Graph Neural Networks go grammatical

Jason Piquenot

Supervisors : Sébastien Adam  Romain Raveaux
Co-supervisors : Maxime Bérar  Pierre Héroux  Jean-Yves Ramel

statistical learning on LARge scale GRaphs (LARGR)

9th march 2023
GNNs expressive power

Weisfeiler-Lehman hierarchy [Xu et al., 2019]

Spectral response [Balcilar et al., 2021b]

Counting power [Chen et al., 2020]
Contributions

New GNN design strategy based on Context Free Grammar (CFG).

Example of 1-WL CFG.

Grammatical Graph Neural Network ($G^2N^2$) a 3-WL GNN.

$G^2N^2$ spectral response and counting power.
In [Brijder et al., 2019] they introduced MATLANG a matrix language.

**Matlang**

ML ($\mathcal{L}$) is a matrix language with an allowed operation set $\mathcal{L} = \{op_1, \ldots, op_n\}$, where $op_i \in \{\cdot, +, ^T, \text{diag}, \text{Tr}, 1, \odot, \times, f\}$.

**Sentence**

$e(X) \in \mathbb{R}$ is a sentence in ML ($\mathcal{L}$) if it consists of any possible consecutive operations in $\mathcal{L}$, operating on a given matrix $X$ and resulting in a scalar value.

As an example, $1^T (X \odot \text{diag}(1)) 1$ is a sentence in $\text{ML} (\text{T}, 1, \odot, \cdot, \text{diag})$ that computes trace of square matrix $X$. 
**ML (ℒ)-equivalent for matrices**

Two matrices $A$ and $B$ in $\mathcal{M}_{m,n}(\mathbb{R})$ are said to be ML (ℒ)-equivalent, denoted by $A \equiv_{ML(ℒ)} B$, if and only if $e(A) = e(B)$ for all sentences in ML (ℒ).

**ML (ℒ)-equivalent for graphs**

Two graphs $\mathcal{G}$ and $\mathcal{H}$ of the same order are said to be ML (ℒ)-equivalent, denoted by $\mathcal{G} \equiv_{ML(ℒ)} \mathcal{H}$, if and only if their adjacency matrices are ML (ℒ)-equivalent.
In [Geerts and Reutter, 2021], they proved the following theorems for the sets of operations $\mathcal{L}_1 = \{\cdot, \mathbf{T}, 1, \text{diag}\}$ and $\mathcal{L}_3 = \{\cdot, \mathbf{T}, 1, \text{diag}, \odot\}$.

**1-WL equivalence**

Two adjacency matrices are indistinguishable by the 1-WL test if and only if $e(A_G) = e(A_H)$ for all $e \in \mathcal{L}_1$. Hence, all possible sentences in $\mathcal{L}_1$ are the same for 1-WL-equivalent adjacency matrices. Thus,

$$A_G \equiv_{1\text{-WL}} A_H \iff A_G \equiv_{\text{ML}(\mathcal{L}_1)} A_H$$

**3-WL equivalence**

Two adjacency matrices are indistinguishable by the 3-WL test if and only if $e(A_G) = e(A_H)$ for all $e \in \mathcal{L}_3$. Hence, all possible sentences in $\mathcal{L}_3$ are the same for 3-WL-equivalent adjacency matrices. Thus,

$$A_G \equiv_{3\text{-WL}} A_H \iff A_G \equiv_{\text{ML}(\mathcal{L}_3)} A_H$$
For $\mathcal{L}_1 = \{\cdot, T, 1, \text{diag}\}$, define

$G_{\mathcal{L}_1} = (V, \Sigma, R, S_t)$.

$V = \{M, V_c, V_l, S\}$

$\Sigma = \{A, \text{diag}, 1, T, (, )\}$

$S_t = S$

where the rules in $R$ are

$S \rightarrow (V_l)(V_c) \mid \text{diag} (S) \mid SS$

$M \rightarrow MM \mid (M)^T \mid \text{diag} (V_c) \mid (V_c)(V_l) \mid A$

$V_c \rightarrow MV_c \mid (V_l)^T \mid V_cS \mid 1$

$V_l \rightarrow V_lM \mid (V_c)^T \mid SV_l$

$S \rightarrow V_lV_c$

$V_l \rightarrow (V_c)^T, \quad V_c \rightarrow MV_c$

$V_c \rightarrow 1, \quad M \rightarrow MM, \quad V_c \rightarrow 1$

$M \rightarrow A, \quad M \rightarrow \text{diag}(V_c)$

$V_c \rightarrow MV_c$

$M \rightarrow A, \quad V_c \rightarrow 1$
Theorem (ML ($\mathcal{L}_1$) Reduced CFG)

