Carl von Ossietzky
Universität
Oldenburg

meta rep

Quantifying similarities between fMRI processing pipelines for efficient multiverse analysis

Micha Burkhardt
Dr. Carsten Gießing
Prof. Dr. Andrea Hildebrandt
Dr. Daniel Kristanto

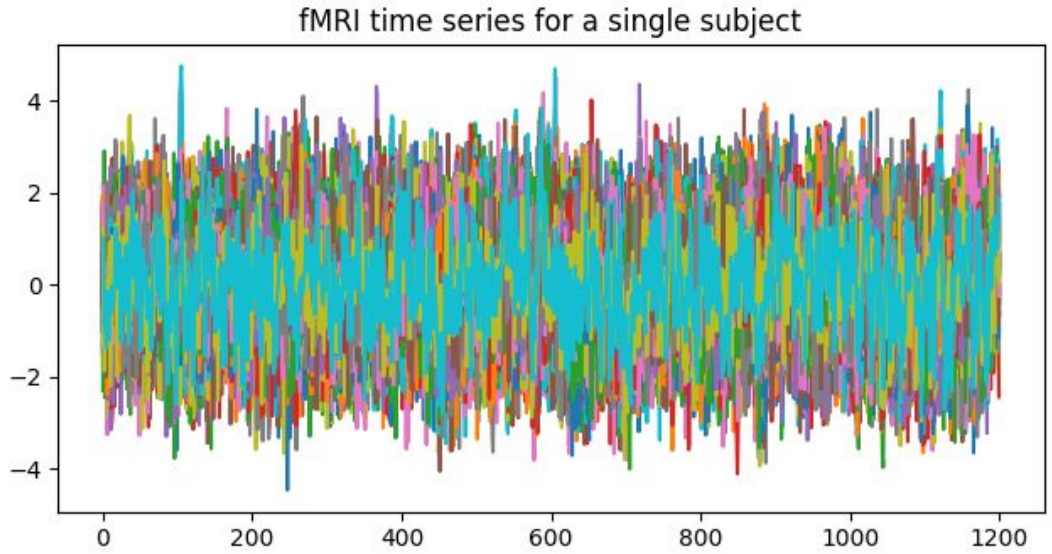
BHAVI
BRAIN HEALTH ALLIANCE
VIRTUAL INSTITUTE

BHAVI 2024 October Symposium:
Guardians 2024 Conference

Problem set-up

fMRI data is **noisy** and requires **long preprocessing**/analysis pipelines

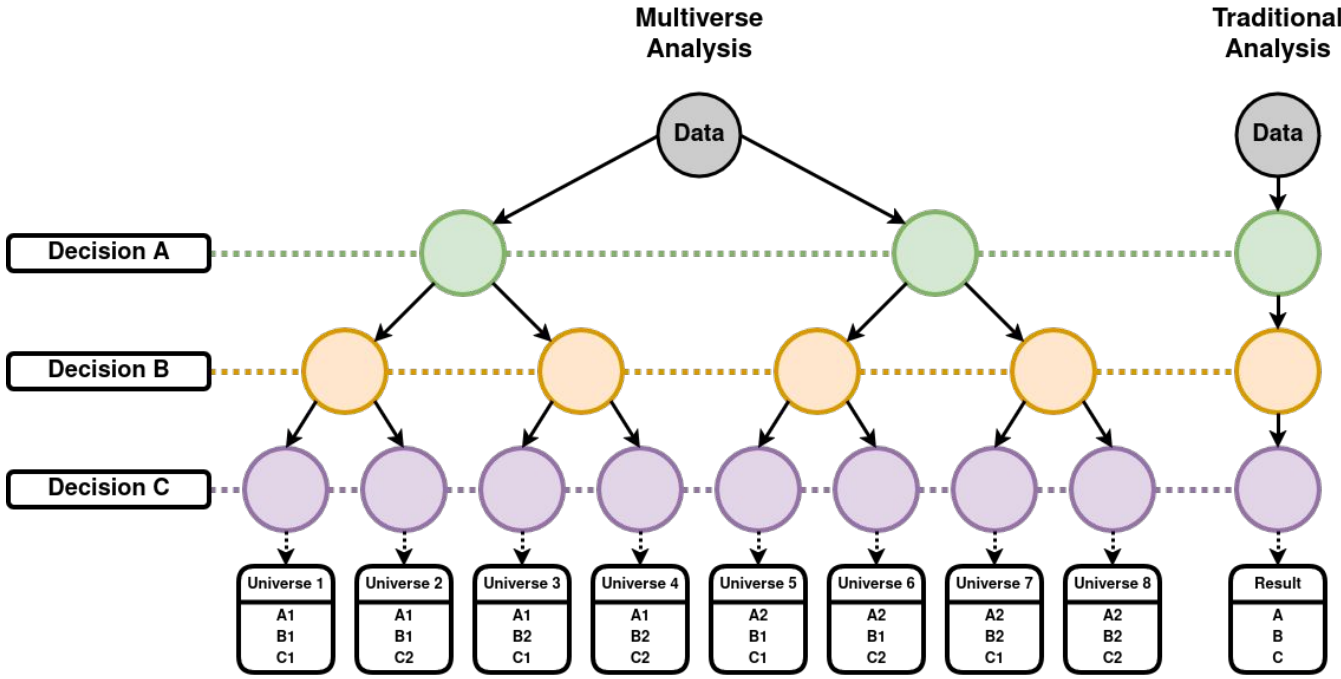
- Researchers are faced with a plethora of arbitrary yet defensible decisions
- Many **researcher degrees of freedom** introduce bias and contribute to the replication crisis



