4.5 Loading OGB datasets using `ogb` package

(中文版)

Open Graph Benchmark (OGB) is a collection of benchmark datasets. The official OGB package ogb provides APIs for downloading and processing OGB datasets into `dgl.data.DGLGraph` objects. The section introduce their basic usage here.

First install ogb package using pip:

```
pip install ogb
```

The following code shows how to load datasets for *Graph Property Prediction* tasks.

```
# Load Graph Property Prediction datasets in OGB
import dgl
import torch
from ogb.graphproppred import DglGraphPropPredDataset
from dgl.dataloading import GraphDataLoader


def _collate_fn(batch):
    # batch is a list of tuple (graph, label)
    graphs = [e[0] for e in batch]
    g = dgl.batch(graphs)
    labels = [e[1] for e in batch]
    labels = torch.stack(labels, 0)
    return g, labels

# load dataset
dataset = DglGraphPropPredDataset(name='ogbg-molhiv')
split_idx = dataset.get_idx_split()
# dataLoader
train_loader = GraphDataLoader(dataset[split_idx["train"]], batch_size=32, shuffle=True,
collate_fn=_collate_fn)
valid_loader = GraphDataLoader(dataset[split_idx["valid"]], batch_size=32, shuffle=False,
collate_fn=_collate_fn)
test_loader = GraphDataLoader(dataset[split_idx["test"]], batch_size=32, shuffle=False,
collate_fn=_collate_fn)
```

Loading *Node Property Prediction* datasets is similar, but note that there is only one graph object in this kind of dataset.

```python
# Load Node Property Prediction datasets in OGB
from ogb.nodeproppred import DglNodePropPredDataset

dataset = DglNodePropPredDataset(name='ogbn-proteins')
split_idx = dataset.get_idx_split()

# there is only one graph in Node Property Prediction datasets
g, labels = dataset[0]
# get split labels
train_label = dataset.labels[split_idx['train']]
valid_label = dataset.labels[split_idx['valid']]
test_label = dataset.labels[split_idx['test']]
```

Link Property Prediction datasets also contain one graph per dataset.

```python
# Load Link Property Prediction datasets in OGB
from ogb.linkproppred import DglLinkPropPredDataset

dataset = DglLinkPropPredDataset(name='ogbl-ppa')
split_edge = dataset.get_edge_split()

graph = dataset[0]
print(split_edge['train'].keys())
print(split_edge['valid'].keys())
print(split_edge['test'].keys())
```