Asymmetric Transitivity Preserving Graph Embedding

Mingdong Ou  Tsinghua U  Peng Cui  Tsinghua U  Jian Pei  Simon Fraser U  Ziwei Zhang  Tsinghua U  Wenwu Zhu  Tsinghua U
Graph Embedding

Graph Data
- Social Network
- Citation/Collaboration
- Image/Video Tag/Caption
- Web Hyperlink
- Etc.

Graph Data Representation

Application
- Similarity Measure
- Link Prediction
- Clustering/Classification
- Visualization
- Etc.
Graph Embedding

- **Graph Embedding**: 

  ![Graph](image)

  Input graph/network  \rightarrow  Low dimensional space

- **Advantages:**
  - Fast computation of nodes similarity
  - Utilization of vector-based machine learning techniques
  - Facilitating parallel computing
Existing graph embedding methods

- **Existing work:**
  - LINE(Tang J, et al. WWW 2015): explicitly preserves first-order and second-order proximity
  - DeepWalk(Perozzi B, et al. KDD 2014): random walk on graphs + SkipGram Model from NLP
  - GraRep(Cao S, et al. CIKM 2015)
  - SDNE(Daixin W, et al. KDD 2016)

- Most methods focus on undirected graph
Directed Graph

Critical property in directed graph: **Asymmetric Transitivity**

- Transitivity is **Asymmetric** in directed graph:

  ![Diagram of directed graph with nodes A, B, and C]

- Key in graph inference.

- Data Validation: Tencent Weibo and Twitter

- **Asymmetric transitivity** is important!
Asymmetric Transitivity $\rightarrow$ Graph Embedding

- Challenge: incorporate asymmetric transitivity in graph embedding
- Problem: metric space is symmetric

![Diagram showing conflict between asymmetric transitivity and metric space](image)
Asymmetric Transitivity → Graph Embedding

- Directed graph embedding: use two vectors to represent each node
  - LINE (Tang J, et al. WWW 2015): second-order proximity is directed
  - PPE (Song H H, et al. SIGCOMM 2009): using sub-block of the proximity matrix

Asymmetric: YES; Transitive: NO!
Similarity metric with asymmetric transitivity

- Asymmetric transitivity:
  - Asymmetry: not symmetric in directed graph
  - Transitivity:
    - More directed paths, larger similarity
    - Shorter paths, larger similarity
  - Compare A -> C similarity:

High-order Proximity!
(E.g. Katz, Rooted PageRank)
**High-Order Proximity**

- Solution: directly model transitivity using **high-order proximity**
  - Example: Katz Index
  - $A$: adjacency matrix, $\beta$: decaying constant
  - $S^{Katz} = \sum_{l=1}^{+\infty} (\beta \cdot A)^l$
  - Example: when $\beta = 1$
Preserve high-order proximity embedding

Naive Solution: SVD?

\[ S^{Katz} = \sum_{l=1}^{+\infty} (\beta \cdot A)^l \]

\[ S^{Katz} = U^s \times U^t \]

- Time and space complexity: \( O(N^3) \), \( N \): node number
High-Order Proximity: a general form

- Katz Index:
  \[ S^{Katz} = \sum_{l=1}^{+\infty} (\beta \cdot A)^l = (I - \beta \cdot A)^{-1} \cdot (\beta \cdot A) \]

- General Form
  \[ M_g^{-1} \cdot M_l \]

where \( M_g, M_l \) are polynomial of adjacency matrix or its variants

General Formulation for High-Order Proximity measurements

<table>
<thead>
<tr>
<th>Proximity Measurement</th>
<th>( M_g )</th>
<th>( M_l )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katz</td>
<td>( I - \beta \cdot A )</td>
<td>( \beta \cdot A )</td>
</tr>
<tr>
<td>Personalized Pagerank</td>
<td>( I - \alpha P )</td>
<td>( (1 - \alpha) \cdot I )</td>
</tr>
<tr>
<td>Common neighbors</td>
<td>( I )</td>
<td>( A^2 )</td>
</tr>
<tr>
<td>Adamic-Adar</td>
<td>( I )</td>
<td>( A \cdot D \cdot A )</td>
</tr>
</tbody>
</table>
The Power of General Form

