GCC: Graph Contrastive Coding for Graph Neural Network Pre-Training

Jiezhong Qiu, Qibin Chen, Yuxiao Dong, Jing Zhang, Hongxia Yang, Ming Ding, Kuansan Wang, Jie Tang
Real-world Graphs

Question:
How to design machine learning models to learn the universal structural patterns across networks?
Pre-training and Fine-tuning

Computer Vision
ResNet
ImageNet

NLP
BERT
Wikipedia + Book corpus

Graph Learning
GCC
Problem

GNN pre-training problem.
The GNN Pre-Training Problem

• Problem:
  • Learn a function $f$ that maps a vertex to a low-dimensional vector
  • **Structural similarity**: map vertices with similar local network topologies close in the vector space
  • **Transferability**: compatible with vertices and graphs from various sources, even unseen during training time.
GCC Framework

Graph Contrastive Coding
Graph Contrastive Coding (GCC)

Hypothesis:
Graph structural patterns are universal and transferable across networks.
GCC Pre-training

• **Pre-training Task:** *Instance* Discrimination
• **InfoNCE Loss:** output *instance representations* that are capable of capturing the *similarities* between instances

$$
\mathcal{L} = - \log \frac{\exp (q^T k_+ / \tau)}{\sum_{i=0}^{K} \exp (q^T k_i / \tau)}
$$

• query instance $x^q$
• query $q$ (embedding of $x^q$), i.e., $q = f(x^q)$
• dictionary of keys $\{k_0, k_1, \ldots, k_K\}$
• key $k = f(x^k)$

• Contrastive learning for graphs?
  • **Q1:** How to define *instances* in graphs?
  • **Q2:** How to define *(dis) similar instance* pairs?
  • **Q3:** What are the proper *encoders*?
GCC Pre-training

- **Q1**: How to define *instances* in graphs?
- **Q2**: How to define *(dis)* similar instance?
- **Q3**: What are the proper *encoders*?

![Graph Encoding Diagram]

\[
\mathcal{L} = - \log \frac{\exp \left( q^\top k_+ / \tau \right)}{\sum_{i=0}^{K} \exp \left( q^\top k_i / \tau \right)}
\]
GCC Pre-training: Learning Algorithms

• Optimizing Contrastive Loss
  • Encoded query $q$
  • $K + 1$ encoded keys $\{k_0, \ldots, k_K\}$

$$L = -\log \frac{\exp (q^T k_+ / \tau)}{\sum_{i=0}^{K} \exp (q^T k_i / \tau)}$$

End-to-end (E2E)  
Momentum Contrast (MoCo)

figure credit: 
Momentum Contrast for Unsupervised Visual Representation Learning  
arxiv.org/abs/1911.05722
GCC Fine-tuning

Node Classification
GCC
US-Airport

Graph Classification
GCC
Reddit

Similarity Search
GCC
GCC
KDD
ICDM

Fine-tuning

Graph Encoder
Classifier
Label y
GCC Fine-tuning: Full v.s. Freezing

Full fine-tuning

Freezing fine-tuning

Graph Encoder Classifier Label

Full Fine-tuning

Freezing Fine-tuning

Feature Extractor
Experiments
GCC Pre-Training / Fine-tuning

• Six real-world information networks for pre-training.

Table 1: Datasets for pre-training, sorted by number of vertices.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Academia</th>
<th>DBLP (SNAP)</th>
<th>DBLP (NetRep)</th>
<th>IMDB</th>
<th>Facebook</th>
<th>LiveJournal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>137,969</td>
<td>317,080</td>
<td>540,486</td>
<td>896,305</td>
<td>3,097,165</td>
<td>4,843,953</td>
</tr>
<tr>
<td></td>
<td>739,384</td>
<td>2,099,732</td>
<td>30,491,458</td>
<td>7,564,894</td>
<td>47,334,788</td>
<td>85,691,368</td>
</tr>
</tbody>
</table>

• Fine-tuning Tasks:
  • Node classification
  • Graph classification
  • Top-k Similarity search
Task 1: Node Classification

• Setup
  • US-Airport
  • AMiner academic graph

<table>
<thead>
<tr>
<th>Datasets</th>
<th>US-Airport</th>
<th>H-index</th>
</tr>
</thead>
<tbody>
<tr>
<td>$</td>
<td>V</td>
<td>$</td>
</tr>
<tr>
<td>$</td>
<td>E</td>
<td>$</td>
</tr>
<tr>
<td>ProNE</td>
<td>62.3</td>
<td>69.1</td>
</tr>
<tr>
<td>GraphWave</td>
<td>60.2</td>
<td>70.3</td>
</tr>
<tr>
<td>Struc2vec</td>
<td>66.2</td>
<td>&gt; 1 Day</td>
</tr>
<tr>
<td>GCC (E2E, freeze)</td>
<td>64.8</td>
<td>78.3</td>
</tr>
<tr>
<td>GCC (MoCo, freeze)</td>
<td>65.6</td>
<td>75.2</td>
</tr>
<tr>
<td>GCC (rand, full)</td>
<td>64.2</td>
<td>76.9</td>
</tr>
<tr>
<td>GCC (E2E, full)</td>
<td>68.3</td>
<td>80.5</td>
</tr>
<tr>
<td>GCC (MoCo, full)</td>
<td>67.2</td>
<td><strong>80.6</strong></td>
</tr>
</tbody>
</table>
Task 2: Graph Classification

