CRWMS: Bipartite Network Embedding based on Constrained Random Walk and Mixed-Skipgram

Yaling Ye¹, Lina Ma², Xia Zhang³, Qixuan Ni¹, Yuyao Wang⁴ and Zhan Bu¹,*

¹Jiangsu Provincial Key Laboratory of E-Business, Nanjing University of Finance and Economics, Nanjing, China
²WinGin Business-Intelligence Academy Nanjing Co., Ltd, Nanjing, China
³School of Artificial Intelligence, Nanjing Vocational College of Information Technology, Nanjing, China
⁴School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China

*buzhan@nuaa.edu.cn

Abstract. A bipartite network is a basic representation model in recommendation systems, in which the explicit links (e.g., representing user-item rating information) only exist between heterogeneous nodes (e.g., users and items). Most traditional bipartite network embedding methods only consider the explicit network structure, but ignore semantic relations and rating information therein. To solve this deficiency, we designed a new bipartite network embedding approach based on the Constrained Random Walk and Mixed-Skipgram (CRWMS). Specifically, we propose a constrained random walk strategy that produces the mixed local sampling sequences, and further adopt the splitting operation to preserve both explicit and implicit structural information. Furthermore, we propose a novel and effective mixed-skipgram model to discern node representation vectors through joint training of explicit and implicit structural sequences. We compare CRWMS with five state-of-the-art network embedding methods on three real-world bipartite networks. Experiments show that our approach improves performance in edge classification, edge clustering, recommendation and link reconstruction tasks.

1. Introduction

Network embedding can be regarded as a dimension-reduction technology. The main idea of network embedding is to find a mapping function that converts each node in a network into low-dimensional vector spaces. Representative papers include Deepwalk [1], Node2vec [2] and Large-scale Information Network Embedding (LINE) [3]. These approaches are unsuitable for bipartite network embedding, in actuality, because they neglect the nodes’ information type. Yet most real networks are heterogeneous in nature. Metapath2vec [4] is a representative vertex embedding method for heterogeneous networks, which cannot preserve both the explicit and implicit structure of bipartite networks. A bipartite network can be seen as a special type of heterogeneous network [5]. A method of bipartite network embedding is based on LINE (published in 2018 [6]), in which they do not take the semantic relations and rating information into consideration.

With such outstanding challenges in mind, here we develop a new bipartite network embedding approach based on the Constrained Random Walk and Mixed-Skipgram (CRWMS). Specifically, we
propose a random walk strategy that preserves both explicit and implicit structural information. It harnesses the explicit structure via the meta path with a rating constraint mechanism, which enriches the semantics of different node types. For example, Figure 1(a) is an example of a movie bipartite network; the given meta path UMU indicates that two users have watched the same movie. After adding a constraint of ‘5,’ the semantics becomes that two users have watched the same movie and evaluated the same rating. This enhances the semantic information in the follow-up walking paths. The network’s comprehensive information integration and rich semantic information make it promising to generate more effective embedding. Finally, we adopt the splitting operation to reserve the implicit structure, as Figure 1(c) shows.

2. Bipartite Network Embedding

2.1. Constrained Random Walk
To generate meaningful paths by random walk and capture the complex structural and semantics correlation between different node types, the key is to design an effective walking strategy. Hence, we propose constrained random walk strategy to generate paths. Given a bipartite network $G = (U, V, R_{u,v})$ and a meta path $\rho : U \xrightarrow{R} V$, obviously, the network has two node types—$T_u, T_v$, the explicit paths generated by the following function:

$$p\left(v^{i+1} | v^i, \rho_i\right) = \begin{cases} \frac{1}{|N_{v^i}(v^i)|}, & R_{v^i,v^{i+1}} = r_i, T_{v^{i+1}} = t_i \\ 0, & \text{otherwise} \end{cases}$$ (1)

Here $p\left(v^{i+1} | v^i, \rho_i\right)$ is the probability of walking from $v_i$ node of step $i$ to node $v$ of step $i+1$ under $\rho_i$ constraint and $t \in T_u \cup T_v$. $N_{v^i}(v^i)$ is the neighbor set of $v^i$. For the explicit path $p_{t}$, we proposed to split $p_t$ into two homogeneous sequences—$p_{t_u}, p_{t_v}$. Figure 1(b) shows an example explicit path “$u_1 \rightarrow m_6 \rightarrow u_2 \rightarrow m_1 \rightarrow u_3$,” which walks from node $u_1$ with meta path of “UUU” (user-item-user) and rating constraint of 5. Figure 1(c) is the two implicit paths after a splitting operation—respectively, “$u_1 \rightarrow u_2 \rightarrow u_1$” and “$m_2 \rightarrow m_1$”.

