work argues for emphasis on rapidly coordinated distributed neuronal networks instead of single-region focus
- Here: direct brain recordings to delineate frontoparietal oscillatory correlates of working memory



3: EEG recordings of a human adult performing a visuospatial working memory task

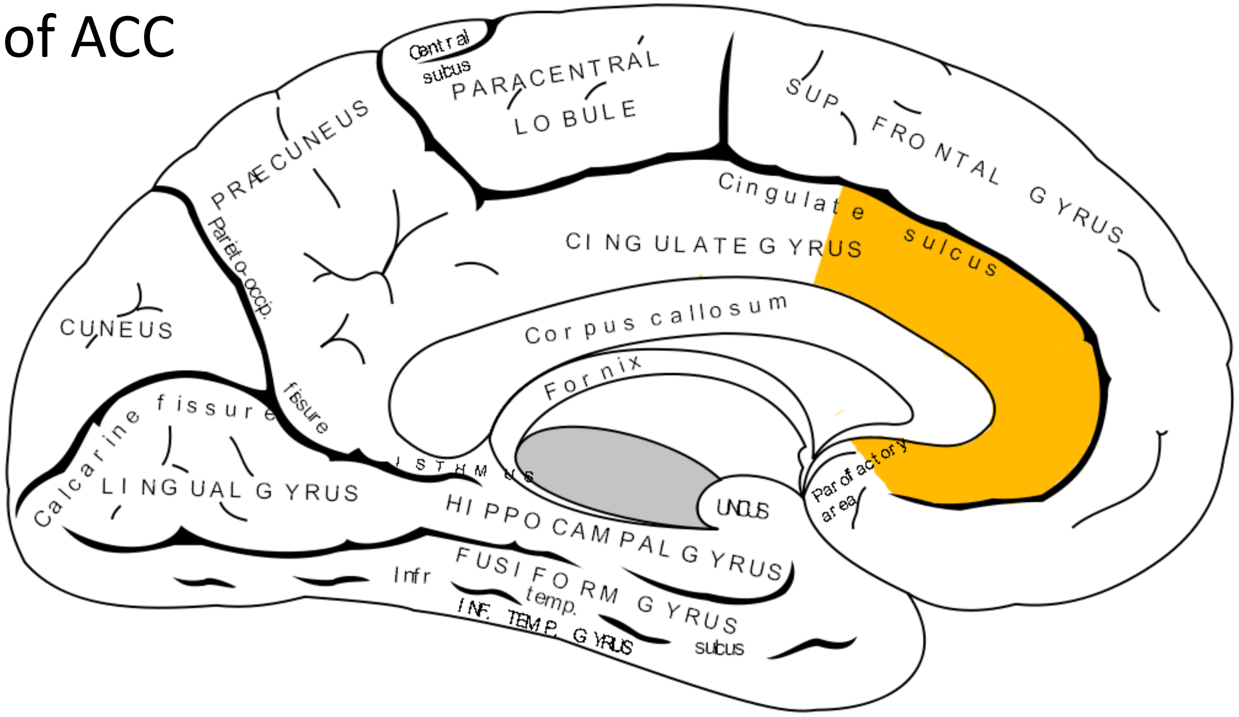
- EEG recordings from the lateral frontal and parietal regions while the epileptic human adult was performing a visuospatial working memory task, including spatial coordinates
- Comparably huge time-series data sets with many parameters
- Based on data in Johnson *et al.*, Front. Syst. Neurosci. 12(65), 2019

- Inspiration for persistent homology usage: Wang, Ombao, Chung, Ann. Appl. Stat. 12(3), 2018
- Inspiration for Mapper usage: Saggar *et al.*, Nature comm. 9, 2018



