Network Representation Learning Share

范敏

2018-10-24

京东

Contents

京东

Network Representation Learning

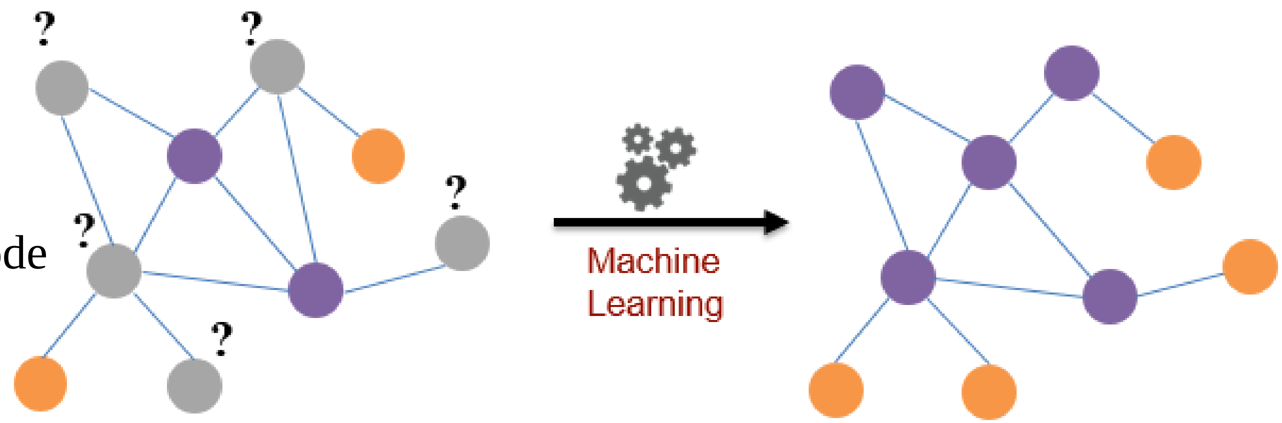
金融场景的反欺诈应用

Network Representation Learning

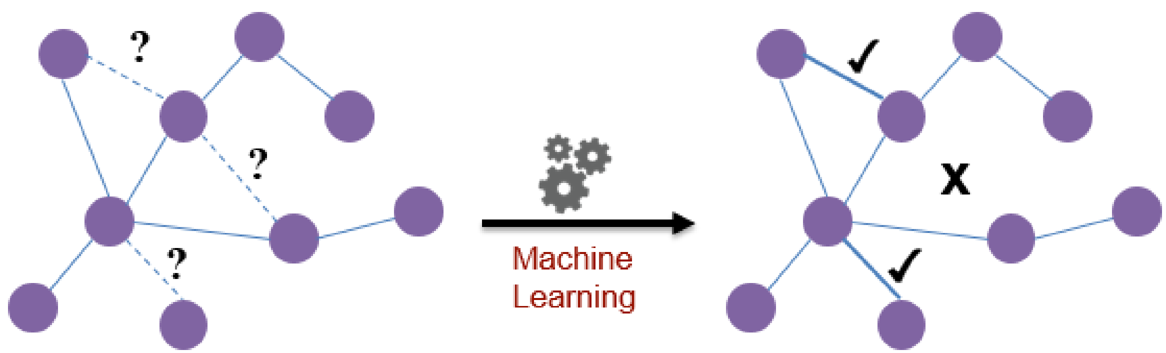
Machine Learning with Network

- Node classification

Predict a type of a given node



- Link prediction

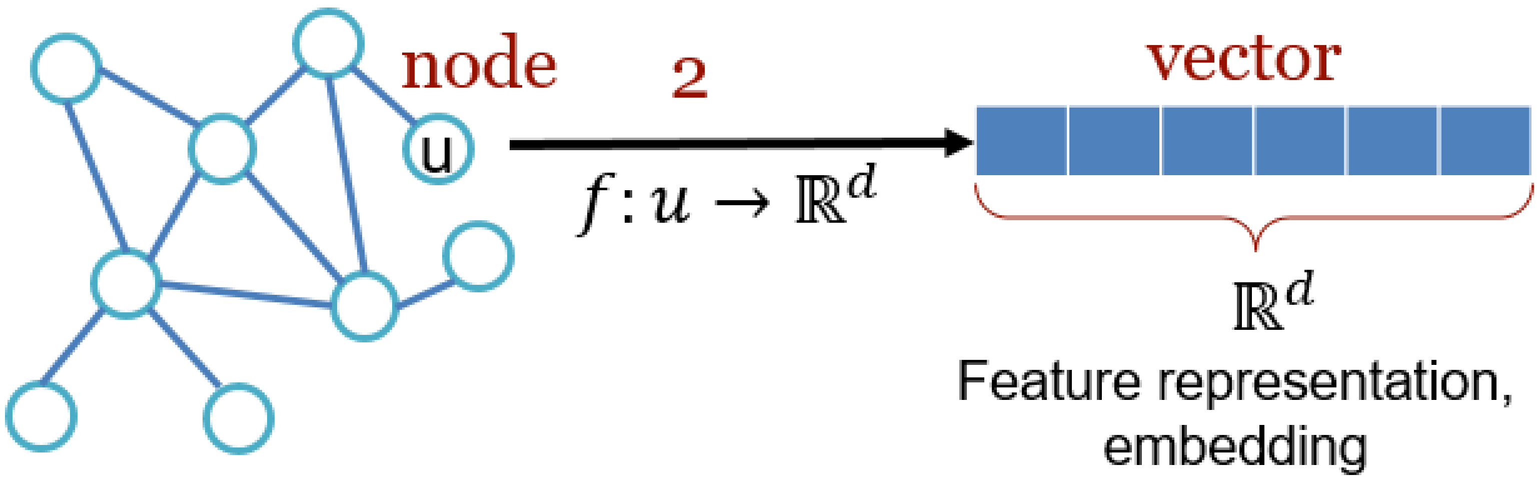


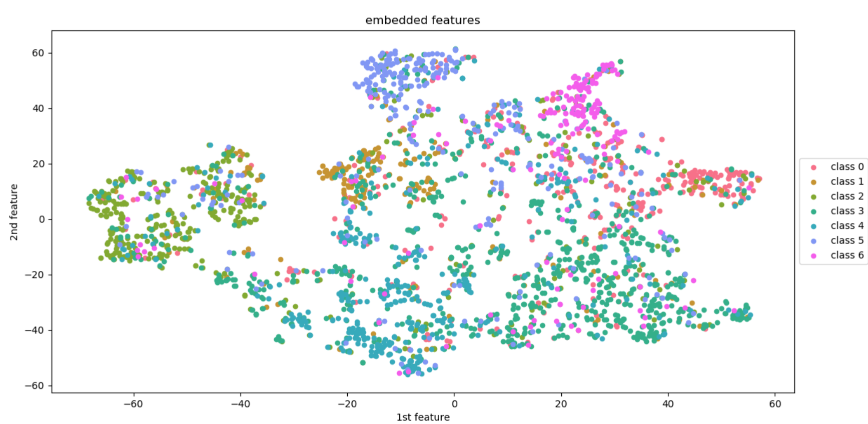
Predict whether two nodes are linked

Network Representation Learning

Machine Learning with Network

Key Goal: Map nodes to low-dimensional embedding



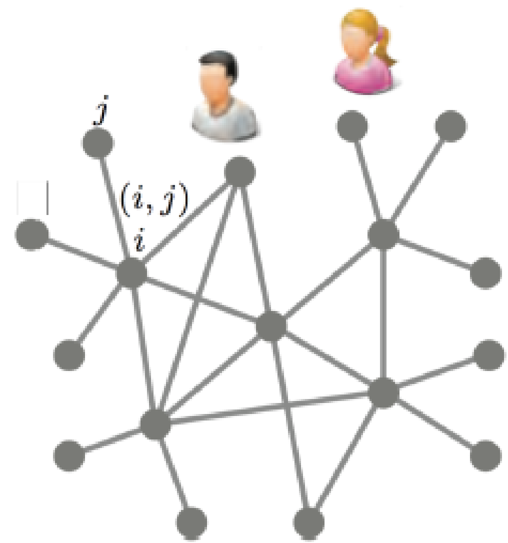


Network Representation Learning

Graphs

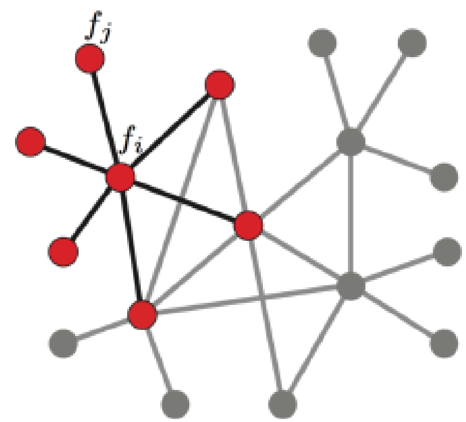
- Graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$
