

# paper2repo: GitHub Repository Recommendation for Academic Papers

Huajie Shao

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
hshao5@illinois.edu

Dachun Sun

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
dsun18@illinois.edu

Jiahao Wu

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
jwu108@illinois.edu

Zecheng Zhang

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
zzhan147@illinois.edu

Aston Zhang

Amazon Web Services Deep Learning  
astonz@amazon.com

Shuochao Yao

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
syao9@illinois.edu

Shengzhong Liu

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
sl29@illinois.edu

Tianshi Wang

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
tianshi3@illinois.edu

Chao Zhang

School of Computational Science and  
Engineering, Georgia Institute of  
Technology  
chaozhang@gatech.edu

Tarek Abdelzaher

Department of Computer Science,  
University of Illinois at Urbana  
Champaign  
zaher@illinois.edu

## ABSTRACT

GitHub has become a popular social application platform, where a large number of users post their open source projects. In particular, an increasing number of researchers release repositories of source code related to their research papers in order to attract more people to follow their work. Motivated by this trend, we describe a novel item-item cross-platform recommender system, *paper2repo*, that recommends relevant repositories on GitHub that match a given paper in an academic search system such as Microsoft Academic. The key challenge is to identify the similarity between an input paper and its related repositories across the two platforms, *without the benefit of human labeling*. Towards that end, *paper2repo* integrates text encoding and constrained graph convolutional networks (GCN) to automatically learn and map the embeddings of papers and repositories into the same space, where proximity offers the basis for recommendation. To make our method more practical in real life systems, labels used for model training are computed *automatically* from features of user actions on GitHub. In machine learning, such automatic labeling is often called *distant supervision*. To the authors' knowledge, this is the first *distant-supervised*

*cross-platform* (paper to repository) matching system. We evaluate the performance of *paper2repo* on real-world data sets collected from GitHub and Microsoft Academic. Results demonstrate that it outperforms other state of the art recommendation methods.

## CCS CONCEPTS

• **Information systems** → **Electronic commerce**; **Recommender systems**.

## KEYWORDS

Recommender system, cross-platform recommendation, constrained graph convolutional networks, text encoding

### ACM Reference Format:

Huajie Shao, Dachun Sun, Jiahao Wu, Zecheng Zhang, Aston Zhang, Shuochao Yao, Shengzhong Liu, Tianshi Wang, Chao Zhang, and Tarek Abdelzaher. 2020. paper2repo: GitHub Repository Recommendation for Academic Papers. In *Proceedings of The Web Conference 2020 (WWW '20)*, April 20–24, 2020, Taipei, Taiwan. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3366423.3380145>

## 1 INTRODUCTION

This paper proposes a novel *item-item recommender system that matches papers with related code repositories* (on GitHub [8]) based on a *joint embedding* of both into the same space, where shorter distances imply a higher degree of relevance. The joint embedding

This paper is published under the Creative Commons Attribution 4.0 International (CC-BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '20, April 20–24, 2020, Taipei, Taiwan

© 2020 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC-BY 4.0 License.

ACM ISBN 978-1-4503-7023-3/20/04.

<https://doi.org/10.1145/3366423.3380145>

is novel in considering both text descriptions (of papers and repositories) as well as their relations (to other papers and repositories) expressed in appropriately defined graphs.

The work is motivated by the growing popularity of GitHub as a platform for collaboration on open source projects and ideas. In recent years, more researchers from academia and industry have shared the source code of their research with others on GitHub. A particularly clear example of that trend in computer science is research on machine learning and data mining, where links are often provided to open source repositories in published papers. This trend is expected to grow, as it is attributed to the increasingly computation-intensive nature of modern scientific discovery. Advances in science are increasingly assisted by complex computing systems and algorithms. Reproducibility, therefore, calls for availing the research community not only of the published scientific results but also of the software responsible for producing them. In the foreseeable future, readers will frequently want to find the source code needed to repeat the experiments when searching for papers of interest on academic search systems, such as Google Scholar and Microsoft Academic<sup>1</sup>. Our new cross-platform recommender system, *paper2repo*, addresses this need by automatically finding repositories relevant to a paper, even if pointers to the code were not included in the manuscript.