2.2. Mixed-Skipgram
Mixed-skipgram is a joint model combining homogeneous and heterogeneous skipgrams, as shown in
Figure 1(d). It aims to get the embedding of explicit and implicit structure of network. Given a constraint of random walk, we easily construct the \( t \) type context set \( N_t(\mathbf{w}) \) and all types of context set \( N(\mathbf{w}) \) for \( t \) type node \( \mathbf{w}_t \). Our model learns effective node representations for bipartite network \( G=(U,V,R_{uv}) \) by maximizing the following function:

\[
\arg \max \sum_{w_t \in \mathcal{W}} \left[ \sum_{c_t \in N_t(w_t)} \log p(c_t | w_t) + \sum_{i \in \mathcal{T}(w_t)} \log p(c_i | w_t) \right]
\]

where \( t, i \in (T_u \cup T_v) \), \( p(c_t | w_t) \) and \( p(c_i | w_t) \) are softmax functions. To optimize the joint model efficiently, we utilize the heterogeneous negative sampling, which considers the type of negative nodes while drawing negative samples from a corpus for the softmax construction. For illustration, consider the paths in Figure 1(b): the center node \( u_2 \) has two context nodes \( m_2 \). For the window size of \( k \).

Such as the positive pair \( (u_2, m_1) \), the type of negative nodes is equal to node type \( m_1 \), which forms negative pairs with the context node. Figure 1(c) has two homogeneous paths after splitting; CRWMS chooses the same path as the \( u_2 \) center node type, which presents two new positive pairs \( (u_2, u_i) \) and \( (u_2, u_j) \). As with the aforementioned method, we easily discern that this part’s negative samples are composed of \( u_i \) node types. To resolve this issue, we optimize every term. Therefore, we apply SGA to optimize these two parts, respectively. Inspired by Metapath2vec [4], we will list the computing process for the homogeneous part, and the heterogeneous part is similar to it, we utilize SGA to maximize:

\[
O_{\text{na}}(X) = \log \sigma(\mathbf{X}_{\mathbf{w}_t}^\top \mathbf{X}_{\mathbf{w}_t}^t) + \sum_{i=1}^{\text{neg}} E_i \log \sigma(-\mathbf{X}_{u_i}^t \cdot \mathbf{X}_{\mathbf{w}_t}^t)
\]

where \( X \) denotes an \((n+m) \times d\) matrix of the latent node embeddings, \( P_i(\mathbf{u}_i) \propto d(\mathbf{u}_i) \) is the pre-defined distribution, and \( d(\mathbf{u}_i) \) is the degree of \( t \) type node \( \mathbf{u}_i \), from which the number of times to extract negative node \( u_i^t \) is \text{neg} when \( i = 0 \), \( u_i^t = c_t \). The aforementioned objective function’s gradient is as follows:

\[
\frac{\partial O_{\text{na}}(X)}{\partial X_{u_i}} = \sum_{u_i^t} L(u_i^t) - \sigma(-\mathbf{X}_{u_i}^t \cdot \mathbf{X}_{\mathbf{w}_t}^t)) \mathbf{X}_{u_i^t} \\
\frac{\partial O_{\text{na}}(X)}{\partial X_{\mathbf{w}_t}} = \sum_{u_i^t} (L(u_i^t) - \sigma(-\mathbf{X}_{u_i}^t \cdot \mathbf{X}_{\mathbf{w}_t}^t)) \mathbf{X}_{u_i^t}
\]

where \( L(u_i^t) \) is an indicator function. Then node vector \( \mathbf{X}_{u_i} \) and \( \mathbf{X}_{\mathbf{w}_t} \) can be updated to:

\[
\mathbf{X}_{u_i} = \mathbf{X}_{u_i} + \eta \frac{\partial O_{\text{na}}(X)}{\partial X_{u_i}} \\
\mathbf{X}_{\mathbf{w}_t} = \mathbf{X}_{\mathbf{w}_t} + \eta \frac{\partial O_{\text{na}}(X)}{\partial X_{\mathbf{w}_t}}
\]

where \( \eta \) is the learning rate.

