CogDL: An Extensive Research Toolkit for Deep Learning on Graphs

Knowledge Engineering Group (KEG)

Tsinghua University, Beijing

What is CogDL

CogDL is a graph representation learning toolkit that allows researchers and developers to easily train and compare baseline or custom models for node classification, link prediction and other tasks on graphs. It provides implementations of many popular models, including: non-GNN Baselines like Deepwalk, LINE, NetMF, GNN Baselines like GCN, GAT, GraphSAGE.

CogDL support these following features:

- **Task-Oriented**: CogDL focuses on tasks on graphs and provides corresponding models, datasets, and leaderboards.
- **Easy-Running**: CogDL supports running multiple experiments simultaneously on multiple models and datasets under a specific task using multiple GPUs.
- **Multiple Tasks**: CogDL supports node classification and link prediction tasks on homogeneous/heterogeneous networks, as well as graph classification.
- **Extensibility**: You can easily add new datasets, models and tasks and conduct experiments for them!

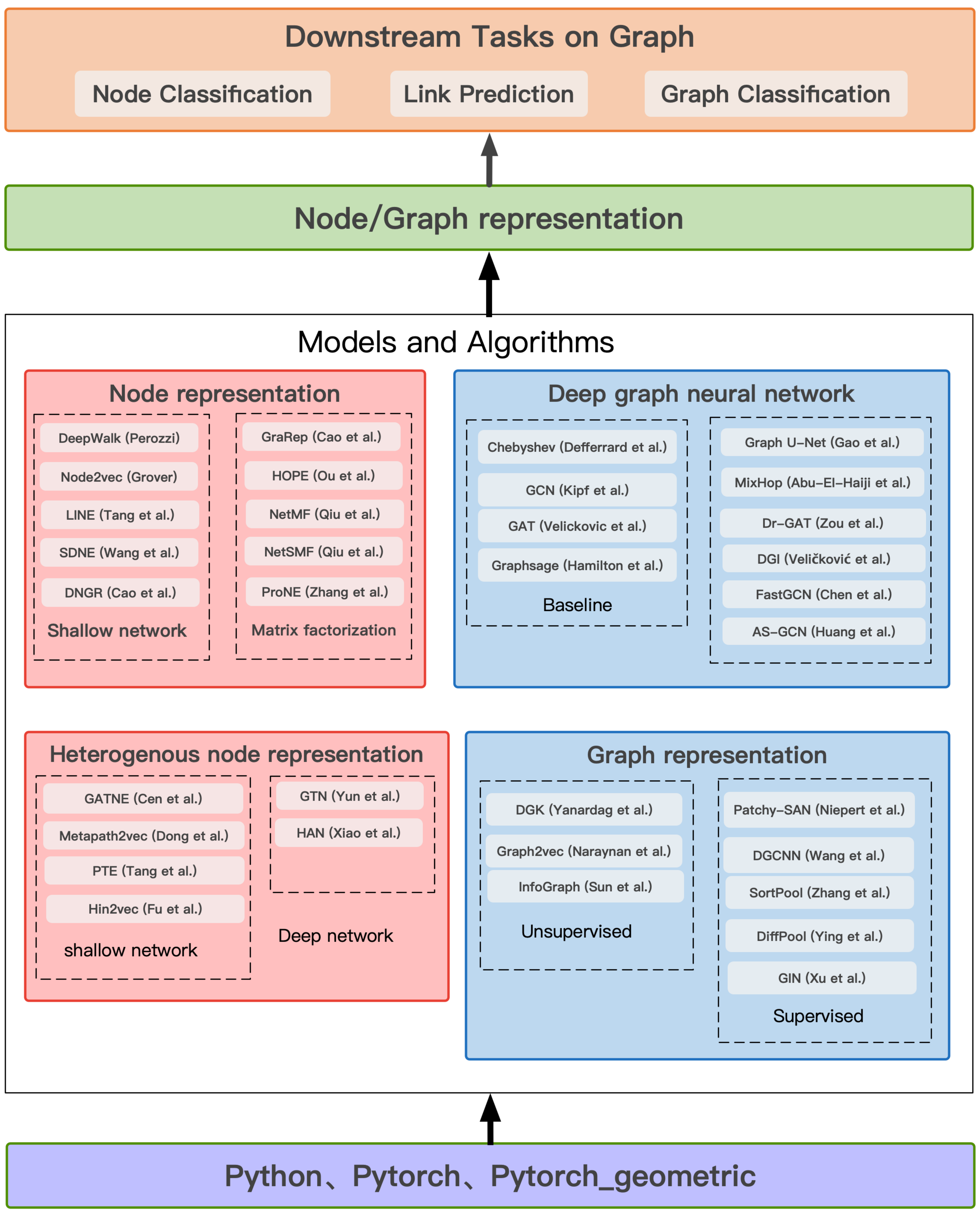

Figure 1: Overall framework of CogDL

Getting started

Basic Usage:

python train.py --task example_task --dataset example_dataset --model example_method to run example_method on example_data and evaluate it via example_task.

