

Knowledge Driven Paper Recommendation Using Heterogeneous Network Embedding Method

Irfan Ahmed¹, Zubair Ahmed Kalhoro²

¹Department of Computer Science and Engineering, Southeast University, Nanjing, China

²Institute of Mathematics and Computer Science, University of Sindh, Jamshoro, Pakistan

Email: irfanabbasi19911@hotmail.com

How to cite this paper: Ahmed, I. and Kalhoro, Z.A. (2018) Knowledge Driven Paper Recommendation Using Heterogeneous Network Embedding Method. *Journal of Computer and Communications*, 6, 157-170.

<https://doi.org/10.4236/jcc.2018.612016>

Received: December 5, 2018

Accepted: December 25, 2018

Published: December 28, 2018

Copyright © 2018 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We search a variety of things over the Internet in our daily lives, and numerous search engines are available to get us more relevant results. With the rapid technological advancement, the internet has become a major source of obtaining information. Further, the advent of the Web2.0 era has led to an increased interaction between the user and the website. It has become challenging to provide information to users as per their interests. Because of copyright restrictions, most of existing research studies are confronting the lack of availability of the content of candidates recommending articles. The content of such articles is not always available freely and hence leads to inadequate recommendation results. Moreover, various research studies base recommendation on user profiles. Therefore, their recommendation needs a significant number of registered users in the system. In recent years, research work proves that Knowledge graphs have yielded better in generating quality recommendation results and alleviating sparsity and cold start issues. Network embedding techniques try to learn high quality feature vectors automatically from network structures, enabling vector-based measurers of node relatedness. Keeping the strength of Network embedding techniques, the proposed citation-based recommendation approach makes use of heterogeneous network embedding in generating recommendation results. The novelty of this paper is in exploiting the performance of a network embedding approach *i.e.*, matapath2vec to generate paper recommendations. Unlike existing approaches, the proposed method has the capability of learning low-dimensional latent representation of nodes (*i.e.*, research papers) in a network. We apply matapath2vec on a knowledge network built by the ACL Anthology Network (all about NLP) and use the node relatedness to generate item (research article) recommendations.

Keywords

Network Embedding, Heterogeneous Representation Learning, Paper-Citation Relations, Recommender System, Learning Latent Representations

1. Introduction

Due to the availability of enormous information on the World Wide Web, it is quite daunting and difficult to get access to relevant information. Information overload and cognitive stress are the core issues users are facing these days when they are in search of relevant items/information. To cope with such issues, recommender systems in different online services are playing a significant role in finding relevant information/item for the user. Traditional approaches such as a matrix factorization attempt try to characterize user-item interaction records (e.g., using user-item rating matrix) by learning an effective prediction function. Due to the availability of web services, recommender systems can consume various kinds of auxiliary information and generate more useful recommendations [1] [2] [3]. However, to model and utilize such complex heterogeneous information is a daunting task. Furthermore, it is more challenging to develop a relatively general approach to model these varying data in different systems or platforms. In recent years, research works have proven that integrating knowledge graphs with state of the art CF methods have yielded quality recommendation results. Additionally, their hybridization helps in resolving issues such as new scarce items, faced by existing recommendation approaches [4]-[9].

Among existing approaches, meta-path2vec has performed well in learning the features of network structure. Research work [10] has exploited its efficiency over other existing approaches in link prediction and node classification scenarios. In light of these studies, our research aims to use metapath2vec in learning knowledge network and find research items that meet user requirements.

In this research, we have proposed a research paper recommendation system using a novel heterogeneous network embedding technique called CN_HER exploiting latent relations in the scholarly paper recommendation. Furthermore, it can adequately amalgamate different kinds of information in HIN to enhance the quality of generated results. In literature, researchers have exploited citation networks in recommending research papers; however, they ignore the latent relations by employing direct relations such as co-citation relationships as accomplished in [11] [12]. Identifying and incorporating the latent relations across research papers could play a significant role and may improve the recommendation performance. The authors mined the hidden relation between a target paper and its references to present utile recommendations.

In a nutshell, the contributions of this paper can be summarized as follows:

- To uncover the semantic and structural information, we use a heterogeneous network embedding approach called metapath2vec.

- To the best of our knowledge, we integrate heterogeneous network embedding approach, *i.e.* CN_HER which can adequately incorporate embedding information into HIN resulting in enhanced recommendation results.
- We perform experiments on real-world dataset showing the effectiveness of our proposed model. Additionally, we examine its capability in resolving various problems, including sparsity, cold start, prediction accuracy and scalability issues, and reveal its improved recommendation results in HINs.

