Node-wise Localization of Graph Neural Networks

Zemin Liu¹, Yuan Fang¹, Chenghao Liu², Steven C.H. Hoi¹,²

¹ Singapore Management University, Singapore; ² Salesforce Research Asia, Singapore

In Proceeding of 30th International Joint Conference on Artificial Intelligence (IJCAI-21)
21st -26th August, 2021
Outline

• Problem
• Proposed model: LGNN
• Experiments
• Conclusions
Problem: graph neural networks

- Graph neural networks (GNNs) [1, 2, 3]

\[ h_v^l = \sigma \left( \text{AGGR} \left( \{ W^l h_u^{l-1} : \forall u \in C_v \} \right) \right) \]

- Node classification

Problem: limitation of GNNs

- Different local context of each node
  - bio: $v_1$
  - bioinf: $v_2$
  - Cs: $v_3$

Can we allow each node to be parameterized by its own weight matrix?
Problem: Our Idea

• Localization
  – Localize the global model for each node

• Significance
  – Global vs. local
  – Node- and edge-level
Outline

• Problem
• Proposed model: LGNN
• Experiments
• Conclusions
Proposed Model: Localization

- General formulation of Localization
  - Localized model
    \[ \Theta_v = f(\Theta, C_v) \]
  - Local context of node \( v \) on graph \( G = (V, E) \)
    \[ C_v = \{ v \} \cup \{ u \in V : \langle v, u \rangle \in E \} \]
Proposed Model: Node-level Localization

- Global model: Conventional GNNs

\[ h^l_v = \sigma \left( \text{AGGR} \left( \{ W^l h^{l-1}_u : \forall u \in C_v \} \right) \right) \]

- Weight matrix
  - Aggregation function
  - Local context

- Node-level localization

\[ W^l_v = W^l \odot \left[ \begin{array}{c} (a^l_v)_{x d_l}^\top \\ (b^l_v)_{x d_l}^\top \end{array} \right] + \left[ \begin{array}{c} (a^l_v)_{x d_l}^\top \\ (b^l_v)_{x d_l}^\top \end{array} \right] \]

\[ c^l_v = \text{MEAN} \left( \{ h^{l-1}_u : \forall u \in C_v \} \right) \]

\[ a^l_v = \sigma \left( M^l_a c^l_v \right) + 1, \quad b^l_v = \sigma \left( M^l_b c^l_v \right) \]
Proposed Model: Edge-level Localization

• Localization of GNNs
  – Edge-level localization

\[
c_{u,v}^l = \text{CONCAT} \left( h_{v}^{l-1}, h_{u}^{l-1} \right)
\]

\[
h_{v}^{l} = \sigma \left( \text{AGGR} \left( \{ W_{v} h_{u}^{l-1} \odot a_{u,v}^{l} + b_{u,v}^{l} : \forall u \in C_v \} \right) \right)
\]

\[
a_{u,v}^{l} = \sigma \left( N_{a} c_{u,v}^{l} \right) + 1, \quad b_{u,v}^{l} = \sigma \left( N_{b} c_{u,v}^{l} \right)
\]
Proposed Model: Loss

• Semi-supervised node classification

\[ z_{v,k} = \text{SOFTMAX} \left( h_{v,k}^\ell \right) = \frac{\exp(h_{v,k}^\ell)}{\sum_{k'=1}^{K} \exp(h_{v,k'}^\ell)} \]

– Overall loss

\[ -\sum_{v \in V_Y} \sum_{k=1}^{K} Y_{v,k} \ln z_{v,k} + \lambda_G \| \Theta_G \|_2^2 + \lambda_L \| \Theta_L \|_2^2 \]
\[ + \lambda \left( \frac{\| A - 1 \|_2^2}{|A|} + \frac{\| B \|_2^2}{|B|} \right) \]
Outline

• Problem
• Proposed model: LGNN
• Experiments
• Conclusions
Datasets, evaluation and baselines

- **Datasets**

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Nodes</th>
<th># Edges</th>
<th># Classes</th>
<th># Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cora</td>
<td>2,708</td>
<td>5,429</td>
<td>7</td>
<td>1,433</td>
</tr>
<tr>
<td>Citeseer</td>
<td>3,327</td>
<td>4,732</td>
<td>6</td>
<td>3,703</td>
</tr>
<tr>
<td>Amazon</td>
<td>13,381</td>
<td>245,778</td>
<td>10</td>
<td>767</td>
</tr>
<tr>
<td>Chameleon</td>
<td>2,277</td>
<td>36,101</td>
<td>5</td>
<td>2,325</td>
</tr>
</tbody>
</table>

