

I. IDENTIFICATION DATA

Thesis title:	Task-Aware Relation Type Selection for Machine Learning Algorithms on Graphs
Author's name:	Michal Mareš
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Computer Science
Thesis reviewer:	Ing. Ondřej Lukáš
Reviewer's department:	Department of Computer Science

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment	challenging
<i>How demanding was the assigned project?</i>	
The thesis explores relation between graph representation and the performance of the model trained with graphs as input. While the problem is motivated a specific use-case in CISCO, it is general and not-well explored topic applicable in a variety of domains which use graph-representation for data.	

Fulfilment of assignment	fulfilled
<i>How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.</i>	
The thesis covers the tasks in the assignment. The research of the related work and state of the art is well covered as well as the problem formalization and proposed novel solution. Implementation of the method is fulfilled together with evaluation on both public and proprietary CISCO datasets.	

Methodology	outstanding
<i>Comment on the correctness of the approach and/or the solution methods.</i>	
The technical level of the thesis is very good. The methodology is valid with proper analysis of the current approaches and their limitations, theoretical background of the problem and proposed solution. The solution is tested and results critically evaluated.	

Technical level	B - very good.
<i>Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?</i>	
The technical level of the thesis is very good. The author utilizes the theoretical background from graph theory in the proposed solution of relation-type selection in graph datasets. The proposed solution is sound and provides an innovative approach to graph-based learning. The evaluation of the method and the analysis of the results, while extensive, is challenging for the reader. Despite evaluating the method on public datasets to complement for the proprietary CISCO network traffic dataset, none of the public datasets are related to network security. While this selection provides insights to the generalization of the algorithm to other domains, I would expect evaluation using dataset with similar properties to the original dataset. Related to question, the evaluation of the proposed method to GraphSAGE and the proprietary data is not provided.	

Formal and language level, scope of thesis	B - very good.
<i>Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?</i>	
The thesis is well-organized and structured. The language level of the thesis is very high. While there are several parts of the thesis which would benefit from better clarity (either in the text or in the form of a figure), overall the thesis is well-written and I do not find any problems with its readability.	

Selection of sources, citation correctness**A - excellent.**

Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?

The thesis is properly referenced and the selection of sources is well chosen both in quality and scope. The contributions of the thesis are clearly stated and separated from existing work.

Additional commentary and evaluation (optional)

Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.

The thesis focuses on not well-explored problem in graph representation learning. Despite the rise of popularity and focus in this domain in the recent years, the importance of the graph structure to the learning process is somewhat scarcely examined. The thesis proposes a novel method suitable for selection of graph features and hyperparameters of the graph algorithm. The student provided extensive evaluation of existing work in this domain and used his theoretical knowledge to propose a regression meta-model as a way of finding the most suitable graph features and hyperparameter configuration for a given task.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading. Pose questions that should be answered during the presentation and defense of the student's work.

The overall level of this theses is very good both in theoretical and practical parts. Despite minor shortcomings in the experiment design and results evaluation, I recommend presented thesis for defense with a **grade B – very good**. For the thesis defense I suggest following questions:

1. Can you comment on benefits of applying SHAP to the meta-model in comparison of applying it (or other XAI methods) directly on the graph-algorithm?
2. Have you considered other explanation methods for evaluation of suitable graph features / model hyperparameters?
3. Can you explain why in 6.2 some features (*nodes_mod* & *edges_bip_pos*) are not excluded despite high correlation?
4. Can you comment on possible adaptation of the proposed method to models not based on graph-features?

The grade that I award for the thesis is **B - very good**.

Date: **13.6.2023**

Signature: