HDMI: High-order Deep Multiplex Infomax

Presenter: Baoyu Jing
Contact: baoyuj2@illinois.edu

Baoyu Jing (UIUC)
Chanyoung Park (KAIST)
Hanghang Tong (UIUC)
Outline

• Introduction
• Preliminary
• Methodology
• Experiments
• Conclusion
Outline

- Introduction
  - Preliminary
  - Methodology
  - Experiments
  - Conclusion
Introduction: Ubiquity of Network

• Network in various applications.
  • Nodes are connected by various relations.

Social Network
  • Friendship
  • Colleague
  • Classmates

Product Network
  • Also View
  • Also Buy
  • Bought Together

Paper Network
  • Citation
  • Same Author
  • Same Keyword

https://www.lebow.drexel.edu/news/relational-networking-not-just-collecting-contacts
https://www.cwts.nl/blog?article=n-r2r294
Introduction: Self-supervised Learning

• Self-supervised learning
  • Train models without external training signals.
    • Do not need human labeling.
  • Pre-trained models perform well for down-stream tasks.
    • E.g., classification and clustering etc.
  • Key challenge: how to build the training signal?

• Deep Graph Infomax (DGI) for graphs
  • Mutual Information (MI) based training signal
  • Key idea:
    • Maximize the MI between node embedding $h_n$ and summary vector $s$.

Introduction: DGI Limitation #1

1. It only considers the **extrinsic** (global) information.
   - **Intrinsic** (node attribute) information is also important.
   - **✗** Existing methods:
     - Use reconstruction error
     - Reconstruction error doesn’t imply high quality!
   - **✓** In our work:
     - Maximize MI between node embedding and attributes
     - We propose to use **High-order Mutual Information** to jointly capture both extrinsic and intrinsic signals.
     - We propose a novel **High-order Deep Infomax (HDI)** as the training signal.
2. DGI assumes a single type of relations among nodes.
   • Nodes are connected by **multiple** relations (**Multiplex Graph**).
     • Each relation is a layer of the graph.
     • Common strategy:
       1) Separately consider each layer.
       2) Combine embedding from different layers.
   ✗ Simplest way to combine embeddings:
     • Average pooling
   ✓ In our work:
     • Attention based fusion module.
Outline

• Introduction

➢ Preliminary

• Methodology

• Experiments

• Conclusion
Preliminary: DGI

• Key steps:
  1) Generate a corrupted network via corruption function $\mathcal{C}$
  2) Use the encoder $\mathcal{E}$ to obtain node embeddings $h_n$ and $\tilde{h}_n$.
  3) Use the readout function $\mathcal{R}$ to obtain the summary vector $s$.
  4) Use the discriminator $\mathcal{D}_E$ to discriminate $h_n$ and $\tilde{h}_n$.
  5) Maximize $I(h_n; s)$ via: $\mathcal{L} = \sup_{\Theta} \mathbb{E}[\log \mathcal{D}(h_n; s)] + \mathbb{E}[\log(1 - \mathcal{D}(\tilde{h}_n; s))]$
Outline

• Introduction
• Preliminary
  ➢ Methodology
• Experiments
• Conclusion
Methodology: High-order Deep Infomax

- High-order Deep Infomax (HDI):
  - Capture extrinsic and intrinsic signals via high-order mutual information $I(h_n; s; f_n)$.

- High-order mutual information:
  $I(h_n; s; f_n) = I(h_n; s) + I(h_n; f_n) - I(h_n; s, f_n)$

- If directly maximize $I(h_n; s; f_n)$
  - Must max $I(h_n; s) + I(h_n; f_n)$ and min $I(h_n; s, f_n)$
  - Maximizing $I(h_n; s, f_n)$ improves performance

- Jointly maximize three mutual information:
  - $\mathcal{L} = \lambda_E I(h_n; s) + \lambda_I I(h_n; f_n) + \lambda_J I(h_n; s, f_n)$
  - Final objective function will be:
    - $\mathcal{L} = \lambda_E \mathcal{L}_E + \lambda_I \mathcal{L}_I + \lambda_J \mathcal{L}_J$

$\mathcal{L}_E$, $\mathcal{L}_I$ and $\mathcal{L}_J$ are the BCE losses of the discriminators.
Methodology: Fusion Module

- Extend HDI to multiplex graphs.
  - How to combine different layers?
- Fusion module is attention-based.
  - Different layers have different weights.
- Training the fusion module:
  - Apply HDI on top of the fused embedding.