- Vertices $\mathcal{V} = \{1, \ldots, n\}$
- Edges $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$
- Vertex weights $b_i > 0$ for $i \in \mathcal{V}$
- Edge weights $a_{ij} \geq 0$ for $(i, j) \in \mathcal{E}$
- Vertex fields $L^2(\mathcal{V}) = \{f : \mathcal{V} \rightarrow \mathbb{R}^h\}$
  Represented as $\mathrm{f} = (f_1, \ldots, f_n)$
- Hilbert space with inner product

$$\langle f, g \rangle_{L^2(\mathcal{V})} = \sum_{i \in \mathcal{V}} a_i f_i g_i$$



Network Representation Learning

Graph Laplacian



- Laplacian operator $\Delta : L^2(\mathcal{V}) \to L^2(\mathcal{V})$

  $$(\Delta f)_i = \frac{1}{b_i} \sum_{j:(i,j)\in\mathcal{E}} a_{ij}(f_i - f_j)$$

  difference between f and its local average (2nd derivative on graphs)

- Core operator in spectral graph theory.

- Represented as a positive semi-definite $n \times n$ matrix

  - Unnormalized Laplacian $\quad \boldsymbol{\Delta} = \mathbf{D} - \mathbf{A}$
  - Normalized Laplacian $\quad \boldsymbol{\Delta} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$
  - Random walk Laplacian $\quad \boldsymbol{\Delta} = \mathbf{I} - \mathbf{D}^{-1}\mathbf{A}$

  where $\mathbf{A} = (a_{ij})$ and $\mathbf{D} = \mathrm{diag}(\sum_{j\neq i} a_{ij})$

京东

Network Representation Learning

Intuition:

一张graph可以利用Laplacian矩阵表达，将矩阵进行分解，求得特征值和特征向量，目前的最优解是GraphWave求得特征值后，利用heat kernel进行采样。

但是，Laplacian的约束条件太苛刻，导致矩阵分解在实际中无法计算以及 node feature learning, inductive learning问题

