Graph Representation Learning (GRL)

Zekarias Tilahun

Department of Electrical Engineering and Computer Science

KTH Royal Institute of Technology

Networks/Graphs

Social Networks

Facebook

Twitter

…

Biological Networks

Brain

Protein-Protein

Drug-Target

…

Product Networks

User-Item

Product-Product

..

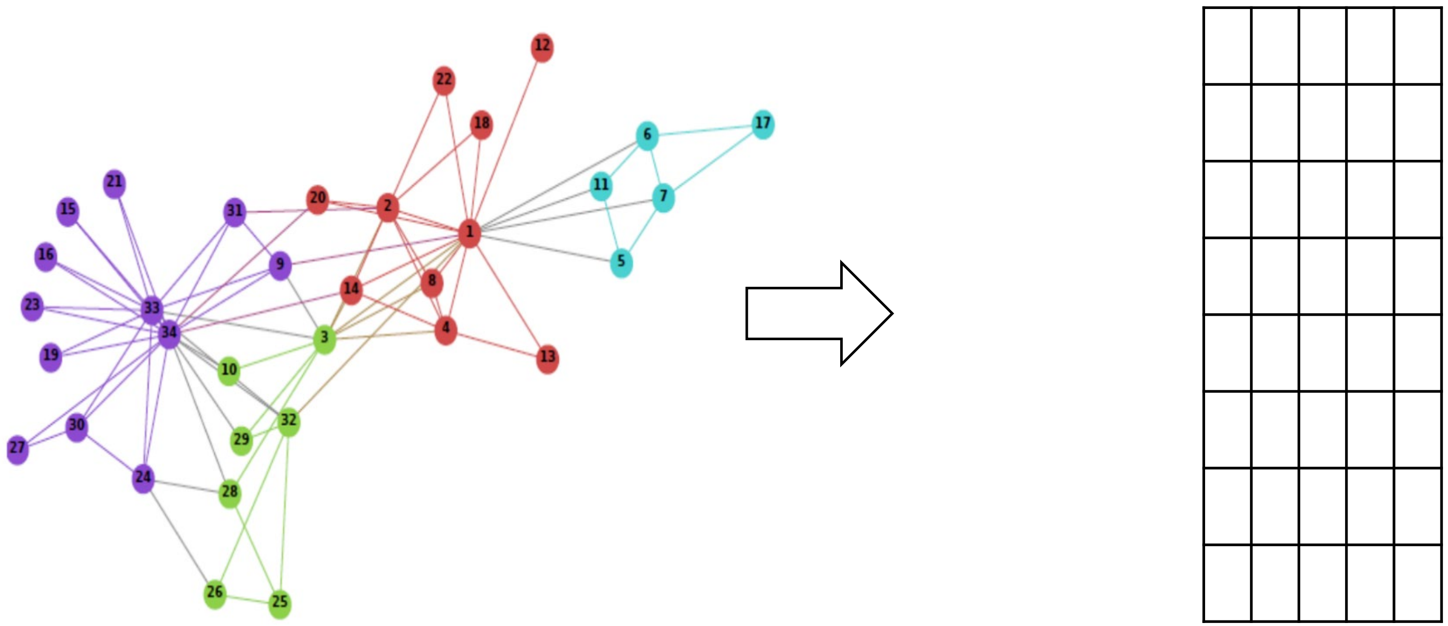
Basic Notations

- $G = (V, E)$
  - $N = |V|, M = |E|$
- $\boldsymbol{A} \in \{0, 1\}^{N \times N}$ – Adjacency matrix notation
  - $\boldsymbol{A}[i, j] = \begin{cases} 1, & (i, j) \in E \\ 0, & (i, j) \notin E \end{cases}$
- $\boldsymbol{X} \in \mathbb{R}^{N \times D}$ – Node feature matrix
  - $\boldsymbol{X}[i] = \boldsymbol{x}_i$- Node i's feature vector

What is GRL?

$N \times N$

$N \times d, d \ll N$



Input Space

Representation Space

What is GRL?

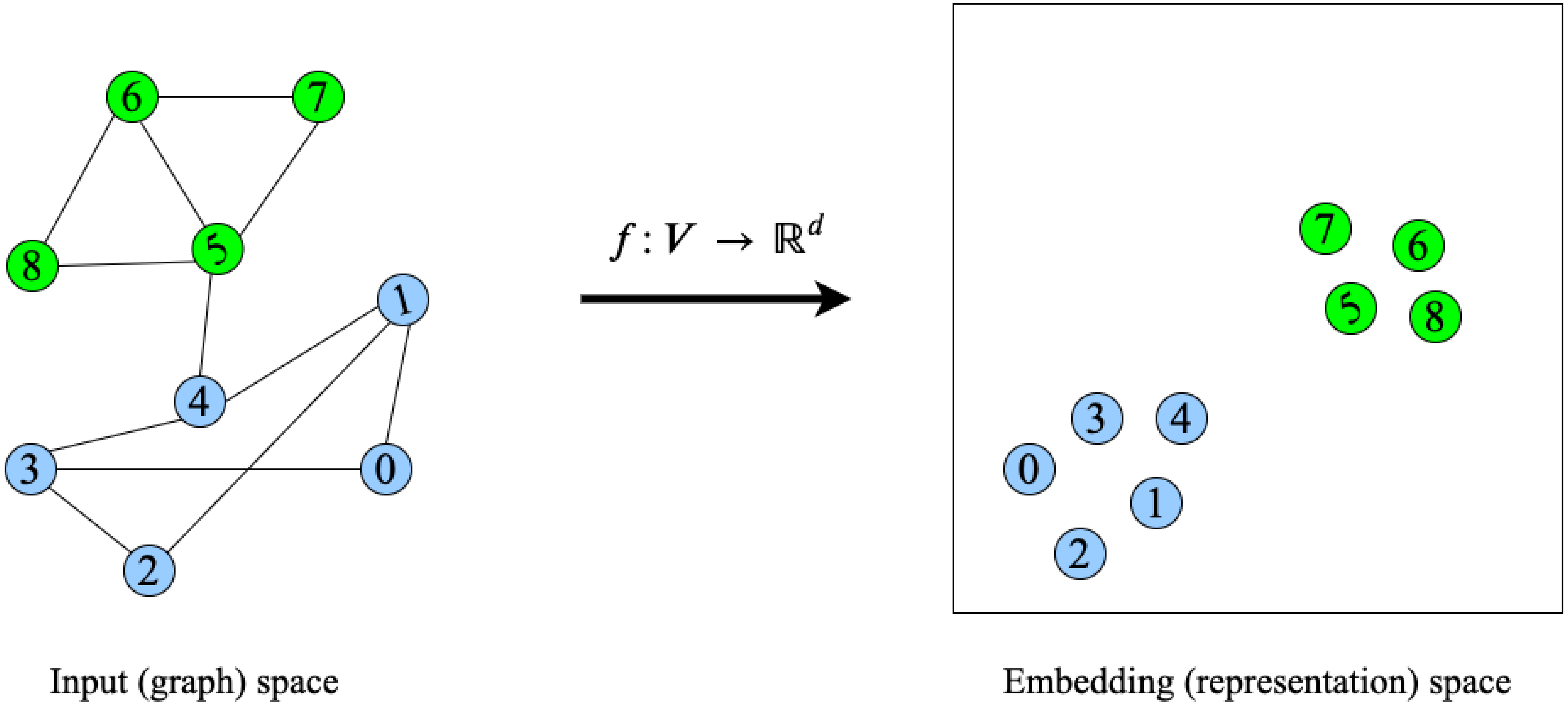
What makes a representation a good representation?