$$\min_{U_s, U_t} \|S - U_s \cdot U_t^T\|_F^2$$

$$S = M_g^{-1} \cdot M_l$$

Generalized SVD (Singular Value Decomposition) theorem

If we have the singular value decomposition of the general formulation

$$M_g^{-1} \cdot M_l = V^s \Sigma V^t\top$$

where \(V^t\) and \(V^s\) are two orthogonal matrices,

$$\Sigma = \text{diag}(\sigma_1, \sigma_2, \cdots, \sigma_N)$$

Then, there exists a nonsingular matrix \(X\) and two diagonal matrices, i.e. \(\Sigma^l\) and \(\Sigma^g\), satisfying that

$$V^t\top M^l\top X = \Sigma^l \quad V^s\top M^g\top X = \Sigma^g$$

, where

$$\Sigma^l = \text{diag}(\sigma^l_1, \sigma^l_2, \cdots, \sigma^l_N) \quad \sigma^l_1 \geq \sigma^l_2 \geq \cdots \geq \sigma^l_K \geq 0$$

$$\Sigma^g = \text{diag}(\sigma^g_1, \sigma^g_2, \cdots, \sigma^g_N) \quad 0 \leq \sigma^g_1 \leq \sigma^g_2 \leq \cdots \leq \sigma^g_K$$

$$\forall i \quad \sigma_i^{l2} + \sigma_i^{g2} = 1$$
The Power of General Form

\[
\min_{U_s, U_t} \| S - U_s \cdot U_t^T \|^2_F
\]

\[
S = M_g^{-1} \cdot M_l
\]

- Generalized SVD: decompose $S$ **without** actually calculating it
- JDGSVD: Time Complexity

\[
O(K^2L \cdot m)
\]

- Embedding Dimension (constant)
- Iteration (constant)
- Edge number

- **Linear complexity** w.r.t. the volume of data (i.e. edge number)

--> **Scalable** algorithm, suitable for **large-scale** data
Theoretical Guarantee

Approximation Error Upper Bound:

**Theorem 2.** Given the proximity matrix, $S$, of a directed graph, and the embedding vectors, $U^s$ and $U^t$, learned by HOPPE. Then the approximation error is

$$\|S - U^s \cdot U^t\|_F^2 = \sum_{i=K+1}^{N} \sigma_i^2$$

and the relative approximation error is:

$$\frac{\|S - U^s \cdot U^t\|_F^2}{\|S\|_F^2} = \frac{\sum_{i=K+1}^{N} \sigma_i^2}{\sum_{i=1}^{N} \sigma_i^2}$$

(22)

where $\{\sigma_i\}$ are the singular values of $S$ in descend order.
HOPE: HIGH-ORDER PROXIMITY PRESERVED EMBEDDING

**HOPE:** High-Order Proximity preserved Embedding

Algorithm framework:

---

**Algorithm 1** High-order Proximity preserved Embedding

**Require:** adjacency matrix $A$, embedding dimension $K$, parameters of high-order proximity measurement $\theta$.

**Ensure:** embedding source vectors $U^s$ and target vectors $U^t$.

1: calculate $M_g$ and $M_l$.
2: perform JDGSVD with $M_g$ and $M_l$, and obtain the generalized singular values $\{\sigma_1^l, \cdots, \sigma_K^l\}$ and $\{\sigma_1^g, \cdots, \sigma_K^g\}$; and the corresponding singular vectors, $\{v^s_1, \cdots, v^s_K\}$ and $\{v^t_1, \cdots, v^t_K\}$.
3: calculate singular values $\{\sigma_1, \cdots, \sigma_K\}$ according to Equation (21).
4: calculate embedding matrices $U^s$ and $U^t$ according to Equation (19) and (20).
Experiment Setting: Datasets

Datasets:

- Synthetic (Syn): generate using Forest Fire Model
- Cora\(^1\): citation network of academic papers
- SN-Twitter\(^2\): Twitter Social Network
- SN-TWeibo\(^3\): Tencent Weibo Social Network

Statistics of datasets. \(|V|\) denotes the number of vertexes and \(|E|\) denotes the number of edges.

