

POLITECNICO DI TORINO

DEPARTMENT OF CONTROL AND COMPUTER ENGINEERING

Master of Science in Computer Engineering

Master Degree Thesis

Deep Learning on Academic Knowledge Graphs

Predicting new facts in a novel semantic graph built on top of the Politecnico di Torino scholarly data



Supervisors

Prof. Antonio Vetrò
Prof. Juan Carlos De Martin

Candidato

Giovanni GARIFO

ACADEMIC YEAR 2018-2019

Abstract

The publication and sharing of new research results is one of the main goal of an academic institution. In recent years, many efforts have been made to collect and organize the scientific knowledge through new, comprehensive data repositories. To achieve such goal new tools that are able not only to store data, but also to describe them are needed.

Knowledge graphs are a particular class of graphs that are used to semantically describe the human knowledge in a specific domain by linking semantic entities through labeled and directed edges.

In this work we are going to present a novel semantic graph built on top of the scholarly data produced by the Politecnico di Torino, and how we employed state-of-the-art machine learning techniques for the prediction of new facts in the knowledge base. Such graph, built by leveraging Semantic Web technologies, connects publications, researchers, fields of study and scientific journals in order to build a knowledge base that describes the Politecnico di Torino scientific community. We decided to call such new academic graph the *Polito Knowledge Graph*.

The prediction of non-existent links between graph entities is one of the most challenging tasks in the field of statistical relation learning for graph data, mainly because, in order to obtain meaningful predictions, the vector representations of the graph entities must embed their semantic characteristics. To accomplish such goal, we decided to employ Deep Learning architectures derived from the image recognition field and specifically adapted to the task of representation learning for graph data. Such architectures allowed us to obtain representations that have been directly learnt from the graph structure itself, without requiring any prior knowledge or feature engineering.

Using such learnt representations, we have been able to obtain meaningful predictions about new, unseen facts in the knowledge base. We

used such predictions to complete the information present in the knowledge graph and to build a recommendation system for the suggestion of useful insights to the Politecnico di Torino researchers.