

torch_geometric.datasets

`class KarateClub(transform=None) [source]`

Zachary's karate club network from the “An Information Flow Model for Conflict and Fission in Small Groups” paper, containing 34 nodes, connected by 154 (undirected and unweighted) edges. Every node is labeled by one of two classes.

Parameters: `transform (callable, optional)` – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)

`class TUDataset(root, name, transform=None, pre_transform=None, pre_filter=None, use_node_attr=False) [source]`

A variety of graph kernel benchmark datasets, .e.g. “IMDB-BINARY”, “REDDIT-BINARY” or “PROTEINS”, collected from the [TU Dortmund University](#).

Parameters:

- `root (string)` – Root directory where the dataset should be saved.
- `name (string)` – The [name](#) of the dataset.
- `transform (callable, optional)` – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
- `pre_transform (callable, optional)` – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
- `pre_filter (callable, optional)` – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)
- `use_node_attr (bool, optional)` – If `True`, the dataset will contain additional continuous node features (if present). (default: `False`)

`class Planetoid(root, name, transform=None, pre_transform=None) [source]`

The citation network datasets “Cora”, “CiteSeer” and “PubMed” from the “[Revisiting Semi-Supervised Learning with Graph Embeddings](#)” paper. Nodes represent documents and edges represent citation links. Training, validation and test splits are given by binary masks.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **name** (*string*) – The name of the dataset ("Cora" , "CiteSeer" , "PubMed").
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)

class `CoraFull`(*root, transform=None, pre_transform=None*) [\[source\]](#)

The full Cora citation network dataset from the "["Deep Gaussian Embedding of Graphs: Unsupervised Inductive Learning via Ranking"](#) paper. Nodes represent documents and edges represent citation links.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)

class `Coauthor`(*root, name, transform=None, pre_transform=None*) [\[source\]](#)

The Coauthor CS and Coauthor Physics networks from the "["Pitfalls of Graph Neural Network Evaluation"](#) paper. Nodes represent authors that are connected by an edge if they co-authored a paper. Given paper keywords for each author's papers, the task is to map authors to their respective field of study.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **name** (*string*) – The name of the dataset ("CS" , "Physics").
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)

class `Amazon`(*root, name, transform=None, pre_transform=None*) [\[source\]](#)

The Amazon Computers and Amazon Photo networks from the “[Pitfalls of Graph Neural Network Evaluation](#)” paper. Nodes represent goods and edges represent that two goods are frequently bought together. Given product reviews as bag-of-words node features, the task is to map goods to their respective product category.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **name** (*string*) – The name of the dataset (`"Computers"` , `"Photo"`).
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)

class `PPI(root, split='train', transform=None, pre_transform=None, pre_filter=None)` [\[source\]](#)

Protein-protein interaction networks from the “[Predicting Multicellular Function through Multi-layer Tissue Networks](#)” paper, containing positional gene sets, motif gene sets and immunological signatures as features (50 in total) and gene ontology sets as labels (121 in total).

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **split** (*string*) – If `"train"` , loads the training dataset. If `"val"` , loads the validation dataset. If `"test"` , loads the test dataset. (default: `"train"`)
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable, optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

class `QM7b(root, transform=None, pre_transform=None, pre_filter=None)` [\[source\]](#)

The QM7b dataset from the “[MoleculeNet: A Benchmark for Molecular Machine Learning](#)” paper, consisting of 7,211 molecules with 14 regression targets.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable, optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

class `QM9(root, transform=None, pre_transform=None, pre_filter=None)` [\[source\]](#)

The QM9 dataset from the “[MoleculeNet: A Benchmark for Molecular Machine Learning](#)” paper, consisting of about 130,000 molecules with 13 regression targets. Each molecule includes complete spatial information for the single low energy conformation of the atoms in the molecule. In addition, we provide the atom features from the “[Neural Message Passing for Quantum Chemistry](#)” paper.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable, optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

class `Entities(root, name, transform=None, pre_transform=None)` [\[source\]](#)