$\boldsymbol{z_i}$ is a d dimensional representation vector of node i

$\boldsymbol{Z}$ is the representation (embedding) matrix

Quality of Representation

- Preserve a certain property of the graph



Quality of Embedding

- We model the property we want to preserve using a similarity function in the input space

$$sim_g: V \times V \rightarrow \mathbb{R}$$

- Adjacency Matrix as a similarity function

$$sim_g(i,j) = \begin{cases} 1, & if\ \boldsymbol{A}[i,j] = 1 \\ 0, & if\ \boldsymbol{A}[i,j] = 0 \end{cases}$$

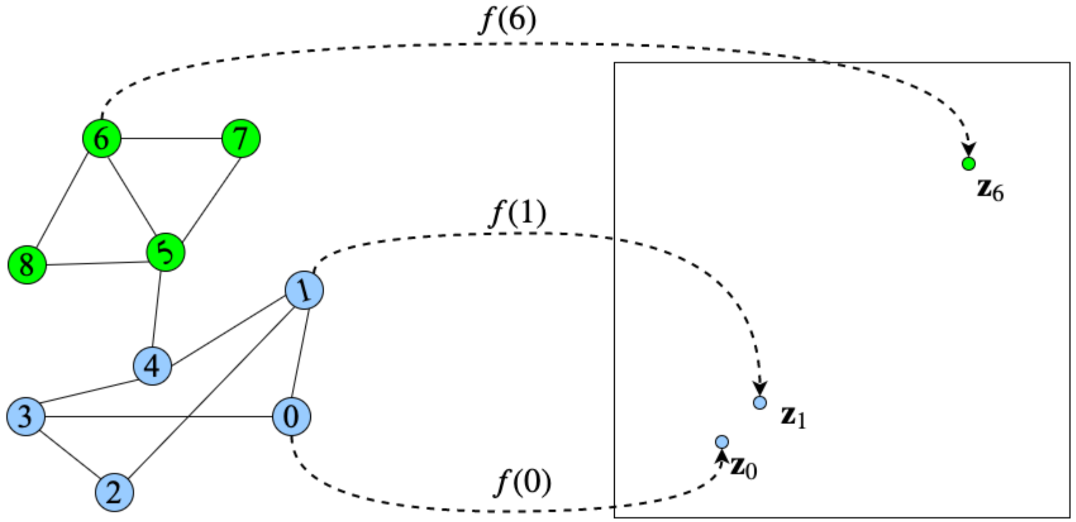
- Common Neighborhood
  - $sim_g(u,v) = |\mathcal{N}_u \cap \mathcal{N}_v|$
  - $\mathcal{N}_i = \{j: (i,j) \in E\}$
- Rooted Page Rank
  - $sim_g(u,v) = p$
  - p is the probability of reaching node v via a random walk starting from node u
- Generally, we use $\boldsymbol{W} \in [0,1]^{N \times N}$

$$\boldsymbol{W}[i,j] = sim_g(i,j)$$

Learning

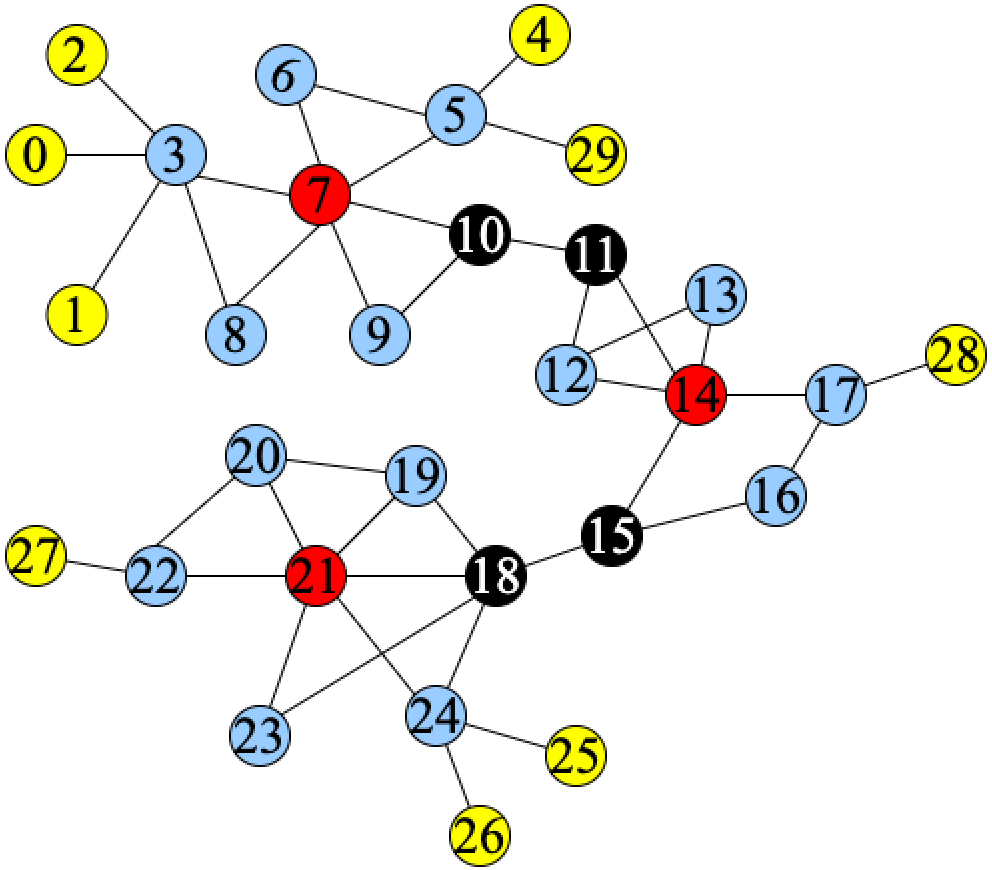
- Given a similarity function sim_g in the input space
- Learn a representation $\mathbf{z}_u = f(u) \in \mathbb{R}^d, u \in V$ that preserves sim_g
  - Similarity in the representation space, $sim_r: \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$
  - sim_r Should approximate sim_g