因此需要对约束条件进行放松，比如利用空间和谱阈的采样来近似原始Laplacian矩阵

$$L = D - A = U\Lambda U^{T} \qquad g_\theta \star x = U g_\theta U^\top x,$$

Classic Model

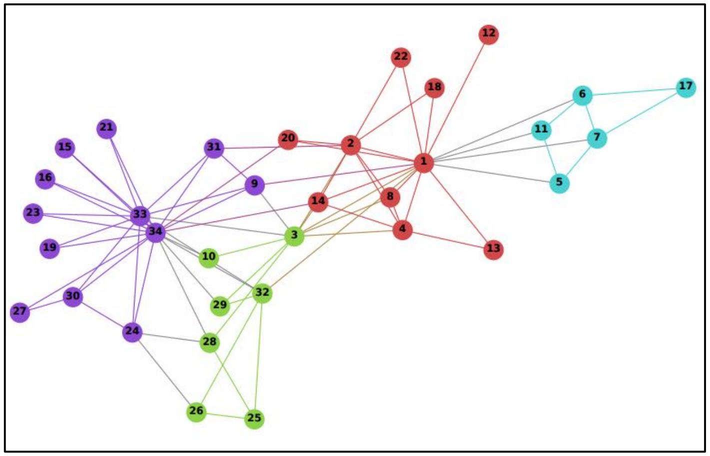
- Spatial Sampling
  - Deepwalk (KDD2014)
  - Node2vec (KDD2016)
  - Struct2vec (KDD2017)
- Spectral Sampling
  - Graphwave (KDD2018)
  - GCN (ICLR2017)
  - GAT (ICLR2018)

京东

Classic Model

Deepwalk KDD2014

Just run fixed-length, unbiased random walks starting from each node



V1, V3, V6......
V4, V1, V33...
V1, V3, V6......
V4, V1, V33...
V1, V3, V6......
V4, V1, V33...
V1, V3, V6......
......

WORD2VEC

V1: (0.22, 0.33, 0.44.....)
V2:(0.45,-0.1, 0.88....)
.....

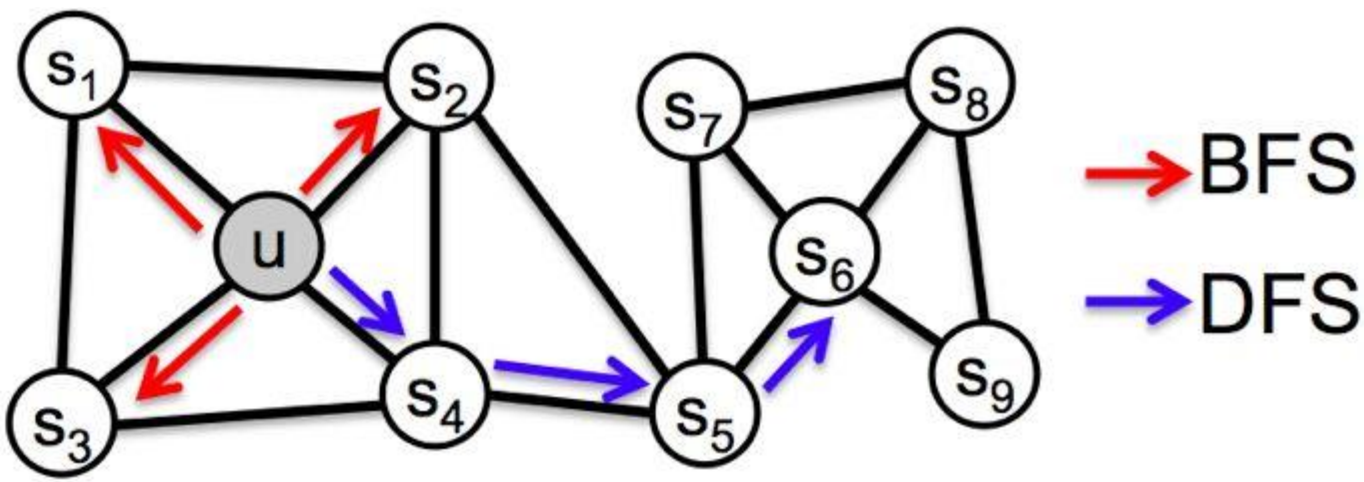
Skip-gram negative sampling

Classic Model

Node2vec KDD2016

Use flexible, biased random walks that can trade off between **local** and **global** views of the network