- Full model: High-order Deep Multiplex Infomax (HDMI)
  - Objective: $\mathcal{L} = \lambda_M \mathcal{L}_M + \sum_r \lambda_r \mathcal{L}_r$

\[ \begin{aligned}
\text{fusion module} & \quad \text{different layers}
\end{aligned} \]
Outline

• Introduction
• Preliminary
• Methodology
  ➢ Experiments
• Conclusion
## Experiments: Setup

### Datasets

<table>
<thead>
<tr>
<th>Datasets</th>
<th># Nodes</th>
<th>Relation Types</th>
<th># Edges</th>
<th># Attributes</th>
<th># Labeled Data</th>
<th># Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>3,025</td>
<td>Paper-Subject-Paper (PSP) &lt;br&gt; Paper-Author-Paper (PAP)</td>
<td>2,210,761 &lt;br&gt; 29,281</td>
<td>1,830 &lt;br&gt; (Paper Abstract)</td>
<td>600</td>
<td>3</td>
</tr>
<tr>
<td>IMDB</td>
<td>3,550</td>
<td>Movie-Actor-Movie (MAM) &lt;br&gt; Movie-Director-Movie (MDM)</td>
<td>66,428 &lt;br&gt; 13,788</td>
<td>1,007 &lt;br&gt; (Movie plot)</td>
<td>300</td>
<td>3</td>
</tr>
<tr>
<td>DBLP</td>
<td>7,907</td>
<td>Paper-Author-Paper (PAP) &lt;br&gt; Paper-Paper-Paper (PPP) &lt;br&gt; Paper-Author-Term-Author-Paper (PATAP)</td>
<td>144,783 &lt;br&gt; 90,145 &lt;br&gt; 57,137,515</td>
<td>2,000 &lt;br&gt; (Paper Abstract)</td>
<td>80</td>
<td>4</td>
</tr>
<tr>
<td>Amazon</td>
<td>7,621</td>
<td>Item-AlsoView-Item (IVI) &lt;br&gt; Item-AlsoBought-Item (IBI) &lt;br&gt; Item-BoughtTogether-Item (IOI)</td>
<td>266,237 &lt;br&gt; 1,104,257 &lt;br&gt; 16,305</td>
<td>2,000 &lt;br&gt; (Item description)</td>
<td>80</td>
<td>4</td>
</tr>
</tbody>
</table>

### Questions
- How will HDI and HDMI improve the quality?
- Will fusion module assign appropriate attention scores?

### Tasks & Metrics
- Node classification: Macro-F1 & Micro-F1
- Node clustering: NMI, Similarity Search (Sim@5)

### Baselines
- Network embedding: Deepwalk, DGI etc.
- Multiplex network embeddings: HAN, DMGI etc.

---

Experiments: Node Classification

- HDMI performs the best.
- HDI is better than baselines.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ACM</th>
<th>IMDB</th>
<th>DBLP</th>
<th>Amazon</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Macro-F1</td>
<td>Micro-F1</td>
<td>Macro-F1</td>
<td>Micro-F1</td>
</tr>
<tr>
<td>DeepWalk</td>
<td>0.739</td>
<td>0.748</td>
<td>0.532</td>
<td>0.550</td>
</tr>
<tr>
<td>node2vec</td>
<td>0.741</td>
<td>0.749</td>
<td>0.533</td>
<td>0.550</td>
</tr>
<tr>
<td>GCN/GAT</td>
<td>0.869</td>
<td>0.870</td>
<td>0.603</td>
<td>0.611</td>
</tr>
<tr>
<td>DGI</td>
<td>0.881</td>
<td>0.881</td>
<td>0.598</td>
<td>0.606</td>
</tr>
<tr>
<td>ANRL</td>
<td>0.819</td>
<td>0.820</td>
<td>0.573</td>
<td>0.576</td>
</tr>
<tr>
<td>CAN</td>
<td>0.590</td>
<td>0.636</td>
<td>0.577</td>
<td>0.588</td>
</tr>
<tr>
<td>DGCN</td>
<td>0.888</td>
<td>0.888</td>
<td>0.582</td>
<td>0.592</td>
</tr>
<tr>
<td>CMNA</td>
<td>0.782</td>
<td>0.788</td>
<td>0.549</td>
<td>0.566</td>
</tr>
<tr>
<td>MNE</td>
<td>0.792</td>
<td>0.797</td>
<td>0.552</td>
<td>0.574</td>
</tr>
<tr>
<td>mGCN</td>
<td>0.858</td>
<td>0.860</td>
<td>0.623</td>
<td>0.630</td>
</tr>
<tr>
<td>HAN</td>
<td>0.878</td>
<td>0.879</td>
<td>0.599</td>
<td>0.607</td>
</tr>
<tr>
<td>DMGI</td>
<td>0.898</td>
<td>0.898</td>
<td>0.648</td>
<td>0.648</td>
</tr>
<tr>
<td>DMGI_{attn}</td>
<td>0.887</td>
<td>0.887</td>
<td>0.602</td>
<td>0.606</td>
</tr>
<tr>
<td>HDI</td>
<td><strong>0.901</strong></td>
<td><strong>0.901</strong></td>
<td><strong>0.634</strong></td>
<td><strong>0.638</strong></td>
</tr>
<tr>
<td>HDMI</td>
<td><strong>0.901</strong></td>
<td><strong>0.901</strong></td>
<td><strong>0.650</strong></td>
<td><strong>0.658</strong></td>
</tr>
</tbody>
</table>
Experiments: Node Clustering