$$sim_r(f(u), f(v)) \approx sim_g(u, v)$$



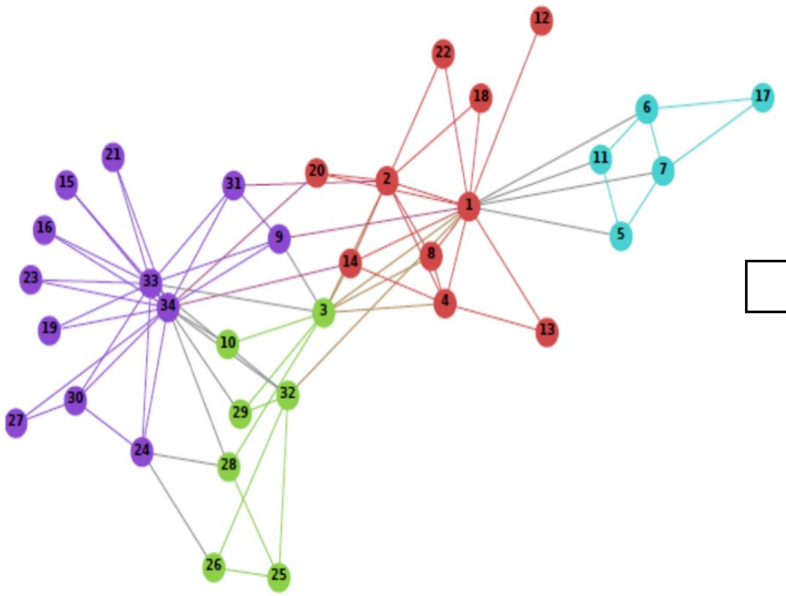
$$sim_r(f(0), f(1)) = sim_r(\mathbf{z}_0, \mathbf{z}_1) \approx sim_g(0,1)$$

Structural similarity



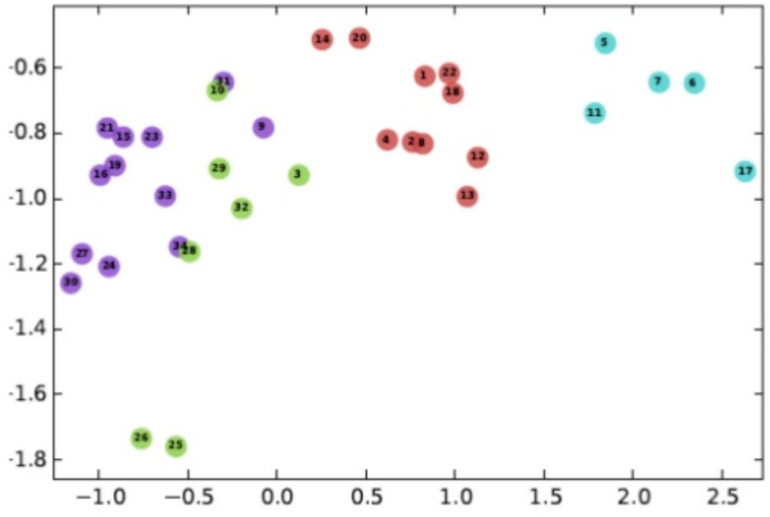
Why GRL?

$N \times N$



Input Space

$N \times d, d \ll N$



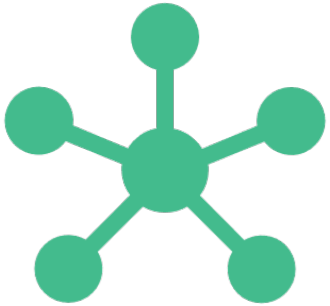
Representation Space

Network Analysis

- Link Prediction
- Node classification
- Clustering
- Recommendation
- ...

