

Proposal - Comparison And Evolution Of Comic Hero Networks

For the final project I want to work with Knowledge Graphs. More specific I would like to analyze and compare two knowledge graphs and use them for *(external) link prediction*.

On *Kaggle* I found two datasets which might be interesting for this task. Both datasets contain information about a possible network between *Comic Characters*. The first dataset is the *Marvel Universe Social Network* [1] and the second dataset is the *Super Hero Dataset (Comic Super Hero)* [2]. The structure of the first dataset directly represents a network. The second dataset contains not only additional information/attributes about a certain *Character* but more importantly information about e.g. *work base* or *connection relatives* from which a network can be build.

When creating these two knowledge graphs it should be possible to compare them. In theory connections like *work base* or *connections group-affiliation* from the second dataset might correlate to the *comic appearance* of characters in the first dataset. Simple properties like centralities might allow a first comparison between entities. More interesting could be the comparison with descriptive community detection possibly with *COMODO* or *MinerLSD*. Analyzing the graphs regarding their structure in general makes more sense than focusing on contained attributes. The analysis should conclude in the following research question: **To what extent can those different datasets result, especially regarding their structure, in similar or comparable networks?**

In addition to that it may be interesting to try to add connections / links and entities in these two graphs based on information from each other as an example of *external refinement*. As a link prediction method I would like to use Answer Set Programming (ASP) with *Clingo* across both graphs. In contrast to the first dataset the knowledge graph for the second dataset has to be mined out of attributes of the characters (e.g. *work base*, *connections group-affiliation*). For that link prediction might also be applicable. Measurements from [3] under the collective name of *Efficient Feature Set* may be useful for explainability since they focus on structure properties if only internally. Regarding link prediction / refinement a research question is: **To what extent can link prediction help creating a more detailed network out of the two datasets?**

References

- [1] Claudio Sanhueza. The marvel universe social network. <https://www.kaggle.com/csanhueza/the-marvel-universe-social-network>, 2021.
- [2] Aakash Verma. Super hero dataset (comic super hero). <https://www.kaggle.com/aakashverma8900/superhero-api-dataset>, 2021.
- [3] Jesper E van Engelen, Hanjo D Boekhout, and Frank W Takes. Explainable and efficient link prediction in real-world network data. In *International Symposium on Intelligent Data Analysis*, pages 295–307. Springer, 2016.