The paper is organized as follows. Section 2 mainly introduces the related work closely related to the research of this paper. Section 3 introduces Proposed Description and problem formulation of our model. Section 4 is devoted to experimental study, and section 5 concludes the paper.

2. Related Work

2.1. Knowledge Graph Embeddings Based Item Recommendation Using Node2vec

In a graph, node2vec [10] is a learning representation of nodes using the word2vec model on node sequences experimented through random walks. Node2vec is the concept of random walk exploration that is easygoing to represent several of connectivity patterns in a graph, node2vec brings this innovation. This proposed work shows that node2vec can be efficiently used to learn Knowledge network embeddings to perform in recommendation system. Node2vec is applied in a knowledge network, including user feedback on items, modelled by the special relation “feedback”, and item relations to other entities; then, using the similarity between users and items in the vector space model for producing the recommendations results. In this study, how learning knowledge network embedding’s using node2vec can be used to generate items. To use node2vec on a knowledge graph built from the MovieLens 1M dataset and DBpedia and use the node similarity for generating item recommendations. Given a knowledge network K consisting of users, items (I, U) (the object of the recommendations, e.g. a movie) and other entities E (objects connected to items, e.g. the director of a movie), vector representations of the users x_u and of the items x_i (and of other entities x_e) are generated by node2vec. Thus, propose to use as a ranking function the similarity between the user and the item vectors: $m(u; i) = c(x_u; x_i)$ where c is the cosine similarity in this work.

2.2. Recommendation Model Based on Knowledge Graph Representation Learning Technology

The learning technology expresses the semantic information of the research object as dense low dimensional real value vector by machine learning method, which solves the problem that the former single heat (one-hot) vector cannot represent the semantic information of the research object. By means of knowledge network representation learning technology, the low dimensional dense vectors of entities and relationships in the Knowledge Graph can be obtained, and the semantic information of the entities under the structure of knowledge

network is expressed. F. Zhang *et al.* [13], the idea of adding the information in the knowledge graph to the recommendation model by means of learning is presented in the following ways: first, through TRANSR [14] learning the vectors of entities and relationships in the knowledge network. Then, the potential feature model is used to decompose the user's object matrix to obtain the latent eigenvector of the user and the object, and finally the vector of the object and the potential eigenvector of the object, which is embedding with the structure information of the knowledge network is added to obtain the final vector of the object. The recommended results are ultimately determined by the score ranking obtained by the user vector and the object vector. The distributed representation of goods is based on the structure learning of the graph in the knowledge network, so the direct addition and form cannot directly describe the structural information that is learned from the graph, we propose a new method to combine the Knowledge Network Embeddings technology with the recommendation model. In response to the above problems, our approach uses a method similar to [13], using embedding technology to learn the vector of entities and relationships in the knowledge network, and then, by multiplying the product vectors to get the relationship matrix of the object. The user-object preference matrix is obtained through the user-object feedback matrix and the product-item relationship matrix products, we introduce the idea of knowledge network information similar to [15], pass the user's preference through the structure of the knowledge network. Finally, the user-object preference matrix is decomposed to obtain the potential eigenvector of users and objects.

We propose a citation based research paper recommender system using Heterogeneous Information Network Embedding method for Recommendation (CN_HER). Given a Heterogeneous Information Network Embeddings, the task is to learn the low-dimensional latent representations (embeddings) $e_v \in R^d$ for every object $v \in V$. The learned low-dimensional latent representations are expected to highly encapsulate informative characteristics, which are probable to be beneficial in paper recommender systems represented on Heterogeneous Information Network Embeddings. Nevertheless, most of the current network embedding approaches mostly focus on homogeneous networks, which are not capable to efficiently model heterogeneous information networks. For example, a node2vec and deepwalk, etc.; to generate node sequences are used by the innovative study [10] [16], which cannot contradistinguish nodes and edges with different object and relation types. Therefore, it needs a more righteous way to traverse the Heterogeneous Information Network Embeddings and generate meaningful node sequences. For the first time, we integrate heterogeneous network embedding model (metapath2vec) in recommendation, which has been enhanced the recommendation results.

3. Proposed Description and Problem Formulation

3.1. Problem Definition

Definition 1. A Graph G defines a Heterogeneous Information Network

$G = (V, E, A)$ consisting of node v and edge e . A Heterogeneous Information Network is associated with its mapping functions $\Phi(v): V \rightarrow A_v$ and $X \in R^{|V| \times d}, d \ll |V|$, respectively. The sets of object and relation types are denoted as A_v and A_e , where $|A_v| + |A_e| > 2$. For instance, the information network can be one represented in **Figure 1** with authors (V), papers (T), venues (A), terms (P) as an object set V (nodes), where, publish (A-P, P-V), co-author (A-A), terms (P-T) relationships are indicated as a link set E (edges). The schema-level description is determined as the complexity of the heterogeneous network of understanding the object types and relation types better in the network. Therefore, to describe the Meta-structure of a network proposes the concept of network schema.

