Graph Contrastive Learning Automated
(Appendix)

Yuning You¹ Tianlong Chen² Yang Shen¹ Zhangyang Wang²

A. Alternating Gradient Descent for JOAOv2

We adapt alternating gradient descent (AGD) in Algorithm 1 to optimize (10) in main text, executed as Algorithm S1, with the following modified upper-level minimization and lower-level maximization.

Algorithm S1 AGD for optimization (10) in main text

Input: initial parameter \(\theta^{(0)}\), sampling distribution \(\mathbb{P}^{(0)}(A_1,A_2)\), \(\mathbb{P}^{(0)}(\theta′,\theta′′) = \mathbb{P}^{(0)}(A_1,A_2)\), optimization step \(N\).

for \(n = 1 \) to \(N\) do

1. Upper-level minimization: fix \(\mathbb{P}^{(n-1)}(A_1,A_2)\), and call equation (1) to update \(\theta^{(n)}\).
2. Lower-level maximization: fix \(\theta^{(n)}\), call equation (3) to update \(\mathbb{P}^{(n)}(A_1,A_2)\), and set \(\mathbb{P}^{(n)}(\theta′,\theta′′) = \mathbb{P}^{(n)}(A_1,A_2)\).

end for

Return: Optimized parameter \(\theta^{(N)}\).

Upper-level minimization. The upper-level minimization w.r.t. \(\theta\) given the sampling distribution \(\mathbb{P}(A_1,A_2)\), setting \(\mathbb{P}(\theta′,\theta′′) = \mathbb{P}(A_1,A_2)\), is represented as:

\[
\theta^{(n)} = \theta^{(n-1)} - \alpha′ \nabla_\theta \mathcal{L}_2(G,A_1,A_2,\theta′,\Theta′′_1,\Theta′′_2), \tag{1}
\]

where \(\alpha′ \in \mathbb{R}_{>0}\) is the learning rate.

Lower-level maximization. To calculate the gradient of the lower-level objective w.r.t. \(\mathbb{P}(A_1,A_2)\), we make similar efforts to approximate the contrastive loss as:

\[
\mathcal{L}_2(G,A_1,A_2,\theta′,\Theta′′_1,\Theta′′_2) \\
\approx \sum_{i=1}^{\mid A \mid} \sum_{j=1}^{\mid A \mid} p_{ij} \ell_2(G,A^i,A^j,\theta′,\theta′′_i,\theta′′_j) \\
= \sum_{i=1}^{\mid A \mid} \sum_{j=1}^{\mid A \mid} p_{ij} \left\{ - \mathbb{E}_{\mathcal{P}_c} \text{sim}(\Theta′_i,\Theta′_j,G) + \mathbb{E}_{\mathcal{P}_c} \log(\text{exp}(\text{sim}(\Theta′_i,G,T^{i}_{v2,\theta}(G)),T^{j}_{v2,\theta}(G'))) \right\}, \tag{2}
\]

where \(T^{i}_{v2,\theta}(G) = A^i \circ f_\theta \circ g_{\mathcal{P}_c}\), \(i = 1, ..., 5\). Thus, projected gradient descent is performed as:

\[
b = \mathbb{P}^{(n-1)} + \alpha'' \nabla_\mathbb{P} \mathcal{L}_2(G,A_1,A_2,\theta′,\Theta′′_1,\Theta′′_2), \tag{3}
\]

where \(p = [p_{ij}]\), \(i, j = 1, ..., \mid A \mid\), \(\mathbb{P}(p) = \sum_{i=1}^{\mid A \mid} \sum_{j=1}^{\mid A \mid} p_{ij}\), \(\mathbb{P}(p) = \sum_{i=1}^{\mid A \mid} \sum_{j=1}^{\mid A \mid} p_{ij} - \frac{\gamma}{\mid A \mid} \), \(\alpha'' \in \mathbb{R}_{>0}\) is the learning rate, \(\mu\) is the root of the equation \(1^T(b - \mu 1) = 1\), and \((\cdot)_+\) is the element-wise non-negative operator.

B. Dataset Statistics

Dataset statistics can be found in Table S1, S2 and S3.

Table S1: Statistics for datasets of diverse nature from the benchmark TUDataset.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Graph Count</th>
<th>Avg. Node</th>
<th>Avg. Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCTI</td>
<td>4,110</td>
<td>29.87</td>
<td>1.08</td>
</tr>
<tr>
<td>PROTEINS</td>
<td>1,113</td>
<td>39.06</td>
<td>1.86</td>
</tr>
<tr>
<td>DD</td>
<td>1,178</td>
<td>284.32</td>
<td>715.66</td>
</tr>
<tr>
<td>MUTAG</td>
<td>188</td>
<td>17.93</td>
<td>19.79</td>
</tr>
<tr>
<td>COLLAB</td>
<td>5,000</td>
<td>74.49</td>
<td>32.99</td>
</tr>
<tr>
<td>RDT-B</td>
<td>2,000</td>
<td>429.63</td>
<td>1.15</td>
</tr>
<tr>
<td>RDB-M</td>
<td>2,000</td>
<td>429.63</td>
<td>497.75</td>
</tr>
<tr>
<td>GITHUB</td>
<td>4,999</td>
<td>508.52</td>
<td>594.87</td>
</tr>
<tr>
<td>IMDB-B</td>
<td>1,000</td>
<td>19.77</td>
<td>96.53</td>
</tr>
</tbody>
</table>

¹Texas A&M University ²The University of Texas at Austin.
Correspondence to: Yang Shen <yshen@tam.edu>, Zhangyang Wang <atlaswang@utexas.edu>.
C. Augmentation Sampling Rules for GraphCL

GraphCL uniformly samples augmentations from a pre-defined pool. Augmentation pools for datasets are presented in Table S4.

D. JOAOv2 Selected Augmentation-Pairs

Alignment with Previous “Best Practices”

Schematic diagram of JOAOv2 is drawn in Figure S2. \( P(A_1, A_2) \) optimized by JOAOv2 is plotted in the top row of Figure S1 along with GraphCL performance gains of different augmentation pairs in the bottom row, following...
the same procedure as described in Sec. 3.2.1 of main text.

**Figure S2:** Schematic diagram of GraphCL with multiple augmentation-aware projection heads where $P(g_{\Theta_1}, g_{\Theta_2}) = P(A_1, A_2)$.

**E. Comparison between JOAO w/ and w/o the Prior**

See Table S5 for comparison between JOAO w/ and w/ the prior in the semi-supervised learning setting.

**Table S5:** Semi-supervised performance (%) of JOAO w/ and w/o prior.

<table>
<thead>
<tr>
<th></th>
<th>w/o prior</th>
<th>w/ prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>NCI1</td>
<td>61.51±0.32</td>
<td>61.97±0.72</td>
</tr>
<tr>
<td>PROTEINS</td>
<td>71.78±0.70</td>
<td>72.13±0.92</td>
</tr>
</tbody>
</table>