In the past few years, cross-platform recommender systems [3, 9, 29] attracted a lot of attention. Cross-platform recommender systems refer to those that leverage information from multiple different platforms to recommend items for users. Existing work mainly adopts transfer learning to enrich users' preference models with auxiliary information from multiple platforms. For example, Elkahky et al. [3] developed a multi-view deep learning model to jointly learn users' features from their preferred items in different domains of Microsoft service. Such prior work typically uses *supervised learning* to infer user preferences from their past behavior. In contrast, this paper designs a new item-item cross-platform recommender system, *paper2repo*, that matches repositories on GitHub to papers in the academic search system, *without the benefit of prior access history or explicit labeling*. Instead, we use *joint embedding* to quickly find repositories related to papers of interest, thereby seamlessly connecting Microsoft Academic with GitHub (acquired by Microsoft in June, 2018<sup>2</sup>).

A simple candidate solution to our matching problem might be to search for keywords from paper titles in the database of repositories on GitHub. However, it may not recommend sufficiently diverse items to users. For instance, a user who is interested in the textCNN repository<sup>3</sup> may like the Transformer repository<sup>4</sup> as well, because both of them are related to natural language processing. Such a comprehensive set of relevant repositories is hard to infer from paper titles and, in fact, will often not be included in the text of the manuscript either.

Deep neural networks have recently been used to implement recommender systems. Among the recent advances, a *semi-supervised* graph convolutional network (GCN) [6, 26, 27] has been widely

used for recommendation because it can exploit both the graph structure and input features together. However, past work applied the approach to *single platforms only*. Since papers and repositories live on two different platforms, their embeddings generated by traditional GCN are *not* in the same space. Our contribution lies in extending this approach to *joint embedding* that bridges the two platforms. The embedding requires constructing appropriate graphs for papers and repositories as inputs to the GCN, as well as selecting the right text features for each. We address these challenges with the following core techniques:

**Context graphs:** Relations between papers can be described by a citations graph. Relations between repositories are less obvious. In this paper, we leverage the tags and user starring to connect repositories together. There exists an edge between two repositories if there are overlapped keywords between them, or if they are starred by the same users. In addition, term frequency-inverse document frequency (TF-IDF) [16] is adopted to extract the important tags to construct the repository-repository context graph.

**Constrained GCN model with text encoding.** To map the embeddings of papers and repositories to the same space, we develop a joint model that incorporates a text encoding technique into the constrained GCN model based on their contents and graph structures. Since some research papers explicitly name (in the manuscript) their corresponding repositories of source code on GitHub, these papers could be used as *bridge papers* to connect the two cross platforms. Accordingly, we propose a constrained GCN model to minimize the distance between the embeddings of bridge papers and their original repositories. Moreover, we leverage a text encoding technique with convolutional neural networks (CNN) to encode information from the abstracts of papers, and the descriptions and tags of repositories as input features of the constrained GCN model. As a result, the proposed constrained GCN with text encoding can jointly learn and map the embeddings of papers and repositories from the two platforms into the same space.

We conduct experiments to evaluate the performance of the proposed *paper2repo* system on real-world data sets collected from GitHub and Microsoft Academic. The evaluation results demonstrate that the proposed *paper2repo* outperforms other compared recommendation methods. It can achieve about 10% higher recommendation accuracy than other methods for the top 10 recommended candidates. To sum up, we make the following contributions to applications, methodology, and experimentation, respectively:

- **Application:** We develop a novel item-item cross-platform recommender system, *paper2repo*, that can automatically recommend repositories on GitHub to a query paper in the academic search system.
- **Methodology:** We propose a joint model that incorporates the text encoding technique into the constrained GCN model to jointly learn and map the embeddings of papers and repositories from two different platforms into the same space.
- **Experimentation:** Evaluation results demonstrate that the proposed *paper2repo* produces better recommendations than other state of the art methods.