Learning Algorithms

Random Walk Based



Message Passing Neural Networks

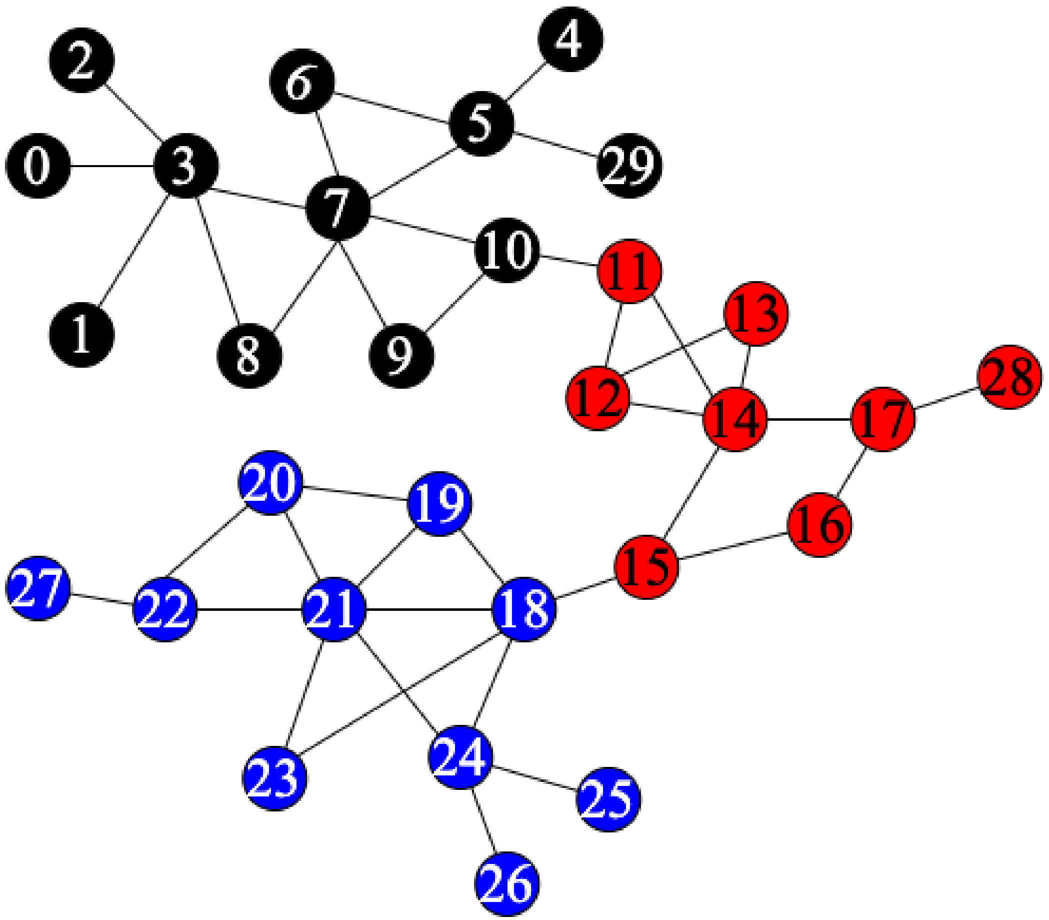
Random Walk (RW) Based

WALK SAMPLING

LEARNING

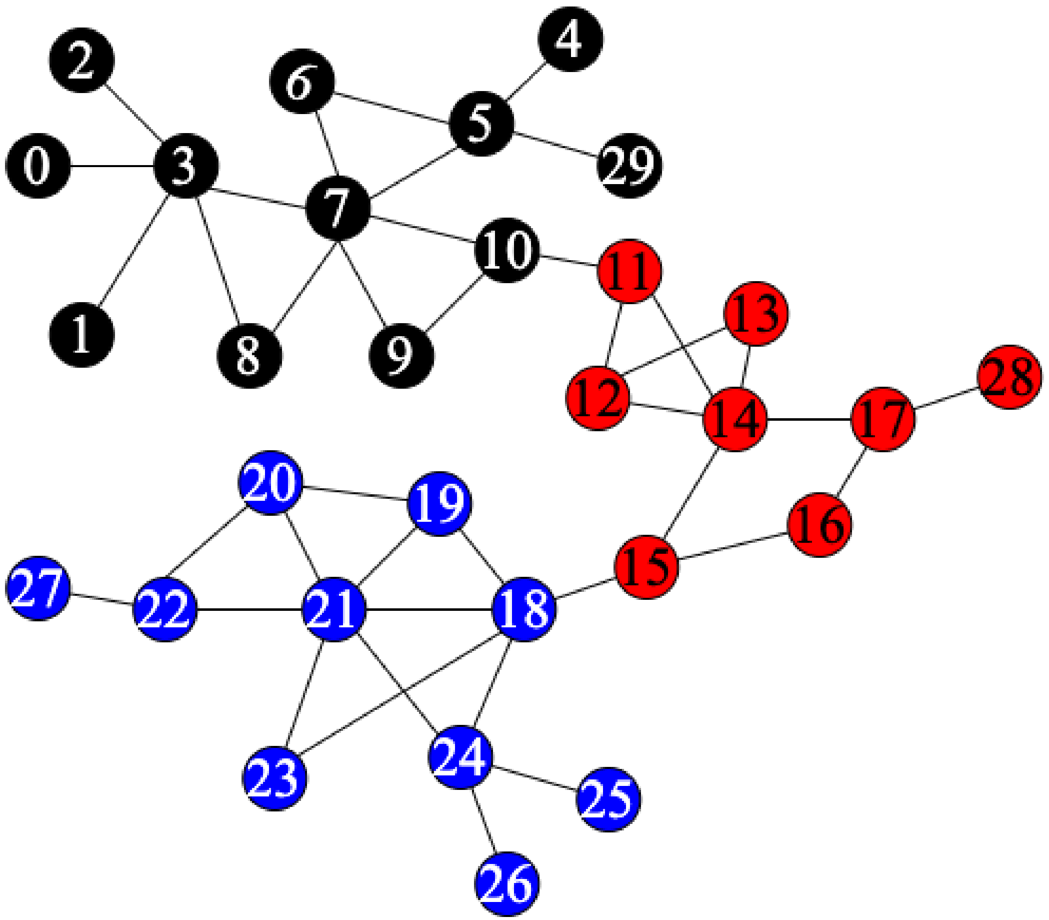
RW: Property

- Preserve Homophily: Similar nodes appear in the same neighborhood



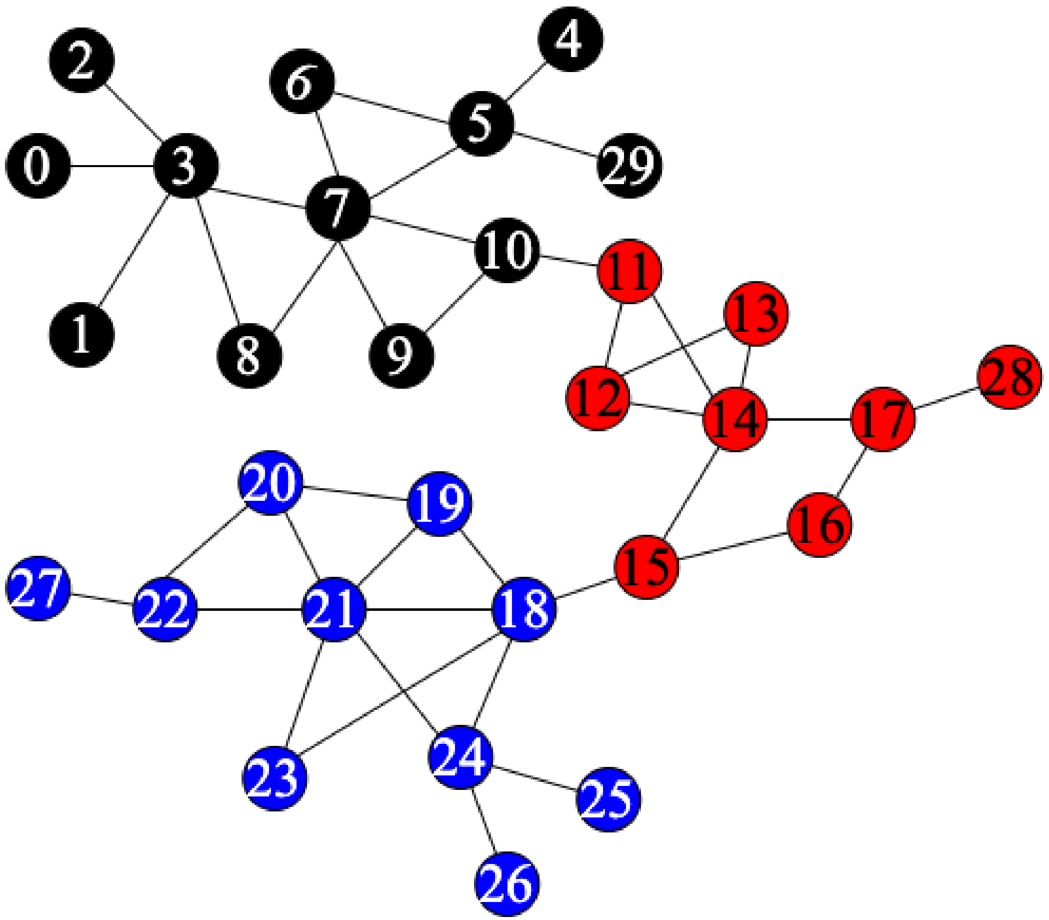
Challenge

- Graphs are arbitrary data structures
- GRL directly on top of graphs is difficult
- We need a different data structure that capture the homophily property (E.g., sequences)
    - Truncated random walks



RW Sampling

- Sample walks starting from node 21
  - 21, 24, 18, 23, 18
  - 21, 18, 19, 18, 15
  - 21, 19, 20, 22, 21
  - 21, 23, 18, 24, 21
- Sample walks starting from node 15
  - 15, 14, 12, 13, 12
  - 15, 14, 11, 14, 12
  - 15, 18, 15, 16, 17
  - 15, 14, 17, 28, 17
- …



