Graph Neural Networks

Graph Neural Networks

What is a Graph?

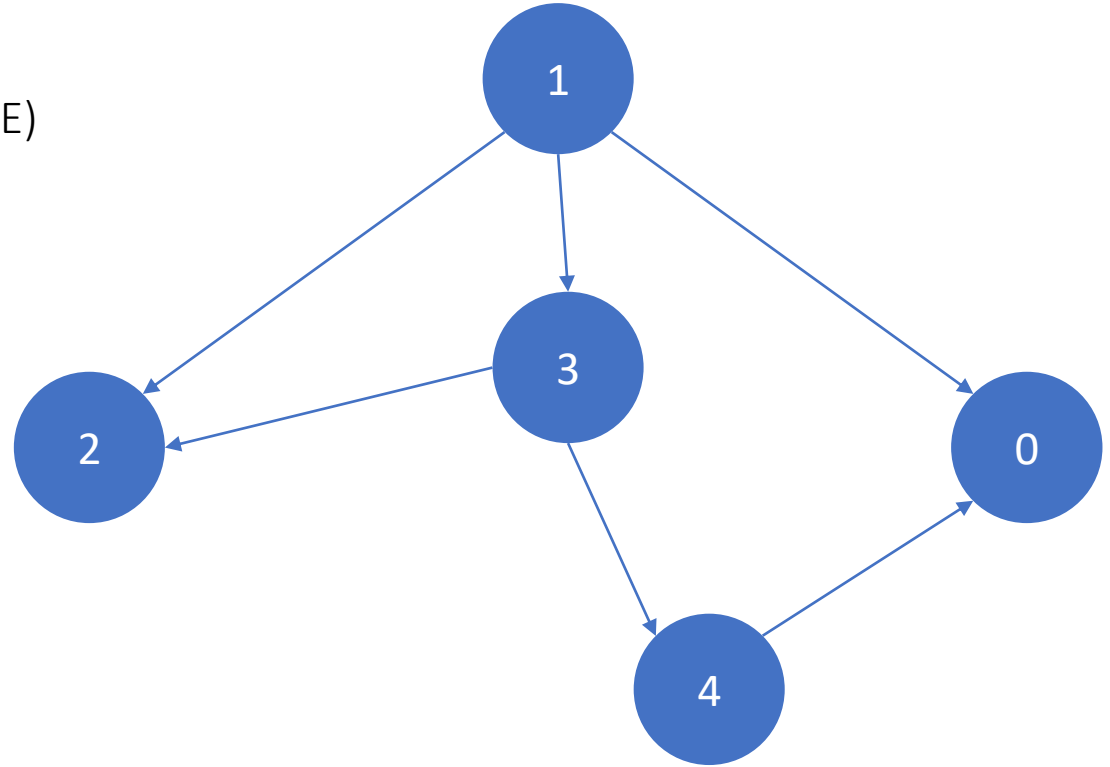
Graph

- data structure
- consists of nodes (N) and edges (E)
- G = (N, E)
- directed vs. undirected Graph

node

edge (directed)

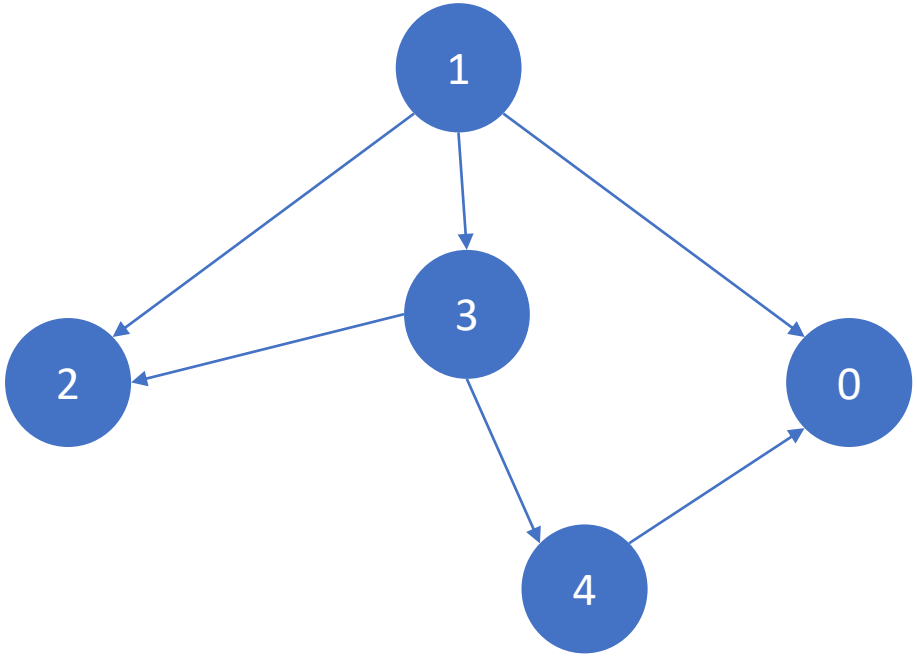
edge (undirected)



