BRIGHT: A Bridging Algorithm for Network Alignment

Yuchen Yan (UIUC)
Si Zhang (UIUC)
Hanghang Tong (UIUC)
Outline

• Introduction
• Theoretical Analysis
• Proposed Model
• Experimental Results
• Conclusions
Network Alignment

• Networks are often multi-sourced
• To find node correspondence across networks
Prob. Def.: Semi-Supervised Attributed Network Alignment

- **Given:** (1) two attributed networks \( \mathcal{G}_1 = \{A_1, X_1\} \), \( \mathcal{G}_2 = \{A_2, X_2\} \); (2) a set of anchor node pairs \( L \).
- **Output:** an \( n_2 \times n_1 \) alignment/similarity \( S \).
- **Scenario variants:**
  - Semi-supervised plain network alignment (without \( X_1, X_2 \))
  - Unsupervised attributed network alignment (without \( L \))
Existing Methods: Limitation #1

• Consistency optimization based methods
  —Consistency assumption violation:
    (1) Attribute change
    (2) Local topology change

Existing Methods: Limitation #2

• Embedding based methods
  — Introduce the space disparity issue

Outline

• Introduction ✓
• Theoretical Analysis
• Proposed Model
• Experimental Results
• Conclusions
Theoretical Analysis #1

Main Claim/Insight: Consistency optimization based methods are essentially random walk propagation of anchor links.

• Drawbacks
  — Exactly same steps
  — Equal weight for all anchor links

\[
O_c(S) = \alpha \sum_{a,b,x,y} \left[ \frac{S(x, a)}{\sqrt{d_2(x)d_1(a)}} - \frac{S(y, b)}{\sqrt{d_2(y)d_1(b)}} \right]^2 A_1(a, b)A_2(x, y) + (1 - \alpha)\|S - H\|_F^2
\]

1. Topology Consistency
2. Known Anchor Links

\[
s = (1 - \alpha)(I - \alpha\hat{W})^{-1}h
\]

\[
W = A_1 \otimes A_2
\]

\[
s(i) = (1 - \alpha) \sum_{j=0}^{n_2 \times n_1 - 1} \sum_{t=0}^{\infty} \alpha^t \hat{W}^t(i, j) \mathbb{1}(h(j))
\]


Main Claim/Insight: Embedding based methods relax the objective function of consistency optimization based methods.

- Drawback
  - Space disparity

In a simple $K$-regular graph:

\[ O_{\text{in}} = \sum_a (d(a, b) - d(a, c)) \]
\[ O_{\text{cross}} = \sum_{(l_1, l_2) \in \mathcal{L}} d(l_1, l_2) \]

Consistency method loss

Embedding method loss

Relaxation relation

\[ d(b, y) \leq d(b, a) + d(a, x) + d(x, y) \]


Outline

• Introduction
• Theoretical Analysis
• Proposed Model
• Experimental Results
• Conclusions
Key Idea #1: RWR for Flexible Propagation

- Drawbacks for consistency optimization based methods
  - Exactly same steps
    Random walk with restart allows **restart**.
  - Equal weight for all anchor links
    Linear layer trains **different weights**.
Key Idea #2: Build a Unified Space

- Drawbacks for embedding based methods
  - Space disparity
    - (1) Anchor links as basis
    - (2) Weight sharing
Part #1: BRIGHT-U (Plain Network)

- RWR from anchor links
  \[ r_{l_1} = (1 - \beta)\hat{W}_1 r_{l_1} + \beta e_{l_1} \quad \hat{W}_1 = (D^{-1}A_1)^T \]
  \[ r_{l_1} = \beta(I - (1 - \beta)\hat{W}_1)^{-1}e_{l_1} \]
- Put all \( r_{l_i} \) together as RWR embedding matrix \( \hat{W}_{R_1} \)
- Use a shared linear layer to adjust anchor link weights
Part #2: BRIGHT-A (Attributed Network)

- Compute RWR embedding same as BRIGHT-U
- Use a shared two-layer GCN to capture attribute
- Combine RWR embedding and GCN embedding

\[ E_1 = \text{COMBINE}([E_1^{\text{RWR}} \left| X_1^2]) \]
Part #3: Model Training

- Ranking loss

\[ J_i = \frac{1}{|L|} \sum_{l \in L} \frac{1}{|U_l|} \sum_{u \in U_l} \max\{0, \gamma + d(l_1, l_2) - d(l_i, u)\} \]

\[ J = J_1 + J_2 \]

- Advanced negative sampling
  — Sort-then-select

- Construct the alignment matrix

\[ S(x, a) = e^{-d(x, a)} \]
Outline

• Introduction
• Theoretical Analysis
• Proposed Model ✓
• Experimental Results
• Conclusions
Experimental Setup