What is multiverse analysis?

Multiverse Analysis: Running multiple (equally defensible) analysis pipelines

- Proposed to help with the issue of flexibility in data processing/analysis
- **Reduces bias** from selective reporting, **improves replicability and robustness**



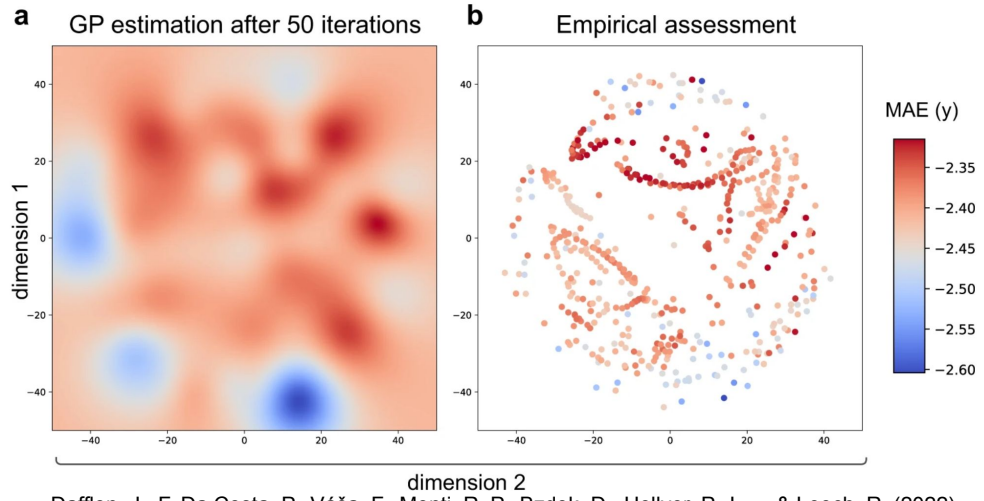
Why do we need multiverse analysis to be more efficient?

- In neuroimaging studies, the **space of decisions is large**
- The size of the multiverse scales with the cartesian product of its decisions

→ It can quickly become computationally infeasible to run an entire multiverse

Why do we need multiverse analysis to be more efficient?

- Dafflon et al. (2022) proposed **subsampling the multiverse**
  - Builds search space from output similarities (e.g., functional connectivity)
  - Uses embedding algorithms to map relationships in a 2D space
  - Proximity in space reflects similar pipeline outputs
- We propose that we can **efficiently subsample** the space of decisions just by the analysis pipelines themselves
  - Embeddings based on pipeline configurations, not output similarity
  - Capture structural differences in the pipeline design