- **Evaluation**
  - Accuracy, micro-F

- **Baselines**
  - Embedding models: DeepWalk [1], Planetoid [2]
  - GNN models: GCN [3], GAT [4], GIN [5]
  - GNN-FiLM [6]: GCN-FiLM, GAT-FiLM, GIN-FiLM

**Node classification**

- LGNN consistently achieves significant performance boosts
- GAT-based models generally attain better performance than GCN- and GIN-based models
- Increasing the number of parameters alone cannot achieve the effect of localization

<table>
<thead>
<tr>
<th>Methods</th>
<th># Params (Cora)</th>
<th>Cora Accuracy</th>
<th>Cora Micro-F</th>
<th>Citeeseer Accuracy</th>
<th>Citeeseer Micro-F</th>
<th>Amazon Accuracy</th>
<th>Amazon Micro-F</th>
<th>Chameleon Accuracy</th>
<th>Chameleon Micro-F</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepWalk</td>
<td>693K</td>
<td>73.8±0.3</td>
<td>74.9±0.1</td>
<td>61.6±0.2</td>
<td>60.5±1.0</td>
<td>80.1±1.6</td>
<td>77.3±1.3</td>
<td>41.2±1.3</td>
<td>40.1±1.1</td>
</tr>
<tr>
<td>Planetoid</td>
<td>345K</td>
<td>66.1±0.4</td>
<td>64.5±0.5</td>
<td>64.5±0.3</td>
<td>62.9±0.4</td>
<td>69.8±1.7</td>
<td>64.5±1.5</td>
<td>39.3±1.8</td>
<td>37.7±1.7</td>
</tr>
<tr>
<td>GCN</td>
<td>11K</td>
<td>81.5±0.7</td>
<td>80.8±0.5</td>
<td>70.4±0.5</td>
<td>68.3±0.7</td>
<td>81.9±0.5</td>
<td>81.0±0.8</td>
<td>46.7±4.3</td>
<td>46.4±2.4</td>
</tr>
<tr>
<td>GCN-64</td>
<td>92K</td>
<td>82.0±0.3</td>
<td>80.9±0.3</td>
<td>71.1±0.3</td>
<td>69.2±0.4</td>
<td>82.1±0.5</td>
<td>81.2±0.8</td>
<td>48.3±3.3</td>
<td>46.3±1.8</td>
</tr>
<tr>
<td>GCN-96</td>
<td>138K</td>
<td>81.9±0.2</td>
<td>80.8±0.3</td>
<td>71.3±0.4</td>
<td>69.4±0.5</td>
<td>82.2±0.4</td>
<td>81.5±0.7</td>
<td>45.5±2.4</td>
<td>43.8±2.5</td>
</tr>
<tr>
<td>GCN-FiLM</td>
<td>35K</td>
<td>78.1±0.6</td>
<td>76.9±0.5</td>
<td>69.8±1.1</td>
<td>67.9±1.0</td>
<td>79.2±1.0</td>
<td>77.1±1.5</td>
<td>42.8±1.1</td>
<td>39.9±1.3</td>
</tr>
<tr>
<td>LGCN</td>
<td>104K</td>
<td>83.5±0.3</td>
<td>82.1±0.4</td>
<td>72.2±0.4</td>
<td>70.2±0.4</td>
<td>83.7±1.5</td>
<td>82.3±2.0</td>
<td>50.9±1.1</td>
<td>49.7±0.7</td>
</tr>
<tr>
<td>(improv.)</td>
<td>-</td>
<td>(1.8%)</td>
<td>(1.5%)</td>
<td>(1.3%)</td>
<td>(1.2%)</td>
<td>(1.8%)</td>
<td>(1.0%)</td>
<td>(5.4%)</td>
<td>(7.1%)</td>
</tr>
<tr>
<td>GAT</td>
<td>92K</td>
<td>82.9±0.6</td>
<td>82.0±0.6</td>
<td>72.4±0.7</td>
<td>70.4±0.8</td>
<td>82.4±1.3</td>
<td>80.1±1.9</td>
<td>47.2±1.1</td>
<td>46.2±2.1</td>
</tr>
<tr>
<td>GAT-64</td>
<td>738K</td>
<td>83.1±0.4</td>
<td>81.9±0.6</td>
<td>71.6±1.5</td>
<td>69.8±1.6</td>
<td>83.0±0.9</td>
<td>81.2±1.4</td>
<td>51.2±1.5</td>
<td>50.2±1.3</td>
</tr>
<tr>
<td>GAT-96</td>
<td>1108K</td>
<td>83.2±0.6</td>
<td>81.9±0.6</td>
