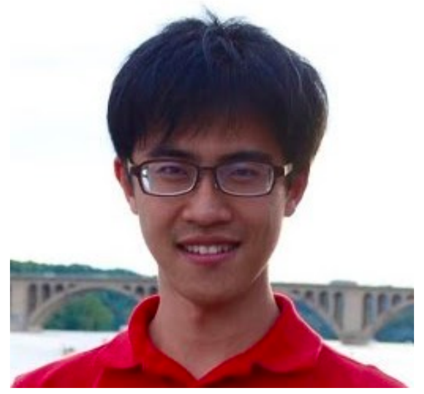
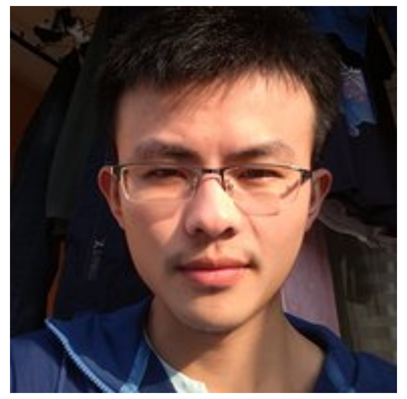
Simplifying Graph Convolutional Networks








Felix Wu* | Tianyi Zhang* | Amauri Holanda de Souza Júnior* | Christopher Fifty | Tao Yu | Kilian Q. Weinberger

*: Equal contribution

Graph Convolutional Networks Applications
(Kipf and Welling, ICLR 2017)

- Social Networks & Citation Networks
- Applied Chemistry
- Natural Language Processing
- Computer Vision
- …

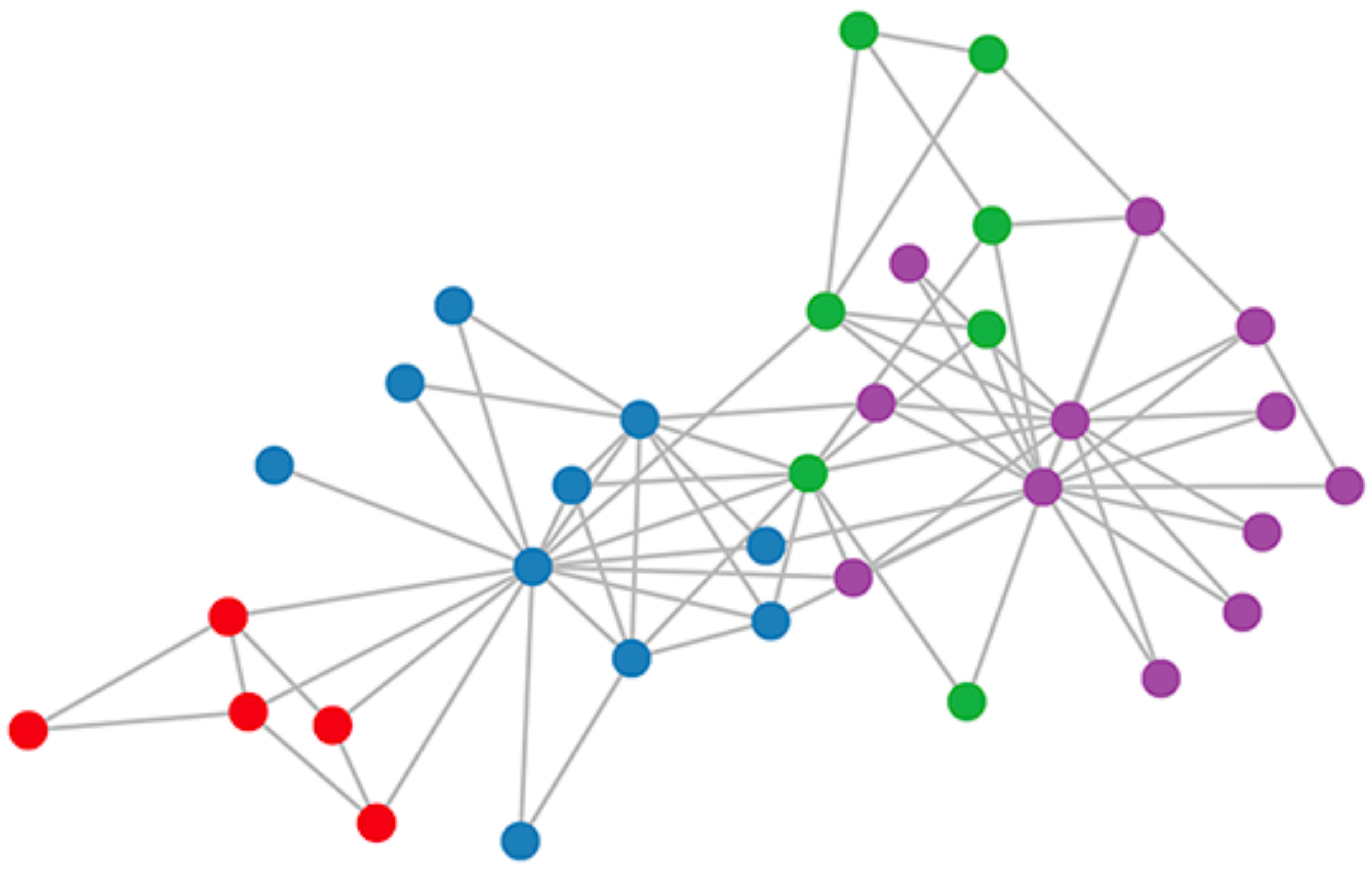
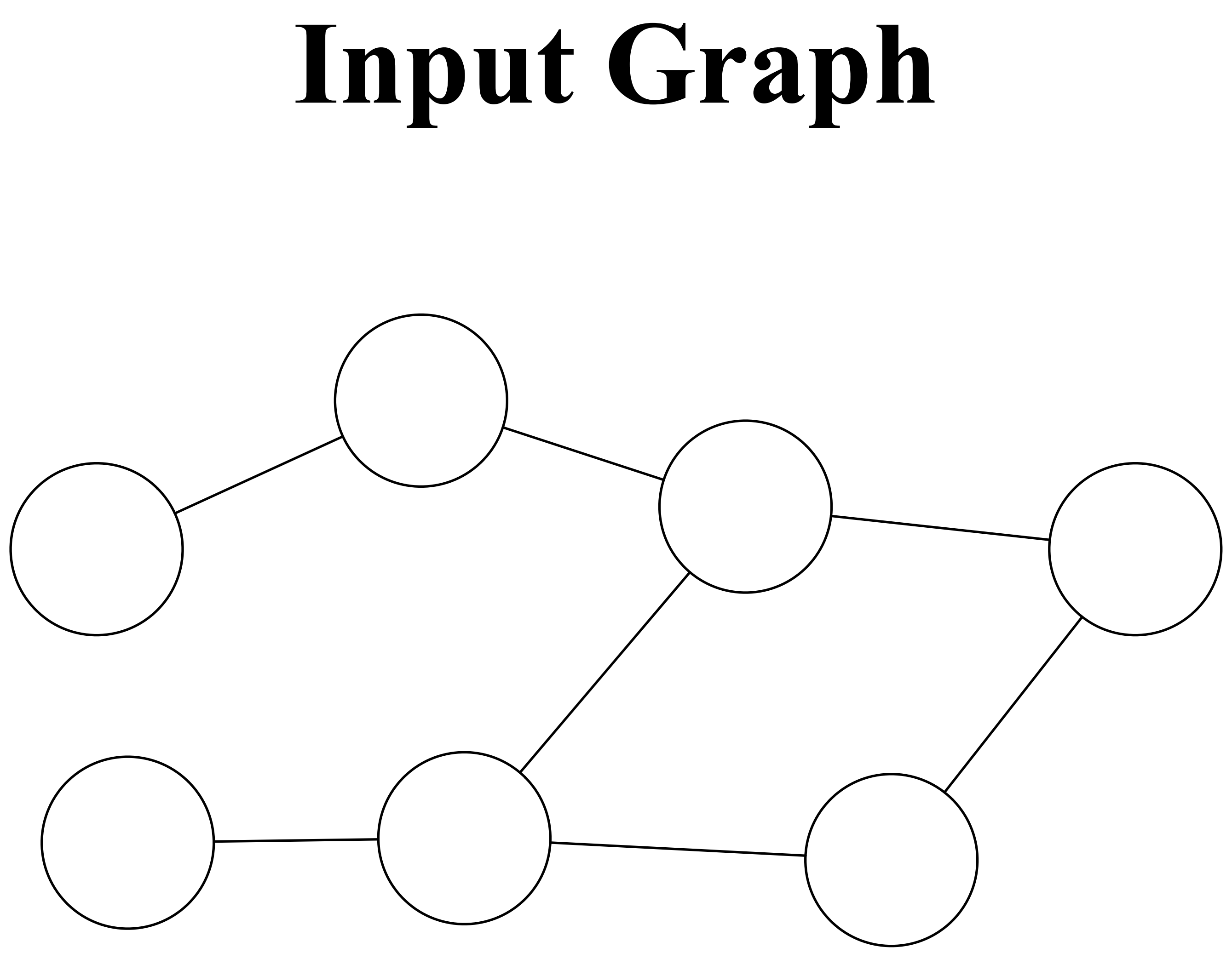


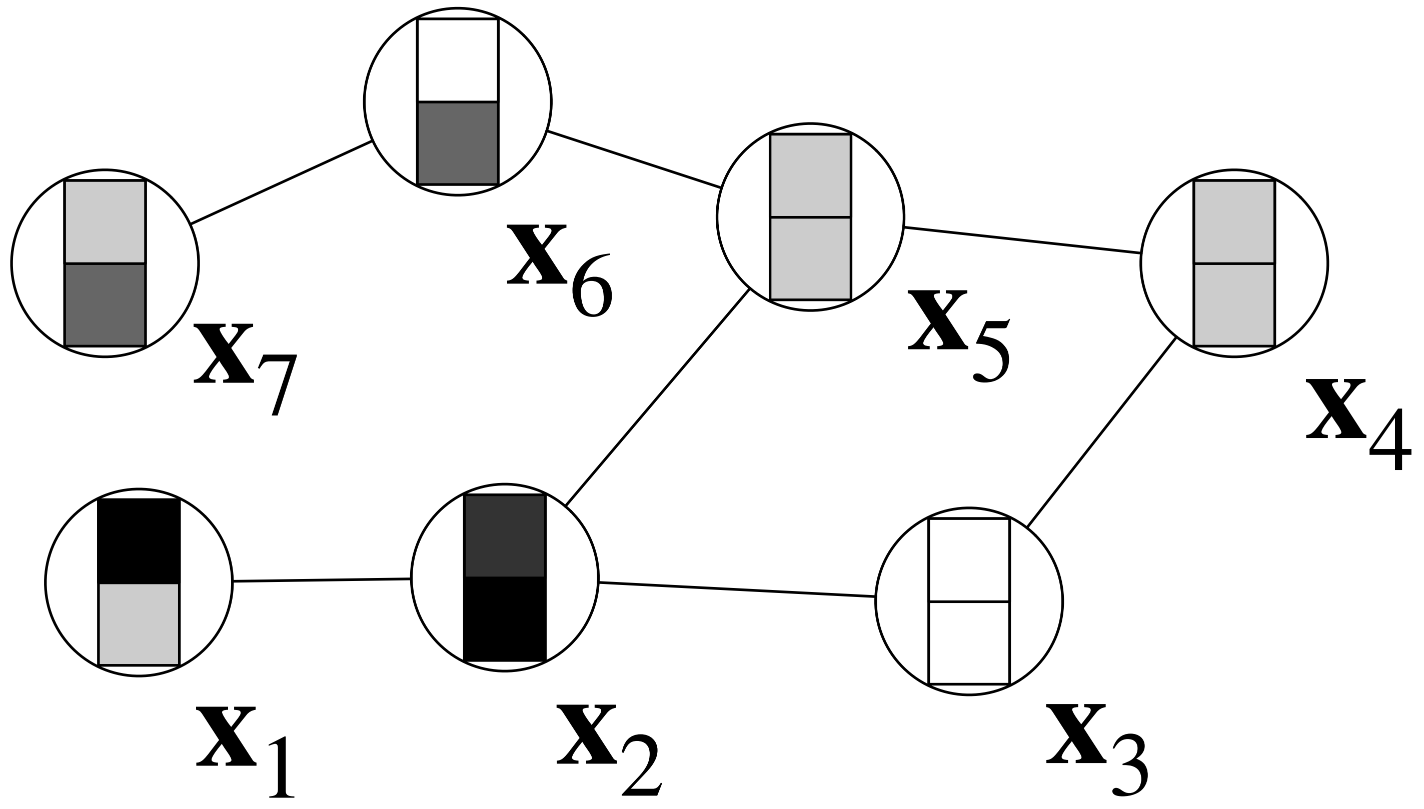
Image source: https://tkipf.github.io/graph-convolutional-networks/images/karate.png

