Table 1: From [5], the matrices that are implicitly approximated and factorized by DeepWalk [4], LINE [6], and node2vec [3].

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Matrix Factorization</th>
</tr>
</thead>
<tbody>
<tr>
<td>DeepWalk</td>
<td>( \log</td>
</tr>
<tr>
<td>LINE</td>
<td>( \log</td>
</tr>
<tr>
<td>node2vec</td>
<td>( \log</td>
</tr>
</tbody>
</table>

Random-Walk Matrix-Polynomial Sparsification

Definition Suppose \( G = (V,E,A) \) and \( G = (V,E,A) \) are two weighted undirected networks. Let \( L = D - A \), \( L = D - A \) be their Laplacian matrices, respectively. We define \( G \) and \( G \) as \( (1+\epsilon) \)-spectrally similar if
\[
\forall x \in \mathbb{R}^n, (1+\epsilon) x^T L x \leq x^T L x \leq (1+\epsilon) x^T L x .
\]

Theorem [1, 2] For random-walk matrix polynomial
\[
L = D - \sum_{i=1}^T \alpha_i D^{i} A^{i-1}, \quad \forall \alpha_i > 0, \quad \sum_{i=1}^T \alpha_i = 1
\]
and \( \alpha_i > 0 \), one can construct, in time \( O(T \epsilon m^{-2} \log n) \), \( (1+\epsilon) \)-spectral sparsifier, \( L \), with \( O(n \log n \log \epsilon) \) non-zeros. For unweighted graphs, the complexity can be reduced to \( O(T \epsilon m^{-2} \log n) \).

NetSMF — Algorithm

- Construct a random walk monoplynomial sparsifier.
- Construct NetMF matrix sparsifier.
- Truncated randomized singular value decomposition.

Algorithm 1: NetSMF

Input: A social network \( G = (V,E,A) \) which we want to learn network embedding. The number of non-zeros \( M \) in the sparsifier. The dimension of embedding \( d \).

Output: An embedding matrix of size \( n \times d \), each row corresponding to a vertex.

1. \( G = (V,E,A) \)
2. for \( i = 1 \) to \( M \) do
3. Uniformly pick an edge \( e = (u,v) \in E \)
4. Uniformly pick an integer \( r \in [1] \)
5. \( U_i, V_i, Z_i \leftarrow \text{PathSampling} (e, r) \)
6. Add an edge \( (u,v) \) to \( G \)
7. end
8. Compute \( L \) to be the unnormalized graph Laplacian of \( G \)
9. Compute \( D \) to \( \text{D} = D - L \) to \( D \)
10. \( U_c, V_c, Z_c \leftarrow \text{RandomizedSVD} (\text{time} = \log b \sum_{k=0}^{d-1} \log b \sum_{l=0}^{d-1} \log b \) \)
11. return \( \bar{U}_c, \bar{V}_c \) as network embeddings

Figure 1: The System Design of NetSMF. The input comes from a graph engine which stores the network data and provides efficient APIs to graph queries. In Step 1, the system launches several PathSampling workers to handle a subset of samples. Then, a reducer is designed to aggregate the output of the PathSampling algorithm. In Step 2, the system distributes data to several sparsifier constructors to perform the transformation and the truncated element-wise matrix logarithm. In the final step, the system applies truncated randomized SVD on the constructed sparsifier and dumps the resulted embeddings to storage.

Experimental Result and Discussions

Multi-label Classification

![Figure 2: Predictive performance on varying the ratio of training data](image)

References


The Web Conference, San Francisco, May 13-17, 2019