Graph Neural Networks

Graph Formats – Dense Matrix

Graph



Dense Matrix

source \ destination	0	1	2	3	4
0	0	0	0	0	0
1	1	0	1	1	0
2	0	0	0	0	0
3	0	0	1	0	1
4	1	0	0	0	0

Graph Neural Networks

Dense and Sparse Matrix

Sparse Matrix (COO)

source

1	1	1	3	3	4

destination

0	2	3	2	4	0

Dense Matrix

source \ destination	0	1	2	3	4
0	0	0	0	0	0
1	1	0	1	1	0
2	0	0	0	0	0
3	0	0	1	0	1
4	1	0	0	0	0

Dimensions: 2 vectors with # of edges

Dimension: #nodes x #nodes

Graph Neural Networks

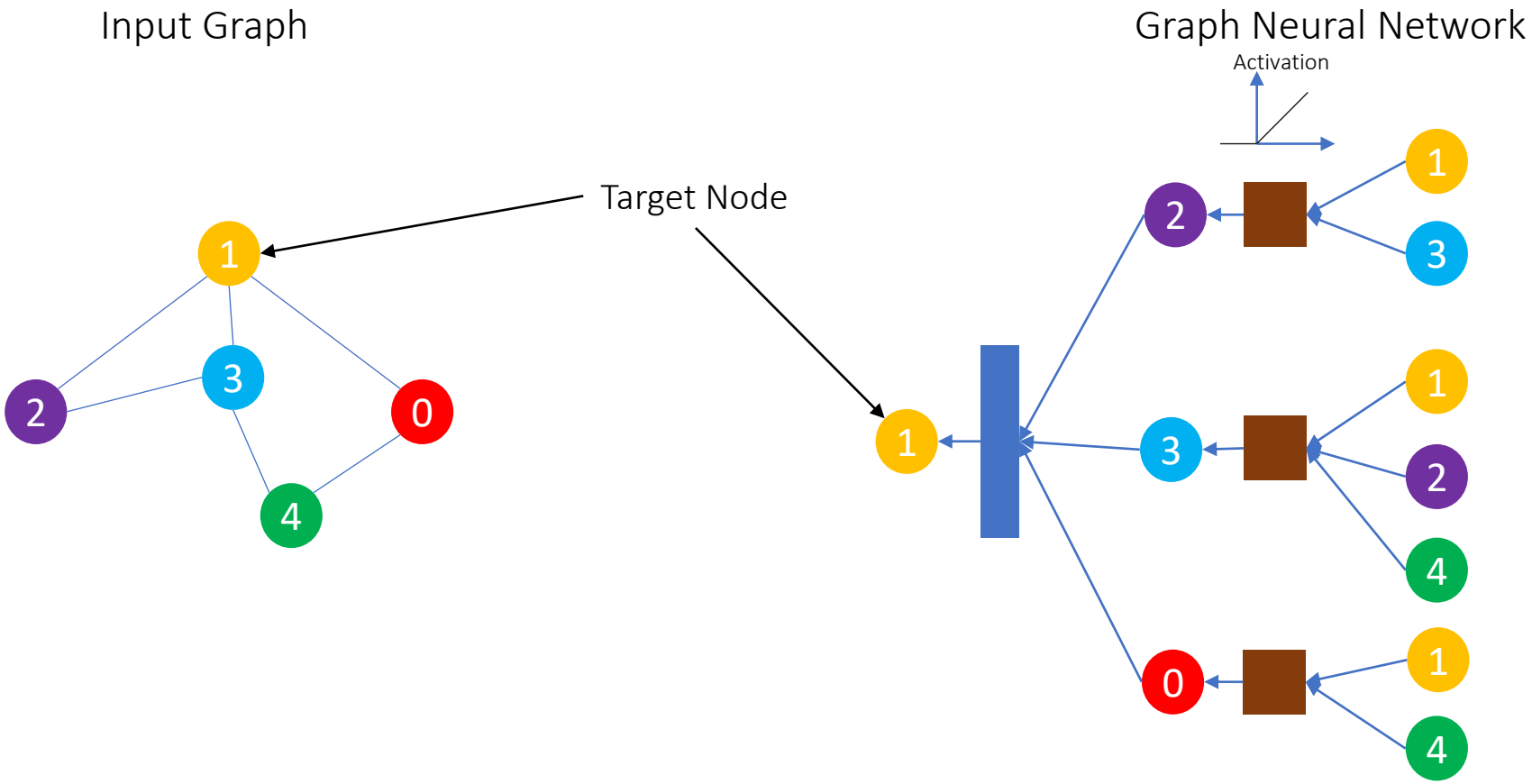
What are GNNs?