Node Classification



Node Classification

Input Graph

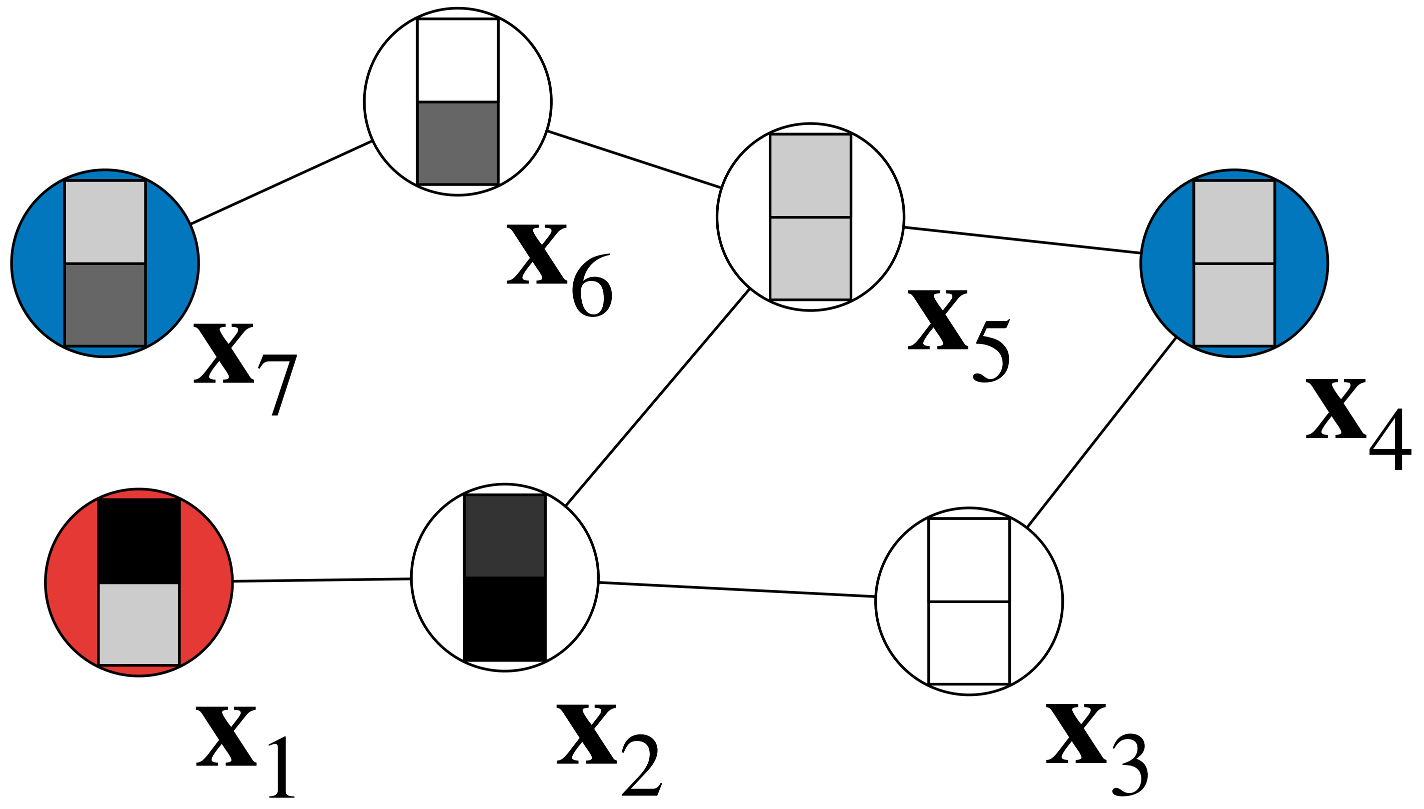


Features: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$

Class +1: Class -1: Feature Vector: Feature Value: -1 0 +1

Node Classification

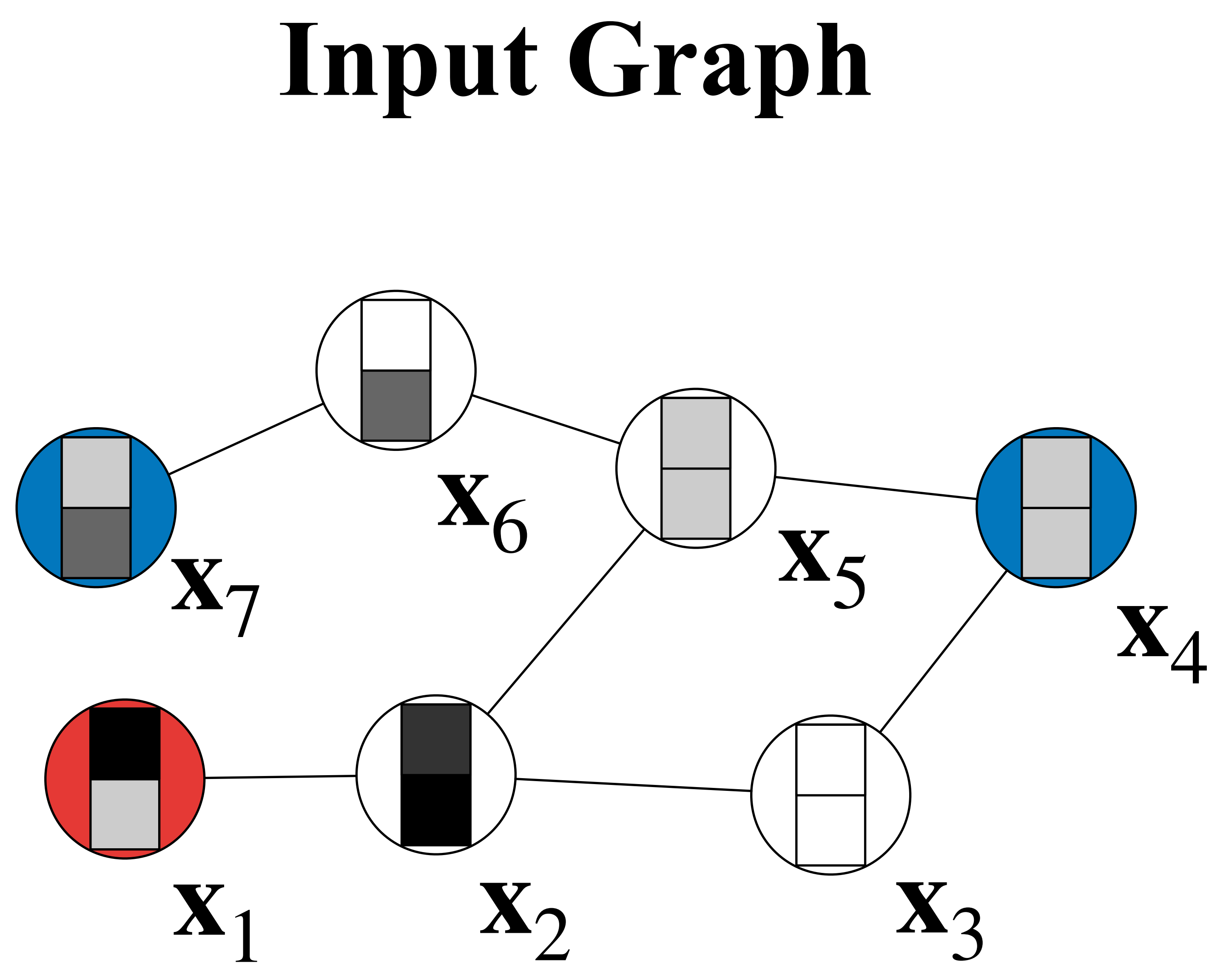
Input Graph