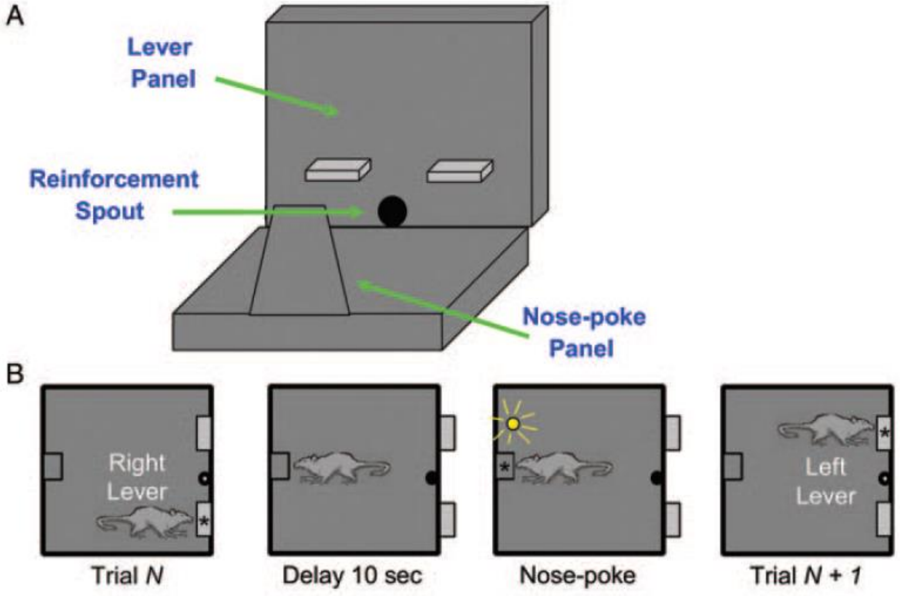
4: Action and outcome activity state patterns of the anterior cingulate cortex of rats

- Theories of anterior cingulate cortex (ACC) function suggest involvement in action or outcome processing
- Study characterized dominant patterns of ACC activity on a given task
- Based on Hyman *et al.*,
  Cerebral Cortex 23(6), 2013



4: Action and outcome activity state patterns of the anterior cingulate cortex of rats

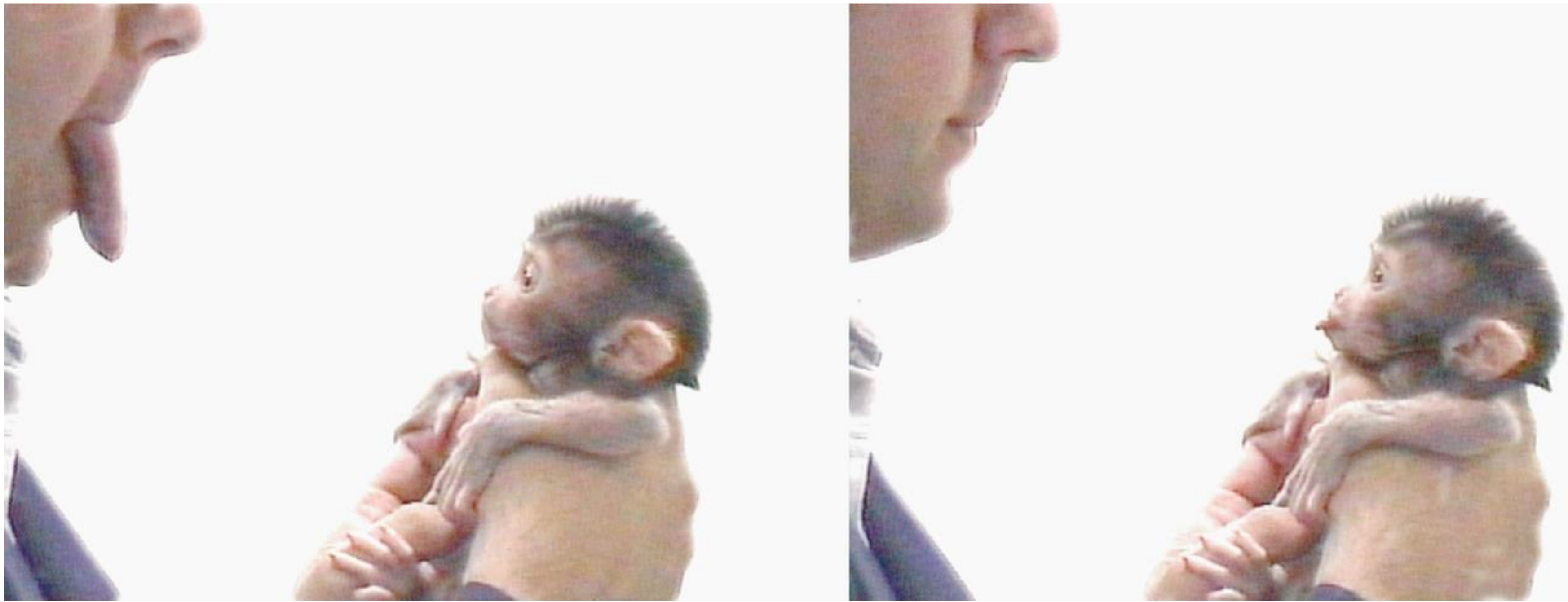
- Tetrodes placed into the vicinity of the ACC of the rat brains to directly record respective neurons
- 22 trials of between 75 and 341 time steps, 10 neurons recorded
- Inspiration for persistent homology usage: Wang, Ombao, Chung, Ann. Appl. Stat. 12(3), 2018
- Inspiration for Mapper usage: Saggar *et al.*, Nature comm. 9, 2018