The following CFG denoted by $r-G_{\mathcal{L}_1}$ is as expressive as 1-WL.

$$V_c \rightarrow \text{diag} (V_c) V_c \mid AV_c \mid 1$$

Corollary (GNNML1 CFG)

The following CFG, as expressive than ML ($\mathcal{L}_1$) represents GNNML1 [Balcilar et al., 2021a].

$$V_c \rightarrow V_c \odot V_c \mid AV_c \mid 1$$

GNNML1 node update:

$$H^{(l+1)} = \sigma(H^{(l)} \cdot W^{(l,1)} + A \cdot H^{(l)} \cdot W^{(l,2)}) + H^{(l)} \cdot W^{(l,3)} \odot H^{(l)} \cdot W^{(l,4)}$$
Proposition (GCN CFG)

The following CFG is strictly less expressive than $\text{ML}(\mathcal{L}_1)$.

$$ V_c \rightarrow C_1 V_c \mid \cdots \mid C_k V_c \mid 1 $$

When $C_s$ is include in $\text{ML}(\mathcal{L}_1)$.

As an example, the following CFG, strictly less expressive than $\text{ML}(\mathcal{L}_1)$ represents GCN [Kipf and Welling, 2017].

$$ V_c \rightarrow CV_c \mid 1 $$

Where $C = \text{diag} \left( (A + I)^{-\frac{1}{2}} (A + I) \text{diag} ((A + I)^{-\frac{1}{2}} \right)$
Theorem (ML ($\mathcal{L}_3$) Reduced CFG)

The following CFG denoted by $r$-$G_{\mathcal{L}_3}$ is as expressive as 3-WL.

$$V_c \to MV_c \mid 1$$
$$M \to (M \odot M) \mid MM \mid \text{diag}(V_c) \mid A$$
From CFG to $G^2N^2$

$G^2N^2$ architecture:

$C^{(0)} \rightarrow C^{(1)} \rightarrow C^{(2)} \ldots \rightarrow C^{(k)}$

$H^{(0)} f_n \rightarrow H^{(1)} \rightarrow H^{(2)} \ldots \rightarrow H^{(k)}$

$G^2N^2$ update equation:

$$C^{(l+1)} = mlp(C^{(l)} | L_1(C^{(l)}) \odot L_2(C^{(l)}) | L_3(C^{(l)}) \cdot L_4(C^{(l)}) | \text{diag}(L_5(H^{(l)}))),$$

$$H^{(l+1)} = \sum_{i=1}^{S(l+1)} C^{(l+1)}_i H^{(l)} W^{(l,i)}.$$
Theorem (G$^2$N$^2$ in the WL hierarchy)

$G^2N^2$ is as expressive as 3-WL.

Table: The accuracy on EXP and SR25 datasets denotes the ratio of pairs of non isomorphic respectively 1-WL equivalent and 3-WL equivalent graphs that are separate by the model.

<table>
<thead>
<tr>
<th>Method</th>
<th>EXP</th>
<th>SR25</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-WL-bounded GNN</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>CHEBNET</td>
<td>87%</td>
<td>0%</td>
</tr>
<tr>
<td>3-WL-bounded GNN</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>PPGN</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>$G^2N^2$</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>$I^2$-GNN</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
Theorem ($G^2N^2$ counting power)

$G^2N^2$ can count chordal cycle and cycle up to length 6 at edge level.

Table: $G^2N^2$ normalised MAE on counting substructures at edge level.

<table>
<thead>
<tr>
<th></th>
<th>triangle</th>
<th>4-cycle</th>
<th>5-cycle</th>
<th>6-cycle</th>
<th>chordal cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.99e-04</td>
<td>4.55e-04</td>
<td>2.93e-03</td>
<td>3.58e-03</td>
<td>1.56e-04</td>
</tr>
</tbody>
</table>
Theorem ($G^2N^2$ spectral response)

$G^2N^2$ can approximate low-pass, high-pass and band-pass filter in the spectral domain.

Table: $R^2$ score on spectral filtering node regression problems. Results are median of 10 different runs.