Features: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$

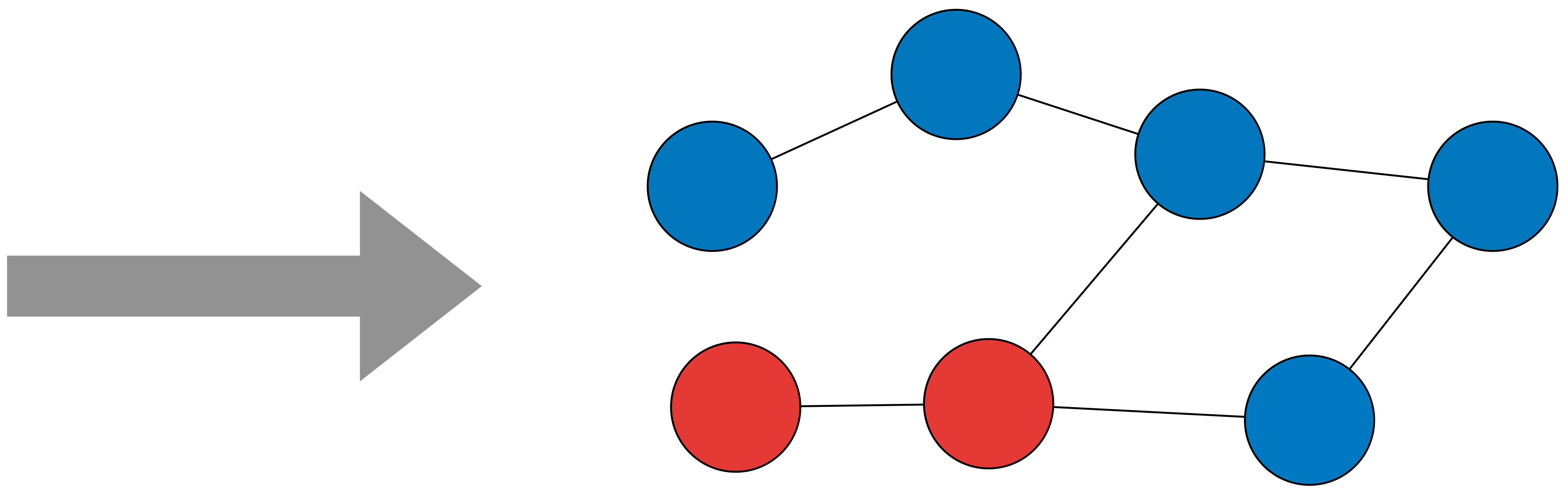
(Partially labelled)

Node Classification



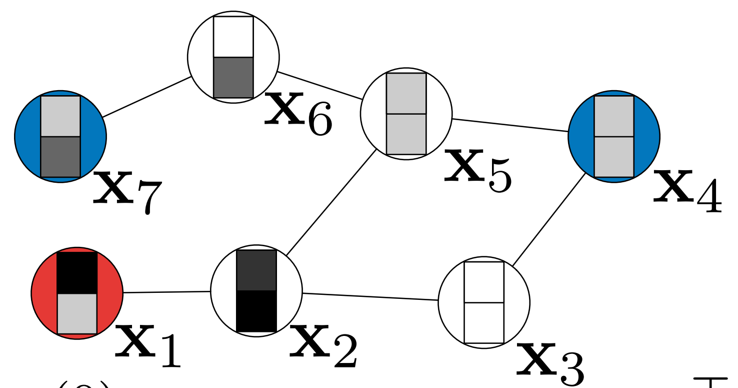
Features: $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$

(Partially labelled)



Graph Convolutional Network (GCN)