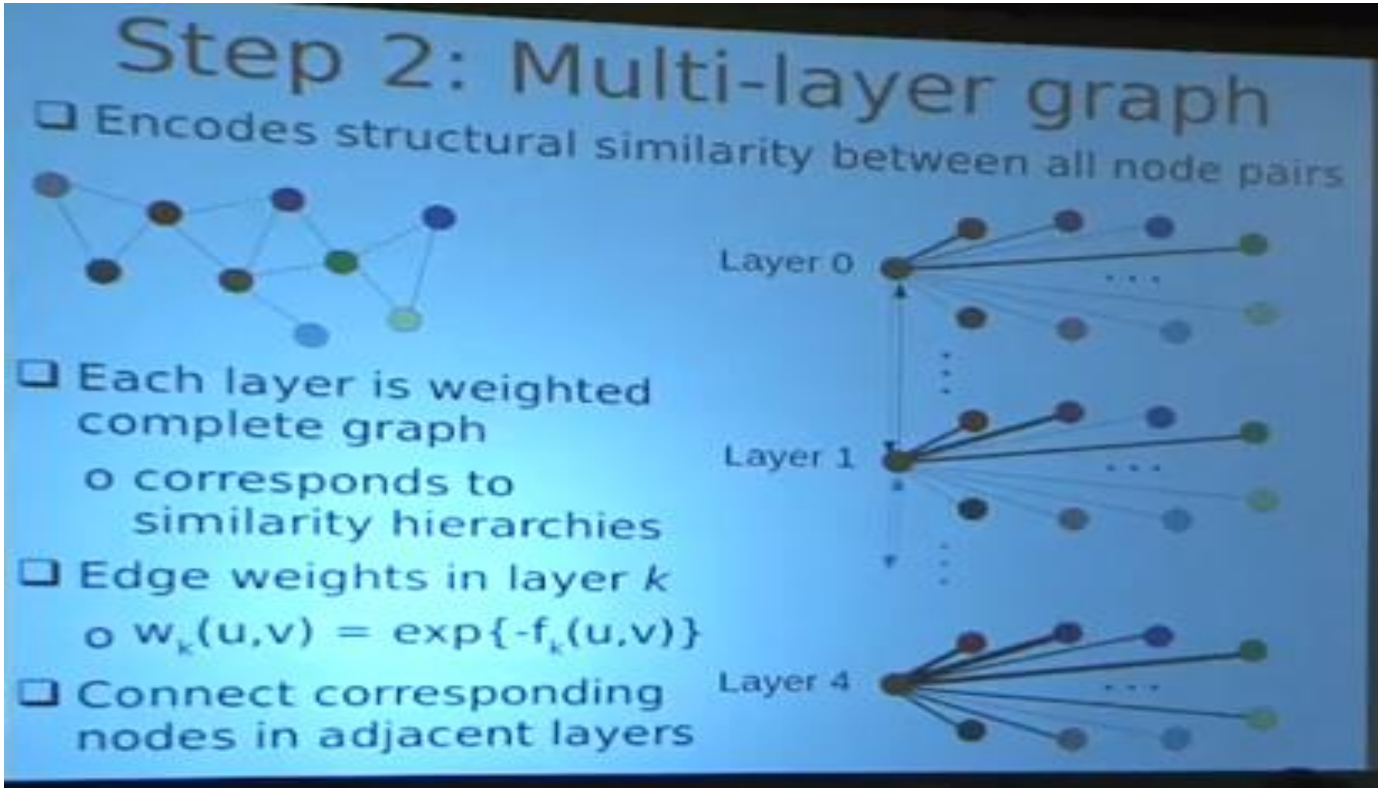
Word2vec Skip-gram negative sampling



Classic Model

Struct2vec KDD2017

Run random walks on modified versions of the original network

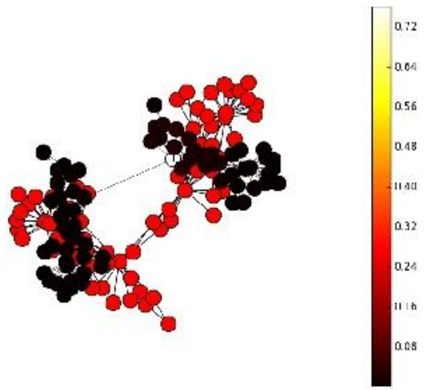


Classic Model

Graphwave KDD2018

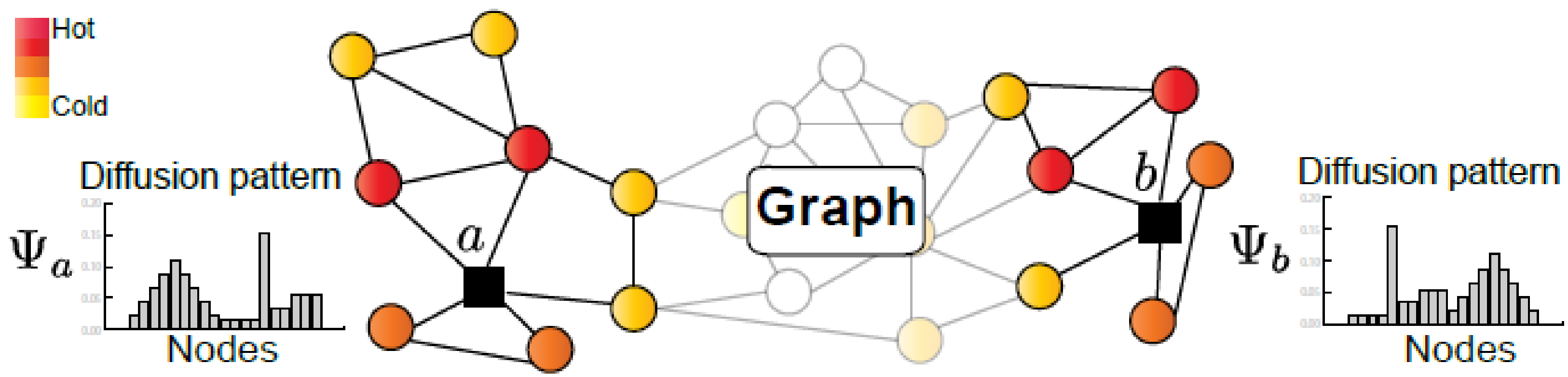
Sampling methods:

- ✓ Random Walk $L = D - A = U\Lambda U^T$