- HDMI performs the best.
- HDI is better than most of the baselines.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ACM NMI</th>
<th>ACM Sim@5</th>
<th>IMDB NMI</th>
<th>IMDB Sim@5</th>
<th>DBLP NMI</th>
<th>DBLP Sim@5</th>
<th>Amazon NMI</th>
<th>Amazon Sim@5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metric</td>
<td>NMI</td>
<td>Sim@5</td>
<td>NMI</td>
<td>Sim@5</td>
<td>NMI</td>
<td>Sim@5</td>
<td>NMI</td>
<td>Sim@5</td>
</tr>
<tr>
<td>DeepWalk</td>
<td>0.310</td>
<td>0.710</td>
<td>0.117</td>
<td>0.490</td>
<td>0.348</td>
<td>0.629</td>
<td>0.083</td>
<td>0.726</td>
</tr>
<tr>
<td>node2vec</td>
<td>0.309</td>
<td>0.710</td>
<td>0.123</td>
<td>0.487</td>
<td>0.382</td>
<td>0.629</td>
<td>0.074</td>
<td>0.738</td>
</tr>
<tr>
<td>GCN/GAT</td>
<td>0.671</td>
<td>0.867</td>
<td>0.176</td>
<td>0.565</td>
<td>0.465</td>
<td>0.724</td>
<td>0.287</td>
<td>0.624</td>
</tr>
<tr>
<td>DGI</td>
<td>0.640</td>
<td>0.889</td>
<td>0.182</td>
<td>0.578</td>
<td>0.551</td>
<td>0.786</td>
<td>0.007</td>
<td>0.558</td>
</tr>
<tr>
<td>ANRL</td>
<td>0.515</td>
<td>0.814</td>
<td>0.163</td>
<td>0.527</td>
<td>0.332</td>
<td>0.720</td>
<td>0.166</td>
<td>0.763</td>
</tr>
<tr>
<td>CAN</td>
<td>0.504</td>
<td>0.836</td>
<td>0.074</td>
<td>0.544</td>
<td>0.323</td>
<td>0.792</td>
<td>0.001</td>
<td>0.537</td>
</tr>
<tr>
<td>DGCN</td>
<td>0.691</td>
<td>0.690</td>
<td>0.143</td>
<td>0.179</td>
<td>0.462</td>
<td>0.491</td>
<td>0.143</td>
<td>0.194</td>
</tr>
<tr>
<td>CMNA</td>
<td>0.498</td>
<td>0.363</td>
<td>0.152</td>
<td>0.069</td>
<td>0.420</td>
<td>0.511</td>
<td>0.070</td>
<td>0.435</td>
</tr>
<tr>
<td>MNE</td>
<td>0.545</td>
<td>0.791</td>
<td>0.013</td>
<td>0.482</td>
<td>0.136</td>
<td>0.711</td>
<td>0.001</td>
<td>0.395</td>
</tr>
<tr>
<td>mGCN</td>
<td>0.668</td>
<td>0.873</td>
<td>0.183</td>
<td>0.550</td>
<td>0.468</td>
<td>0.726</td>
<td>0.301</td>
<td>0.630</td>
</tr>
<tr>
<td>HAN</td>
<td>0.658</td>
<td>0.872</td>
<td>0.164</td>
<td>0.561</td>
<td>0.472</td>
<td>0.779</td>
<td>0.029</td>
<td>0.495</td>
</tr>
<tr>
<td>DMGI</td>
<td>0.687</td>
<td>0.898</td>
<td>0.196</td>
<td>0.605</td>
<td>0.409</td>
<td>0.766</td>
<td>0.425</td>
<td>0.816</td>
</tr>
<tr>
<td>DMGI_{attn}</td>
<td>0.702</td>
<td>0.901</td>
<td>0.185</td>
<td>0.586</td>
<td>0.554</td>
<td>0.798</td>
<td>0.412</td>
<td>0.825</td>
</tr>
<tr>
<td>HDI</td>
<td>0.650</td>
<td>0.900</td>
<td>0.194</td>
<td>0.605</td>
<td>0.570</td>
<td>0.799</td>
<td>0.487</td>
<td>0.856</td>
</tr>
<tr>
<td>HDMI</td>
<td>0.695</td>
<td>0.898</td>
<td>0.198</td>
<td>0.607</td>
<td>0.582</td>
<td>0.809</td>
<td>0.500</td>
<td>0.857</td>
</tr>
</tbody>
</table>
**Experiments: Ablation Study**

