Improving data mining of PPI networks by combining deep learning methods with knowledge graphs

Laura Balbi, Cátia Pesquita

LASIGE, Faculdade de Ciências da Universidade Lisboa

Abstract

Many bioinformatics problems pertain to large, highly complex amounts of biological data, that are often modelled in a graph-like arrangement to allow for a systemic-level analysis of data. A graph provides a data structure for knowledge representation, useful for the description and analysis of relationships in all kinds of knowledge domains, including biological systems [1]. It can model a multitude of biological processes, like transportation and metabolic processes, and capture associations between any type of biological entity – e.g. between proteins, by building up a Protein-Protein Interaction Network.

Several approaches have been made to apply machine learning (ML) models to data represented on graphs [2], either through the use of dense low-dimension vector representations of the different graph elements, or by building the models as representation-learning approaches that can receive and analyse such graph-structured data – prompting graph neural networks [3,4].

A form of knowledge representation that allows for the conceptualization and specification of domains of interest is by means of the use of ontologies [5]. With the increased use of biomedical ontologies [6] for representing and structuring the existing biological knowledge by their meaning and relationships, large volumes of biological entities have been represented and organized into Knowledge Graphs (KG) [7].

Usually, for machine learning models to be employed in prediction tasks over knowledge graphs, their learning of the KG's components' semantic information is dependent on it being depicted through vector-based representations that can be processed by the models. However, when both biological data and biological knowledge are represented as graphs, an opportunity to directly enrich the data graph with knowledge about its entities arises [8]. Thus, we aim to explore how complementing a data graph with knowledge of its biological data may allow different graph mining approaches to leverage the additional information to improve their predictive performance, in particular over PPI networks.

Acknowledgements

This work is funded by the FCT through the LASIGE Research Unit, ref. UIDB/00408/2020 and ref. UIDP/00408/2020.

Preferred Presentation: Poster Presentation

References

[1] – Georgios A. Pavlopoulos, Maria Secrier, Charalampos N. Moschopoulos, Theodoros G. Soldatos, Sophia Kossida, Jan Aerts, Reinhard Schneider, & Pantelis G. Bagos. 2011. Using graph theory to analyze biological networks. BioData Mining, 4(1), 1–27. https://doi.org/10.1186/1756-0381-4-10

[2] – William L. Hamilton, Rex Ying & Jure Leskovec. 2017. Inductive representation learning on large graphs. Advances in Neural Information Processing Systems, 2017-December(Nips), 1025–1035. https://dl.acm.org/doi/10.5555/3294771.3294869

[3] – Marco Gori, Gabriele Monfardini & Franco Scarselli. 2005. A new model for learning in graph domains. Proceedings. 2005 IEEE International Joint Conference on Neural Networks. 2, (2005), 729-734.

[4] – Franco Scarselli, Marco Gori, Ah Chung Tsoi, Markus Hagenbuchner & Gabriele Monfardini. 2008. The Graph Neural Network Model. IEEE Transactions on Neural Networks, 20, 1 (2009), 61-80. https://doi.org/10.1109/TNN.2008.2005605

[5] – Manuel de Sousa Ribeiro & João Leite. 2021. Aligning Artificial Neural Networks and Ontologies towards Explainable AI. Proceedings of the AAAI Conference on Artificial Intelligence. 35 (6), 4932-4940.

[6] – Darren A. Natale, Cecilia N. Arighi, Winona C. Barker, Judith Blake, Ti-Cheng Chang, Zhangzhi Hu, Hongfang Liu, Barry Smith, & Cathy H. Wu. 2007. Framework for a protein ontology. BMC bioinformatics 8, (2007), S1. https://doi.org/10.1186/1471-2105-8-S9-S1

[7] – Chang Su, Jie Tong, Yonjun Zhu, Peng Cui & Fei Wang. 2018. Network embedding in biomedical data science. Briefings in Bioinformatics, 21 (1), 182–197. https://doi.org/10.1093/bib/bby117

[8] – Xiaojun Chen, Shengbin Jia, Yang Xiang. 2020. A review: Knowledge reasoning over knowledge graph. Expert Systems with Applications, Volume 141, 112948. https://doi.org/10.1016/j.eswa.2019.112948.