Definition 2. $S:(A_v, A_e)$ Denotes the network meta. Network meta [17] [18]. Network meta is a meta-template for a heterogeneous information network $G = (V, E, A)$ is associated with the sets of object type mapping $f(v): V \rightarrow A_v$ and the relation type mapping $\phi(e): E \rightarrow A_e$, it is an object type A_v , with edges as relations from A_e was defined by a directed graph.

Definition 3. Heterogeneous Network Representation Learning: To learn the d -dimensional latent representations $X \in R^{|V| \times d}, d \ll |V|$ are the task in a given to capture the structure and semantic associations between them. The output of the Heterogeneous Network Representation Learning is the low-dimensional matrix X , with the V^{th} row—a d -dimensional vector X corresponds to the representation of node v . Notice that, even though the representations of different types of nodes in V mapped into the same latent space. The node representations are learned can benefit heterogeneous research papers network recommendation task. For instance, as the feature input in similarity search tasks can use the embedding vector of each node.

Definition 4. Heterogeneous Information Network based recommendation. A Heterogeneous citation network can model the various kinds of information in a

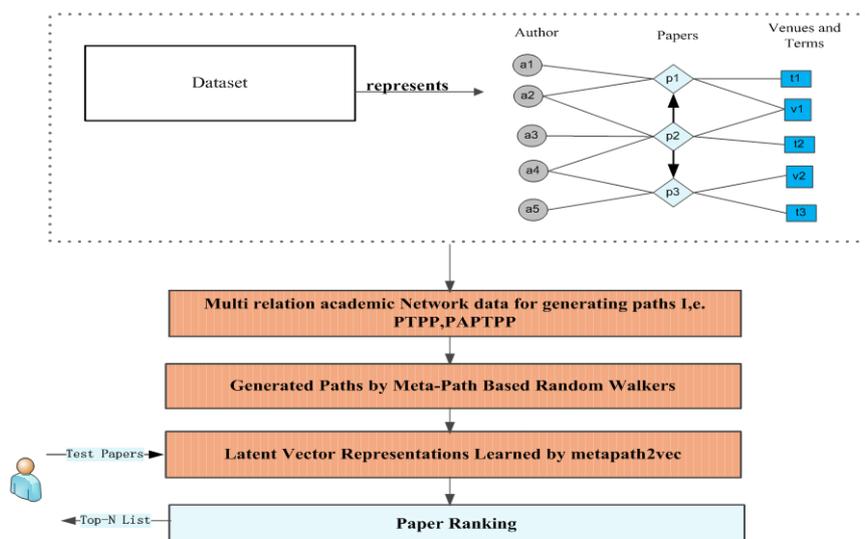


Figure 1. The graphic diagram of the proposed CN_HER method.

recommender system. In this work, we formulate the citation based heterogeneous networks, recommendation problem as the problem of learning a relevance score list $s(x, y): X \times Y \rightarrow R$ for a manuscript $x \in X$ and a training paper $y \in Y$ based on the heterogeneous paper's network of scientific papers. The learned score list $s(x, y)$ is used to rank the scores between manuscript and all the scientific papers in the dataset to produce a recommendation.

3.2. Architecture and Working of Proposed Solution

In this paper, we investigate a citation based research paper recommender system using Heterogeneous Information Network Embedding model for Recommendation to recommend the relevant top- N research paper. In recommending papers, the proposed work goes through different phases.

3.2.1. System Overview

As introduced earlier, our Network-based citation recommendation framework includes the following three main stages: Training data preparation, Representation learning and Network-based citation recommendation (See **Figure 1**).

3.2.2. Heterogeneous Information Network Construction

For given information, our multi-layered network model can be defined as follows: A Graph G defines a Heterogeneous Information Network $G = (V, E, A)$ consisting of node v and edge e . A Heterogeneous Information Network is associated with their mapping functions $f(v): V \rightarrow A_v$ and $X \in R^{|V| \times d}$, $d \ll |V|$, respectively. The sets of object and relation types are denoted as a A_v and A_e , where $|A_v| + |A_e| > 2$.

For instance, the information network can be one represented in **Figure 1** with authors (V), papers (T), venues (A), terms (P) as an object set V (nodes), where, publish (A-P, P-V), co-author (A-A), terms (P-T) relationships are indicated as a link set E (edges).

3.2.3. Heterogeneous Network Embedding: Metapath2vec

The recent progress on network embedding [10] [16] encouraged. We propose a CN_HER approach in order to extract and represent useful information on Heterogeneous Information Networks for the paper recommendation.