- ✓ Heat Kernel $g_s(\lambda) = e^{-\lambda s}$



谱图小波的定义：

$$\Psi_a = \underline{U \, \mathrm{Diag}(g_s(\lambda_1), \ldots, g_s(\lambda_N)) U^T \delta_a},$$

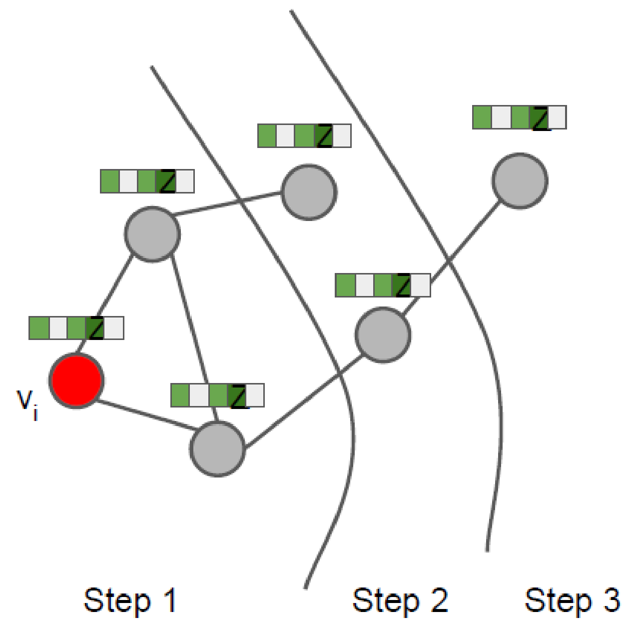


Classic Model

GCN (ICLR2017)

Graph View of GCN Model

At every iteration, the model aggregates information from one hop deeper.