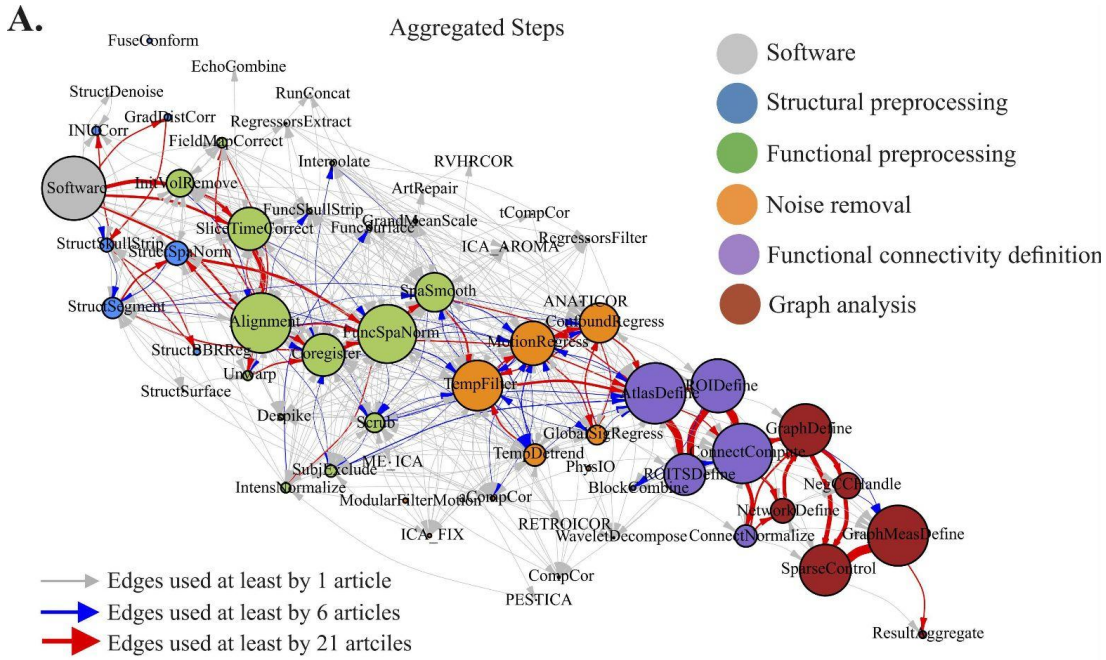


Dafflon, J., F. Da Costa, P., Váša, F., Monti, R. P., Bzdok, D., Hellyer, P. J., ... & Leech, R. (2022). A guided multiverse study of neuroimaging analyses. *Nature Communications*, *13*(1), 3758.

Data: fMRI analysis pipelines

The METEOR Project

- 220 fMRI analysis pipelines derived from literature (Kristanto et al., 2024)
  - 61 distinct data preprocessing and analysis steps

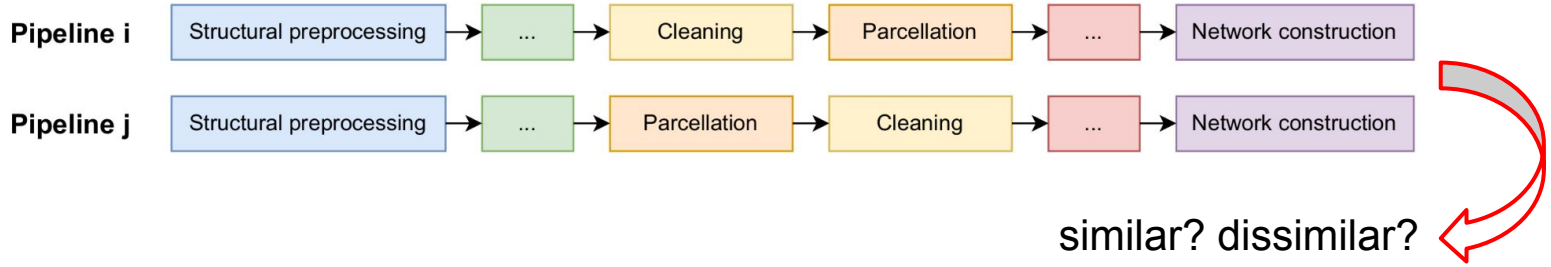


Kristanto, D., Burkhardt, M., Thiel, C., Debener, S., Gießing, C., & Hildebrandt, A. (2024). The multiverse of data preprocessing and analysis in graph-based fMRI: A systematic literature review of analytical choices fed into a decision support tool for informed analysis. *Neuroscience and biobehavioral reviews*, *165*