Learning: Algorithm

- SkipGram: Used for word representation
- Key idea: Similar words frequently appear together in similar context
- E.g.:
  - Stockholm vs. Sweden
  - Vaccine vs. AstraZeneca
  - Darwin vs. Evolution
  - Arthur Conan Doyle vs. Sherlock Holmes
- Learning:
  - Given a target work, w_t
  - Learn its embedding, $\boldsymbol{z}_t$ by predicting its context words $w_c \in ctx(w_t)$

$$The, capital, of, Sweden, is, Stockholm$$

$$w_t = Sweden$$

$$ctx(w_t) = \{capital, stocholm\}$$

$$\max_{w_c \in ctx(w_t)} P(w_c | \boldsymbol{z}_t)$$

SkipGram for GRL

- Given a target node n_t from a random walk sequence
- Learn $\boldsymbol{z}_t$ by predicting its context nodes $n_c \in ctx(n_t)$

$ctx(18)$

$walk$

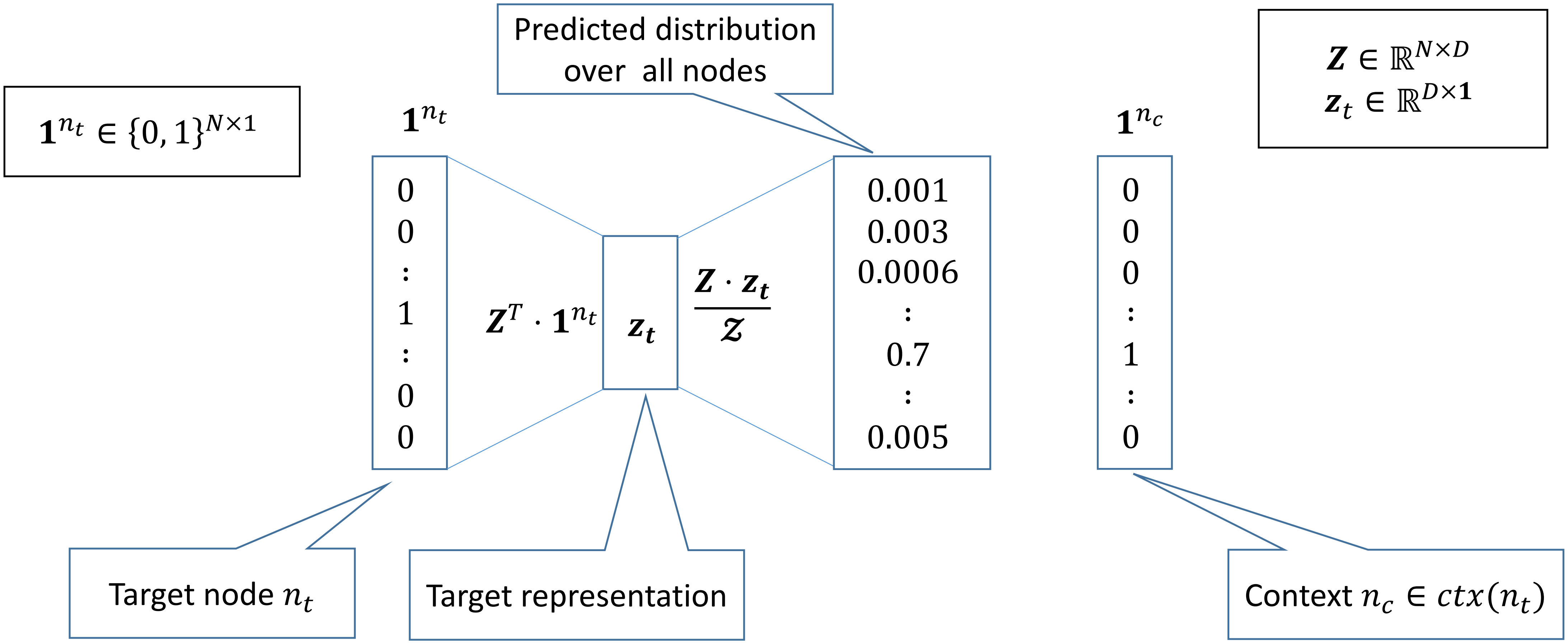
Finally update $\boldsymbol{z}_t$ using SGD according to the prediction errors

Current Context node

nt Target node=18

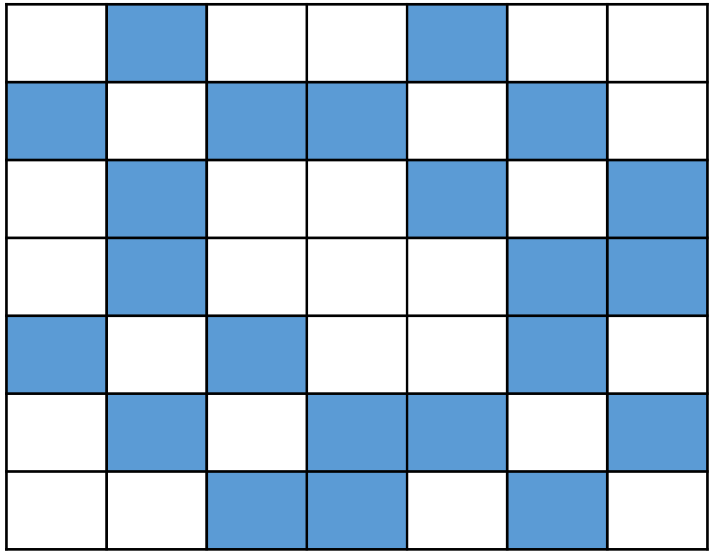
Pre	$Predict($	Pre	$Predict(24 \mid \boldsymbol{z}_{18})$

SkipGram Architecture

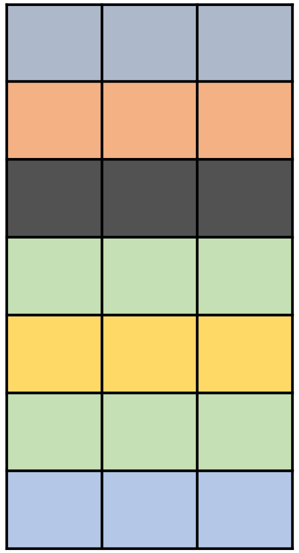


Message Passing Neural Networks (MPNN)