Given a Heterogeneous Information Network Embeddings, The task is to learn the low-dimensional latent representations (embedding) $e_v \in R^d$ for every object $v \in V$. The learned low-dimensional latent representations are expected to highly encapsulate informative characteristics, which are probable to be beneficial in paper recommender systems represented on Heterogeneous Information Network Embeddings. Nevertheless, most of the current network embedding approaches mostly focus on homogeneous networks, which are not capable to efficiently model heterogeneous information networks. For example, a random walk to generate node sequences are used by the innovative study deepwalk [16], which cannot contradistinguish nodes and edges with different object and relation types. Therefore, it needs a more righteous way to traverse

the Heterogeneous Information Network Embeddings and generate meaningful node sequences.

3.2.4. Meta-Path Based-Random Walks

To design an effective walking strategy, it is essential to generate meaningful node sequences, that are capable to capture the complex semantics in Heterogeneous Information Network Embeddings. In the literature of Heterogeneous Information Network Embeddings, to characterize the semantic patterns for Heterogeneous Information Network Embeddings [19], the meta-path is an important concept. Hence, to generate the node sequence, meta-path based random walk method proposed. As shown in **Figure 2**, giving a heterogeneous network, and a meta-path $m_p : A_1 \xrightarrow{R_1} \dots A_t \xrightarrow{R_t} A_{t+1} \dots \xrightarrow{R_i} A_{t+1}$ wherein $R = R_1, R_2, \dots, R_{t-1}$ denotes the composite relations between node types $A1$ and $A1$ [20]. We can generate a sampled sequence " $p_1 \rightarrow a_1 \rightarrow p_2 \rightarrow a_2$ " according to Equation (2). The following distribution generates the walk path:

$$P(m_{t+1} = x | m_t = v, m_p) = \begin{cases} \frac{1}{|N^{A_{t+1}(v)}|}, & (v, x) \in E \text{ and } \mathcal{O}(x) = A_{t+1}; \\ 0, & \text{otherwise,} \end{cases} \quad (1)$$

where v has the type of A_p , m_p is the t -th node in the walk, and the first-order neighbor set for node v with the type of A_{t+1} is $N^{A_{t+1}(v)}$. The pattern of a meta-path will be followed continuously by a walk, until Walk has reached the pre-defined length.

3.2.5. CN_HER Model Citation Recommendation Algorithm

Given a manuscript P , we propose a heterogeneous scientific papers network representation citation recommendation approach, which aims to return top-ranked scientific papers as reference papers by measuring the similarity scores between the manuscript and all the scientific papers in the dataset. In our work, we formulate the manuscript P as a manuscript author a , term t , venue v *i.e.* $P^t = \{a, t, v\}$. We suppose P_t as a testing paper, P_a as an author of the manuscript and all the scientific papers in the dataset P^T as training papers. **Algorithm 1**

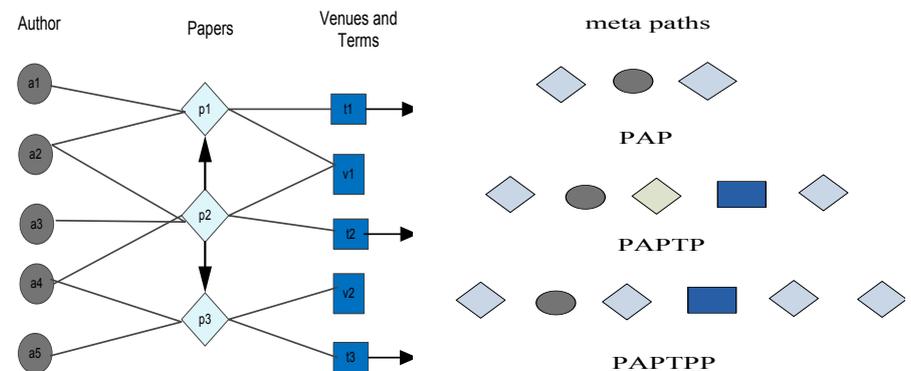


Figure 2. A demonstrative example of the proposed meta-path based random walk. We perform random walks directed by some selected meta-paths.

Input: the training papers $P^T = \{a, t, v, p\}$, author a , term t , venue v , citation p ,
A meta-path scheme p , # walks/node w , Walk/length l , vector dimension d ,
neighborhood size ks , test papers to $P^t = \{a, t, v\}$

Output: Top N Recommendation list.

Step1: Build heterogeneous papers network $G = \langle V, E, A \rangle$

Step2: Using metapath2vec to obtain the representation vectors.