1. **Intrinsic (I.) & Joint (J.) MI** significantly improve over **Extrinsic (E) MI**.
2. **Fusion (HDMI)** improves over simple average pooling.
3. **Reconstruction Error (R.)** does not imply high quality embedding!

### Table 1: Performance Comparison

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ACM</th>
<th>PAP</th>
<th>IMDB</th>
<th>MAM</th>
<th>DBLP</th>
<th>PAP</th>
<th>MMM</th>
<th>PATAP</th>
<th>Amazon</th>
<th>IOI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Layer</strong></td>
<td><strong>Metric</strong></td>
<td><strong>FSP</strong></td>
<td><strong>PAP</strong></td>
<td><strong>MDM</strong></td>
<td><strong>MAM</strong></td>
<td><strong>PAP</strong></td>
<td><strong>PPP</strong></td>
<td><strong>PATAP</strong></td>
<td><strong>IVI</strong></td>
<td><strong>IBI</strong></td>
</tr>
<tr>
<td>E.</td>
<td>MaF1</td>
<td>0.663</td>
<td>0.688</td>
<td>0.573</td>
<td>0.586</td>
<td>0.804</td>
<td>0.796</td>
<td>0.240</td>
<td>0.380</td>
<td>0.569</td>
</tr>
<tr>
<td>E. + R.</td>
<td>MaF1</td>
<td>0.668</td>
<td>0.673</td>
<td>0.590</td>
<td>0.597</td>
<td>0.809</td>
<td>0.801</td>
<td>0.240</td>
<td>0.392</td>
<td>0.579</td>
</tr>
<tr>
<td>E. + I.</td>
<td>0.719</td>
<td>0.732</td>
<td>0.617</td>
<td>0.624</td>
<td>0.593</td>
<td>0.600</td>
<td>0.803</td>
<td>0.792</td>
<td>0.240</td>
<td>0.517</td>
</tr>
<tr>
<td>E. + I. + J.</td>
<td>0.742</td>
<td>0.744</td>
<td>0.626</td>
<td>0.631</td>
<td>0.600</td>
<td>0.606</td>
<td>0.812</td>
<td>0.803</td>
<td>0.241</td>
<td>0.524</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ACM</th>
<th>PAP</th>
<th>IMDB</th>
<th>MAM</th>
<th>DBLP</th>
<th>PAP</th>
<th>MMM</th>
<th>PATAP</th>
<th>Amazon</th>
<th>IOI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Layer</strong></td>
<td><strong>Metric</strong></td>
<td><strong>FSP</strong></td>
<td><strong>PAP</strong></td>
<td><strong>MDM</strong></td>
<td><strong>MAM</strong></td>
<td><strong>PAP</strong></td>
<td><strong>PPP</strong></td>
<td><strong>PATAP</strong></td>
<td><strong>IVI</strong></td>
<td><strong>IBI</strong></td>
</tr>
<tr>
<td>Metric</td>
<td>MaF1</td>
<td>MiF1</td>
<td>MaF1</td>
<td>MiF1</td>
<td>MaF1</td>
<td>MiF1</td>
<td>MaF1</td>
<td>MiF1</td>
<td>MaF1</td>
<td>MiF1</td>
</tr>
<tr>
<td>HDI</td>
<td>0.901</td>
<td>0.901</td>
<td>0.634</td>
<td>0.638</td>
<td>0.814</td>
<td>0.800</td>
<td>0.804</td>
<td>0.806</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDMI</td>
<td>0.901</td>
<td>0.901</td>
<td>0.650</td>
<td>0.658</td>
<td>0.820</td>
<td>0.811</td>
<td>0.808</td>
<td>0.812</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Additional Performance Comparison

