Between the interaction of Graph Neural Networks and Semantic Web

Xavier Sumba
Concordia University

José Ortiz
University of Cuenca

1 Introduction

Relational data containing relationships and entities are all over the Web, and one of the main goals of the Semantic Web [1] has been to exploit the web more intelligently by creating associations. This relationships can be represented as graphs. More recently, Graph Neural Networks (GNN) have emerged and have shown promising results in graph data, being used in different tasks such as node classification, link prediction, community detection or network similarity. As a result of this, many new GNN architectures have been proposed (to name a few [2, 3, 4, 5, 6, 7]). We believe that the interaction of these two fields can play an important role in the advancement of both fields.

Due that there is an overwhelming quantity of heterogeneous data on the web. Integration of data silos provided by the Linked Open Data community can provide information not only to build effective machine learning models but to curate this data and boost the Semantic Web field to its true potential. This data has been compiled over several years and commonly manually curated. Nevertheless, even the largest graphs, for example DBpedia, suffer from incompleteness, and GNN have demonstrated to be effective for identifying entities or relationships [8]. So far, this has been achieved by learning good embedding models that capture representations on graphs.

However, there are other open problems that are worth of attention and need further investigation or experimentation. For example, semantic reasoners contain logical rules, but some part of reasoning and understanding can be automated using GNN, and as a result of this, new facts can be derived (see [9, 10] for some recent work). Additionally, over the years, a variety of ontologies have been created for a variety of domains, for example *biol* for biology or *sioc* for social communities. Thus, graph representations are helpful not only for entity comparison or graph comparison, but also for ontology matching, solving the limitations and scalability issues of current approaches [11].

Having content rich knowledge graphs, it can allow us to build better question answering systems, perform rule mining by navigating a set facts, find optimal patterns, automatically create knowledge graphs for certain tasks, improve recommender systems or search engines. For areas where data is scarce, these set of facts can allow to build powerful models by enriching data. GNN specifically are important for the study of RDF graphs since can allow to extract important representations taking advantage of the natural structure of a graph in contrast with other approaches [12]. Finally in areas where of high-risk domain, for example healthcare, this additional information can be used for decision making. Then many of the named problems are important and need further investigation.

Finally, we are currently populating an ontology from the open academic dataset[1] that contains information about authors and its publications [13], so far containing 15M authors, 53M publications and 269M relationships between them. Moreover, we look forward to improve our ontology by enriching it through an entity recognition and linkage process with other well-known Linked Data platforms such as the Springer-Nature SciGraph initiative[2]. To our knowledge, this will be the largest open academic knowledge graph, and we aim this dataset can be used by the scientific community to obtain insights about new proposed methods.

[1] `https://www.openacademic.ai/oag/`
[2] `https://www.springernature.com/gp/researchers/scigraph`

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

References

[1] Christian Bizer, Tom Heath, and Tim Berners-Lee. Linked data: The story so far. In *Semantic services, interoperability and web applications: emerging concepts*, pages 205–227. IGI Global, 2011.

[2] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[3] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

[4] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*, 2016.

[5] Zhilin Yang, William W Cohen, and Ruslan Salakhutdinov. Revisiting semi-supervised learning with graph embeddings. *arXiv preprint arXiv:1603.08861*, 2016.

[6] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852, 2016.

[7] Federico Monti, Davide Boscaini, Jonathan Masci, Emanuele Rodola, Jan Svoboda, and Michael M Bronstein. Geometric deep learning on graphs and manifolds using mixture model cnns. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5115–5124, 2017.

[8] Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pages 593–607. Springer, 2018.

[9] Will Hamilton, Payal Bajaj, Marinka Zitnik, Dan Jurafsky, and Jure Leskovec. Embedding logical queries on knowledge graphs. In *Advances in Neural Information Processing Systems*, pages 2026–2037, 2018.

[10] Koustuv Sinha, Shagun Sodhani, Jin Dong, Joelle Pineau, and William L Hamilton. Clutrr: A diagnostic benchmark for inductive reasoning from text. *arXiv preprint arXiv:1908.06177*, 2019.

[11] Alina Petrova, Evgeny Sherkhonov, Bernardo Cuenca Grau, and Ian Horrocks. Entity comparison in rdf graphs. In *International Semantic Web Conference*, pages 526–541. Springer, 2017.

[12] Michael Glass, Alfio Gliozzo, Oktie Hassanzadeh, Nandana Mihindukulasooriya, and Gaetano Rossiello. Inducing implicit relations from text using distantly supervised deep nets. In *International Semantic Web Conference*, pages 38–55. Springer, 2018.

[13] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, li Zhang, and Zhong Su. Arnetminer: Extraction and mining of academic social networks. pages 990–998, 08 2008.