Step3: Use vectors to compute similarity score r_p for P^t and P^T

Step4: Choose top N for recommendation.

Algorithm 1. The algorithm of CN_HER.

summarizes the whole process of the heterogeneous scientific papers network representation citation recommendation approach.

The similarity scores are represented as,

$$r_p = \{r_{pP_1}, r_{pP_2}, \dots, r_{pP_n}\}, P_n \in P (i = 1, 2, \dots, n) \quad (2)$$

The input to the recommendation system is the word sequence of training papers and testing papers, all the authors, terms, venues and citation papers of the training papers and authors, terms, venues of the test papers, as well as the adjacency matrix based on the network. All these papers and, all the authors, terms, venues and citation papers are mapped into vectors based on our proposed CN-HER model. Thus the similarity scores can be calculated as:

$$r_p = V_{P^T R} V_{P_{txt}}^T + V_{AR} V_{P_a}^T + V_{TR} V_{P_t}^T + V_{VR} V_{P_v}^T + V_{PR} V_{P_p}^T \quad (3)$$

where $P^T R$ is the vector representation of training papers, P_{txt} is the vector representation of the manuscript text, V_{AR} is the vector representation of authors related to training papers, V_{P_a} is the vector representation of the manuscript author, V_{P_t} is the vector representation of the manuscript terms, V_{P_v} is the vector representation of the manuscript venues, V_{P_p} is the vector representation of the manuscript citation papers. Training papers are ranked according to the similarity scores, the top-ranked ones, are selected as the final recommendation list.

4. Experimental Evaluation

The aim of this section is to evaluate the performance of the recommendation using the metapath2vec methodology on the dataset for determining its accuracy and efficiency.

4.1. Data Preparation

To evaluate the quality of our proposed model, we conduct experiments on the heterogeneous networks bibliographic dataset: AAN (ACL Anthology Network) dataset, which is established by Radev and Muthukrishnan (2009), it consists of conference papers and journal papers in the field of Natural Language Processing.

In our work, we extract the title and abstract of the papers in the dataset as document content. We remove the papers which have missed titles or abstracts in the dataset. Then, we use the remaining 13,929 papers published from 1975 to 2013 as an experimental dataset. Through the following steps each paper pre-processed 1) extract paper abstract and title; 2) remove the words that consist less than 3 characters; 3) remove stop words. For reducing the noise, we removed those words that seemed less than ten in the dataset. We got a total of 4397 different candidate words. Using a naive approach TF-IDF as an indicator [21], we recognized keywords from the set of candidate words for constructing the paper-word graph. If the TF-IDF of a word was greater than a TF-IDF threshold of 0:03, this word was selected as a keyword. A total of 3804 different keywords were identified in the set of 13,939 papers at a TF-IDF threshold frequency 0:03. For assessment purposes, we divide the entire data set into two separate sets, the papers 2015 are assessed as a training set (12,738 papers) and the remaining papers fall into the testing published before set (1211 papers).

4.2. Evaluation Metrics

In order to assess and evaluate recommendation results, we have used *Recall* [20], precision and NDCG [22]. These matrices are used to check the performance of the proposed method in terms of recommendation accuracy and predicted ranks quality. In literature, they are commonly used for evaluating recommendation results.

- **Recall.** Recall measure the ratio between the number of total cited papers to the number of articles system has presented in a Top-N recommendation list. The ratio represents the number of hits divided by the size of each user test data. The formula for recall is given as follows:

$$\text{Recall} = \frac{\sum(\text{relevant papers}) \cap \sum(\text{retrieved_papers})}{\sum(\text{relevant_papers})} \quad (4)$$

- **NDCG.** The recommendation results generated by the paper recommender system are sensitive to the positions of the relevant reference papers, using recall we cannot evaluate that intuitively. Therefore, it is indispensable for closely related references to appear higher in the Top-N list. We use (NDCG) to measure the ranked recommendation list. The NDCG value can be calculated using the following formula: calculated as

$$\text{NDCG} = \sum_{i=1}^n \frac{2^{r_i} - 1}{\log_2(i+1)} \quad (5)$$

where $r(i)$ is the rating value of the i -th paper in the ranking list. $r(i) = 1$ means the article is relevant otherwise $r(i) = 0$ if it is not.

