PRUNE: Preserving Proximity and Global Ranking for Network Embedding (Supplementary Material)

Yi-An Lai \( ^\dagger\)†
National Taiwan University
b99202031@ntu.edu.tw

Chin-Chi Hsu \( ^\ddagger\)
Academia Sinica
chinchi@iis.sinica.edu.tw

Wen-Hao Chen \( ^\ast\)
National Taiwan University
b02902023@ntu.edu.tw

Mi-Yen Yeh \( ^\dagger\)
Academia Sinica
miyen@iis.sinica.edu.tw

Shou-De Lin \( ^\ast\)
National Taiwan University
sdlin@csie.ntu.edu.tw

1 Notation introduction

Table 1: Commonly used notations

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( G = (V, E) )</td>
<td>Input directed network (or graph)</td>
</tr>
<tr>
<td>( A \in {0,1}^{N \times N} )</td>
<td>Adjacency matrix of network ( G )</td>
</tr>
<tr>
<td>( V )</td>
<td>Set of nodes or vertices</td>
</tr>
<tr>
<td>( E = {(i,j) : a_{ij} = 1} )</td>
<td>Set of links or edges</td>
</tr>
<tr>
<td>( N =</td>
<td>V</td>
</tr>
<tr>
<td>( M =</td>
<td>E</td>
</tr>
<tr>
<td>( P_i )</td>
<td>Set of direct predecessors of node ( i )</td>
</tr>
<tr>
<td>( S_i )</td>
<td>Set of direct successors of node ( i )</td>
</tr>
<tr>
<td>( m_i =</td>
<td>P_i</td>
</tr>
<tr>
<td>( n_i =</td>
<td>S_i</td>
</tr>
<tr>
<td>( z_i \in [0, \infty)^D )</td>
<td>Latent ( D )-community distribution vector of node ( i )</td>
</tr>
<tr>
<td>( W \in [0, \infty)^{D \times D} )</td>
<td>Shared matrix of community interactions</td>
</tr>
<tr>
<td>( \pi_i \geq 0 )</td>
<td>Global ranking score of node ( i )</td>
</tr>
</tbody>
</table>

\( ^\dagger \)Department of Computer Science and Information Engineering
\( ^\ddagger \)Institute of Information Science
\( ^\ast \)These authors contributed equally to this paper.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
2 Proof for the closed-form solution of binary classification

The objective function of our binary classification is shown below:
\[
\arg\max_{z,W} E_{(i,j) \in E} \left[ \log \sigma(z_i^\top W z_j) \right] + \alpha E_{(i,j) \in E} \left[ \log (1 - \sigma(z_i^\top W z_k)) \right] \\
= E_{i,j \in S_i} \left[ \log \sigma(z_i^\top W z_j) \right] + \alpha E_{k \in V} \left[ \log (1 - \sigma(z_i^\top W z_k)) \right] \\
= \sum_{i \in V} \sum_{j \in S_i} \sigma \left( p_s(i) p_t(j|i) \log \sigma(z_i^\top W z_j) \right) + \alpha \sum_{i \in V} \sum_{k \in V} \sigma \left( p_s(i) p_t(k) \log (1 - \sigma(z_i^\top W z_k)) \right) \\
= \sum_{i \in V} \sum_{j \in S_i} \frac{n_i}{M} \frac{1}{n_j} \log \sigma(z_i^\top W z_j) + \alpha \sum_{i \in V} \sum_{k \in V} \frac{n_k}{M} \frac{1}{M} \log (1 - \sigma(z_i^\top W z_k)).
\]

Given source node \( i \), one of linked target node \( j \in S_i \) enjoys a conditional distribution proportional to \( \frac{1}{n_j} \). Since \( S_i \subseteq V \) implies \( k \) including \( j \), for specific positive example \((i, j)\), we have:
\[
\arg\max_{z,W} L_{ij} = \frac{1}{M} \log \sigma(z_i^\top W z_j) + \alpha \frac{n_k}{M} \frac{1}{M} \log (1 - \sigma(z_i^\top W z_k)).
\]