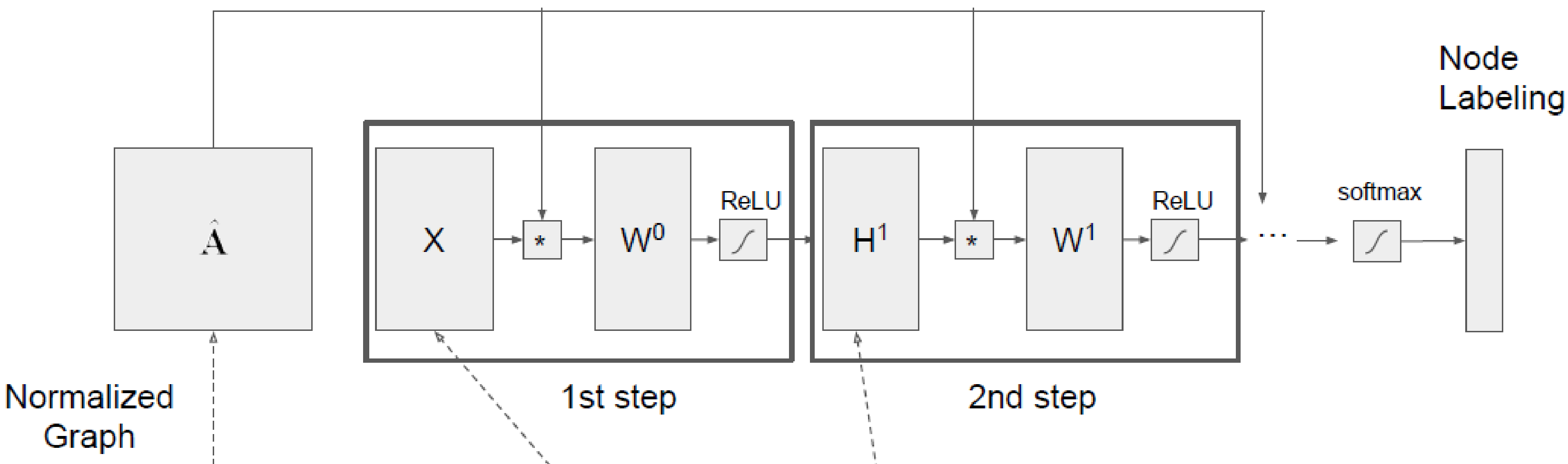


京东

Classic Model

GCN (ICLR2017)