3. Experiments
In this section, we empirically evaluate the performances of CRWMS by comparing it with same homogeneous and heterogeneous network embedding methods: DeepWalk [1], Node2vec [2], LINE [3], Metapath2vec [4], BiNE [5] on multiclass edge classification, edge clustering, recommendation and link reconstruction. We conducted experiments on three rating bipartite networks’ different scales and densities, namely MI-100k, CiaoDVD, and MI-1M. Each network includes the user, item, rating of user to item, time stamp, and rating scale from 1 to 5.
3.1 Multi-class Link Classification
In the multiclass link classification task, each link’s representation embedding contains the vector of two nodes. For example, the link \( e_i = (u_i, m_i) \) consists of node \( u_i \) and \( m_i \), and the representation vector of \( e_i \) is \( X_e = X_u \oplus X_m \), where \( X_u \) and \( X_m \) denotes the embedding of node \( u_i \) and \( m_i \) trained by a bipartite network, and \( \oplus \) is a symbol for series’ connection. Figure 2 shows the result of edge classification, from which we observe that the CRWMS model significantly outperforms all baselines. Particularly, the CiaoDVD dataset’s classification task shows the best performance, in which Micro-F1 shows a corresponding 24.42%-30.30% increase, and Macro-F1 gains 36.11%-41.10%. The least improvement is in the MI-1M dataset; however, the best performance of all baselines is 5.37%-6.89% worse than the proposed CRWMS in Micro-F1 and 10.20%-11.39% in Macro-F1.

![Figure 2. Multi-class edge classification results on the three bipartite networks.](image)

3.2 Link Clustering
In this experiment, we illustrate how the latent representations learned by embedding methods can help the link clustering task in bipartite networks and we employ the same 5 categories in the classification task above. The Link representation vectors are generated by a series connection of nodes, and we leverage the k-means algorithm [7] to cluster the edge and evaluate the clustering results in terms of normalized mutual information (NMI) [8]. We conducted all clustering experiments using 10-fold cross validation and then we averaged the performance. Table 1 indicates that CRWMS’ performance outstrips all comparative methods on all datasets, and the improvement obtained by CRWMS over the best baseline (BiNE) is more significant-around 10%.

|                | MI-100k | CiaoDVD | MI-1M |
|----------------|---------|---------|-------|
| Deepwalk       | 0.0295  | 0.0033  | 0.0160|
| Node2vec       | 0.0152  | 0.0080  | 0.0195|
| LINE           | 0.0166  | 0.0209  | 0.0188|
| BiNE           | 0.0212  | 0.0226  | 0.0290|
| Metapath2vec   | 0.0167  | 0.0079  | 0.0183|
| CRWMS          | 0.0504  | 0.1231  | 0.0861|

3.3 Recommendation
For recommended tasks, we rank all the links in chronological order, taking the first 80% of the links as the training set and the last 20% as the test set. We adopt the traditional user-based collaborative filtering to prove that our proposed embedding method’s performance is more effective than other traditional
methods. Table 2 shows the performance of baselines and our CRWMS. Based on these results, we observe that our model consistently and significantly outperform all baselines in terms of mean average error (MAE) metrics.

Table 2. Recommendation on three bipartite networks.

|                  | MI-100k   | CiaoDVD   | MI-1M     |
|------------------|-----------|-----------|-----------|
| Deepwalk         | 0.7258    | 1.2522    | 0.5788    |
| Node2vec         | 0.7214    | 1.2493    | 0.5796    |
| LINE             | 0.7211    | 1.2470    | 0.5742    |
| BiNE             | 0.7224    | 1.2449    | 0.5786    |
| Metapath2vec     | 0.7202    | 1.2512    | 0.5786    |
| CRWMS            | **0.7182**| **1.2437**| **0.5621**|

Table 3. Link reconstruction on three bipartite networks.

