Masked Graph Convolutional Network

Liang Yang[1,2] , **Fan Wu**[1,2] , **Yingkui Wang**[3] , **Junhua Gu**[1,2] and **Yuanfang Guo**[4,*]
[1]School of Artificial Intelligence, Hebei University of Technology, China
[2]Hebei Province Key Laboratory of Big Data Calculation, Hebei University of Technology, China
[3]College of Intelligence and Computing, Tianjin University, China
[4]School of Computer Science and Engineering, Beihang University, China
yangliang@nextseason.cc, wufanslient@outlook.com, ykwang@tju.edu.cn, jhgu@hebut.edu.cn, andyguo@buaa.edu.cn

项雅丽

2019.10.10

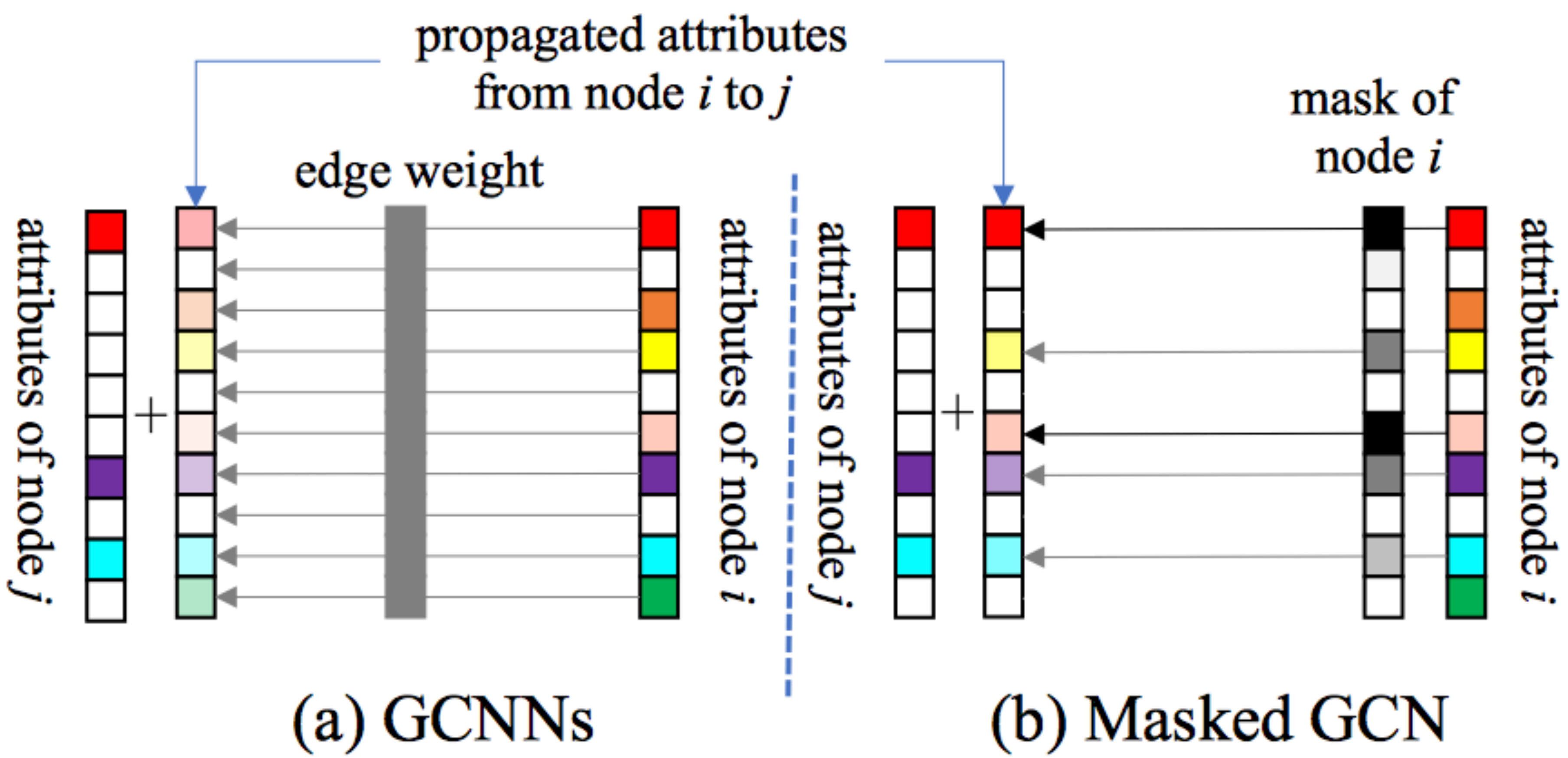
Motivation

很多方法都基于 **相邻的节点相似** 这个假设，
这个假设对节点标签成立，但是对节点特征只是**部分**成立

例如，粉丝和idol可能会有一些兴趣上的相似点，但是他们的性别可能不一样，因为在传播过程中，性别这个特征不应该传递下去。

因此，论文思想就是学习mask，筛选需要传递下去的特征，过滤不重要的特征。

Framework


propagated attributes
from node i to j
edge weight
mask of
node i
attributes of node j
attributes of node i
attributes of node j
attributes of node i
(a) GCNNs
(b) Masked GCN

Related work

Traditional Attribute-graph based Semi-supervised Classification

- Label Propagation
- Label Spreading

Graph Convolutional Neural Networks

- Graph Convolutional Network
- PageRank GCN
- Graph Attention Network

Related work

Traditional Attribute-graph based Semi-supervised Classification

$$w_{ij} = \exp\left(-\sum_d \frac{(x_{id}-x_{jd})^2}{\sigma_d^2}\right) \qquad P = \left[\begin{array}{c|c} P_{ll} & P_{lu} \\ \hline P_{ul} & P_{uu} \end{array}\right]$$

Label Propagation

$$h_i^{(k+1)} = \sum_j \frac{w_{ij}}{\sum_j w_{ij}} h_j^{(k)} = \sum_j \frac{w_{ij}}{d_i} h_j^{(k)}$$

$$E(H) = \sum\nolimits_{i,j} w_{ij} ||h_i - h_j||_2^2 = \mathrm{tr}(H^T L H)$$

analytical solution can be expressed as

$$Y_u = (I - P_{uu})^{-1} P_{ul} Y_l,$$

Label Spreading