Similarity of analysis pipelines

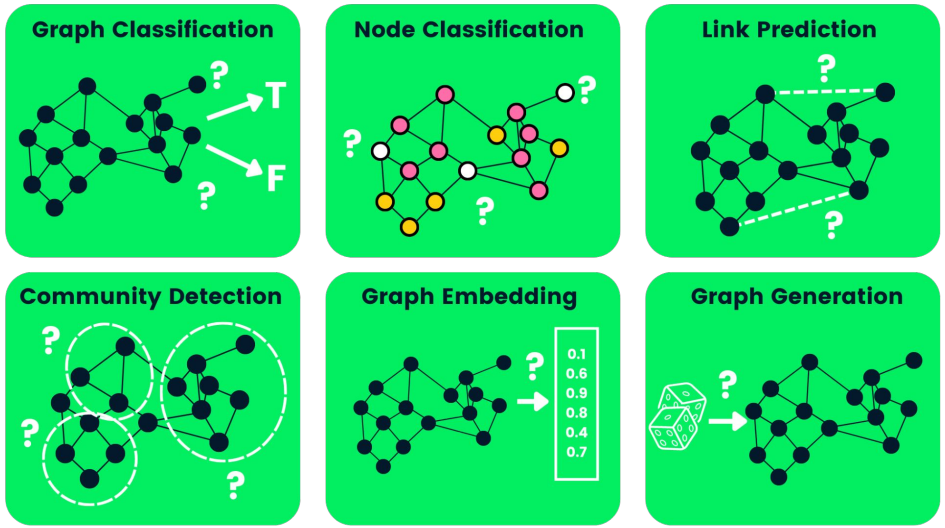
Similarity measures

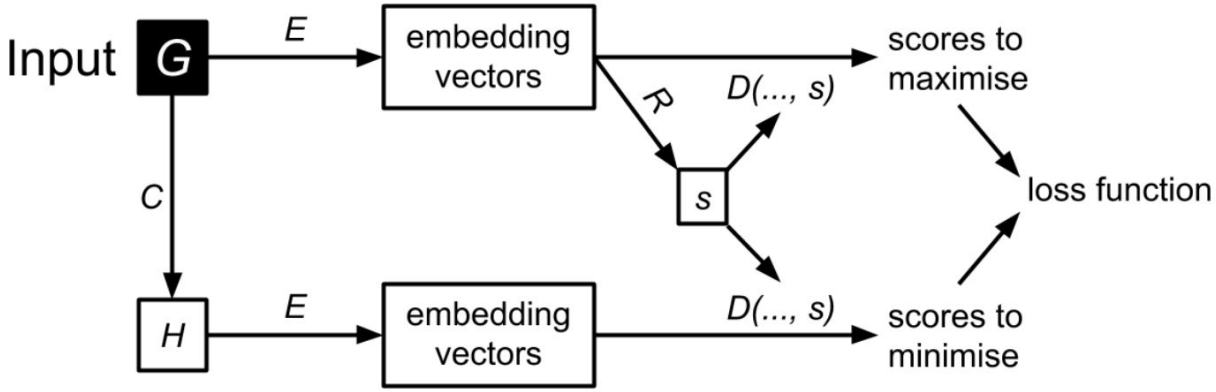
- Classical measures
  - Jaccard index
  - Hamming distance
  - Levenshtein distance
- Graph convolutional network approach
  - Applies Deep Graph Infomax
  - Creates nodal embeddings containing implicit features



Graph convolutional network (GNN) + Deep Graph Infomax (DGI)

- **GNNs**: Learn node/graph embeddings by aggregating neighbor info
- **DGI**: Unsupervised, maximizes info between node and graph embeddings
  - Training: Compares real vs. corrupted graphs to enhance structure
  - Also uses global features (frequency in literature, degree, type, …)
- **Result**: Produces robust embeddings for downstream tasks
  - Embeddings are then used to assess pipeline similarity

