GATTO: Can Topological Information Improve Node Classification via GAT?

Midterm Report for Learning from Network's project

Francesco Biscaccia Carrara *(2120934)*, Riccardo Modolo *(2123750)*,
Alessandro Viespoli *(2120824)*

Master Degree in Computer Engineering
University of Padova

I. INTRODUCTION

The scope of this document is to:

- **Clarify all points** not properly explained in the proposal paper
- Summarize all **work done**
- Define the **missing work**
- Estimate **time to finish the project**.

II. CLARIFICATION POINTS

One of the most important points of clarification concern the data: in the proposal paper, we did not provide an extensive explanation of how data are built and how we want to use them. The graphs chosen in SNAP[1] are undirected and contain communities, no other features are present in the files. We want to use all communities as label, and a node can only belong to one community. For GAT tuning we want to compute the following feature for each node:

- betweenness centrality
- closeness centrality
- clustering coefficient
- *(other possible features)*

III. WORK DONE

We have already built the **precomputation module** (explained in the project proposal) and the code to automate the testing phase inside cluster.

IV. WORK IN PROGRESS

We nedd to build the **GAT module** and do the tests on **CAPRI**[2]. We must also decide the hyperparameters in node2vec.

V. ESTIMATION

In our estimation, we can finish the paper five or six days before the deadline.

REFERENCES

[1] Jure Leskovec and Andrej Krevl. *SNAP Datasets: Stanford Large Network Dataset Collection*. `http://snap.stanford.edu/data`. June 2014.

[2] DEI University of Padova. *CAPRI: Calcolo ad Alte Prestazioni per la Ricerca e l'Innovazione*. 2017. URL: `https://capri.dei.unipd.it`.