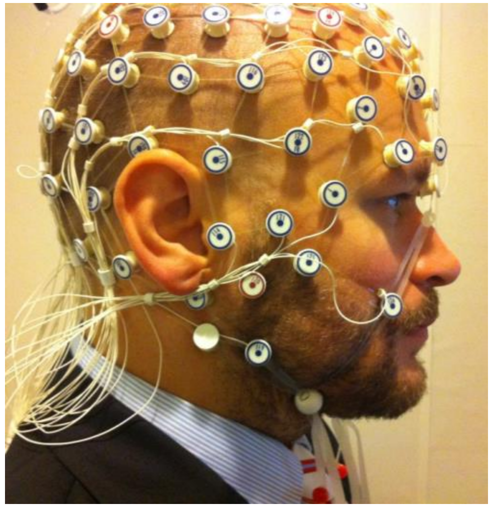


5: EEG measurement data of different empathy tasks

- Provided by Dr. Stephanie Schmidt, U Konstanz

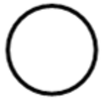
Mirror neurons






MC-B70

Empathy Task

	Cue: 2s	ISI: 2(±1)s	Stim: 3s	Rating: 4s
Distress	How bad do you feel?	+		How bad do you feel? not medium very
Cognitive Empathy	How bad does this person feel?	+		How bad does this person feel? not medium very
Affective Empathy	How much do you feel with this person?	+		How much do you feel with this person? not medium very
Control	How big is this circle?	+		How big is this circle? not medium very

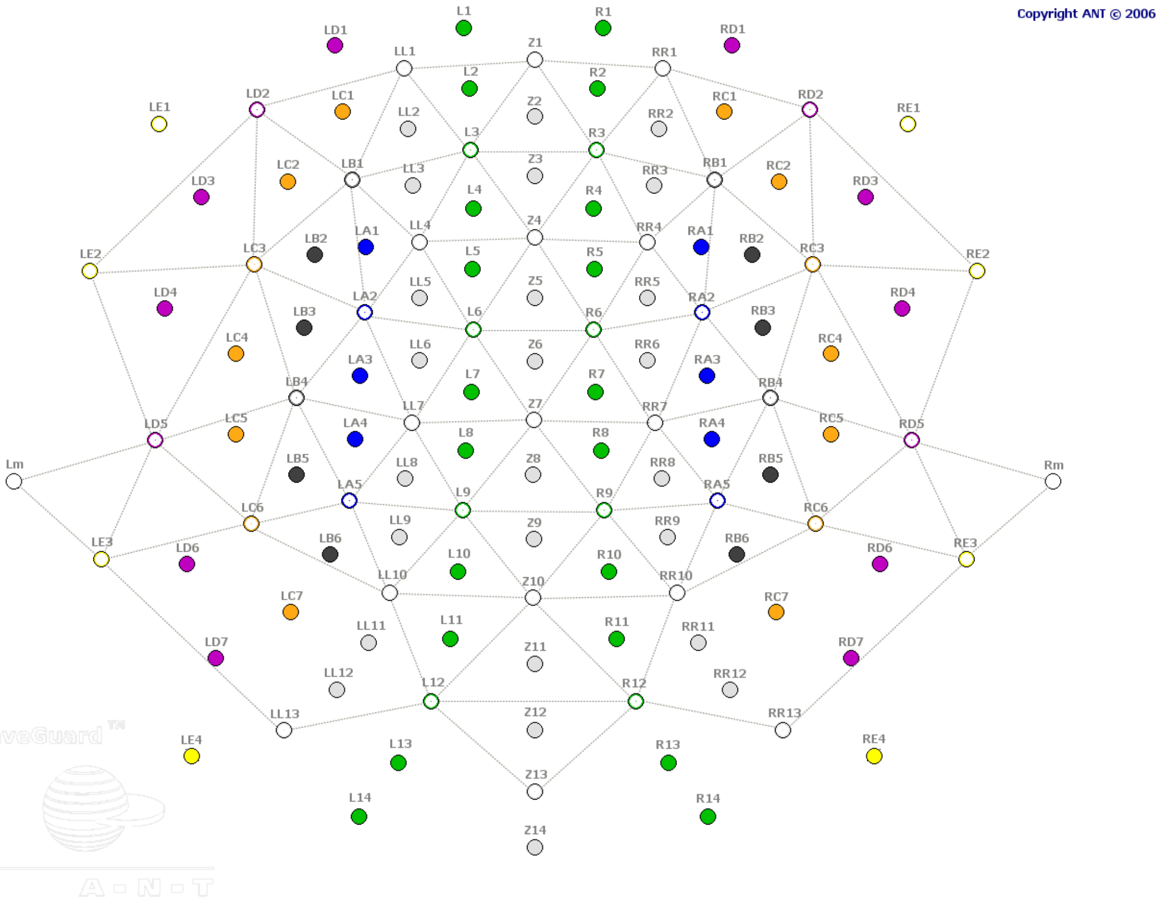
Dataset

EEG-data of one person

- 128 channels
- 2 conditions:

„how much do I feel with this person?“

„how bad do I feel?“



Dataset

- 20 segments
- Each segments lasts 3 seconds
- Sampling rate 256 Hz
- Filters:
  Highpass 0.01, Lowpass 80, Notch 49-51 Hz

Inspirations for TDA usage

- **Robust clustering?** Mapper might be worth a try
- **Connected components, loops or holes?** Persistent homology might be worth a try
- Scientifically, where could topological nontrivialities show up?
- **Continuous field data**
  - How to construct sublevel or superlevel sets?
  - How to connect different grid points?
- If no particular topological structure expected: Use TDA as a **statistical descriptor** beyond fixed-order correlations?

Breakout rooms

- Individual subgroups depending on the data investigated
- If Zoom-version not sufficiently recent: type a room ID in chat to get assigned to it
- If you run into problems while in a room, hit the contact moderator button
- If you want to change rooms, simply get back to the main session and let us know

Room IDs:

- 1: CMB data
- 2: scRNAseq data
- 3: Visuospatial working task EEG data
- 4: Activity patterns in rat brains
- 5: EEG data empathy
- 8: Exercises of the previous workshop days
- 9: Break room