- Now, we consider nodes with features



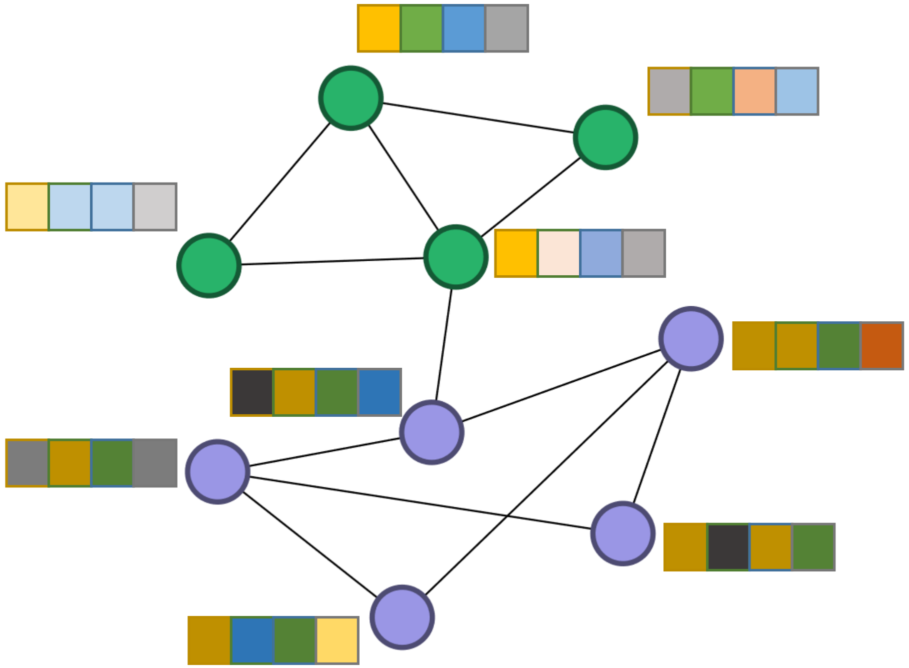
$\boldsymbol{A}$



$\boldsymbol{X}$

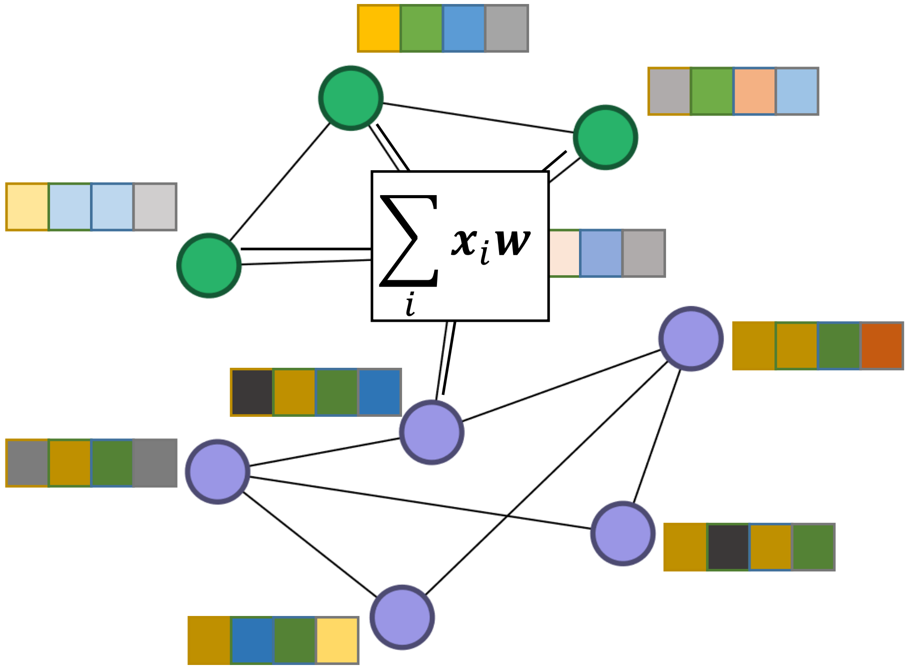
Message Passing Neural Networks (MPNN)