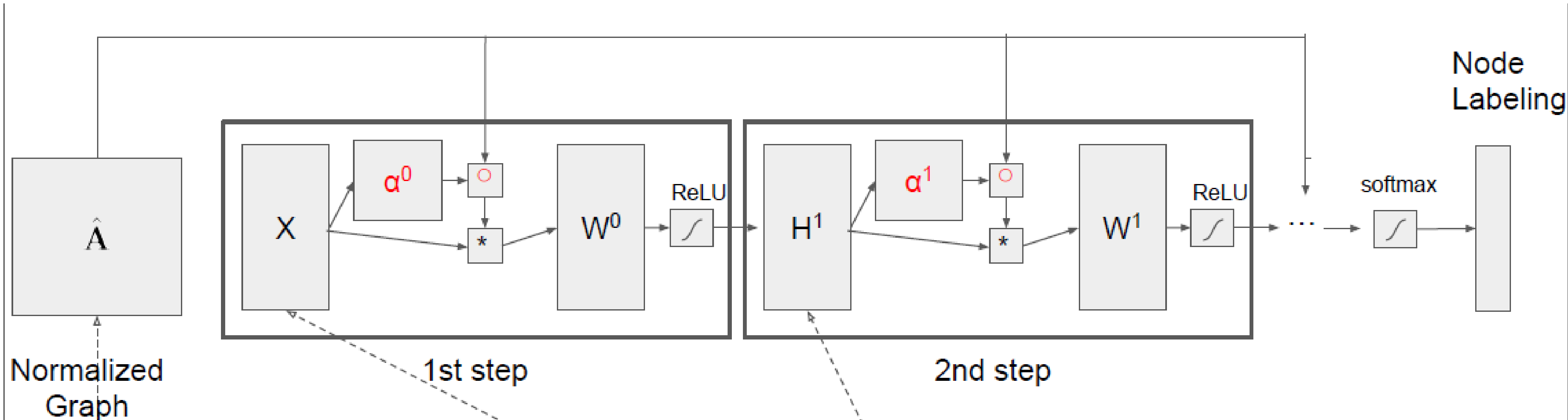
Block View of the GCN Model



$$\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{A}+\mathbf{I}_N)\tilde{\mathbf{D}}^{-\frac{1}{2}} \qquad \mathbf{X} = \mathbf{H}^{(0)} \qquad \mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

Classic Model

GAT(ICLR2018)



$$\hat{\mathbf{A}} = \tilde{\mathbf{D}}^{-\frac{1}{2}}(\mathbf{A} + \mathbf{I}_N)\tilde{\mathbf{D}}^{-\frac{1}{2}}$$

$$\mathbf{X} = \mathbf{H}^{(0)}$$

$$\mathbf{H}^{(l+1)} = \sigma\left(\hat{\mathbf{A}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right)$$

Classic Model

IDEA

Localization in space = smoothness in frequency domain
Convolution = re-weight = sampling

京东

Experiment Result

[实验报告]

Application and Challenge

Application

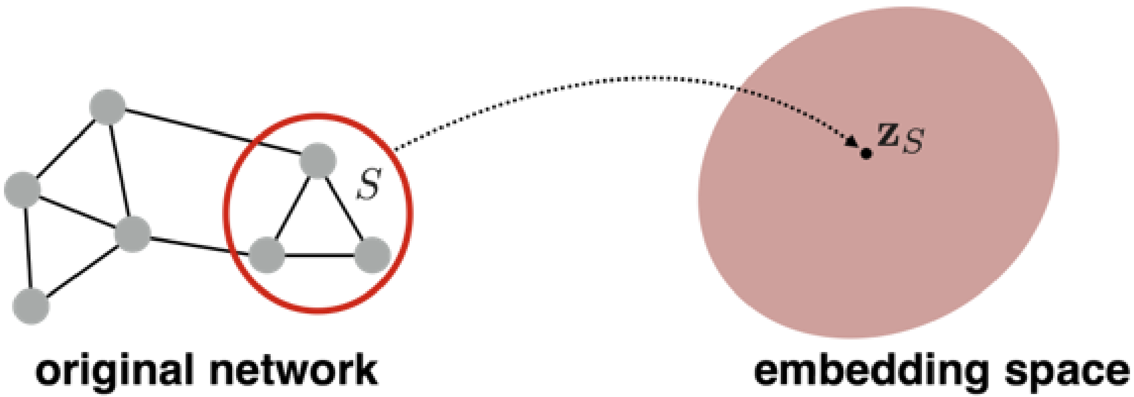
Anti-fraud

Struct2vec： Ant Financial bad/good people classification

Application and Challenge

Challenge

- How to efficiently and incrementally update the learned representations?
- How to do inductive learning?
- Subgraph Embedding?



感谢聆听

By Fan Min