Neural Network Architecture

- combine deep learning and graph theory
- suitable for Graphs
- help analyzing graphs
- understand relationships between objects
- create predictions

Graph Neural Networks

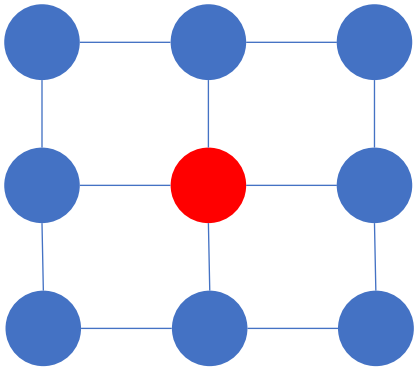
How do GNNs work?



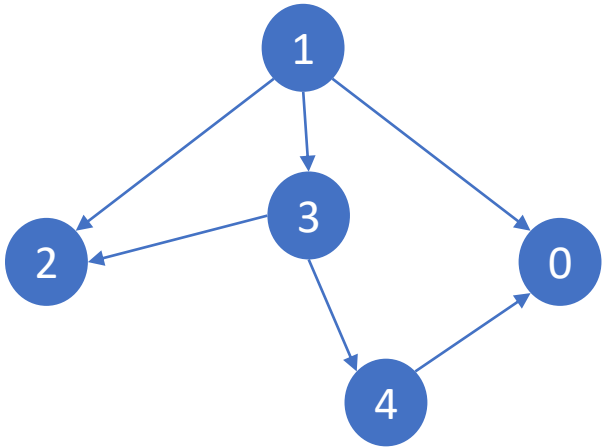
Graph Neural Networks

GNNs and CNNs

- lots of similarities with Convolutional Neural Networks (CNN)
- CNNs very useful for computer vision
- CNNs suitable for images (euclidean space), but not for graphs (non-euclidean space)



Euclidean Space

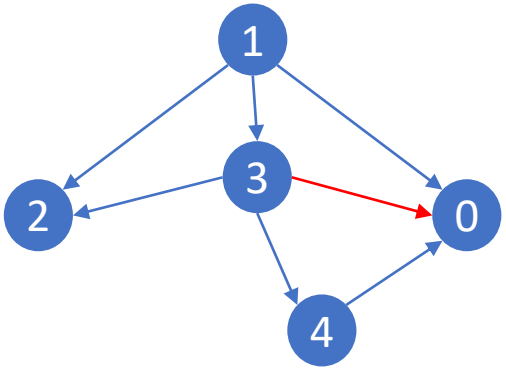


Non-Euclidean Space

Graph Neural Networks

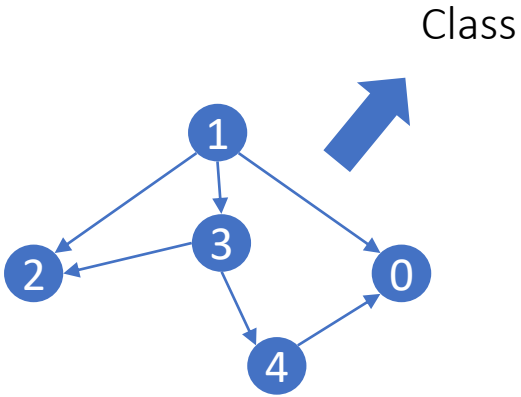
Applications

Link Prediction



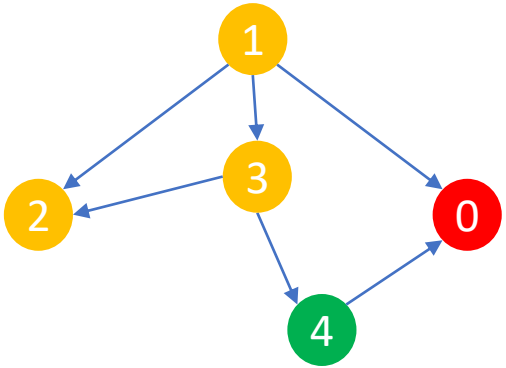
- predict link between nodes

Graph Classification



- predict graph class
- e.g. text classification, social network analysis

Node Classification



- predict node class

Graph Neural Networks

Advantages / Disadvantages

\+

- learn representations of nodes / edges and capture complex relationships
- can be used to predict graphs with varying sizes and structures
- applicable to different tasks like node classification, link prediction, graph classification, and more
- model real-world systems like social networks, recommender systems, or drug discovery

\-

- usually shallow networks and thus not suitable for very large datasets
- changing graph structure might require frequent re-training
- computationally expensive
- require significant training data
- can suffer from overfitting, especially on noisy or very sparse data
- still very new → more standardization required, more architectures, metrics

Graph Neural Networks

Coding

- PubMed dataset
- consists of 19.717 scientific publications
- diabetes classified into three classes
- 88.648 edges
- 500 features
- Task: predict node class

data.x
(19717, 500)

data.edge_indices
(2, 88648)

data.y
(19717)