Input Graph

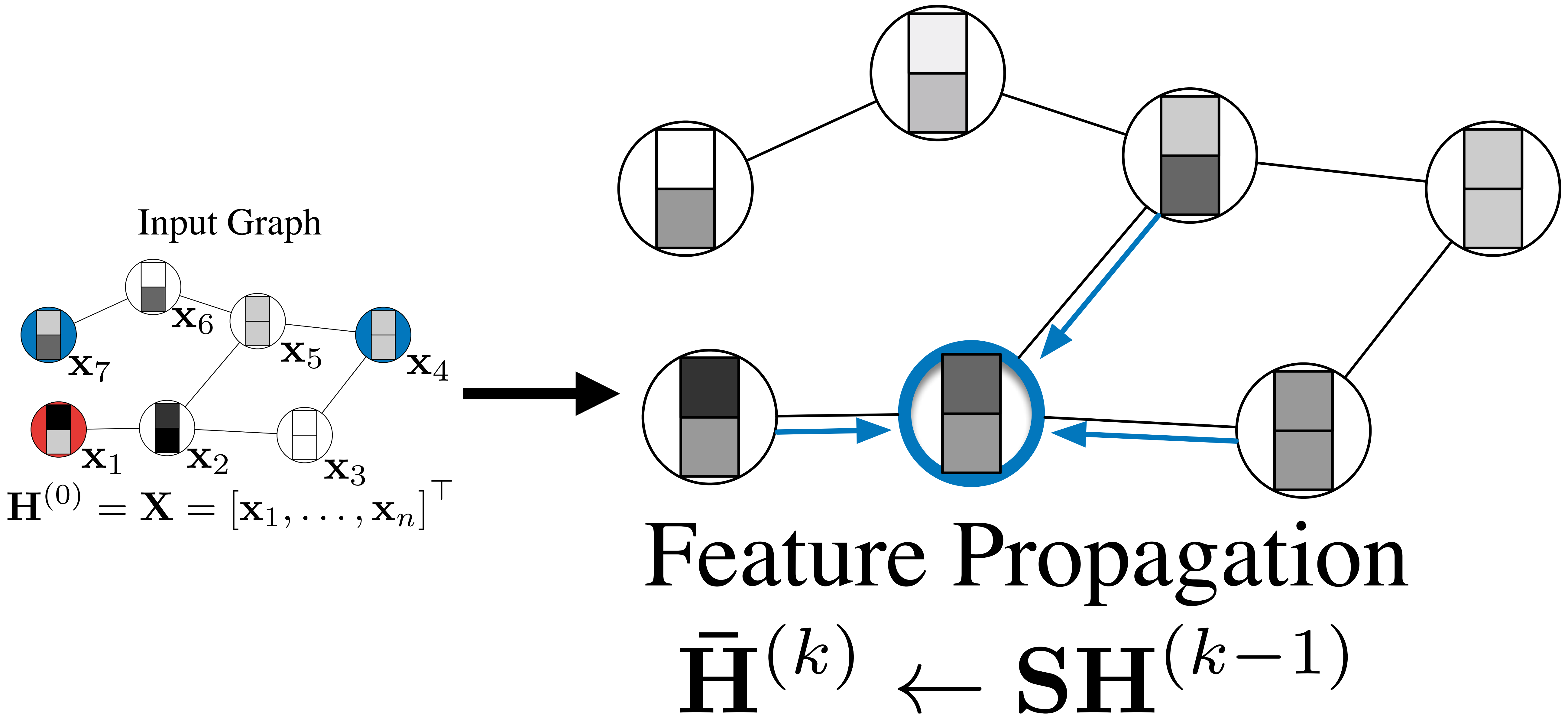


$$\mathbf{H}^{(0)} = \mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$$

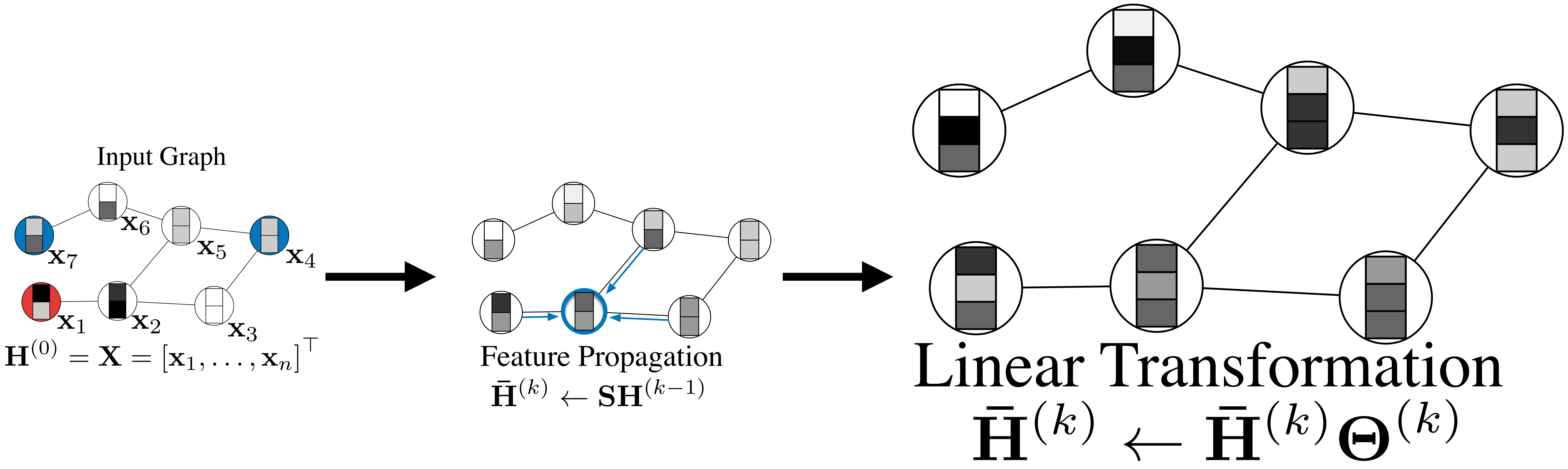
Class +1: Class -1: Feature Vector:

Feature Value: -1 0 +1

Graph Convolutional Network (GCN)



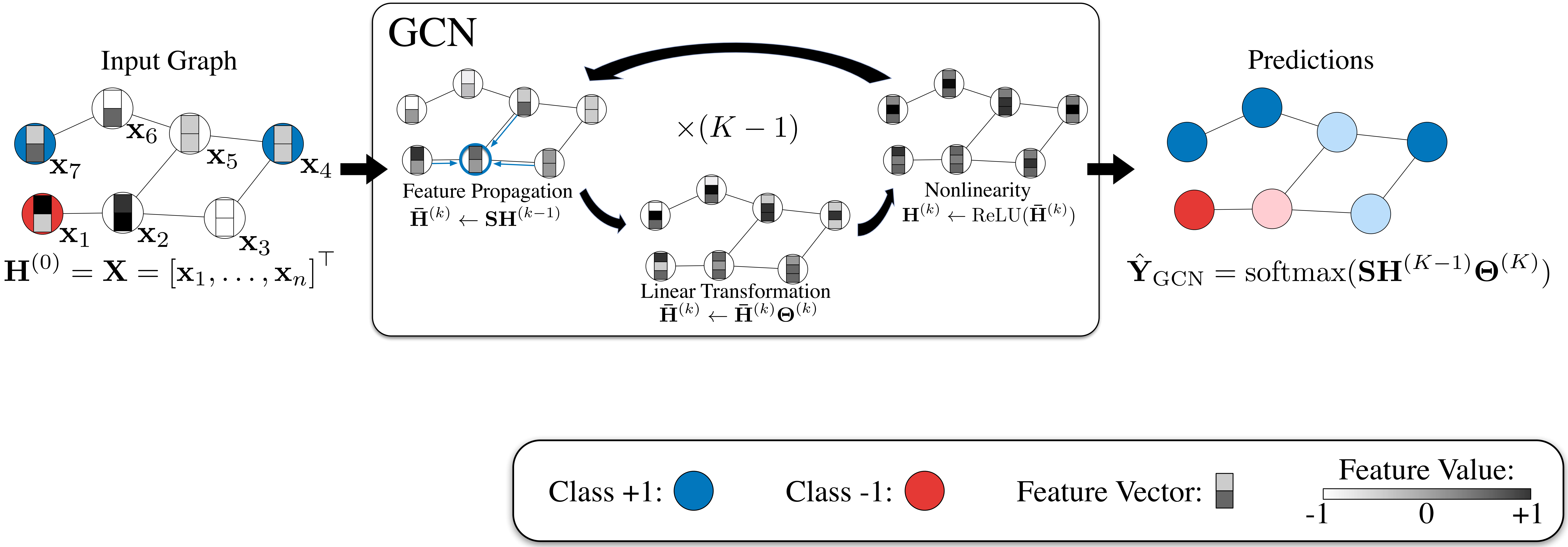
Graph Convolutional Network (GCN)