- Now, we consider nodes with features



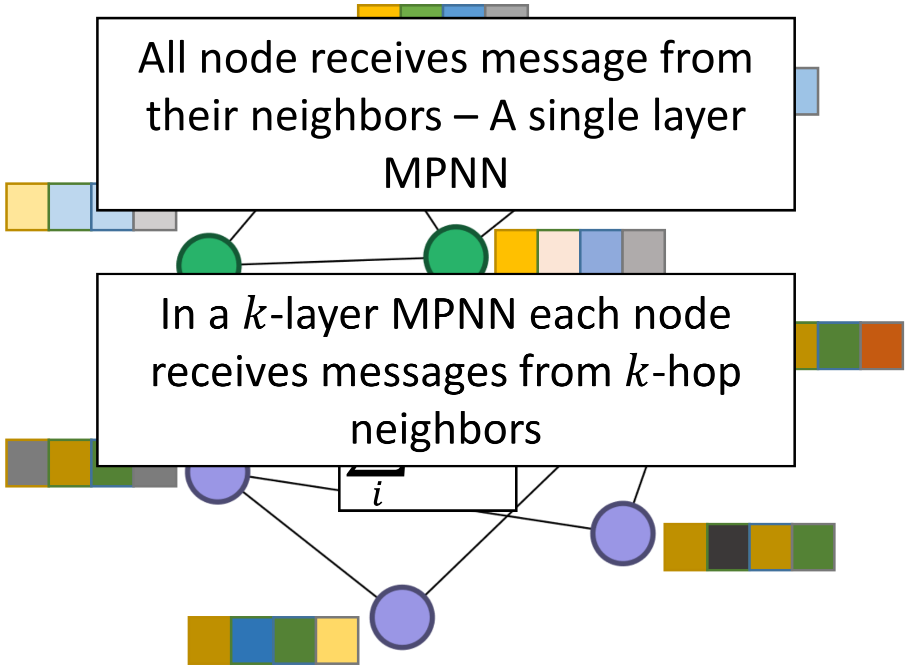
MPNN

- Key Idea: Each node sends and receives messages



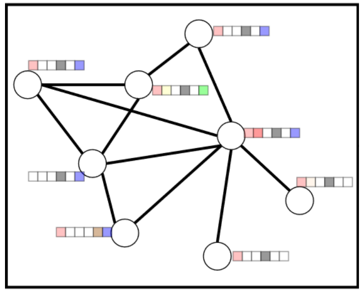
MPNN

- Key Idea: Each node sends and receives messages

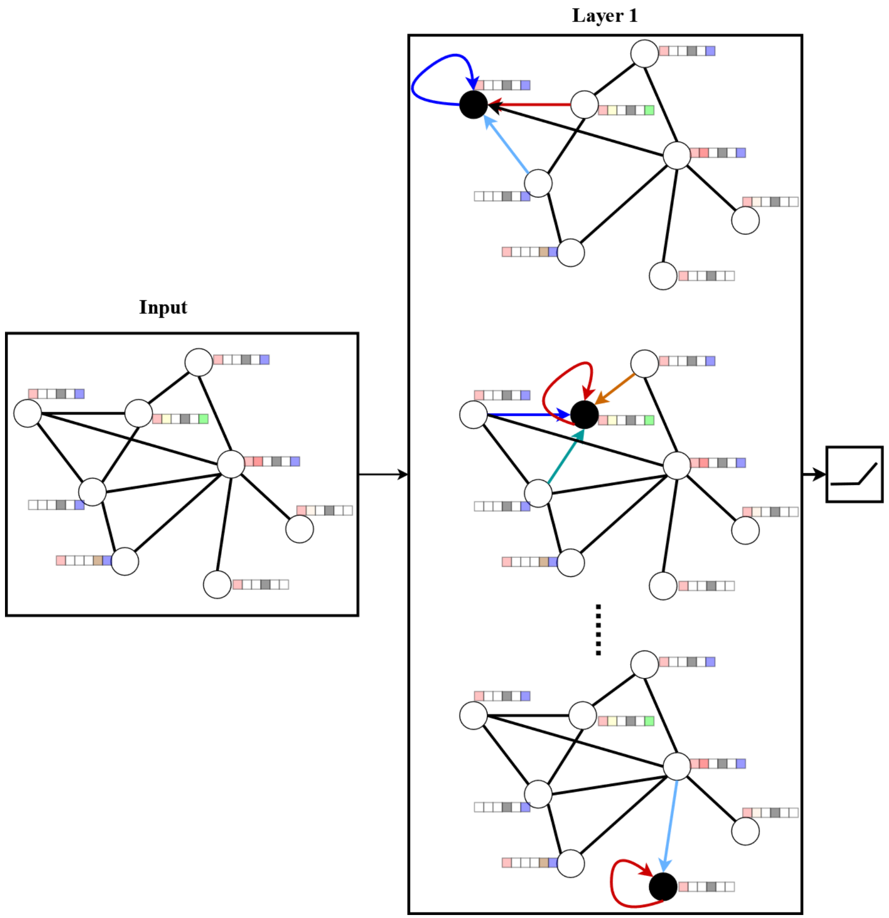


GCN

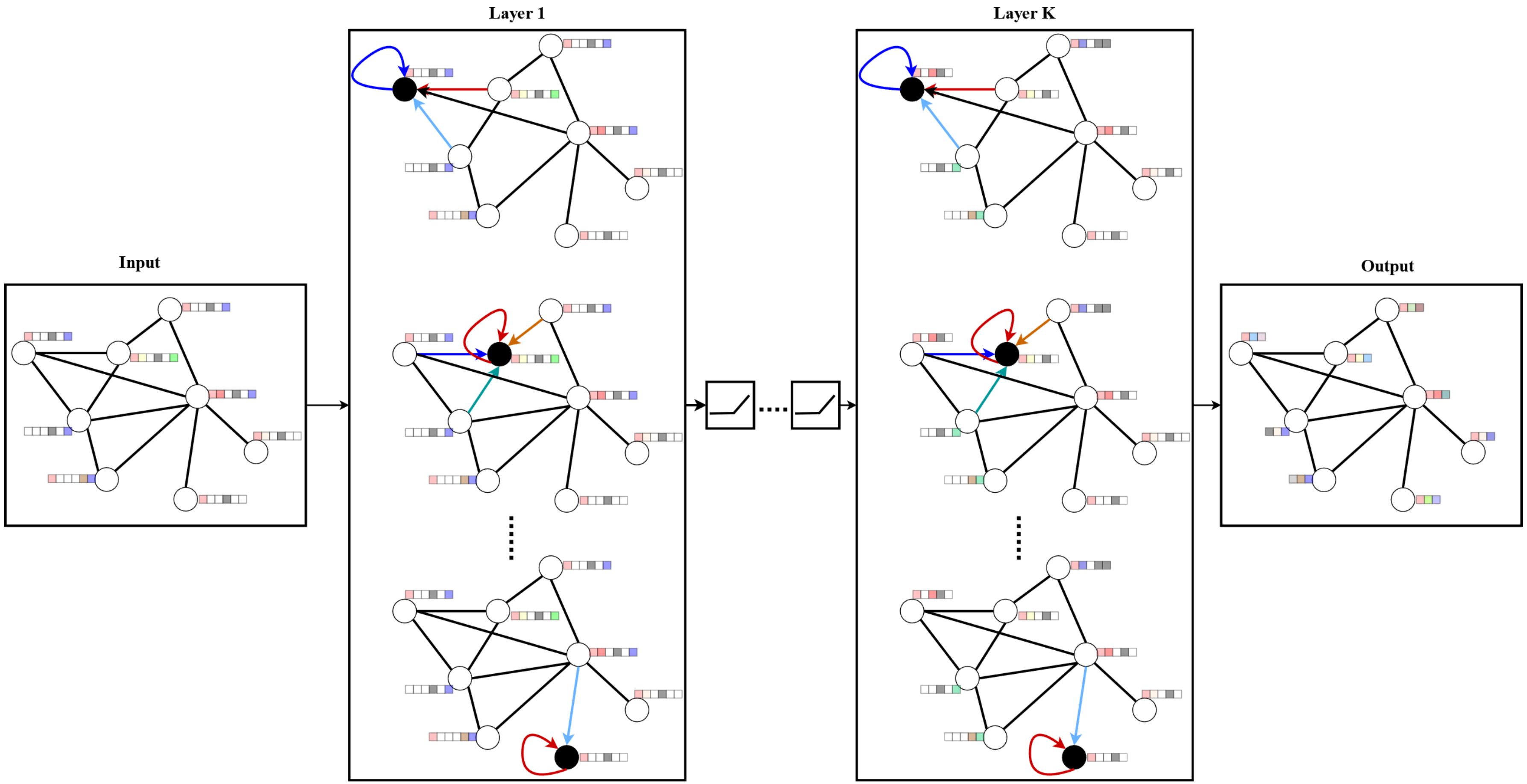
Input



GCN



GCN



Applications of GRL

- Social Networks
  - Friendship recommendation
- Recommendation Systems
  - Content (e.g., book, movie, …) recommendation
- Biomedical Systems
  - Drug discovery
  - Predicting functions of proteins
  - Predicting molecular properties
- …

Referrences

- Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. DeepWalk: online learning of social representations. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '14). ACM, USA
- Aditya Grover and Jure Leskovec. node2vec: Scalable Feature Learning for Networks. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16). ACM, NY, USA.
- Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Y. Chang. 2015. Network representation learning with rich text information. In Proceedings of the 24th International Conference on Artificial Intelligence (IJCAI'15), Qiang Yang and Michael Wooldridge (Eds.). AAAI Press 2111-2117.
- Shirui Pan, Jia Wu, Xingquan Zhu, Chengqi Zhang, and Yang Wang. 2016. Tri-party deep network representation. In Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16), Gerhard Brewka (Ed.). AAAI Press 1895-1901.
- Nasrullah Sheikh, Zekarias Kefato, and Alberto Montresor. GAT2VEC: Representation learning for attributed graphs. Computing Journal, 2018.
- T. N. Kipf, M. Welling, Semi-Supervised Classification with Graph Convolutional Networks (ICLR 2017)
- William L. Hamilton, Rex Ying, and Jure Leskovec. 2017. Inductive representation learning on large graphs. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Ulrike von Luxburg, Isabelle Guyon, Samy Bengio, Hanna Wallach, and Rob Fergus (Eds.). Curran Associates Inc., USA, 1025-1035.
- Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. and Bengio, Y. (2018) Graph Attention Networks. *The 6th International Conference on Learning Representations (ICLR 2018)*
- Alessandro Epasto and Bryan Perozzi. 2019. Is a Single Embedding Enough? Learning Node Representations that Capture Multiple Social Contexts. In The World Wide Web Conference (WWW '19), Ling Liu and Ryen White (Eds.). ACM, New York, NY, USA, 394-404.