

# Exercise Sheet 5

Due date: Monday, July 4th, 23:59

- The content of your master branch at the time of the deadline is your code submission.
- The zip file uploaded to moodle before the deadline contains your submitted predictions.
- Rules:
  - All learning algorithms are allowed, as long as you implement them yourself.
  - All Deep Learning must be done with PyTorch.
  - Graph Learning Libraries (PyTorch Geometric, DGL, etc.) are strictly forbidden.
  - No restrictions to hardware. Feel free to use the RWTH HPC Cluster.
  - You can only use the data contained within the datasets for training. Pre-training on other datasets is not allowed.
  - The datasets are anonymized, sub-sampled versions of public datasets. Attempts to recover the test labels are strictly forbidden.

**Exercise 1 (HOLU)**

**10 points**

Develop a graph learning approach to predict missing graph labels in the HOLU dataset.

**Exercise 2 (CITE)**

**10 points**

Develop a graph learning approach to predict missing node labels in the CITE dataset.

**Exercise 3 (LINK)**

**10 points**

Develop a graph learning approach to predict missing edge labels in the LINK dataset.

**Exercise 4 (Upload)**

**0 points**

Save your predictions as pickle files, as described in the lecture. Make sure the results are all in the correct order!

Upload a zip file with your predictions to moodle before **Monday, July 4th, 23:59**.

### Exercise 5 (Code Quality, Comments and Presentation)

20 points

Clean your code and add useful comments. Your repository must contain a `README.md` file. For each of the three datasets, the readme should contain the following information:

- A detailed description of your learning method. The exact configuration of your final algorithm should be specified. For a GNN this would include activation functions, type of aggregation, hidden dimension, etc.
- How did you split the data for training and validation? What results did you achieve on these splits?
- A motivation for your method. Which alternatives did you consider/try? Why was this method chosen over the alternatives? How were the configuration details selected?

The readme should also specify how to run every executable script in your repository.

Prepare a short presentation (~5 min) for your group meeting (held in the week after the submission deadline). It should briefly provide the following information:

- The methods you implemented and how you chose them.
- How the work was split.
- The results you obtained on your train and validation splits.