Now let \( y_{ij} = z_i^\top W z_j \). We first derive the closed-form solution of zero first-order derivative over \( \sigma(y_{ij}) \):
\[
\frac{\partial L_{ij}}{\partial \sigma(y_{ij})} = \frac{1}{M} \frac{1}{M} \frac{1}{M} \frac{1}{M} \frac{1}{M} \frac{1}{M} \frac{1}{M} \frac{1}{M} \sigma(y_{ij}) = 0
\]
\[
\implies \sigma(y_{ij}) = \frac{\frac{1}{M} + \alpha \frac{n_k}{M} \frac{1}{M}}{M + \alpha n_i m_j} = \frac{M}{M + \alpha n_i m_j}.
\]

Next We obtain \( y_{ij} \) after calculations:
\[
\frac{1}{1 + e^{-y_{ij}}} = \frac{M}{M + \alpha n_i m_j} \\
\implies y_{ij} = \log \frac{M}{\alpha n_i m_j} = \log \frac{\frac{M}{M}}{\alpha \frac{n_i m_j}{M}} = \log \frac{\frac{M}{M}}{\alpha \frac{n_i m_j}{M}} = \log \frac{p_s(i|j)}{p_s(i|j)} - \log \alpha.
\]

3 Proof for matrix tri-factorization supporting the second-order proximity

The second-order proximity implies high similarity between two representation vectors \( z_i, z_j \) if nodes \( i, j \) have similar sets of direct predecessors or direct successors.

Consider the non-missing entries of the \( i \)-th and \( j \)-th column \( a_i^{PMI}, a_j^{PMI} \) in our derived PMI matrix \( A^{PMI} \). Since all the non-missing entries are in link set \( E \), the two columns represent the sets of direct predecessors of node \( i \) and \( j \) where the links are weighted by PMI. Based on our matrix tri-factorization \( Z^\top W Z \approx A^{PMI} \), we have:
\[
a_i^{PMI} \approx Z^\top W z_i, \\
\]
\[
a_j^{PMI} \approx Z^\top W z_j
\]
where \( z_i \) is the \( i \)-th column of representation matrix \( Z \). As the predecessor sets are similar \( a_i^{PMI} \approx a_j^{PMI} \), then their corresponding representation vector must be similar \( z_i \approx z_j \) due to the same weight matrix \( Z^\top W \). Similarly, when modeling the matrix tri-factorization for the rows in \( A^{PMI} \), we also obtain \( z_i \approx z_j \) if nodes \( i, j \) have similar successor sets.
4 Proof for the expectation of community interactions

Let $W \in [0, \infty)^{D \times D}$ be the community interaction matrix where each entry $w_{cd}$ denotes the expected number of interactions from community $c$ to $d$. $c = d$ implies the number of internal interactions within a community. We assume that the existence of link $(i, j)$ is determined by the expected value of $W$ with community distributions of $i$ and $j$:

$$E_{(i,j)}[W] = \sum_{c=1}^{D} \sum_{d=1}^{D} \Pr(i \in C_c, j \in C_d) w_{cd}$$

where $C_c$ is the set of nodes in community $c$. Let $z_i$ be an unnormalized distribution vector where each dimension $0 \leq z_{ic} \propto \Pr(i \in C_c)$. Under the independence assumption between $\Pr(i \in C_c)$ and $\Pr(j \in C_d)$, we have:

$$\sum_{c=1}^{D} \sum_{d=1}^{D} \Pr(i \in C_c, j \in C_d) w_{cd} = \sum_{c=1}^{D} \sum_{d=1}^{D} \Pr(i \in C_c) \Pr(j \in C_d) w_{cd}$$

