CogDL: An Extensive Research Platform 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 train baseline or custom models for node classification, link prediction and other tasks on graphs. It provides implementations of several models, including: non-GNN Baselines like Deepwalk, LINE, NetMF, GNN Baselines like GCN, GAT, GraphSAGE.

CogDL support these following features:

- **Arbitrary**: Dealing with different graph structures attributed, multiplex and heterogeneous networks
- **Parallel**: Parallel training on multiple GPUs
- **Extensible**: Easily add new datasets, models, criterions and tasks
- **Sparsification**: Fast network embedding on large-scale networks with tens of millions of nodes

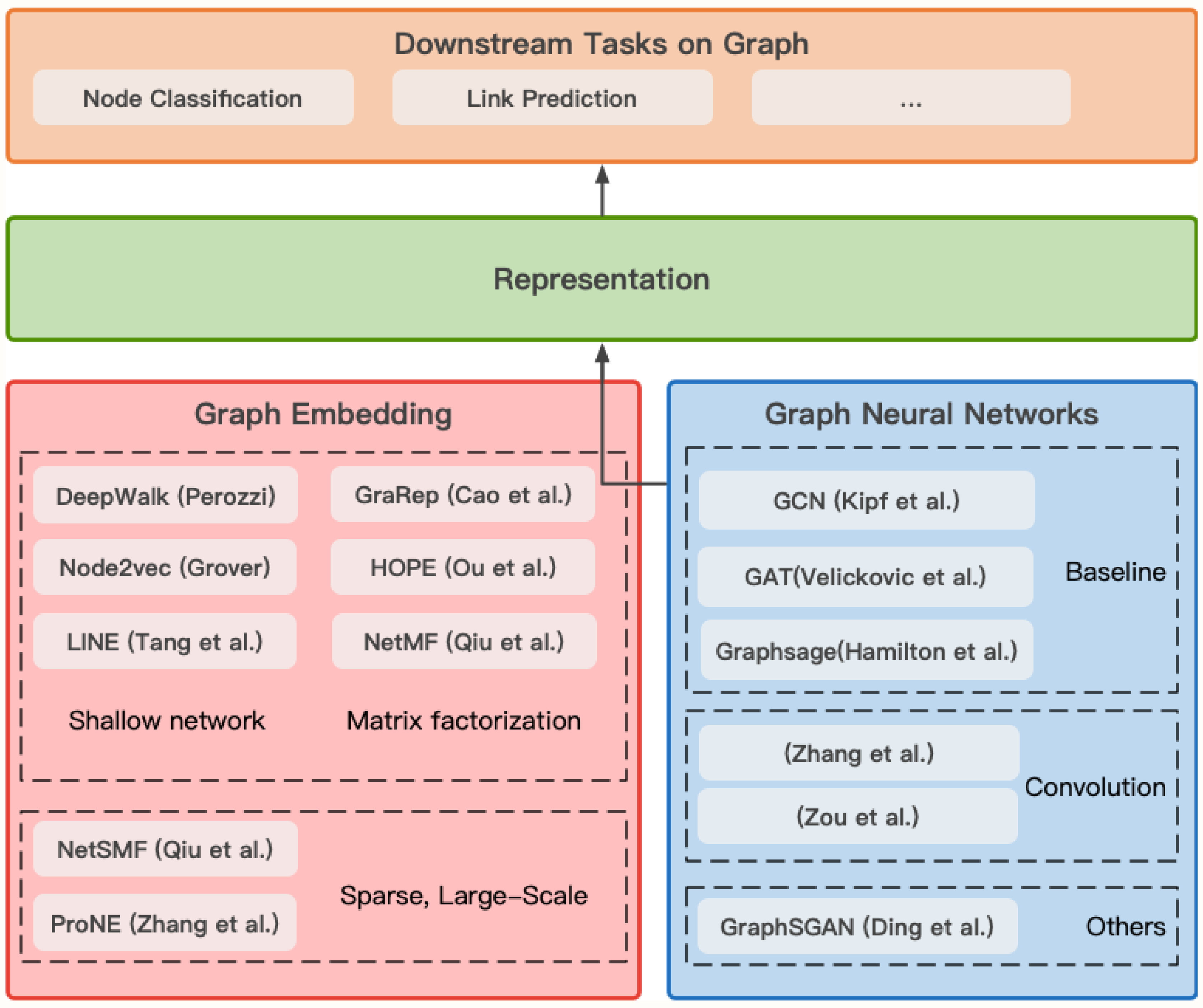

Figure 1: Overall framework of CogDL

How to use CogDL

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, Downsteam tasks to evaluate representation like node_classification, unsupervised_node_classification, link_prediction or multiplex_link_prediction.
- –datasets. Supported datasets including 'cora', 'citeseer', 'pumbed', 'PPI', 'wikipedia', 'blogcatalog', 'dblp', 'flickr'.
- –models including 'gcn', 'gat', 'graphsage', 'deepwalk', 'line' 'node2vec', 'hope', 'grarep', 'netmf', 'netsmf', 'prone'.

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

Customization

If you have a well-perform algorithm or unique dataset and are willing to add it into our leaderboard or toolkit , you can submit your implementation or dataset in following ways via opening an issue in our repository or join our slack group.

1. Submit Your State-of-the-Art Into The Leaderboard.
2. Add Your Own Dataset Into The Leaderboard.
3. Implement Your Own Model Into The Toolkit

Datasets

This section maintains a list of datasets supported by CogDL, including datasets with/without attributes.

Dataset	PPI	Wikipedia	Blogcatalog	DBLP
#Nodes	3,890	4,777	10,312	51,264
#Edges	76,584	184,812	333,983	127,968
#Labels	50	40	39	60

Table 1: Datasets without attributes for multi-label node classification

Dataset	Cora	Citeseer	Pubmed	PPI
Task	Transductive	Transductive	Transductive	Inductive
#Nodes	2,708	3,327	19,717	56,944(24 graphs)
#Edges	5,429	4,732	44,338	818,736
#Features	1,433	3,703	500	50
#Classes	7	6	3	121(multilabel)
#Training Nodes	140	120	60	44,906(20 graphs)
#Validation Nodes	500	500	500	6,514(2 graphs)
#Test Nodes	1,000	1,000	1,000	5,524(2 graphs)

Table 2: Datasets with attributes for multi-class node classification

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	GraRep(Cao et al., CIKM'15):	20.96	34.35	51.84

Node Classification with Attributes

Here is an example leaderboard built from supervised node classification setting including several popular graph neural network methods.

Rank	Method	Cora	Citeseer	Pubmed
1	NSGCN (Zhang et al., 2019)	$\mathbf{84.0 \pm 0.5}$	72.7 ± 0.4	$\mathbf{79.2 \pm 0.3}$
2	DR-GAT (Zou et al., 2019)	83.6 ± 0.5	$\mathbf{72.8 \pm 0.8}$	79.1 ± 0.3
3	DR-GCN (Zou et al., 2019)	81.6 ± 0.1	71.0 ± 0.6	79.2 ± 0.4
4	GAT (Velikovi et al., ICLR'18)	83.0 ± 0.7	72.5 ± 0.7	79.0 ± 0.3
5	GCN (Kipf et al., ICLR'17)	81.4 ± 0.5	70.9 ± 0.5	79.0 ± 0.3
6	FastGCN (Chen, Ma et al., ICLR'18)	81.4 ± 0.5	68.8 ± 0.9	77.6 ± 0.5
7	GraphSAGE (Hamilton et al., NIPS'17)	78.9 ± 0.8	67.4 ± 0.7	77.8 ± 0.6

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