$$h_i^{(k+1)} = (1-\alpha)\sum_j \frac{w_{ij}}{\sqrt{d_i d_j}} h_j^{(k)} + \alpha y_i$$

$$\begin{aligned} E(H) &= \sum_{i,j} w_{ij} \left\| \frac{h_i}{\sqrt{d_i}} - \frac{h_j}{\sqrt{d_j}} \right\|_2^2 + \mu \sum_i ||h_i - y_i||_2^2 \\ &= \mathrm{tr}(H^T \hat{L} H) + \mu ||H - Y||_F^2, \end{aligned}$$

the analytical solution as Eq. (7) shows.

$$H = (1-\alpha)(I - \alpha\hat{L})^{-1} Y \tag{7}$$

Related work-Graph Convolutional Neural Networks

- **GAT**

$$H_{GAT}^{(k+1)} = OH_{GAT}^{(k)}$$

$$o_{nk} = \frac{\exp\left(c(x_n^T W, x_k^T W)\right)}{\sum_{k \in N(n)} \exp\left(c(x_n^T W, x_k^T W)\right)}$$

$$Q = HW$$

$$\mathcal{L} = -\sum_{n \in V_l} \sum_{f=1}^{F} Y_{nf} \log(Q_{nf})$$

- **GCN**

$$H_{GCN}^{(k+1)} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H_{GCN}^{(k)}$$

- **PageRank GCN**

$$H_{PR\_GCN}^{(k+1)} = (1-\alpha) \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H_{PR\_GCN}^{(k)} + \alpha X$$

Related work-Comparisons

Symmetric Propagation

- **Label Spreading**

$$h_i^{(k+1)} = (1-\alpha)\sum_j \frac{w_{ij}}{\sqrt{d_i d_j}} h_j^{(k)} + \alpha y_i$$

- **GCN**

$$H_{GCN}^{(k+1)} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} H_{GCN}^{(k)}$$

- **PageRank GCN**

$$H_{PR\_GCN}^{(k+1)} = (1-\alpha)\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} H_{PR\_GCN}^{(k)} + \alpha X$$

Asymmetric Propagation

- **Label Propagation**

$$h_i^{(k+1)} = \sum_j \frac{w_{ij}}{\sum_j w_{ij}} h_j^{(k)} = \sum_j \frac{w_{ij}}{d_i} h_j^{(k)}$$

- **GAT**

$$H_{GAT}^{(k+1)} = O H_{GAT}^{(k)}$$

Related work-Comparisons

SP loss

$$E(H) = \sum_{i,j} w_{ij} \left\| \frac{h_i}{\sqrt{d_i}} - \frac{h_j}{\sqrt{d_j}} \right\|_2^2 + \mu \sum_i ||h_i - y_i||_2^2$$
$$= \mathrm{tr}(H^T \hat{L} H) + \mu ||H - Y||_F^2,$$

