

Training Camp on “Knowledge Graph Completion”

— *Sapienza University, M.Sc. Degree in Data Science* —

Fabio Galasso, Laura Laurenti, Alessio Sampieri
Sapienza University of Rome

Ilaria Bordino, Francesco Gullo, Lorenzo Severini
UniCredit Services
“AI, Data & Analytics ICT” Department
“Applied Research & Innovation” unit

<https://www.kaggle.com/c/unicredittrainingcamp/overview/lectures>

June 30th – July 2nd, 2021

Day 1: Introduction to the Training Camp

Who are we: One Bank, One UniCredit

- UniCredit is a simple successful pan-European Commercial Bank, with a fully plugged in CIB, delivering a unique Western, Central and Eastern European network to its extensive and growing client franchise
- We meet real client needs with real solutions which harness synergies between our businesses: CIB, Commercial Banking and Wealth Management. The way in which these solutions are developed and provided is as important as the solutions themselves, which is why everything we do is based on our two core values: Ethics and respect
- By focusing on banking that matters, we offer local and international expertise, providing unparalleled access to market leading products and services in 13 core markets* through our European banking network

Who are we: One Bank, One UniCredit

- Leveraging on an international network of representative offices and branches, UniCredit serves clients in another 16 countries worldwide
- UniCredit Services is the Group global services company and oversees activities in 11 countries: Austria, Germany, Italy, Poland, United Kingdom, Czech Republic, Romania, Slovakia, Hungary, New York and Singapore
- UniCredit Services groups all services in the areas of Information and Communication Technology (ICT), Back Office and Middle Office Operations, Real Estate, Security and Procurement

Applied Research and Innovation (ARI)

We are a unit of the **AI, Data & Analytics ICT** Department of **UniCredit Services**, the services company of the **UniCredit Group**



Our Team



3 centres located in Italy

- Milan
- Rome
- Verona



Researchers



Data Scientists



Developers

Math/Computer Science background



Our Mission

- Design and develop cutting-edge IT solutions that are of strategic interest and impact for UniCredit
- Bridge academia and Industry
- Collaborate with universities and research centres
- Share knowledge and know-how
- Publish scientific papers
- Manage open-source contributions
- Supervise interns/students
- "Quasi-production"-stage prototyping

Topics



Broad areas: Artificial Intelligence, Machine Learning, Data Science



Natural language processing



Graph learning/mining



Computer vision



Time-series processing



Cryptography



Privacy



- A **Knowledge Graph** is a powerful semantic model to represent complex relations between real-world entities such as news, people
- Knowledge graphs are nowadays the model of choice for many different application domains that have to deal with the challenge of extracting and structuring knowledge on millions of entities and relations
- Modern enterprise knowledge graphs are built by ingesting vast amounts of data, but typically miss many entities and relations
- The task of **Knowledge Graph Completion** aims at completing such missing information

The Task: Knowledge Graph Completion

- The aim of this camp is to overview the main state-of-the-art techniques for the tasks of link prediction and knowledge graph completion. More specifically:
- Link prediction using Machine Learning
- Knowledge graph embedding
- Rule mining for knowledge graph
- Specifically, the competition consists in **predicting whether a given knowledge-graph triple is true or false**

Training Camp Schedule

Day 1 (Wednesday, June 30th)

Topic	Time	Slides	Lab resources
<i>Basic Methods for (Knowledge) Graph Completion</i>	9:30-11:30	day1 slides	day1 lab resources

Day 2 (Tuesday, July 1st)

Topic	Time	Slides	Lab resources
<i>Knowledge Graph Embedding (KGE)</i>	9:00-11:00	day2 slides	day2 lab resources

Day 3 (Friday, July 2nd)

Topic	Time	Slides	Lab resources
<i>Combining Rule Mining and Embedding Learning for Knowledge Graph Completion</i>	9:00-11:00	day3 slides	day3 lab resources

Training Camp Organization

- [*website*] [Training Camp website on github](#)
- [*slack*] [Use this slack channel to ask questions](#)
- Course format:
 - Brief lecture (30 min)
 - [*kaggle*] [Kaggle Competition](#)
 - Lectures will provide code snippets, closely following the code you will use in your implementation

Kaggle Competition: Rules

- Opens at the end of Lecture 1 on Wed June 30th; Closes on Friday July 2nd, 4PM
- Shortly after closure (around 5PM), the winner team will be announced
- Team Size: Min 2 people, Max 3. Please organize yourself into teams, and send an email to Prof Galasso to register your team
- To participate, a team must achieve a **minimum score of 0.6** (Task and metric will be explained in details during the first lecture)
- To participate, teams must upload their best results on Kaggle (instructions will be provided during the first lecture)
- Upon closure of the competition, teams must also send their code to the TAs
- The winner team will be given the chance to present their solution at UniCredit
- Valuable participants could be considered for *hiring*

Are you ready to start?

- No panic, we will provide all needed material and instructions :-)
- We will be available during the class hours, and always reachable on Slack
- Looking forward to the next 3 days!
- Good luck with the competition!