Beyond Low-frequency Information in Graph Convolutional Networks

Deyu Bo\textsuperscript{1}, Xiao Wang\textsuperscript{1}, Chuan Shi\textsuperscript{1}, Huawei Shen\textsuperscript{2}

\textsuperscript{1}Beijing University of Posts and Telecommunications
\textsuperscript{2}CAS Key Lab of Network Data Science and Technology, Institute of Computing Technology, Chinese Academy of Sciences
Information is conveyed at different frequencies where higher frequencies are encoded with details and lower frequencies are encoded with global structures. 

Background

### Frequency in graph

- Spectral graph theory has been leveraged as a tool to define frequency spectra and expansion bases for graph Fourier transforms\(^2\)

\[
x \ast_{G_f} = \mathcal{U}((\bar{U} f) \odot (\bar{U} x)) = U_{\theta} \bar{U} x
\]

Background

Frequency in GNNs

- Most existing GNNs usually exploit **low-frequency** signals\(^3\), e.g., GCN and GAT

Is the low-frequency information all we need?

---

1 Background

An Experimental Investigation

- **Low-frequency** signals perform better on assortative graphs
- **High-frequency** signals perform better on disassortative graphs

Only use low-frequency information is not always optimal for the complex networks!
Challenges

- How to use signals of different frequencies in GNNs?
- How to make GNNs suitable for different types of networks?

Intuition of FAGCN

Existing GNNs

Frequency Adaptation Graph Convolutional Networks (FAGCN)

Accuracy

Probability of inter-connection
1 Background
2 Method
3 Experiments
4 Conclusions
Compared with previous methods

\[ \mathcal{F}_L = \varepsilon I + D^{-1/2} AD^{-1/2} = (\varepsilon + 1)I - L, \]

\[ \mathcal{F}_H = \varepsilon I - D^{-1/2} AD^{-1/2} = (\varepsilon - 1)I + L, \]

\[ \tilde{h}_i = \alpha^L_{ij} (\mathcal{F}_L \cdot \mathbf{H})_i + \alpha^H_{ij} (\mathcal{F}_H \cdot \mathbf{H})_i \]

Ratio of low-frequency information
Problems of signal combination

\[ \tilde{h}_i = \alpha_{ij}^L (F_L \cdot H)_i + \alpha_{ij}^H (F_H \cdot H)_i \]

Spatial vision of FAGCN

\[ \tilde{h}_i = \alpha_{ij}^L (F_L \cdot H)_i + \alpha_{ij}^H (F_H \cdot H)_i = \varepsilon h_i + \sum_{j \in N_i} \frac{\alpha_{ij}^L - \alpha_{ij}^H}{\sqrt{d_i d_j}} h_j, \]

Coeficients of edges

\[ \alpha_{ij}^G = \alpha_{ij}^L - \alpha_{ij}^H \quad \alpha_{ij}^G = \tanh (g^T [h_i \parallel h_j]) \]

Learning the coefficients of edges is equivalent to learn the ratio of low- and high-frequency signals
FAGCN

Expressive power of FAGCN

**Proposition 1.** Low-pass filtering makes the representations become similar, while high-pass filtering makes the representations become discriminative.

\[
\mathcal{D} = \|h_u - h_v\|_2.
\]

\[
\mathcal{D}_L = \|(\epsilon h_u + h_v) - (\epsilon h_v + h_u)\|_2 = |1 - \epsilon|\mathcal{D}.
\]

\[
\mathcal{D}_H = \|(\epsilon h_u - h_v) - (\epsilon h_v - h_u)\|_2 = |1 + \epsilon|\mathcal{D}.
\]

Connection with CNNs

In CNNs, we actually do not constrain the weight of convolutional kernel to be positive.

How about GNNs?
## Experiments

### Node classification

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Assortivity</th>
<th>Nodes</th>
<th>Edges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cora</td>
<td>0.771</td>
<td>2,708</td>
<td>5,429</td>
</tr>
<tr>
<td>Citeseer</td>
<td>0.671</td>
<td>3,327</td>
<td>4,732</td>
</tr>
<tr>
<td>Pubmed</td>
<td>0.686</td>
<td>19,717</td>
<td>44,338</td>
</tr>
<tr>
<td>Chameleon</td>
<td>0.180</td>
<td>2,277</td>
<td>36,101</td>
</tr>
<tr>
<td>Squirrel</td>
<td>0.018</td>
<td>5,201</td>
<td>217,073</td>
</tr>
<tr>
<td>Actor</td>
<td>0.003</td>
<td>7,600</td>
<td>33,544</td>
</tr>
</tbody>
</table>

### Disassortative graphs

### Assortative graphs

<table>
<thead>
<tr>
<th>Method</th>
<th>Cora</th>
<th>Citeseer</th>
<th>Pubmed</th>
</tr>
</thead>
<tbody>
<tr>
<td>SGC</td>
<td>81.0%</td>
<td>71.9%</td>
<td>78.9%</td>
</tr>
<tr>
<td>GCN</td>
<td>81.5%</td>
<td>70.3%</td>
<td>79.0%</td>
</tr>
<tr>
<td>GWNN</td>
<td>82.8%</td>
<td>71.7%</td>
<td>79.1%</td>
</tr>
<tr>
<td>ChebNet</td>
<td>81.2%</td>
<td>69.8%</td>
<td>74.4%</td>
</tr>
<tr>
<td>GraphHeat</td>
<td>83.7%</td>
<td>72.5%</td>
<td>80.5%</td>
</tr>
<tr>
<td>GIN</td>
<td>77.6%</td>
<td>66.1%</td>
<td>77.0%</td>
</tr>
<tr>
<td>GAT</td>
<td>83.0%</td>
<td>72.5%</td>
<td>79.0%</td>
</tr>
<tr>
<td>MoNet</td>
<td>81.7%</td>
<td>-</td>
<td>78.8%</td>
</tr>
<tr>
<td>APPNP</td>
<td>83.7%</td>
<td>72.1%</td>
<td>79.2%</td>
</tr>
<tr>
<td>GraphSAGE</td>
<td>82.3%</td>
<td>71.2%</td>
<td>78.5%</td>
</tr>
</tbody>
</table>

FAGCN  

| FAGCN   | 84.1±0.5% | 72.7±0.8% | 79.4±0.3% |
3 Experiments Clustering Performance

- **Alleviate over-smoothing**
  - (a) Cora
  - (b) Citeseer
  - (c) Pubmed
  - (d) Chameleon
  - (e) Squirrel
  - (f) Actor

- **Visualization of $\alpha_{ij}$**
  - (a) Cora, Citeseer and Pubmed
  - (b) Chameleon
  - (c) Squirrel
  - (d) Actor
Conclusions

- We study the roles of both low-frequency and high-frequency signals in GNNs and verify that high-frequency signals are useful for disassortative networks.

- We propose a novel graph convolutional networks FAGCN, which can adaptively change the proportion of low-frequency and high-frequency signals without knowing the types of networks.

- We theoretically prove that the expressive power of FAGCN is greater than other GNNs. Moreover, our proposed FAGCN is able to alleviate the over-smoothing problem. Extensive experiments on six real-world networks validate that FAGCN has advantages over state-of-the-arts.
Thank you!
Q&A