- **Precision.** Precision is used to calculate and find the ratio between the relevant papers recommended by the system to the total number of recommended papers. In contrast, recall calculates the relevant articles recommended by the system to total relevant recommendations. Precision measures the exactness of recommendation results generated by the system. Preci-

sion value has an inverse relationship with false positive, greater the precision the less is false positive. Trade-off exists between precision and recall scores as there is an inverse relationship between the two, increase in one metric results decrease in another. For establishing the balance, there should be a trade-off between the two as given following:

$$\text{Precision} = \frac{\sum(\text{relevant_papers}) \cap \sum(\text{retrieved_papers})}{\sum(\text{retrieved_papers})} \quad (6)$$

4.3. Results and Discussions

Evaluating metapath2vec with the various latest network representations, learning methods, DeepWalk [16], and node2vec [10], our original intention to propose the network representation approach consists in hopes to obtain more meaningful vector representation of each vertex in the network, and then perform citation recommendation based on the vector representations of these vertices. For that purpose, so we compare our proposed bibliographic network representation approach with another two network representation approach: DeepWalk, (2) Node2Vec, which learns a mapping of paper vertices to a low dimensional space of features that maximizes the likelihood of preserving paper network neighborhoods of paper vertices. After obtaining network representation with the above different approaches, citation recommendation can then be performed. Which learns paper network representation by utilizing network structure information; the obtained results are given in **Figure 3** and **Figure 4**. These figures show the results of comparisons supported (Recall) and (Precision) severally. The result obtained in **Figure 5** shows that the model has immensely and commonly outperformed the baseline ways for all N recommendations values based on Recall. DeepWalk performs the worst of the three results, whereas for sure, the performance of our model will increase as the amount of N will increase. However, the best results supported Recall are obtained once N = 100 (Top@N).

Then again, the assessment results based on Precision are described in **Figure 4**.

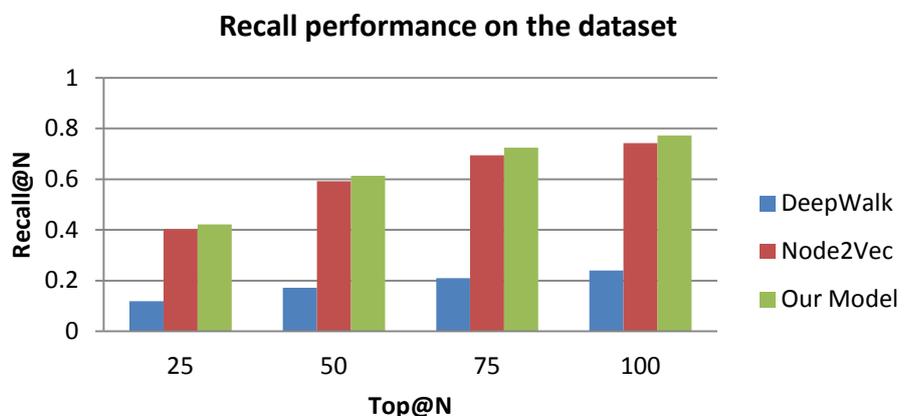


Figure 3. Recall performance on the dataset.

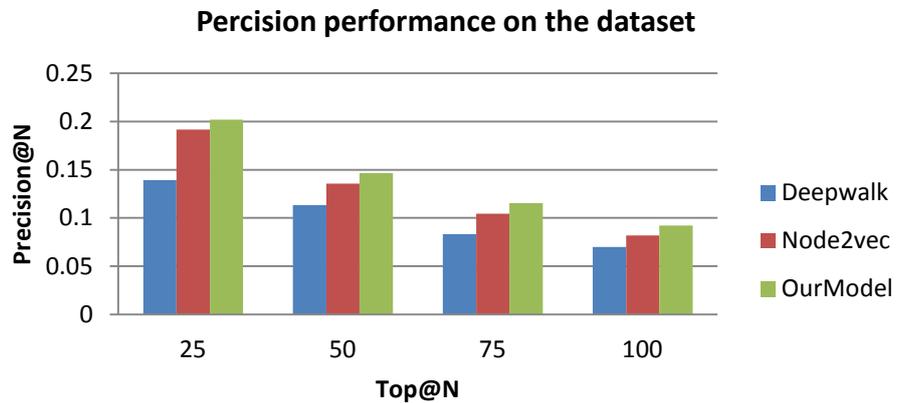


Figure 4. Precision performance on the dataset.

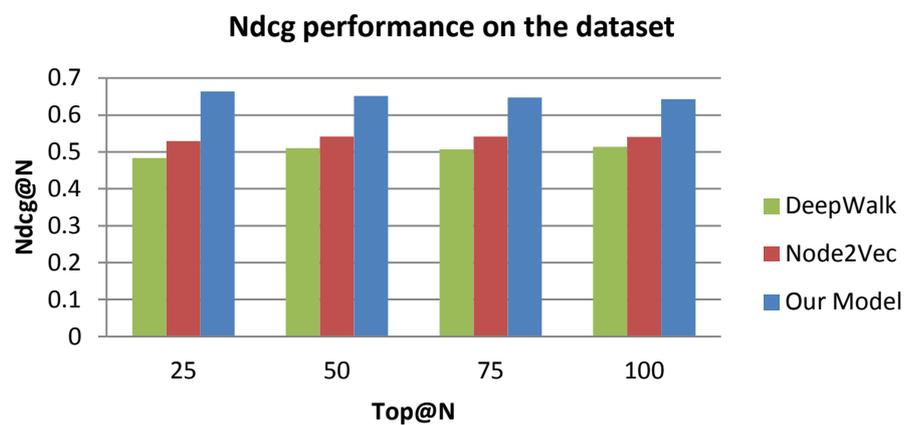


Figure 5. Ndcg performance on the dataset.