Symmetric Propagation

- Label Spreading

$$h_i^{(k+1)} = (1-\alpha) \sum_j \frac{w_{ij}}{\sqrt{d_i d_j}} h_j^{(k)} + \alpha y_i$$

$$\frac{1}{\sqrt{d_i d_j}}$$

传递标签

- PageRank GCN

$$H_{PR\_GCN}^{(k+1)} = (1-\alpha)\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H_{PR\_GCN}^{(k)} + \alpha X$$

$$\frac{1}{\sqrt{(d_i+1)(d_j+1)}}$$

传递特征

- GCN

$$H_{GCN}^{(k+1)} = \tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H_{GCN}^{(k)}$$

$$\mathcal{L} = \sum_{i,j} w_{ij} || \frac{h_i}{\sqrt{d_i}} - \frac{h_j}{\sqrt{d_j}} ||_2^2$$

Related work-Comparisons

Symmetric Propagation

- **Label Spreading**

$$h_i^{(k+1)} = (1-\alpha)\sum_j \frac{w_{ij}}{\sqrt{d_i d_j}} h_j^{(k)} + \alpha y_i$$

- **GCN**

$$H_{GCN}^{(k+1)} = \tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} H_{GCN}^{(k)}$$

- **PageRank GCN**

$$H_{PR\_GCN}^{(k+1)} = (1-\alpha)\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}} H_{PR\_GCN}^{(k)} + \alpha X$$

Asymmetric Propagation

- **Label Propagation**

$$h_i^{(k+1)} = \sum_j \frac{w_{ij}}{\sum_j w_{ij}} h_j^{(k)} = \sum_j \frac{w_{ij}}{d_i} h_j^{(k)}$$

- **GAT**

$$H_{GAT}^{(k+1)} = O H_{GAT}^{(k)}$$

Related work-Comparisons

Asymmetric Propagation

- **Label Propagation**

$$h_i^{(k+1)} = \sum_j \frac{w_{ij}}{\sum_j w_{ij}} h_j^{(k)} = \sum_j \frac{w_{ij}}{d_i} h_j^{(k)}$$

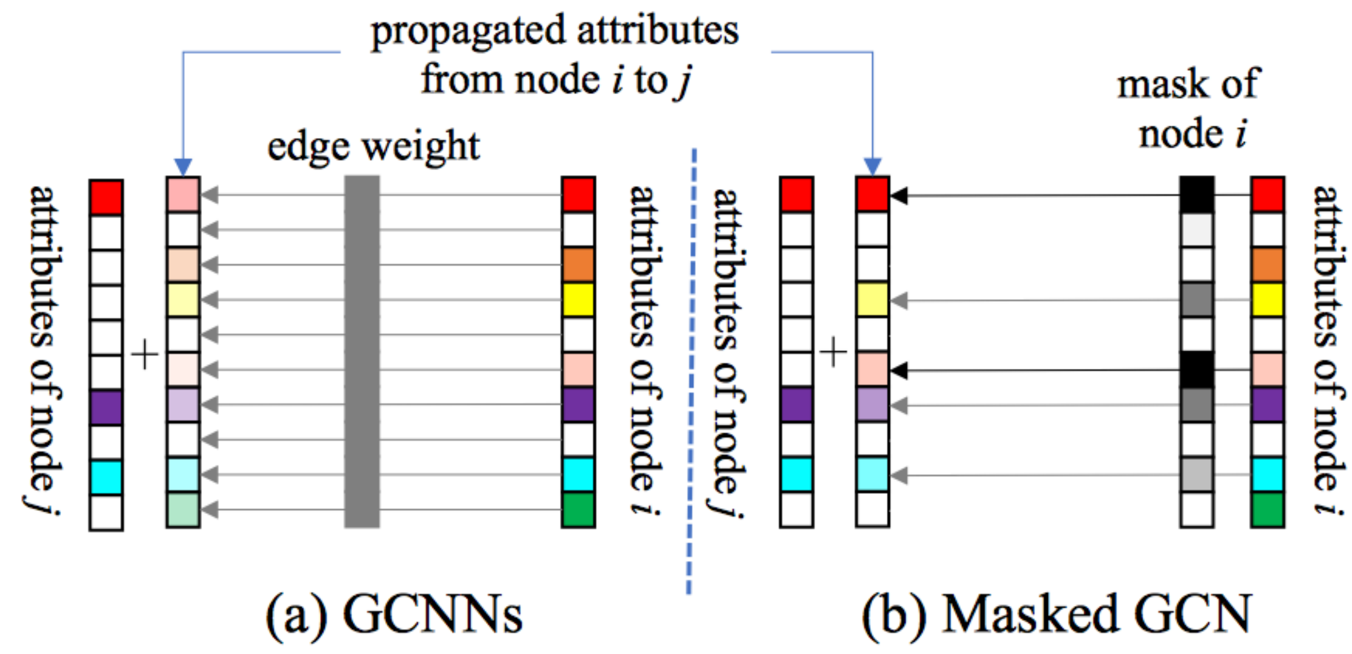
$$w_{ij} = \exp\left(-\sum_d \frac{(x_{id} - x_{jd})^2}{\sigma_d^2}\right)$$