<table>
<thead>
<tr>
<th>Method</th>
<th>Low-pass</th>
<th>High-pass</th>
<th>Band-pass</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLP</td>
<td>0.9749</td>
<td>0.0167</td>
<td>0.0027</td>
</tr>
<tr>
<td>GCN</td>
<td>0.9858</td>
<td>0.0863</td>
<td>0.0051</td>
</tr>
<tr>
<td>GAT</td>
<td>0.9811</td>
<td>0.0879</td>
<td>0.0044</td>
</tr>
<tr>
<td>GIN</td>
<td>0.9824</td>
<td>0.2934</td>
<td>0.0629</td>
</tr>
<tr>
<td>CHEBNET</td>
<td>0.9995</td>
<td>0.9901</td>
<td>0.8217</td>
</tr>
<tr>
<td>PPGN</td>
<td>0.9991</td>
<td>0.9925</td>
<td>0.1041</td>
</tr>
<tr>
<td>GNNML1</td>
<td>0.9994</td>
<td>0.9833</td>
<td>0.3802</td>
</tr>
<tr>
<td>GNNML3</td>
<td>0.9995</td>
<td>0.9909</td>
<td>0.8189</td>
</tr>
<tr>
<td>$G^2N^2$</td>
<td>0.9996</td>
<td>0.9994</td>
<td>0.8206</td>
</tr>
</tbody>
</table>

On the left $G^2N^2$'s prediction, on the center the input and on the right, the target
Table: Results on QM9 dataset focusing on the best methods. The metric is MAE, the lower, the better.

<table>
<thead>
<tr>
<th>Target</th>
<th>$I^2$-GNN</th>
<th>PPGN(1)</th>
<th>PPGN(12sm)</th>
<th>$G^2N^2(12)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>0.428</td>
<td>0.0934</td>
<td>0.485</td>
<td><strong>0.0621</strong></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.230</td>
<td>0.318</td>
<td>0.504</td>
<td><strong>0.119</strong></td>
</tr>
<tr>
<td>$\epsilon_{\text{homo}}$</td>
<td>0.00261</td>
<td><strong>0.00174</strong></td>
<td>0.126</td>
<td>0.0312</td>
</tr>
<tr>
<td>$\epsilon_{\text{lumo}}$</td>
<td>0.00267</td>
<td><strong>0.0021</strong></td>
<td>0.138</td>
<td>0.0333</td>
</tr>
<tr>
<td>$\Delta\epsilon$</td>
<td>0.0038</td>
<td><strong>0.0029</strong></td>
<td>0.182</td>
<td>0.0427</td>
</tr>
<tr>
<td>$R^2$</td>
<td>18.64</td>
<td>3.78</td>
<td>17.17</td>
<td><strong>0.992</strong></td>
</tr>
<tr>
<td>ZPVE</td>
<td><strong>0.00014</strong></td>
<td>0.000399</td>
<td>0.0105</td>
<td>0.00426</td>
</tr>
<tr>
<td>$U_0$</td>
<td>0.211</td>
<td><strong>0.022</strong></td>
<td>14.41</td>
<td>1.71</td>
</tr>
<tr>
<td>$U$</td>
<td>0.206</td>
<td><strong>0.0504</strong></td>
<td>14.41</td>
<td>1.72</td>
</tr>
<tr>
<td>$H$</td>
<td>0.269</td>
<td><strong>0.0294</strong></td>
<td>14.41</td>
<td>1.71</td>
</tr>
<tr>
<td>$G$</td>
<td>0.261</td>
<td><strong>0.024</strong></td>
<td>14.41</td>
<td>1.72</td>
</tr>
<tr>
<td>$C_v$</td>
<td>0.0730</td>
<td>0.144</td>
<td>0.192</td>
<td><strong>0.0581</strong></td>
</tr>
</tbody>
</table>

Table: Results of $G^2N^2$ on TUD dataset compared to the best competitor. The metric is accuracy, the higher, the better.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$G^2N^2$</th>
<th>rank</th>
<th>Best GNN competitor</th>
</tr>
</thead>
<tbody>
<tr>
<td>MUTAG</td>
<td>92.0±4.3</td>
<td>2</td>
<td>92.2±7.5</td>
</tr>
<tr>
<td>PTC</td>
<td>71.8±6.7</td>
<td>1</td>
<td>68.2±7.2</td>
</tr>
<tr>
<td>Proteins</td>
<td>77.8±3.2</td>
<td>1</td>
<td>77.4±4.9</td>
</tr>
<tr>
<td>NCI1</td>
<td>80.2±2.1</td>
<td>8</td>
<td>83.5±2.0</td>
</tr>
<tr>
<td>IMDB-B</td>
<td>76.8±2.8</td>
<td>2</td>
<td>77.8±3.3</td>
</tr>
<tr>
<td>IMDB-M</td>
<td>54.0±2.93</td>
<td>2</td>
<td>54.3±3.3</td>
</tr>
</tbody>
</table>
G$^2$N$^2$ assets

- 3-WL equivalent
- Large spectral response
- Counting power at edge level
- Direct from CFG

G$^2$N$^2$ drawbacks

- Time complexity $O(n^3)$
- Memory complexity $O(n^2)$


Thank you for your attention.

Our paper is available here.

Graph Neural Network go Grammatical

https://arxiv.org/abs/2303.01590