Graph Convolutional Network (GCN)

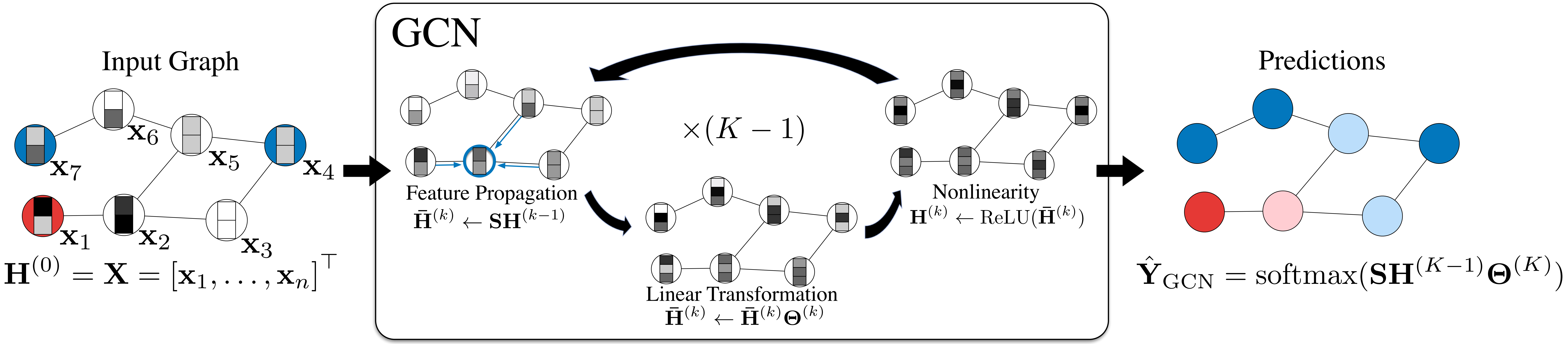
Graph Convolutional Network (GCN)

Graph Convolutional Network (GCN)



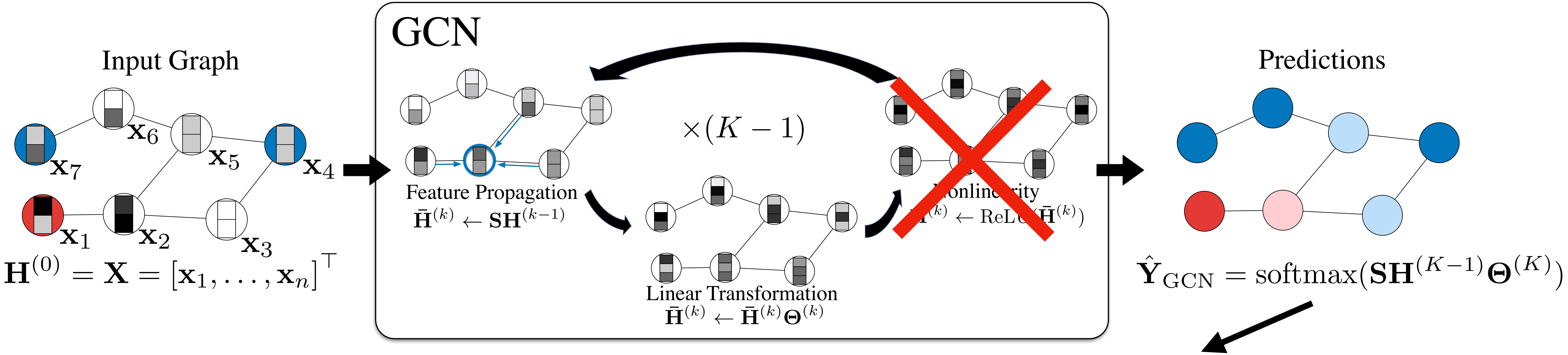
Graph Convolutional Network (GCN)

Is the nonlinearity necessary?



Graph Convolutional Network (GCN)

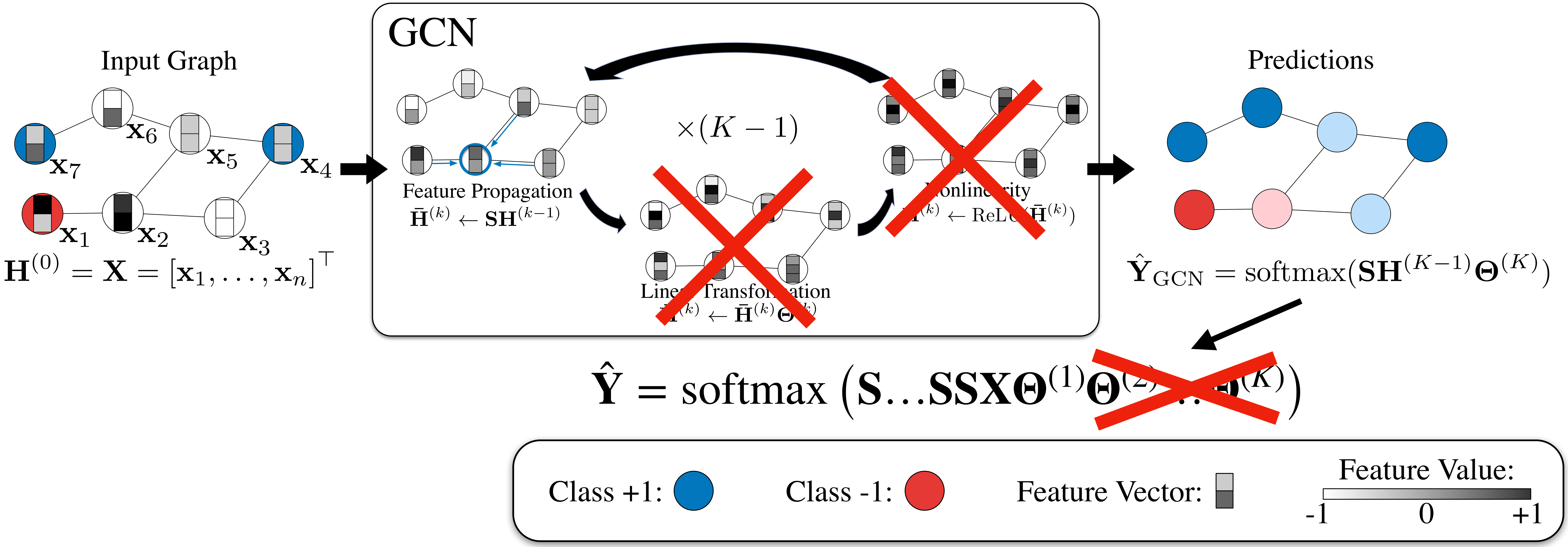
Is the nonlinearity necessary?