The relational entities networks “AIFB”, “MUTAG”, “BGS” and “AM” from the “[Modeling Relational Data with Graph Convolutional Networks](#)” paper. Training and test splits are given by node indices.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **name** (*string*) – The name of the dataset ("AIFB" , "MUTAG" , "BGS" , "AM").
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)

```
class MNISTSuperpixels(root, train=True, transform=None, pre_transform=None, pre_filter=None)
```

[\[source\]](#)

MNIST superpixels dataset from the “[Geometric Deep Learning on Graphs and Manifolds Using Mixture Model CNNs](#)” paper, containing 70,000 graphs with 75 nodes each. Every graph is labeled by one of 10 classes.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **train** (*bool, optional*) – If `True` , loads the training dataset, otherwise the test dataset. (default: `True`)
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable, optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

```
class FAUST(root, train=True, transform=None, pre_transform=None, pre_filter=None) [source]
```

The FAUST humans dataset from the “[FAUST: Dataset and Evaluation for 3D Mesh Registration](#)” paper, containing 100 watertight meshes representing 10 different poses for 10 different subjects.

⚠ Note

Data objects hold mesh faces instead of edge indices. To convert the mesh to a graph, use the `torch_geometric.transforms.FaceToEdge` as `pre_transform` . To convert the mesh to a point cloud, use the `torch_geometric.transforms.SamplePoints` as `transform` to sample a fixed number of points on the mesh faces according to their face area.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **train** (*bool*, *optional*) – If `True`, loads the training dataset, otherwise the test dataset. (default: `True`)
 - **transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable*, *optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

```
class ShapeNet(root, category, train=True, transform=None, pre_transform=None, pre_filter=None)
[source]
```

The ShapeNet part level segmentation dataset from the “[A Scalable Active Framework for Region Annotation in 3D Shape Collections](#)” paper, containing about 17,000 3D shape point clouds from 16 shape categories. Each category is annotated with 2 to 6 parts.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **category** (*string*) – The category of the CAD models (one of `"Airplane"`, `"Bag"`, `"Cap"`, `"Car"`, `"Chair"`, `"Earphone"`, `"Guitar"`, `"Knife"`, `"Lamp"`, `"Laptop"`, `"Motorbike"`, `"Mug"`, `"Pistol"`, `"Rocket"`, `"Skateboard"`, `"Table"`).
 - **train** (*bool*, *optional*) – If `True`, loads the training dataset, otherwise the test dataset. (default: `True`)
 - **transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable*, *optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

```
class ModelNet(root, name='10', train=True, transform=None, pre_transform=None, pre_filter=None)
[source]
```

The ModelNet10/40 datasets from the “[3D ShapeNets: A Deep Representation for Volumetric Shapes](#)” paper, containing CAD models of 10 and 40 categories, respectively.

Note

Data objects hold mesh faces instead of edge indices. To convert the mesh to a graph, use the `torch_geometric.transforms.FaceToEdge` as `pre_transform`. To convert the mesh to a point cloud, use the `torch_geometric.transforms.SamplePoints` as `transform` to sample a fixed number of points on the mesh faces according to their face area.

- Parameters:**
- `root` (`string`) – Root directory where the dataset should be saved.
 - `name` (`string, optional`) – The name of the dataset ("10" for ModelNet10, "40" for ModelNet40). (default: "10")
 - `train` (`bool, optional`) – If `True`, loads the training dataset, otherwise the test dataset. (default: `True`)
 - `transform` (`callable, optional`) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - `pre_transform` (`callable, optional`) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - `pre_filter` (`callable, optional`) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

`class CoMA(root, train=True, transform=None, pre_transform=None, pre_filter=None)` [\[source\]](#)

The CoMA 3D faces dataset from the “[Generating 3D faces using Convolutional Mesh Autoencoders](#)” paper, containing 20,466 meshes of extreme expressions captured over 12 different subjects.