Between our model and the Node2vec results differences are not much noteworthy. While as estimated, the performance of our model decreases when the number of @N increases. This is because as the number of @N increases, the affinity of retrieving inappropriate results also increases and thereby affecting the cumulative Precision results. Though, both the two approaches considerably outclassed the Deepwalk method. This is because, the two approaches can leverage which learns a mapping of paper vertices to a low dimensional space of features that maximizes the likelihood of preserving paper network neighborhoods of paper vertices, and different from the Deepwalk method which learns paper network representation by utilizing network structure information.

Figure 5 respectively represents the results, comparisons based on Ndcg. The Ndcg result of our model is significant over the baseline methods for all N recommendations values. However, the Node2vec approach outperformed the proposed approach when $N = 50$ ($N@50$). Our model results start with encouraging results, specifically when $N = 25$ ($N@25$), but becomes less noteworthy when the number of recommendations (N) increases. However, both methodologies statistically outperformed the Deep walk method based on Ndcg.

In this section, we have compared our network embeddings method to the two state-of-the-art methods node2vec and deepwalk. These methods proposed,

based on word2vec-based network embeddings learning frameworks. However, these methods have thus far focused on homogeneous network representation learning of a single type of nodes and relation types. Our network embeddings method (metapath2vec) performs considerably better heterogeneous network embedding's more than current high-tech methods, as measured by evaluation metrics. The plus point of our proposed method lies in their proper thought of the network heterogeneity challenge the existence of more than one type of nodes and relation types.

5. Conclusion

In this paper, to uncover the semantic and structural information of research articles, we use a heterogeneous network embedding approach called metapath2vec. The proposed citation-based recommendation approach makes use of heterogeneous network embedding in generating recommendation results. The novelty of this paper is in exploiting the performance of a network embedding approach *i.e.*, matapath2vec to generate paper recommendations. Unlike existing approaches, the proposed method has the capability of learning low-dimensional latent representation of nodes (*i.e.*, research papers) in a network. We performed experiments on real-world dataset showing the effectiveness of our proposed model. Additionally, we examine its capability in resolving various problems, including sparsity, cold start, prediction accuracy and scalability issues, and reveal its improved recommendation results in HINs.

Acknowledgements

Time flies, two years of graduate study are coming to an end, and I will leave the familiar southeast university and embark on a new journey of life. I would like to express my heartfelt thanks to my teachers, classmates who have helped me and my family who have been supported me.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

- [1] Dias, M.B., Locher, D., Li, M., El-Dereby, W. and Lisboa, P.J. (2008) The Value of Personalized Recommender Systems to E-Business: A Case Study. *Proceedings of the 2nd ACM Conference on Recommender Systems*, Lausanne, 23-25 October 2008, 291-294.
- [2] Koren, Y. and Bell, R. (2015) Advances in Collaborative Filtering. In: Ricci, F., Rokach, L. and Shapira, B., Eds., *Recommender Systems Handbook*, Springer, Boston, 77-118. https://doi.org/10.1007/978-1-4899-7637-6_3
- [3] Schafer, J.B., Frankowski, D., Herlocker, J. and Sen, S. (2007) *The Adaptive Web*. Springer, Berlin, Heidelberg, 291-324.
- [4] Catherine, R. and Cohen, W. (2016) Personalized Recommendations Using Know-