• Setup
  • COLLAB, RDT-B, RDT-M, & IMDB-B, IMDB-M

<table>
<thead>
<tr>
<th>Datasets</th>
<th>IMDB-B</th>
<th>IMDB-M</th>
<th>COLLAB</th>
<th>RDT-B</th>
<th>RDT-M</th>
</tr>
</thead>
<tbody>
<tr>
<td># graphs</td>
<td>1,000</td>
<td>1,500</td>
<td>5,000</td>
<td>2,000</td>
<td>5,000</td>
</tr>
<tr>
<td># classes</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Avg. # nodes</td>
<td>19.8</td>
<td>13.0</td>
<td>74.5</td>
<td>429.6</td>
<td>508.5</td>
</tr>
<tr>
<td>DGK</td>
<td>67.0</td>
<td>44.6</td>
<td>73.1</td>
<td>78.0</td>
<td>41.3</td>
</tr>
<tr>
<td>graph2vec</td>
<td>71.1</td>
<td>50.4</td>
<td>–</td>
<td>75.8</td>
<td>47.9</td>
</tr>
<tr>
<td>InfoGraph</td>
<td>73.0</td>
<td>49.7</td>
<td>–</td>
<td>82.5</td>
<td>53.5</td>
</tr>
<tr>
<td>GCC (E2E, freeze)</td>
<td>71.7</td>
<td>49.3</td>
<td>74.7</td>
<td>87.5</td>
<td>52.6</td>
</tr>
<tr>
<td>GCC (MoCo, freeze)</td>
<td>72.0</td>
<td>49.4</td>
<td>78.9</td>
<td>89.8</td>
<td>53.7</td>
</tr>
<tr>
<td>DGCNN</td>
<td>70.0</td>
<td>47.8</td>
<td>73.7</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>GIN</td>
<td>75.6</td>
<td>51.5</td>
<td>80.2</td>
<td>89.4</td>
<td>54.5</td>
</tr>
<tr>
<td>GCC (rand, full)</td>
<td>75.6</td>
<td>50.9</td>
<td>79.4</td>
<td>87.8</td>
<td>52.1</td>
</tr>
<tr>
<td>GCC (E2E, full)</td>
<td>70.8</td>
<td>48.5</td>
<td>79.0</td>
<td>86.4</td>
<td>47.4</td>
</tr>
<tr>
<td>GCC (MoCo, full)</td>
<td>73.8</td>
<td>50.3</td>
<td>81.1</td>
<td>87.6</td>
<td>53.0</td>
</tr>
</tbody>
</table>
Task 3: Top-\(k\) Similarity Search

- Setup
  - AMiner academic graph

---

| \(|V|\) | 2,867 | 2,607 | 2,851 | 3,548 | 2,616 | 2,559 |
| \(|E|\) | 7,637 | 4,774 | 6,354 | 7,076 | 8,304 | 6,668 |
| # ground truth | 697 | 874 | 874 | 898 |
| \(k\) | 20 | 40 | 20 | 40 | 20 | 40 |
| Random | 0.0198 | 0.0566 | 0.0223 | 0.0447 | 0.0221 | 0.0521 |
| RoIX | 0.0779 | 0.1288 | 0.0548 | 0.0984 | 0.0776 | 0.1309 |
| Panther++ | 0.0892 | 0.1558 | 0.0782 | 0.1185 | 0.0921 | 0.1320 |
| GraphWave | 0.0846 | 0.1693 | 0.0549 | 0.0995 | 0.0947 | 0.1470 |
| GCC (E2E) | **0.1047** | 0.1564 | 0.0549 | **0.1217** | 0.0835 | 0.1336 |
| GCC (MoCo) | 0.0904 | 0.1521 | 0.0652 | 0.1178 | 0.0846 | 0.1425 |
We study the pre-training of GNN with the goal of characterizing and transferring structural representations in social and information networks. We present Graph Contrastive Coding, which is a graph-based contrastive learning framework to pre-train GNN. The pre-trained GNN achieves competitive performance to its supervised trained-from-scratch counterparts in 3 graph learning tasks on 10 graph datasets.
Thanks.

Q&A

https://github.com/THUDM/GCC
Find us at KDD 2020

https://github.com/THUDM/GCC