Graph Deep Learning

Carlos Muniz Cuza, Pablo Lopez Estigarribia

Big Data Management and Analytics Master Programme. eBISS 2019

Introduction & Motivation

- Deep Learning (DL) is a machine learning technique that has evolved the way current research and applications are carryout in the area.

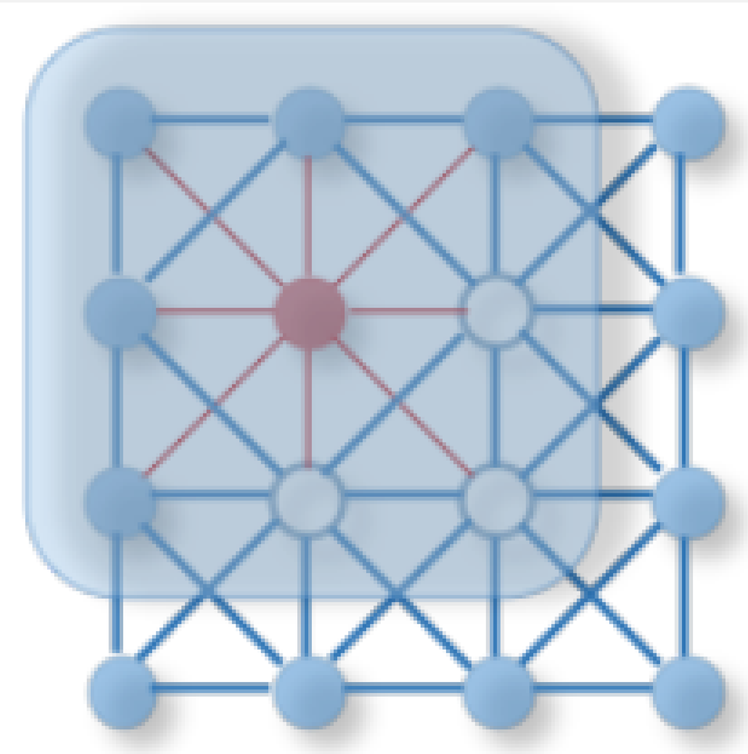

Figure: taken from [1].

- From the ideas of Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN) and more recently Generative Adversarial Networks researches have produced new techniques fine-tuned to consider the specificities of graphs.

Challenges

- **Irregular domain:** Graphs lie in an irregular domain, making it hard to generalize some basic mathematical operations.
- **Varying structures and tasks:** Graph itself can be complicated with diverse structures. Graphs also vary greatly, ranging from node-focused problems such as node classification and link prediction.
- **Scalability and parallelization:** Real graphs can easily have millions of nodes and edges. As a result, how to design scalable models, with a linear time complexity, is a problem.

Graph Neural Network

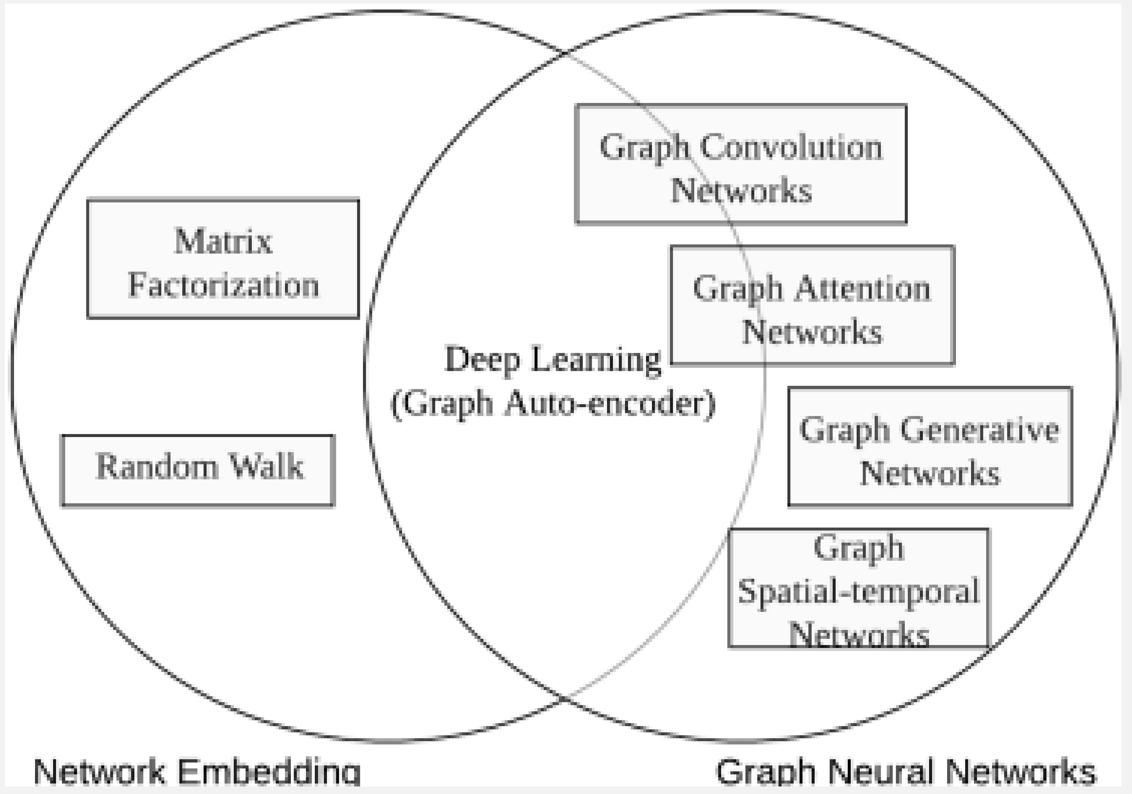

Figure: taken from [1].