|                  | MI-100k   | CiaoDVD   | MI-1M     |
|------------------|-----------|-----------|-----------|
| PR               | Recall    | $f_1$     | PR        | Recall    | $f_1$     | PR        | Recall    | $f_1$     |
| Deepwalk         | 0.5644    | 0.6418    | 0.6005    | 0.4960    | 0.5514    | 0.5199    | 0.5634    | 0.6329    | 0.5960    |
| Node2vec         | 0.5645    | 0.6088    | 0.5857    | 0.4953    | 0.5372    | 0.5134    | 0.5632    | 0.6218    | 0.5909    |
| LINE             | 0.5704    | 0.6533    | 0.6090    | 0.5419    | 0.5111    | 0.5249    | 0.5628    | 0.6303    | 0.5946    |
| BiNE             | **0.5759**| 0.5970    | 0.5862    | **0.5716**| 0.4154    | 0.4744    | **0.5704**| 0.5819    | 0.5759    |
| Metapath2vec     | 0.5605    | 0.6400    | 0.5976    | 0.4905    | 0.5094    | 0.4982    | 0.5618    | 0.6286    | 0.5933    |
| CRWMS            | 0.5688    | **0.6657**| **0.6134**| 0.5283    | **0.5995**| **0.5596**| 0.5627    | **0.6469**| **0.6018**|

3.4 Link Reconstruction

We generate a negative link for each positive link, labelled as 1 and 0, respectively. The positive link is the real existing link in the network, and the non-existing link—that is, the virtual link we created later—is a negative link. Next, we generate the link representation of all positive and negative links, and put it into the logical regression classifier. Table 3 compares the results of our proposed model and baselines on three datasets. CRWMS achieves 0.44%-2.72% improvement in terms of $f_1$ over LINE, Deepwalk, Node2vec, Metapath2vec and BiNE on MI-100k. The best result is on CiaoDVD, where the gain obtained by CRWMS over the best baselines (Deepwalk) is more significant—around 3.47%-8.52%.

4. Conclusion

In this paper, we propose a constrained random walking strategy to preserve both the explicit and implicit structure of the bipartite networks, which we use the constrained random walk based on a meta path to preserve the explicit structure and semantic relationship between different node types. Then we adopt a splitting operation, which directly extracts two node types for each walking path, and separate the heterogeneous path into two homogeneous paths. To jointly train the explicit and implicit structure we preserved, we design a mixed-skipgram model to learn the bipartite networks’ representation.

Acknowledgments

This research was partially supported by Postgraduate Research & Practice Innovation Program of Jiangsu province of China (Research on Personalized Recommendation Algorithm Based on False Comment Recognition and Bipartite Graph Mining) under Grant KYCX19_1383, in part by the National Natural Science Foundation of China (Research on Formation and Evolution Mechanism of Consumption Communities in Social Media) under Grant 71871109, in part by Educational Reform Project of Nanjing University of Finance and Economics (Research on Teaching Evaluation Model based on Multi-Source Information Fusion) under Grant JGY19051, and in part by the Qing Lan Project of Jiangsu Province.
References

[1] Bryan Perozzi, Rami Al-Rfou' and Steven Skiena: DeepWalk: online learning of social representations. *SIGKDD* 2014: 701-710.

[2] Aditya Grover and Jure Leskovec: Node2vec: Scalable Feature Learning for Networks. *SIGKDD* 2016: 855-864.

[3] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan and Qiaozhu Mei: LINE: Large-scale Information Network Embedding. *WWW* 2015: 1067-1077.

[4] Yuxiao Dong, Nitesh V. Chawla and Ananthram Swami: Metapath2vec: Scalable Representation Learning for Heterogeneous Networks. *SIGKDD* 2017: 135-144.

[5] Yizhou Sun, Brandon Norick, Jiawei Han, Xifeng Yan, Philip S. Yu and Xiao Yu: PathSelClus: Integrating Meta-Path Selection with User-Guided Object Clustering in Heterogeneous Information Networks. *TKDD* 7(3): 11:1-11:23 (2013).

[6] Ming Gao, Leihui Chen, Xiangnan He and Aoying Zhou: BiNE: Bipartite Network Embedding. *SIGIR* 2018: 715-724.

[7] David Arthur and Sergei Vassilvitskii: k-means++: the advantages of careful seeding. *SODA* 2007: 1027-1035.

[8] Strehl, Ghosh and Mooney: Impact of similarity measures on web-page clustering. *AAAI*, 2000: 58–64.