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ACM</th>
<th>PAP</th>
<th>IMDB</th>
<th>MAM</th>
<th>DBLP</th>
<th>PAP</th>
<th>MMM</th>
<th>PATAP</th>
<th>Amazon</th>
<th>IOI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Layer</strong></td>
<td><strong>Metric</strong></td>
<td><strong>FSP</strong></td>
<td><strong>PAP</strong></td>
<td><strong>MDM</strong></td>
<td><strong>MAM</strong></td>
<td><strong>PAP</strong></td>
<td><strong>PPP</strong></td>
<td><strong>PATAP</strong></td>
<td><strong>IVI</strong></td>
<td><strong>IBI</strong></td>
</tr>
<tr>
<td>E.</td>
<td>NMI</td>
<td>0.526</td>
<td>0.698</td>
<td>0.651</td>
<td>0.872</td>
<td>0.145</td>
<td>0.549</td>
<td>0.089</td>
<td>0.495</td>
<td>0.547</td>
</tr>
<tr>
<td>E. + R.</td>
<td>NMI</td>
<td>0.525</td>
<td>0.728</td>
<td>0.659</td>
<td>0.874</td>
<td>0.150</td>
<td>0.552</td>
<td>0.079</td>
<td>0.490</td>
<td>0.564</td>
</tr>
<tr>
<td>E. + I.</td>
<td>0.527</td>
<td>0.708</td>
<td>0.656</td>
<td>0.882</td>
<td>0.193</td>
<td>0.595</td>
<td>0.143</td>
<td>0.527</td>
<td>0.569</td>
<td>0.802</td>
</tr>
<tr>
<td>E. + I. + J.</td>
<td>0.528</td>
<td>0.716</td>
<td>0.662</td>
<td>0.886</td>
<td>0.194</td>
<td>0.592</td>
<td>0.143</td>
<td>0.527</td>
<td>0.562</td>
<td>0.805</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>ACM</th>
<th>PAP</th>
<th>IMDB</th>
<th>MAM</th>
<th>DBLP</th>
<th>PAP</th>
<th>MMM</th>
<th>PATAP</th>
<th>Amazon</th>
<th>IOI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Layer</strong></td>
<td><strong>Metric</strong></td>
<td><strong>FSP</strong></td>
<td><strong>PAP</strong></td>
<td><strong>MDM</strong></td>
<td><strong>MAM</strong></td>
<td><strong>PAP</strong></td>
<td><strong>PPP</strong></td>
<td><strong>PATAP</strong></td>
<td><strong>IVI</strong></td>
<td><strong>IBI</strong></td>
</tr>
<tr>
<td>Metric</td>
<td>NMI</td>
<td>Sim@5</td>
<td>NMI</td>
<td>Sim@5</td>
<td>NMI</td>
<td>Sim@5</td>
<td>NMI</td>
<td>Sim@5</td>
<td>NMI</td>
<td>Sim@5</td>
</tr>
<tr>
<td>HDI</td>
<td>0.650</td>
<td>0.900</td>
<td>0.194</td>
<td>0.605</td>
<td>0.570</td>
<td>0.799</td>
<td>0.487</td>
<td>0.856</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HDMI</td>
<td>0.695</td>
<td>0.898</td>
<td>0.198</td>
<td>0.607</td>
<td>0.582</td>
<td>0.809</td>
<td>0.500</td>
<td>0.857</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Experiments: T-SNE Visualization

• Comparison of different signals: IOI layer of the Amazon network
  • Intrinsic (I.) and Joint (J.) MI improve the quality.
  • Reconstruction error (R.) does not significantly improve the quality.

• Comparison of fusion mechanism: ACM network
  • Proposed fusion module is better than average pooling.

  Different Layers
  (a) PSP  (b) PAP  (c) Average  (d) Fusion

  Different Fusing Methods

Better
Experiments: Attention Scores

- Appropriate attention scores are assigned to different layers.  
  - Higher F1 scores -> Higher attention scores.
Outline

• Introduction
• Preliminary
• Methodology
• Experiments

➢ Conclusion
Conclusion

- MI based self-supervised learning for graphs
  - **Challenge 1**: Jointly capture extrinsic and intrinsic information for graphs.
    - **Solution 1**: High-order Deep Infomax (HDI)
  - **Challenge 2**: Extend HDI to multiplex graphs.
    - **Solution 2**: Attention based fusion
      - High-order Deep Multiplex Infomax (HDMI)

- **Results:**
  - HDI significantly improves the quality of embeddings.
  - HDMI further improves HDI.

(a) E.  (d) E. + I. + J.  (a) PSP  (b) PAP  (d) Fusion
Thank you!