$$\propto \sum_{c=1}^{D} \sum_{d=1}^{D} z_{ic} z_{jd} w_{cd}$$

$$= z_i^\top W z_j.$$

5 Proof for community interactions following Poisson distribution

Based on the proof in the previous section, for specific link $(i, j)$, the expected number of interactions from community $c$ to $d$ is

$$\Pr(i \in C_c) \Pr(j \in C_d) w_{cd} \propto z_{ic} z_{jd} w_{cd}.$$

Here we model discrete random variable $X_{cd}^{(i,j)}$ as the number of interactions from community $c$ to $d$ for link $(i, j)$, following Poisson distribution $X_{cd}^{(i,j)} \sim \mathcal{P}(\mu = z_{ic} z_{jd} w_{cd})$. Using the properties of Poisson distribution, the overall number of interactions among community pairs is

$$X^{(i,j)} = \sum_{c=1}^{D} \sum_{d=1}^{D} X_{cd}^{(i,j)} \sim \mathcal{P} \left( \mu = \sum_{c=1}^{D} \sum_{d=1}^{D} z_{ic} z_{jd} w_{cd} = z_i^\top W z_j \right).$$

Assume that node $i$ and $j$ belong to at least one community. Link $(i, j)$ exists due to at least one interaction between the communities that $i$ and $j$ belong to, which is

$$\mathcal{P}(X^{(i,j)} > 0) = 1 - \mathcal{P}(X^{(i,j)} = 0) = 1 - \exp(-z_i^\top W z_j).$$
6 Proof for PageRank upper-bound objective function

Let \( P_j \) be the set of direct predecessors of node \( j \), and \( n_i \) be the out-degree of node \( i \). Then we have:

\[
\arg \min_\pi \sum_{j \in V} \left( \sum_{i \in P_j} \frac{\pi_i}{n_i} - \pi_j \right)^2 = \sum_{j \in V} \left( \left( \sum_{i \in P_j} \frac{\pi_i}{n_i} \right)^2 - 2 \pi_j \sum_{i \in P_j} \frac{\pi_i}{n_i} + \pi_j^2 \right) \\
\leq \sum_{j \in V} \left( \sum_{i \in P_j} 1^2 \right) \left( \sum_{i \in P_j} \left( \frac{\pi_i}{n_i} \right)^2 \right) - 2 \pi_j \sum_{i \in P_j} \frac{\pi_i}{n_i} + \pi_j^2 \\
= \sum_{(i,j) \in E} \left( \frac{m_j}{n_i} - \frac{\pi_j}{m_j} \right)^2.
\]

Since \( \left( \sum_{i \in P_j} 1^2 \right) \left( \sum_{i \in P_j} \left( \frac{\pi_i}{n_i} \right)^2 \right) \geq 0 \), we constrain \( \pi_i \geq 0 \) for all node \( i \) to make the upper bound tighter.

7 Proof for PageRank sufficient condition

For each node \( j \in V \), let \( P_j \) be the set of direct predecessors of node \( j \). We denote node \( i \in P_j \). Then for each node \( j \), we show a sufficient condition:

\[
\frac{\pi_i}{n_i} = \frac{\pi_j}{m_j} \quad \forall \; i \in P_j, \; j \in V
\]

where \( m_j = |P_j| \), \( n_i \) is respectively the in-degree of node \( j \) and the out-degree of node \( i \). Now we calculate the sum of the left-hand-side for all the direct predecessors \( i \) of each node \( j \):

\[
\sum_{i \in P_j} \frac{\pi_i}{n_i} = \sum_{i \in P_j} \frac{\pi_j}{m_j} \\
= \frac{1}{m_j} \sum_{i \in P_j} \pi_j \\
= \frac{1}{m_j} m_j \pi_j \\
= \pi_j \quad \forall \; j \in V.
\]

The equation is just the PageRank assumption: \( \sum_{i \in P_j} \frac{\pi_i}{n_i} = \pi_j \forall \; j \in V \) (here we omit the damping factor).