- **GAT**

$$H_{GAT}^{(k+1)} = O H_{GAT}^{(k)}$$

$$w_{ij} = \exp\left(b^T [h_i || h_j]\right) = \exp\left(b_1^T h_i + b_2^T h_j\right)$$

Methods



$$\boxed{\begin{aligned} w_{ij} &= \exp\left(b^T [h_i || h_j]\right) \\ &= \exp\left(b_1^T h_i + b_2^T h_j\right) \end{aligned}}$$

$$\mathcal{L}_{Masked\_GCN\_Asym} = \sum_{i,j} w_{ij} || h_i - M^{(j)} h_j ||_2^2$$

$$h_i^{(k+1)} = \sum_j \frac{\boxed{w_{ij}}}{\sum_j w_{ij}} M^{(j)} h_j^{(k)}$$

$$M^{(j)} = \mathrm{diag}(m_1^{(j)}, m_2^{(j)}, ..., m_T^{(j)})$$

$$m_t^{(j)} = \exp\left(-\frac{1}{d_j} \sum_{p \in N(j)} \frac{\boxed{w_{jp}(h_{pt} - h_{jt})^2}}{\boxed{\sigma_t^2}}\right)$$

local

global

$$T_{Masked\_GCN} = \mathrm{softmax}(H^{(2)}_{Masked\_GCN} W)$$

Parameters to learn

$W, \quad \sigma_1 \ldots \sigma_T, \quad b$ **in** w_{ij}

Methods

- only propagates part of the attributes
- transductive & inductive semi-supervised node classification
- can be expressed to symmetric methods

$$\mathcal{L}_{Masked\_GCN\_Sym} = \sum_{i,j} w_{ij} \left\| \frac{h_i}{\sqrt{d_i}} - M^{(j)} \frac{h_j}{\sqrt{d_j}} \right\|_2^2$$

Experiments

Methods	Cora	Citeseer	Pubmed	NELL
MLP	55.1%	46.5%	71.4%	22.9%
ManiReg [Belkin *et al.*, 2006]	59.5%	60.1%	70.7%	21.8%
SemiEmb [Weston *et al.*, 2012]	59.0%	59.6%	71.7%	26.7%
LP [Zhu *et al.*, 2003]	68.0%	45.3%	63.0%	26.5%
DeepWalk [Perozzi *et al.*, 2014]	67.2%	43.2%	65.3%	58.1%
ICA [Lu and Getoor, 2003]	75.1%	69.1%	73.9%	23.2%
Planetoid [Yang *et al.*, 2016]	75.7%	64.7%	77.2%	61.9%
Chebyshev [Defferrard *et al.*, 2016]	81.2%	69.8%	74.4%	-
MoNet [Monti *et al.*, 2017]	81.7%	69.9%	78.8%	64.2%
Random-scheme Mask GCN	14.8%	17.2%	35.1%	4.8%
GCN [Kipf and Welling, 2017]	81.5%	70.3%	79.0%	66.0%
Masked GCN (Sym)	**82.7%**	**72.0%**	**79.3%**	**68.2%**
GAT [Veličković *et al.*, 2018]	83.0%	72.5%	79.0%	-
Masked GCN (Asym)	**84.4%**	**73.8%**	**80.2%**	**68.9%**

Table 3: Transductive learning results.

Methods	PPI
Random	0.396
Logistic Regression	0.422
GraphSAGE-mean	0.598
GraphSAGE-LSTM	0.612
GraphSAGE-pool	0.600
Inductive GCN	0.500
Masked GCN (Sym)	**0.892**
GAT	0.934
Masked GCN (Asym)	**0.952**

Table 4: Inductive learning results.

谢谢 :)