$$\hat{\mathbf{Y}} = \mathrm{softmax}\left(\mathbf{S}\ldots\mathbf{S}\mathbf{S}\mathbf{X}\mathbf{\Theta}^{(1)}\mathbf{\Theta}^{(2)}\ldots\mathbf{\Theta}^{(K)}\right)$$

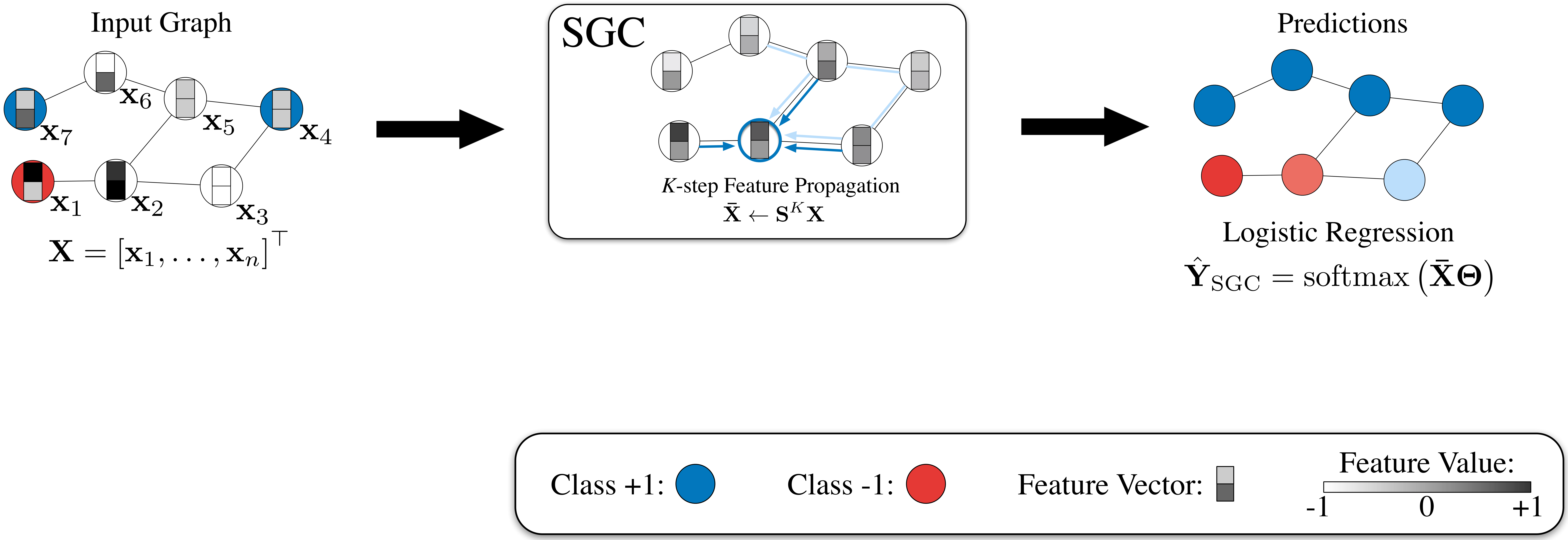
Graph Convolutional Network (GCN)

Is the nonlinearity necessary?



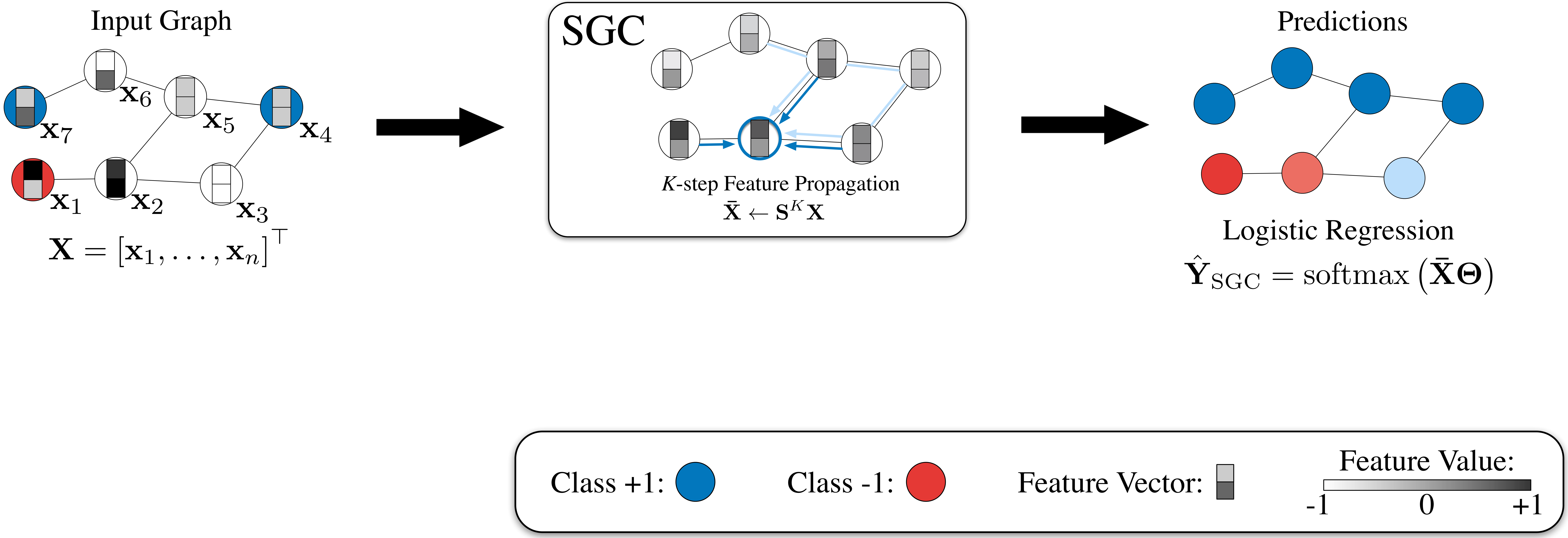
$$\hat{\mathbf{Y}} = \mathrm{softmax}\left(\mathbf{S}\ldots\mathbf{S}\mathbf{S}\mathbf{X}\boldsymbol{\Theta}^{(1)}\cancel{\boldsymbol{\Theta}^{(2)}\ldots\boldsymbol{\Theta}^{(K)}}\right)$$

