An algorithmic framework for heterogeneous multigraph embedding

Giorgio Valentini,* Mauricio Soto-Gomez, Alberto Cabri, Jessica Gliozzo, Emanuele Cavalleri, Dario Malchiodi, Marco Mesiti, and Elena Casiraghi

AnacletoLab, Dept. of Computer Science, University of Milan

Carlos Cano

Dept. of Computer Science and Artificial Intelligence, University of Granada

(Dated: June 30, 2023)

The representation of interrelated concepts in complex graphs is a common task on a large spectrum of disciplines, including computer science, physics, biology, medicine, social science, and economy. In particular, the highly informative representation provided by graphs that explicitly model different types of entities and relationships, sometimes referred as multigraph-heterogeneous networks and also including Knowledge Graphs, stimulated the development of graph embedding methods aware of the enriched semantic content of the underlying heterogeneous graph [1]. In this context, two main research lines, both inspired by homogeneous Graph Representation Learning techniques, recently emerged: Random-Walk-based (RW) and Graph Neural Networks-based (GNN) heterogeneous graph embedding. A common limitation to most of the proposed methods is that they require human intervention and non-automatic pre-processing steps for designing the meta-paths and the overall network scheme. Moreover, in most cases they handle heterogeneous networks by separately extracting each type of their homogeneous component, and are not able to focus on specific types of nodes or edges that constitute the objective of the underlying prediction task (e.g., prediction of a specific edge type).

To overcome these drawbacks, we propose a general framework to deal with complex heterogeneous networks able to embed heterogeneous multigraphs, and to scale up with big networks, due to its intrinsic parallel nature and its efficient GRAPE implementation [2]. The proposed algorithmic framework does not require manual exploration of heterogeneous structures and meta-paths, because it directly models the heterogeneous graph as a whole, without splitting the heterogeneous graph into its homogeneous components. It can focus on specific edges or nodes of the heterogeneous graph, thus introducing a sort of "attention" mechanism, conceptually borrowed from the deep neural network literature, but realized in an original and simple way in the world of random walk visits of heterogeneous graphs. Within this framework, we propose a set of algorithms to learn embeddings that are aware of the topology and of the different types of nodes and edges of the overall heterogeneous network.

---

[1] F. Xia et al. IEEE Trans. Art. Int. (2021). doi:10.1109/TAI.2021.3076021

[2] L. Cappelletti et al. Nat. Comput. Sci, 2023. doi:10.1038/s43588-023-00465-8

* valentini@di.unimi.it