Working session in subgroups

Brief presentations of subgroups

1: CMB data

2: scRNAseq data

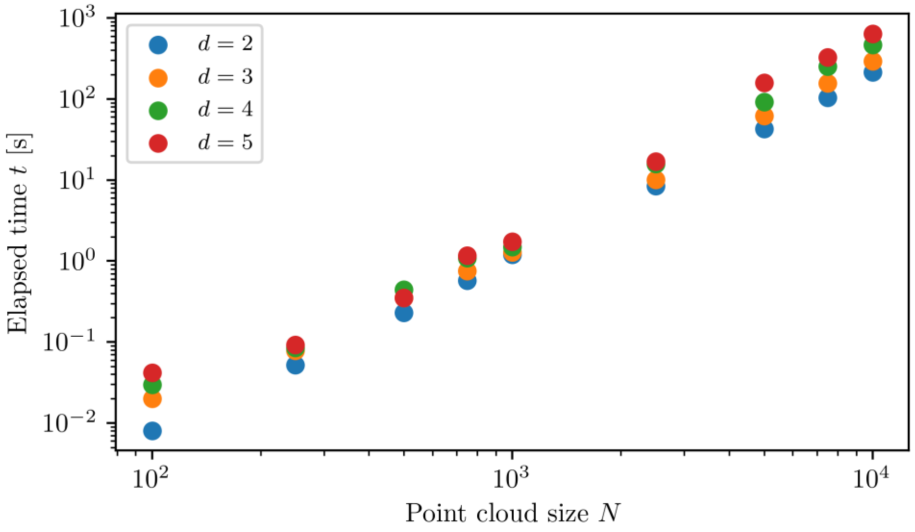
3: Visuospatial working task EEG data

4: Activity patterns in rat brains

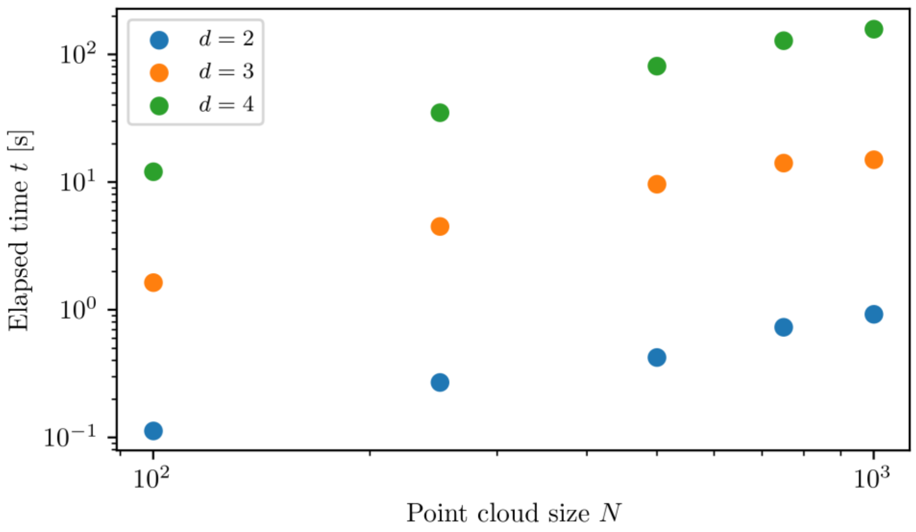
5: EEG data empathy

Further material

Benchmarking and further TDA software

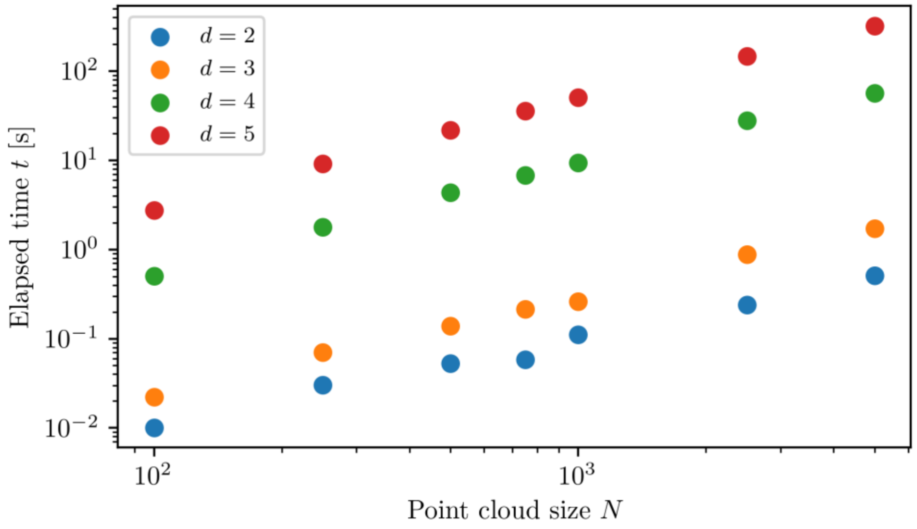


Ripser benchmark on simple laptop



Cechmate benchmark on simple laptop

Benchmarking and further TDA software



GUDHI benchmark on simple laptop

Other TDA software solutions:

- GUDHI (C++, Python interface)
- JavaPlex (Java)
- PHAT (C++)
- Many more…

More info: Otter *et al.*, EPJ Data Science 17, 2017

Further reading

- Example references **applications**:
  - Spreading processes on networks: Taylor *et al.*, Nature comm. 6, 2015
  - Medium-range order in the glass: Nakamura *et al.*, Nanotechnology 26(30), 2015
  - Dynamical organization of the brain: Saggar *et al.*, Nature comm. 9, 2018
  - Ultracold atom simulations: Spitz, Berges, Oberthaler, Wienhard, arXiv:2001.02616

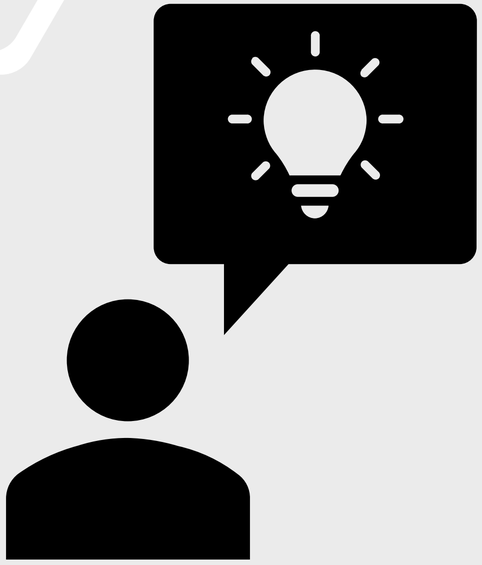
- Example references **math and methods**:
  - Mapper: Singh, Mémoli, Carlsson, SPBG proceedings, 2007
  - Mathematical introduction: Carlsson. *Topology and data,* AMS 2009
  - Book: Edelsbrunner & Harer. *Computational Topology*, AMS 2010
  - Computational roadmap: Otter *et al.*, EPJ Data Science 17, 2017

Journal Club

- STRUCTURES EP Math and Data: (bi-)weekly Journal Club sessions
- Persistent homology topics from an application- and math-perspective
- Mondays, 11-13am
- Starting with an organizational meeting on November 9

- **Online Wiki:** wiki.structures.mathi.uni-heidelberg.de

Final remarks



Recap

- **Mapper**
  - Versatile method to map data to a graph, giving new insights into clusters within high-dimensional data
- **Persistent homology**
  - Detection scheme for topological structures of arbitrary dimension within data
  - Robust to noise
- **Employed TDA to real data**
  - TDA provides new sophisticated tools beyond traditional analysis techniques
  - TDA can reveal fully new structures in data
  - Analysis toolchain needs to be tailored to scientific needs

Concluding remarks

- TDA: versatile toolbox of methods, not fully exploited so far
- Available software makes sophisticated TDA tools widely accessible and available
- The shape of data needs to allow for a proper analysis of its topology

Organizational remarks:

- Google form for feedback will follow
- Certificates mostly via post

Thank you for attending!