Simple Graph Convolution (SGC)



Simple Graph Convolution (SGC)

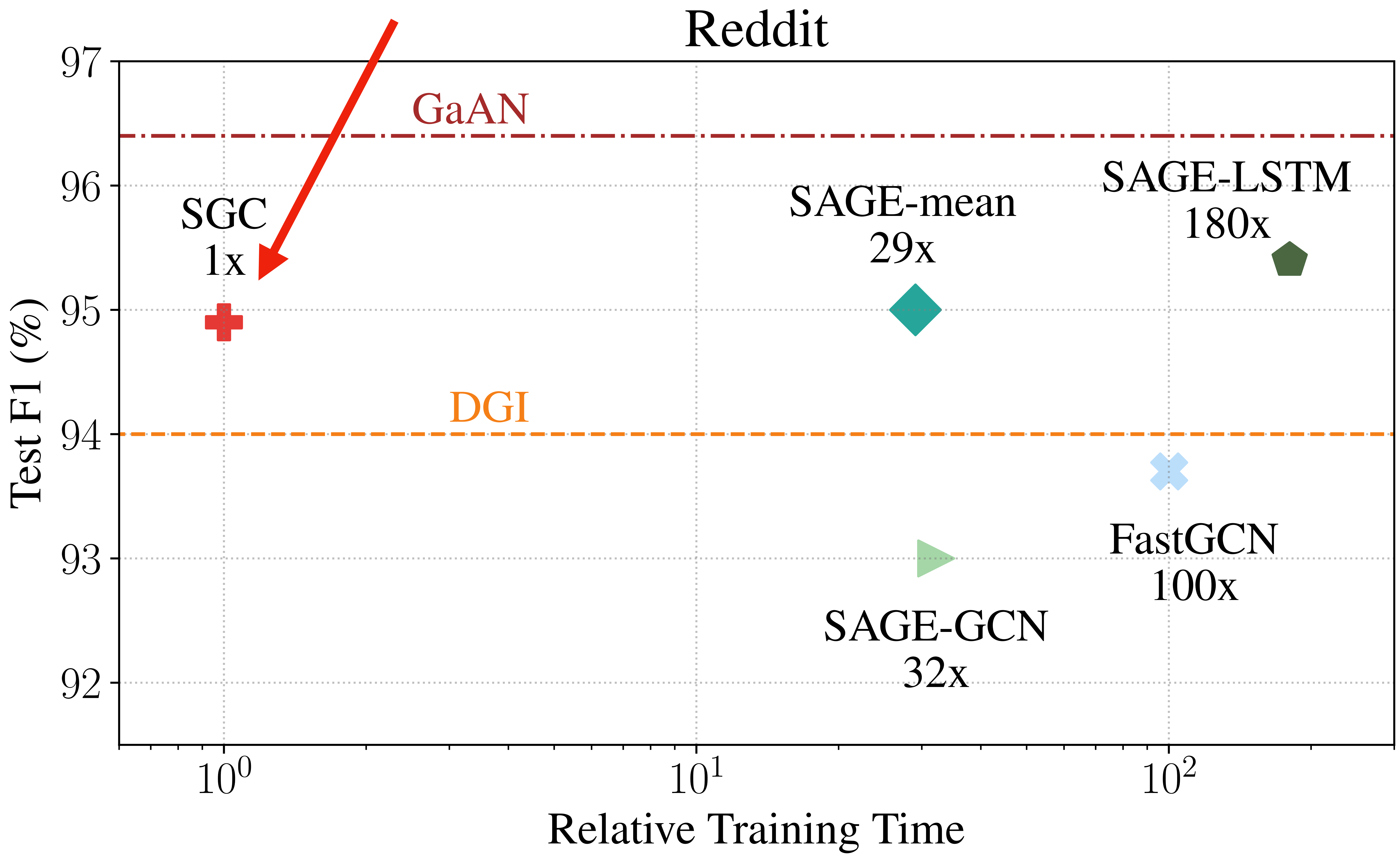
1. Interpretability
2. Mini-batch training on large graphs
3. Second-order opimization methods



Performance vs. Training Time on Reddit Dataset

(Hamilton *et al.*, NeurIPS 2017)

Within 2.7 seconds.



Reddit dataset contains 233K nodes and 11.6M edges

Comparable or Better Results

- SGC performs on par with or better than GCN across **5 tasks** including **14 datasets.**
  - Graph classification on citation and social networks: Cora, Citeseer, Pubmed, and Reddit datasets
  - Text classification: 20NG, R8, R52, Ohsumed, MR
  - Semi-supervised user geolocation: GEOTEXT, TWITTER-US, TWITTER-WORLD
  - Relation extraction: TACRED
  - Zero-shot image classification: ImageNet

Comparable or Better Results

- SGC performs on par with or better than GCN across **5 tasks** including **14 datasets.**
  - Graph classification on citation and social networks: Cora, Citeseer, Pubmed, and Reddit datasets
  - Text classification: 20NG, R8, R52, Ohsumed, MR
  - Semi-supervised user geolocation: GEOTEXT, TWITTER-US, TWITTER-WORLD
  - Relation extraction: TACRED
  - Zero-shot image classification: ImageNet

But not on **graph classification** task! (Nonlinearity is still beneficial on this task)

Thank You!

- A linear model (SGC) is sufficient on many graph tasks.
- Our official code is available at: https://github.com/Tiiiger/SGC
- We thank *Deep Graph Library*, *PyTorch Geometric*, *Spektral,* and *StellarGraph* for including SGC in their library.
- Please feel free to come by our poster section if you have questions:

Tonight 06:30~09:00 PM @ **Pacific Ballroom #267**