Graph Embedding

Graph Embeddings is just a way of representing the latent information present in the graph in a convenient structure more easy to handle and to apply well known mathematical operations than a graph.

- **Matrix Decomposition & Random Walk:** These methods are not properly deep learning
- **Deep Learning:** Inside this are:
  - **Autoencoders:** Learn the vector representation
  - **Variational Autoencoders:** Learn the parameters of a probability distribution to generate samples afterwards.

General Architecture for Graph Convolutional Networks

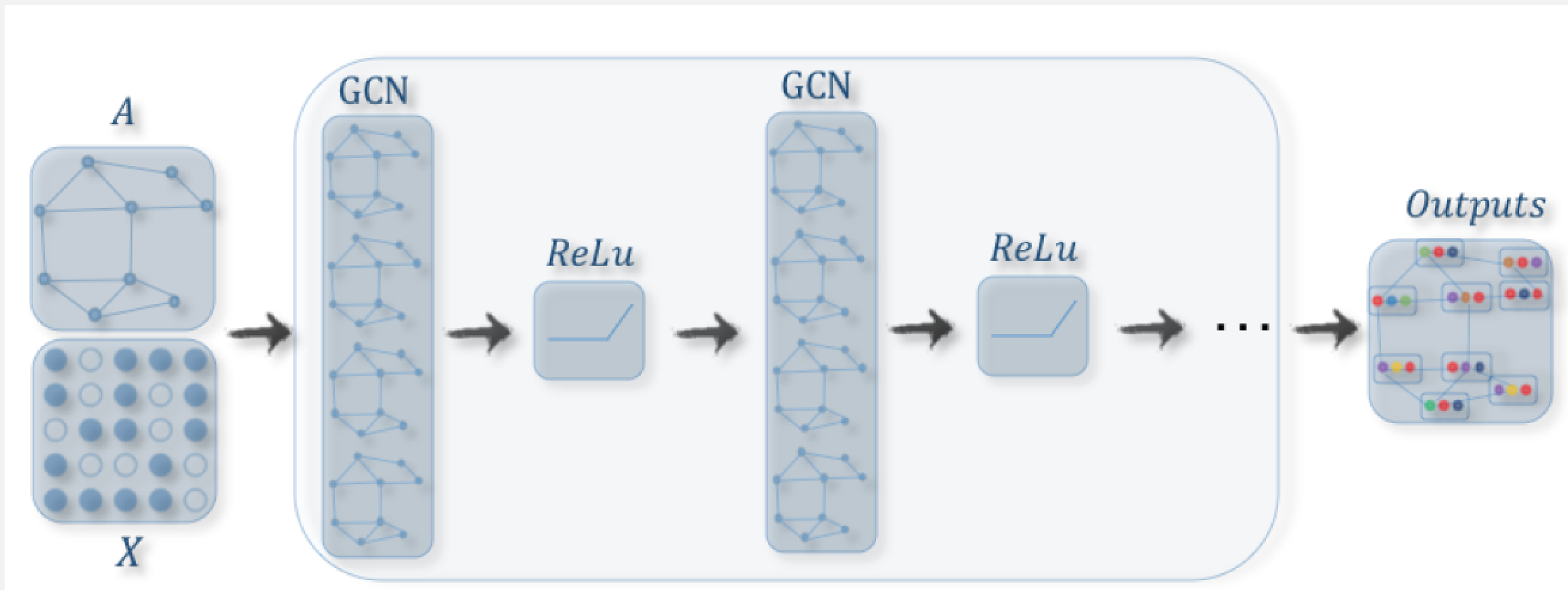

Figure: taken from [1].

Graph Convolutional Networks

Graph Convolution Networks (GCNs) generalize the convolution operations from euclidean (or grid-like) structures to a non-Euclidean domain, such as graphs.

- GCN approaches: Spectral and Spatial. In comparison, spatial approaches outperforms spectral ones in terms of generality, flexibility and efficiency and thus have attracted increasing attention in recent years.
- Different graph analytics task: node-level node regression and classification tasks, edge-level (edge classification and link prediction), graph-level (graph classification).
- Types of training: Semi-supervised learning for node-level classification, Supervised learning for graph-level classification, Unsupervised learning for graph embedding.
- Graph convolution networks is used as building blocks by other methods: graph attention networks, graph auto encoders, graph generative networks and graph spatial temporal networks.

Applications

- **Computer Vision:** Scene graph generation, point clouds classification and segmentation, and action recognition.
- **Chemistry:** In chemistry, study of the graph structure of molecules.
- **Network analysis:** Classification in network analysis can be improved by using information from neighbor vertices.

Datasets & Results

- **Academic Citations:** Cora, Citeseer, Pubmed, HEP-TH and DBLP.
- **Social Networks:** BlogCatalog, Reddit, Epinions, GDELT, Github Dataset, Social Evolution Dataset, Enron and FB-FORUM.

Results are not consistent across the literature with best results ranging from 78% to 98% for different datasets for node classification task.

Conclusions & Future Directions

- The literature is not clear when is defining what kind of graph are dealing.
- Spatial models have proven to outperform spectral approaches.
- Small steps integrating Generative Adversarial Networks and Reinforcement Learning have been already done.
- Dealing with dynamic graph where edges and nodes are removed/inserted is an open field.
- Going deeper will bring scalability problems, sampling strategic and sub-graphs explorations could solve the problems.

References: [1] Wu, Zonghan, et al. "A comprehensive survey on graph neural networks." (2019).