<td>71.4±0.9</td>
<td>69.6±0.9</td>
<td>83.1±1.0</td>
<td>81.5±1.4</td>
<td>51.9±1.2</td>
<td>50.2±1.8</td>
</tr>
<tr>
<td>GAT-FiLM</td>
<td>277K</td>
<td>82.0±0.5</td>
<td>80.6±0.6</td>
<td>71.2±1.0</td>
<td>69.2±1.1</td>
<td>83.3±0.6</td>
<td>81.9±0.8</td>
<td>46.8±5.7</td>
<td>45.1±5.2</td>
</tr>
<tr>
<td>LGAT</td>
<td>836K</td>
<td>83.6±0.4</td>
<td>82.3±0.4</td>
<td>72.8±0.4</td>
<td>70.8±0.5</td>
<td>83.7±0.7</td>
<td>82.3±0.8</td>
<td>52.6±1.0</td>
<td>51.1±0.9</td>
</tr>
<tr>
<td>(improv.)</td>
<td>-</td>
<td>(0.5%)</td>
<td>(0.4%)</td>
<td>(0.6%)</td>
<td>(0.6%)</td>
<td>(0.5%)</td>
<td>(0.5%)</td>
<td>(1.3%)</td>
<td>(1.8%)</td>
</tr>
<tr>
<td>GIN</td>
<td>11K</td>
<td>80.2±0.5</td>
<td>78.8±0.3</td>
<td>68.5±0.7</td>
<td>66.5±1.0</td>
<td>79.6±1.7</td>
<td>78.5±2.6</td>
<td>45.8±3.0</td>
<td>41.2±4.0</td>
</tr>
<tr>
<td>GIN-64</td>
<td>92K</td>
<td>80.3±1.1</td>
<td>79.1±1.0</td>
<td>67.8±1.5</td>
<td>66.1±1.1</td>
<td>79.8±1.1</td>
<td>79.0±1.4</td>
<td>45.7±4.5</td>
<td>40.7±5.7</td>
</tr>
<tr>
<td>GIN-96</td>
<td>138K</td>
<td>79.9±1.1</td>
<td>78.9±1.0</td>
<td>68.6±1.4</td>
<td>66.6±1.6</td>
<td>80.2±2.1</td>
<td>79.0±3.2</td>
<td>45.9±3.5</td>
<td>41.5±4.1</td>
</tr>
<tr>
<td>GIN-FiLM</td>
<td>35K</td>
<td>79.8±0.7</td>
<td>78.5±0.5</td>
<td>67.7±1.4</td>
<td>65.8±1.5</td>
<td>78.6±2.8</td>
<td>77.2±3.3</td>
<td>38.8±2.6</td>
<td>34.2±2.9</td>
</tr>
<tr>
<td>LGIN</td>
<td>126K</td>
<td>82.6±0.8</td>
<td>81.6±0.8</td>
<td>71.3±0.4</td>
<td>69.5±0.5</td>
<td>84.0±1.2</td>
<td>82.7±1.7</td>
<td>48.3±1.9</td>
<td>47.3±1.9</td>
</tr>
<tr>
<td>(improv.)</td>
<td>-</td>
<td>(2.9%)</td>
<td>(3.2%)</td>
<td>(3.9%)</td>
<td>(4.4%)</td>
<td>(4.7%)</td>
<td>(4.7%)</td>
<td>(5.2%)</td>
<td>(14.0%)</td>
</tr>
</tbody>
</table>
Ablation study

- Utilizing only one module consistently outperforms the global model
- The node-level localization tends to perform better than edge-level localization.
- Modeling both jointly results in the best performance
Outline

• Problem
• Proposed model: LGNN
• Experiments
• Conclusions
• Motivation
  – We identified the need to localize GNNs for different nodes

• Proposed model: LGNN
  – Encode graph-level general patterns using a global weight matrix
  – Node-level and edge-level localization

• Experiments
  – Extensive experiments demonstrate that LGNN significantly outperforms state-of-the-art GNNs.
Thanks!

**Node-wise Localization of Graph Neural Networks.**
Zemin Liu, Yuan Fang, Chenghao Liu, Steven C.H. Hoi.

*In Proceeding of 30th International Joint Conference on Artificial Intelligence (IJCAI-21)*
21st -26th August, 2021

Paper, code, data…
www.yfang.site