Note

Data objects hold mesh faces instead of edge indices. To convert the mesh to a graph, use the `torch_geometric.transforms.FaceToEdge` as `pre_transform`. To convert the mesh to a point cloud, use the `torch_geometric.transforms.SamplePoints` as `transform` to sample a fixed number of points on the mesh faces according to their face area.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **train** (*bool*, *optional*) – If `True`, loads the training dataset, otherwise the test dataset. (default: `True`)
 - **transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable*, *optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

```
class PCPNetDataset(root, category, split='train', transform=None, pre_transform=None, pre_filter=None) [source]
```

The PCPNet dataset from the “[PCPNet: Learning Local Shape Properties from Raw Point Clouds](#)” paper, consisting of 30 shapes, each given as a point cloud, densely sampled with 100k points. For each shape, surface normals and local curvatures are given as node features.

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **category** (*string*) – The training set category (one of `"NoNoise"`, `"Noisy"`, `"VarDensity"`, `"NoisyAndVarDensity"` for `split="train"` or `split="val"`, or one of `"All"`, `"LowNoise"`, `"MedNoise"`, `"HighNoise"`, `:obj: "VarDensityStriped"`, `"VarDensityGradient"` for `split="test"`).
 - **split** (*string*) – If `"train"`, loads the training dataset. If `"val"`, loads the validation dataset. If `"test"`, loads the test dataset. (default: `"train"`)
 - **transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable*, *optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable*, *optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

```
class GeometricShapes(root, train=True, transform=None, pre_transform=None, pre_filter=None) [source]
```

Synthetic dataset of various geometric shapes like cubes, spheres or pyramids.

Note

Data objects hold mesh faces instead of edge indices. To convert the mesh to a graph, use the `torch_geometric.transforms.FaceToEdge` as `pre_transform`. To convert the mesh to a point cloud, use the `torch_geometric.transforms.SamplePoints` as `transform` to sample a fixed number of points on the mesh faces according to their face area.

Parameters:

- `root (string)` – Root directory where the dataset should be saved.
- `train (bool, optional)` – If `True`, loads the training dataset, otherwise the test dataset. (default: `True`)
- `transform (callable, optional)` – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
- `pre_transform (callable, optional)` – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
- `pre_filter (callable, optional)` – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

`class BitcoinOTC(root, edge_window_size=10, transform=None, pre_transform=None)` [source]

The Bitcoin-OTC dataset from the “EvolveGCN: Evolving Graph Convolutional Networks for Dynamic Graphs” paper, consisting of 138 who-trusts-whom networks of sequential time steps.

Parameters:

- `root (string)` – Root directory where the dataset should be saved.
- `edge_window_size (int, optional)` – The window size for the existence of an edge in the graph sequence since its initial creation. (default: `10`)
- `transform (callable, optional)` – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
- `pre_transform (callable, optional)` – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)

`class ICEWS18(root, split='train', transform=None, pre_transform=None, pre_filter=None)` [source]

The Integrated Crisis Early Warning System (ICEWS) dataset used in the, e.g., “[Recurrent Event Network for Reasoning over Temporal Knowledge Graphs](#)” paper, consisting of events collected from 1/1/2018 to 10/31/2018 (24 hours time granularity).

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **split** (*string*) – If “train”, loads the training dataset. If “val”, loads the validation dataset. If “test”, loads the test dataset. (default: “train”)
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable, optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)

```
class GDELT(root, split='train', transform=None, pre_transform=None, pre_filter=None) [source]
```

The Global Database of Events, Language, and Tone (GDEL) dataset used in the, e.g., “[Recurrent Event Network for Reasoning over Temporal Knowledge Graphs](#)” paper, consisting of events collected from 1/1/2018 to 1/31/2018 (15 minutes time granularity).

- Parameters:**
- **root** (*string*) – Root directory where the dataset should be saved.
 - **split** (*string*) – If “train”, loads the training dataset. If “val”, loads the validation dataset. If “test”, loads the test dataset. (default: “train”)
 - **transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before every access. (default: `None`)
 - **pre_transform** (*callable, optional*) – A function/transform that takes in an `torch_geometric.data.Data` object and returns a transformed version. The data object will be transformed before being saved to disk. (default: `None`)
 - **pre_filter** (*callable, optional*) – A function that takes in an `torch_geometric.data.Data` object and returns a boolean value, indicating whether the data object should be included in the final dataset. (default: `None`)