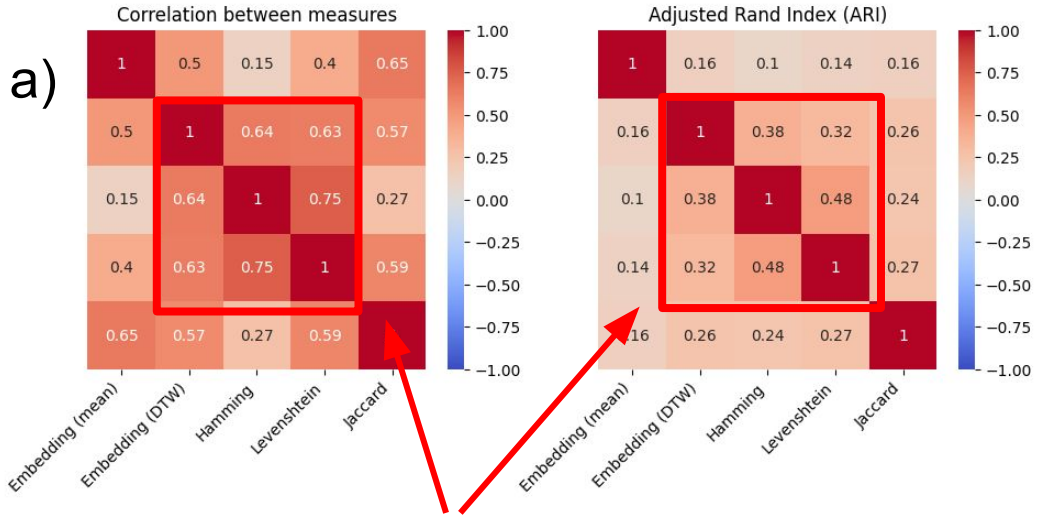




Results

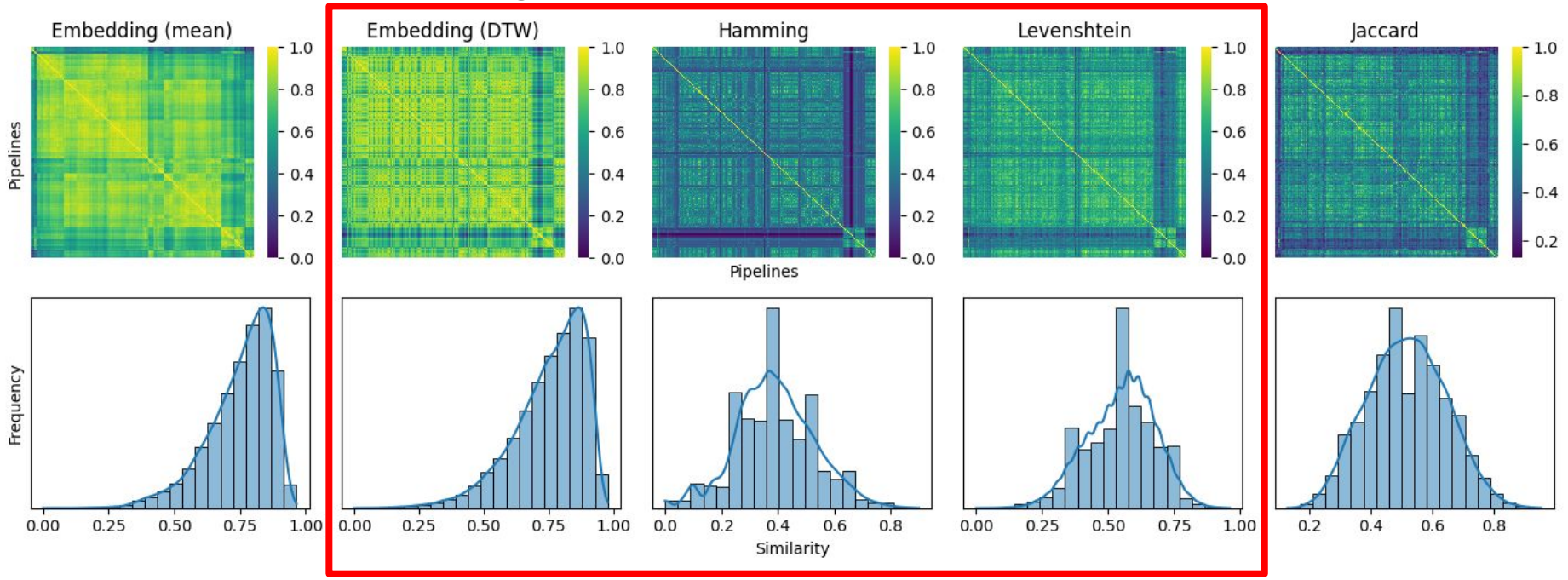
- Reasonable overlap between methods
- Considering order improves ARI (cluster overlap)

a)