- --task, downstream tasks to evaluate representation like node_classification, unsupervised_node_classification, link_prediction. More tasks can be found in the cogdl/tasks.
- --dataset, dataset name to run, can be a list of datasets with space like cora citeseer ppi. Supported datasets include 'cora', 'citeseer', 'pumbed', 'PPI', 'wikipedia', 'blogcatalog', 'dblp', 'flickr'. More datasets can be found in the cogdl/datasets.
- --model, model name to run, can be a list of models like deepwalk line prone. Supported datasets include 'gcn', 'gat', 'graphsage', 'deepwalk', 'node2vec', 'hope', 'prone', 'netmf', 'netsmf', 'prone'. More models can be found in the cogdl/models.

For specific parameters for each algorithms, you can read `https://github.com/THUDM/cogdl/tree/master/cogdl/models`.

Customization

If you have a well-perform algorithm or unique dataset and are willing to publicize it, you can submit your implementation or dataset in following ways via opening an issue in our repository or commenting on our website.

- Submit your state-of-the-art results into our leaderboard
- Add your own dataset into the leaderboard
- Implement your own model into our toolkit

LeaderBoard

CogDL provides several downstream tasks including node classification(with or without node attributes), link prediction(with or without attributes, heterogeneous or not) to evaluate implemented methods.

Multi-label Node Classification

Here is an example leaderboard, which built from unsupervised multi-label node classification setting. We run all algorithms on several real-world datasets and report the sorted experimental results (Micro-F1 score with 90% labels as training data in L2 normalization logistic regression).

Rank	Algorithm	PPI	Blogcatalog	Wikipedia
1	ProNE(Zhang et al, IJCAI'19)	**26.32**	**43.63**	57.64
2	NetMF(Qiu et al, WSDM'18)	24.86	43.49	**58.46**
3	Node2vec(Grover et al, KDD'16)	23.86	42.51	53.68
4	NetSMF(Qiu et at, WWW'19)	24.39	43.21	51.42
5	DeepWalk(Perozzi et al, KDD'14)	22.72	42.26	50.42
6	LINE(Tang et al, WWW'15)	23.15	39.29	49.83
7	Hope(Ou et al, KDD'16)	23.24	35.52	52.96
8	SDNE(Wang et al, KDD'16)	20.14	40.32	48.24
9	GraRep(Cao et al, CIKM'15):	20.96	34.35	51.84
10	DNGR(Cao et al, AAAI'16):	16.45	28.54	48.57

Node Classification with Attributes

This leaderboard reports the semi-supervised node classification under the transductive setting including several popular graph neural network methods.

Rank	Method	Cora	Citeseer	Pubmed
1	Graph U-Net (Gao et al., 2019)	**84.4 ± 0.6**	**73.2 ± 0.5**	79.6 ± 0.2
2	MixHop (Abu-El-Haija, ICML'19)	81.9 ± 0.4	71.4 ± 0.8	**80.8 ± 0.6**
3	DR-GAT (Zou et al., 2019)	83.6 ± 0.5	72.8 ± 0.8	79.1 ± 0.3
4	GAT (Veličković et al., ICLR'18)	83.0 ± 0.7	72.5 ± 0.7	79.0 ± 0.3
5	DGI (Veličković et al., ICLR'19)	82.3 ± 0.6	71.8 ± 0.7	76.8 ± 0.6
6	GCN (Kipf et al., ICLR'17)	81.4 ± 0.5	70.9 ± 0.5	79.0 ± 0.3
7	GraphSAGE (Hamilton, NeurIPS'17)	80.1 ± 0.2	66.2 ± 0.4	76.9 ± 0.7
8	Chebyshev (Defferrard, NeurIPS'16)	79.2 ± 1.4	69.3 ± 1.3	68.5 ± 1.2

Website and code for CogDL, http://keg.cs.tsinghua.edu.cn/cogdl/ and https://github.com/thudm/cogdl