<table>
<thead>
<tr>
<th></th>
<th>Syn</th>
<th>Cora</th>
<th>SN-Twitter</th>
<th>SN-TWeibo</th>
</tr>
</thead>
<tbody>
<tr>
<td>(</td>
<td>V</td>
<td>)</td>
<td>10,000</td>
<td>23166</td>
</tr>
<tr>
<td>(</td>
<td>E</td>
<td>)</td>
<td>144,555</td>
<td>91500</td>
</tr>
</tbody>
</table>

\(^1\)http://konect.uni-koblenz.de/networks/subelj_cora
\(^2\)http://konect.uni-koblenz.de/networks/munmun_twitter_social
\(^3\)http://www.kddcup2012.org/c/kddcup2012-track1/data
Experiment Setting: Task

- Approximation accuracy
  - High-order proximity approximation: how well can embedded vectors approximate high-order proximity

- Reconstruction
  - Graph Reconstruction: how well can embedded vectors reconstruct training sets

- Inference:
  - Link Prediction: how well can embedded vectors predict missing edges
  - Vertex Recommendation: how well can embedded vectors recommend vertices for each node
Experiment Setting: Baseline

- **Graph embedding**
  - **PPE**: approximate high-order proximity by selecting landmarks and using sub-block of the proximity matrix
  - **LINE**: preserves first-order and second-order proximity, called LINE1 and LINE2 respectively
  - **DeepWalk**: random walk on graphs + SkipGram Model

- **Task Specific:**
  - **Common Neighbors**: used for link prediction and vertex recommendation task
  - **Adamic-Adar**: used for link prediction and vertex recommendation task
Experiment result: high-order Proximity Approximation

Conclusion: HOPE achieves much smaller RMSE error

-> generalized SVD achieves a good approximation
Conclusion: HOPE successfully capture the information of training sets
Conclusion: HOPE has good inference ability

-> based on asymmetric transitivity
Experiment result: Vertex Recommendation

<table>
<thead>
<tr>
<th>Method</th>
<th>SN-TWebio</th>
<th></th>
<th></th>
<th>SN-Twitter</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAP@10</td>
<td>MAP@50</td>
<td>MAP@100</td>
<td>MAP@10</td>
<td>MAP@50</td>
<td>MAP@100</td>
</tr>
<tr>
<td>HOPE</td>
<td>0.2295</td>
<td>0.1869</td>
<td>0.169</td>
<td>0.1000</td>
<td>0.0881</td>
<td>0.0766</td>
</tr>
<tr>
<td>PPE</td>
<td>0.0928</td>
<td>0.0845</td>
<td>0.077</td>
<td>0.0061</td>
<td>0.0077</td>
<td>0.0081</td>
</tr>
<tr>
<td>LINE1</td>
<td>0</td>
<td>0</td>
<td>0.005</td>
<td>0.0209</td>
<td>0.0221</td>
<td>0.0221</td>
</tr>
<tr>
<td>LINE2</td>
<td>0.051</td>
<td>0.051</td>
<td>0.048</td>
<td>0.0044</td>
<td>0.0043</td>
<td>0.0035</td>
</tr>
<tr>
<td>DeepWalk</td>
<td>0.0635</td>
<td>0.0583</td>
<td>0.004</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.001</td>
</tr>
<tr>
<td>Common Neighbors</td>
<td>0.1217</td>
<td>0.1031</td>
<td>0.155</td>
<td>0.0394</td>
<td>0.0379</td>
<td>0.0369</td>
</tr>
<tr>
<td>Adamic-Adar</td>
<td>0.1173</td>
<td>0.0990</td>
<td>0.156</td>
<td>0.0455</td>
<td>0.0442</td>
<td>0.0423</td>
</tr>
</tbody>
</table>

Conclusion: HOPE significantly outperforms all state-of-the-art baselines on all these experiments.
Conclusion

- Directed graph embedding:
  - **High-order Proximity → Asymmetric Transitivity**
  - Derivation of a *general form* for high-order proximities, and solution with *generalized SVD*
    - Covering multiple commonly used high order proximity
    - **Time complexity linear** w.r.t. graph size
    - **Theoretically guaranteed** accuracy.

- Extensive experiments on several datasets
  - Outperforming all baselines in various applications.
  - *x4/x10* smaller approximation error for Katz
  - *+50% improvement* in reconstruction and inference
Thanks!

Ziwei Zhang, Tsinghua University
zw-zhang16@mails.tsinghua.edu.com
https://cn.linkedin.com/in/zhangziwei