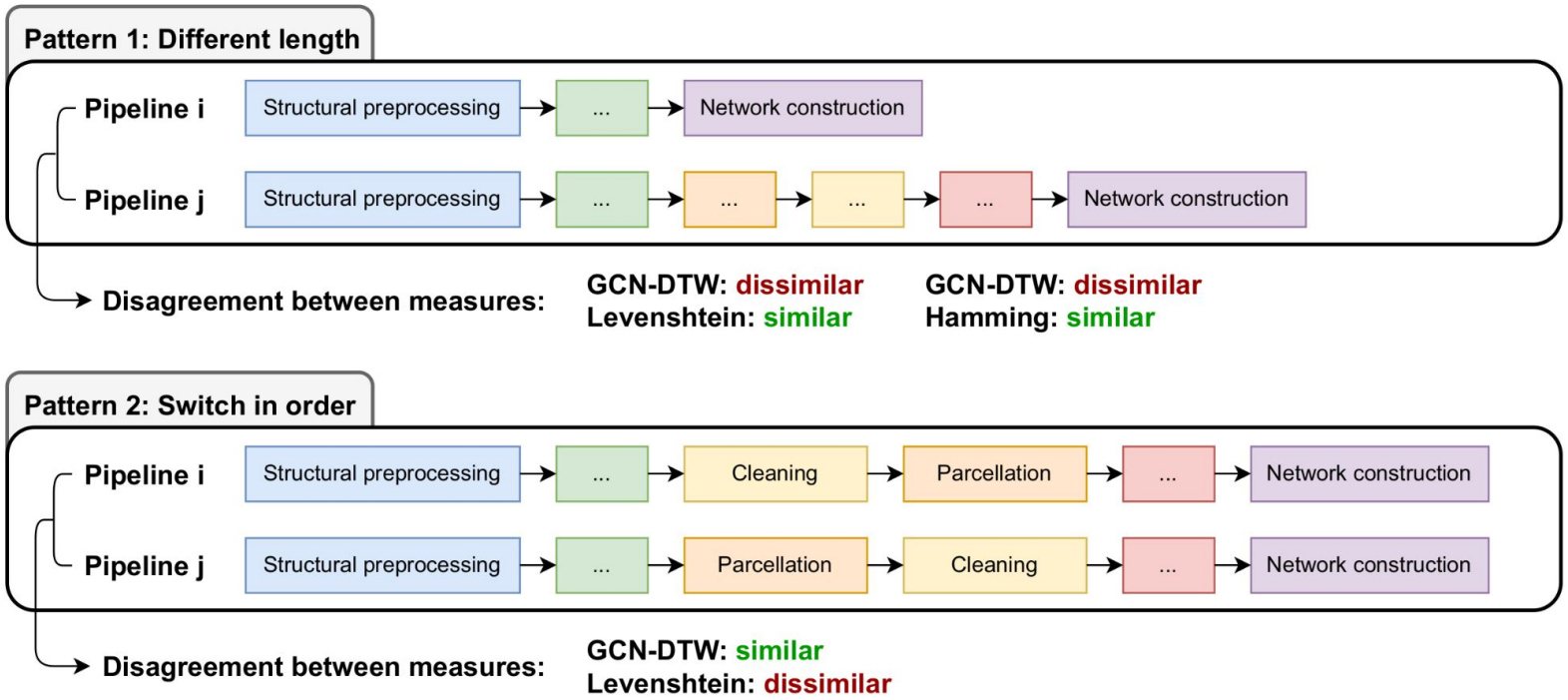
considering order!

b)



Results

- Two distinct patterns:
  - Difference in pipeline length
  - Difference in pipeline ordering



Discussion

Summary

- The GCN approach seems to be promising for capturing relevant information
- Different measures show a general overlap but also distinct behaviours

Advantages

- The GCN approach does not require calculating the entire multiverse
- No need to use (valuable) subjects for creating a search space
- The approach is method-agnostic

Future work

- The performance of the proposed approach needs to be tested on empirical data
- Integrating the GCN approach into efficient multiverse analysis

Carl von Ossietzky
Universität
Oldenburg

Thank you for your attention!

Do you have any further feedback, questions, or ideas? Feel free to reach out! :)

Micha Burkhardt
micha.burkhardt@uol.de

Daniel Kristanto
daniel.kristanto@uol.de