• Datasets

<table>
<thead>
<tr>
<th>Categories</th>
<th>Networks</th>
<th># of Nodes</th>
<th># of Edges</th>
<th># of Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plain Networks</td>
<td>Foursquare</td>
<td>5,313</td>
<td>54,233</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>Twitter</td>
<td>5,120</td>
<td>130,575</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>ACM</td>
<td>9,916</td>
<td>44,808</td>
<td>—</td>
</tr>
<tr>
<td></td>
<td>DBLP</td>
<td>9,872</td>
<td>39,561</td>
<td>—</td>
</tr>
<tr>
<td>Attributed</td>
<td>ACM(A)</td>
<td>9,916</td>
<td>44,808</td>
<td>17</td>
</tr>
<tr>
<td>Networks</td>
<td>DBLP(A)</td>
<td>9,872</td>
<td>39,561</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Cora-1</td>
<td>2,708</td>
<td>5,806</td>
<td>1,433</td>
</tr>
<tr>
<td></td>
<td>Cora-2</td>
<td>2,708</td>
<td>4,547</td>
<td>1,433</td>
</tr>
</tbody>
</table>

• Metrics: Hit@K, MRR

• Baseline methods:
  – Plain network: CrossMNA, IONE, FINAL-P
  – Attributed network: REGAL, FINAL-N, NetTrans

Experimental Results - Alignment

<table>
<thead>
<tr>
<th>Plain Task</th>
<th>DBLP vs. ACM</th>
<th>Foursquare vs. Twitter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Hit@1</td>
<td>Hit@10</td>
</tr>
<tr>
<td>CrossMNA</td>
<td>7.90%</td>
<td>62.53%</td>
</tr>
<tr>
<td>IONE</td>
<td>30.91%</td>
<td>74.25%</td>
</tr>
<tr>
<td>FINAL-P</td>
<td>19.49%</td>
<td>68.75%</td>
</tr>
<tr>
<td>BRIGHT-U</td>
<td>40.45%</td>
<td>81.26%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Attributed Task</th>
<th>DBLP(A) vs. ACM(A)</th>
<th>Cora-1 vs. Cora-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Hit@1</td>
<td>Hit@10</td>
</tr>
<tr>
<td>REGAL</td>
<td>36.26%</td>
<td>60.36%</td>
</tr>
<tr>
<td>FINAL-N</td>
<td>38.18%</td>
<td>79.74%</td>
</tr>
<tr>
<td>NetTrans</td>
<td>11.84%</td>
<td>84.11%</td>
</tr>
<tr>
<td>BRIGHT-A</td>
<td>45.26%</td>
<td>86.76%</td>
</tr>
</tbody>
</table>

**Observation:**
(1) An advantage of 10% in Hit@1 on DBLP vs. ACM;
(2) 99% in Hit@10 on Cora.
Experimental Results - Ablation Study

### Ablation Study for BRIGHT-U

<table>
<thead>
<tr>
<th>plain Task</th>
<th><em>DBLP vs. ACM</em></th>
<th><em>Foursquare vs. Twitter</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Hit@10  MRR</td>
<td>Hit@10  MRR</td>
</tr>
<tr>
<td>BRIGHT-U(SPD)</td>
<td>75.75% 48.78%</td>
<td>6.06% 3.07%</td>
</tr>
<tr>
<td>BRIGHT-U</td>
<td>81.26% 53.85%</td>
<td>25.24% 13.04%</td>
</tr>
</tbody>
</table>

### Ablation Study for BRIGHT-A

<table>
<thead>
<tr>
<th>Attributed Task</th>
<th><em>DBLP(A) vs. ACM(A)</em></th>
<th><em>Cora-1 vs. Cora-2</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Metrics</td>
<td>Hit@10  MRR</td>
<td>Hit@10  MRR</td>
</tr>
<tr>
<td>BRIGHT-A(-RWR)</td>
<td>79.43% 51.61%</td>
<td>99.08% 90.12%</td>
</tr>
<tr>
<td>BRIGHT-A(-RWR:3500)</td>
<td>84.31% 58.01%</td>
<td>99.08% 90.12%</td>
</tr>
<tr>
<td>BRIGHT-A</td>
<td>86.76% 59.87%</td>
<td>99.08% 90.41%</td>
</tr>
</tbody>
</table>

**Observation:**
(1) RWR module performs better than SPD;
(2) Attribute plays an important role in BRIGHT-A.
Experimental Results-Small Training Ratio

Observation: (1) Perform well under small training ratio; (2) Avoid the space disparity issue better.
Experimental Results - Unsupervised Setting

Observation: Good generalization to unsupervised setting.
Outline
• Introduction
• Theoretical Analysis
• Proposed Model
• Experimental Results ✔
• Conclusions
Conclusions

Problem: Network Alignment

Solution:

• BRIGHT-U
  ▪ RWR for flexible propagation
  ▪ Anchor link as basis
  ▪ Weight sharing

• BRIGHT-A
  ▪ Shared GCN to capture attribute

Results:

• Outperform all baselines
• Perform well with small training ratio
• Generalization to unsupervised setting