- ledge Graphs: A Probabilistic Logic Programming Approach. *Proceedings of the 10th ACM Conference on Recommender Systems*, Boston, 1-19 September 2016, 325-332. <https://doi.org/10.1145/2959100.2959131>
- [5] Noia, T.D., Ostuni, V.C., Tomeo, P. and Sciascio, E.D. (2016) SPrank: Semantic Path-Based Ranking for Top-*N* Recommendations Using Linked Open Data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, **8**, Article No. 9. <https://doi.org/10.1145/2899005>
- [6] Ostuni, V.C., Di Noia, T., Di Sciascio, E. and Mirizzi, R. (2013) Top-*N* Recommendations from Implicit Feedback, Leveraging Linked Open Data. *Proceedings of the 7th ACM Conference on Recommender Systems*, Hong Kong, 12-16 October 2013, 85-92. <https://doi.org/10.1145/2507157.2507172>
- [7] Palumbo, E., Rizzo, G. and Troncy, R. (2017) Entity2rec: Learning User-Item Relatedness from Knowledge Graphs for Top-*N* Item Recommendation. *Proceedings of the Eleventh ACM Conference on Recommender Systems*, Como, 27-31 August 2017, 32-36. <https://doi.org/10.1145/3109859.3109889>
- [8] Rosati, J., Ristoski, P., Di Noia, T., Leone, R.D. and Paulheim, H. (2016) RDF Graph Embeddings for Content-Based Recommender Systems. *CEUR Workshop Proceedings*, Vol. 1673, 16 September 2016, Boston, 23-30.
- [9] Yu, X., Ren, X., Sun, Y., Gu, Q., Sturt, B., Khandelwal, U., Norick, B. and Han, J. (2014) Personalized Entity Recommendation: A Heterogeneous Information Network Approach. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, New York, 24-28 February 2014, 283-292. <https://doi.org/10.1145/2556195.2556259>
- [10] Grover, A. and Leskovec, J. (2016) Node2vec: Scalable Feature Learning for Networks. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 855-864. <https://doi.org/10.1145/2939672.2939754>
- [11] McNee, S.M., Albert, I., Cosley, D., Gopalkrishnan, P., Lam, S.K. and Rashid, A.M. (2002) On the Recommending of Citations for Research Papers. *Proceedings of the ACM Conference on Computer Supported Cooperative Work*, New Orleans, 16-20 November 2002, 116-125. <https://doi.org/10.1145/587078.587096>
- [12] Small, H. (1973) Co-Citation in the Scientific Literature: A New Measure of the Relationship between Two Documents. *Journal of the American Society for Information Science*, **24**, 265-269.
- [13] Konstas, I., Stathopoulos, V. and Jose, J.M. (2009) On Social Networks and Collaborative Recommendation. *Proceedings of the 32nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Boston, 19-23 July 2009, 195-202.
- [14] Lin, Y., Liu, Z., Sun, M., Liu, Y. and Zhu, X. (2015) Learning Entity and Relation Embeddings for Knowledge Graph Completion. *Proceedings of the 29th AAAI Conference on Artificial Intelligence*, Austin, 25-30 January 2015, 2181-2187.
- [15] Jamali, M. and Ester, M. (2009) *TrustWalker*: A Random Walk Model for Combining Trust-Based and Item-Based Recommendation. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, 28 June-1 July 2009, 397-406. <https://doi.org/10.1145/1557019.1557067>
- [16] Perozzi, B., Al-Rfou, R. and Skiena, S. (2014) DeepWalk: Online Learning of Social Representations. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, 24-27 August 2014, 701-710. <https://doi.org/10.1145/2623330.2623732>

- [17] Sun, Y. and Han, J. (2013) Mining Heterogeneous Information Networks: A Structural Analysis Approach. *ACM SIGKDD Explorations Newsletter*, **14**, 20-28. <https://doi.org/10.1145/2481244.2481248>
- [18] Sun, Y., Yu, Y. and Han, J. (2009) Ranking-Based Clustering of Heterogeneous Information Networks with Star Network Schema. *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Paris, 28 June-1 July 2009, 797-806.
- [19] Sun, Y., Han, J., Yan, X., Yu, P.S. and Wu, T. (2011) PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. *Proceedings of the VLDB Endowment*, **4**, 992-1003.
- [20] Liu, Q., Chen, E., Xiong, H., Ding, C.H.Q. and Chen, J. (2012) Enhancing Collaborative Filtering by User Interest Expansion via Personalized Ranking. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **42**, 218-233. <https://doi.org/10.1109/TSMCB.2011.2163711>
- [21] Wu, H., He, J., Pei, Y. and Long, X. (2010) Finding Research Community in Collaboration Network with Expertise Profiling. In: Huang, D.S., Zhao, Z., Bevilacqua, V. and Figueroa, J.C., Eds., *Advanced Intelligent Computing Theories and Applications. ICIC 2010. Lecture Notes in Computer Science*, Vol. 6215, Springer, Berlin, Heidelberg, 337-344.
- [22] Totti, L.C., Mitra, P., Ouzzani, M. and Zaki, M.J. (2016) A Query-Oriented Approach for Relevance in Citation Networks. *Proceedings of the 25th International Conference Companion on World Wide Web*, Montréal, Québec, 11-15 April 2016, 401-406. <https://doi.org/10.1145/2872518.2890518>