<sup>1</sup><https://academic.microsoft.com>

<sup>2</sup><https://news.microsoft.com/2018/06/04/microsoft-to-acquire-github-for-7-5-billion/>

<sup>3</sup><https://github.com/dennybritz/cnn-text-classification-tf>

<sup>4</sup><https://github.com/tensorflow/models/tree/master/official/transformer>



Figure 1: An example of the tensorflow repository

The rest of the paper is organized as follows. Section 2 presents the motivation, background, and formulation for our *paper2repo* problem. In Section 3, we introduce the overall architecture of the *paper2repo* recommender system. Section 4 evaluates its performance on real-world data sets. Section 5 summarizes related work. Finally, we conclude the paper in Section 6.

## 2 PRELIMINARIES

In this section, we first introduce relevant features of GitHub, and use a real-world example to motivate our work. We then review traditional graph convolutional networks (GCNs) that constitute the analytical basis for our solution approach. Finally, we formulate the problem of GitHub repository recommendation and offer the main insight behind the adopted solution.

### 2.1 Background and Motivation

GitHub is a social application platform, where users can collaborate on open source projects and share code with others. Users can create different repositories (*i.e.*, digital directories) to store files and source code for their research projects. Fig. 1 illustrates an example of the Tensorflow repository for deep learning. The repository contains basic information such as descriptions, tags (also called topics), and user starring. Descriptions refer to a couple of sentences that describe the repository. Tags are keywords used to classify a repository, including its purpose and subject area. Starring a repository represents an expression of interest from a user in the repository. Much like browser bookmarks, stars give users improved shortcut-like access to repository information.

Consider a simple example to motivate features of our solution approach based on real-world data sets collected from GitHub and Microsoft Academic. Fig. 2 illustrates selected repositories and papers in the two different platforms. Repository  $R_1$  is the original open source code for paper  $P_1$ , ResNeXt [25], which presents a highly modularized network architecture for image classification developed by Facebook Research. The pair of  $R_1$  and  $P_1$  could be used as a bridge to connect the GitHub platform and Microsoft Academic platform. In addition, we find that users who star the repository  $R_1$  also star certain other repositories related to image classification and deep neural networks. We collect 540 users who star repository  $R_1$  and discover that 241 of them (such as user  $U_1$ ) star  $R_4$  as well. Thus, we infer that the two repositories starred by many of the same users are likely related to each other. We can also find repositories in the same subarea or topic that have similar tags. Take repository  $R_2$  and  $R_3$ , for example. These two repositories are related to image classification and they have the same tag “ImageNet”. In this paper, we use both user starring and tags to construct a context graph for repositories.

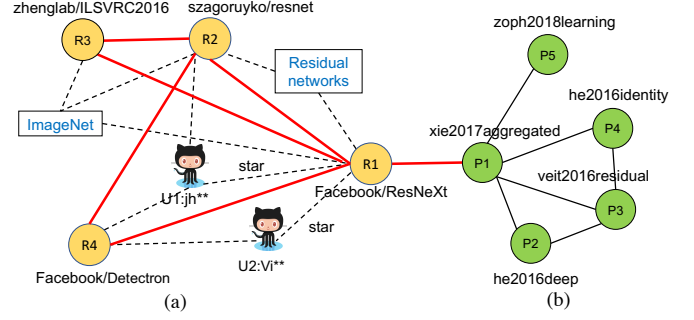


Figure 2: A real-world example of repository-repository context graph on GitHub and paper citation graph on Microsoft Academic. In this figure, each orange circle represents a repository while each green circle represent a paper. In addition, repository  $R_1$  is the original open source code of paper  $P_1$ , which could be used to bridge the two different platforms.

In the papers domain, related papers are connected together via citation relationships. As illustrated in Fig 2, paper  $P_1$ , called ResNeXt, develops a deep residual networks for image classification. It has more than 100 citing/cited papers. Most of them are also involved in the same or similar areas as  $P_1$ . For example, paper  $P_1$  cites paper  $P_2$ ,  $P_3$ , and  $P_4$  that are related to image classification using deep neural networks. Paper  $P_3$  cites  $P_2$  and  $P_4$  because they share the same research topics on deep residual networks. Hence, related papers can be identified from the paper citation graph. We further find that the more edges (connections) exist among papers, the more related they are. These observations form the basis of constructing the context graph for publications.

Finally, we feed both the context graphs (describing relations) and content information (describing nodes), such as raw paper abstracts and repository descriptions, into the graph convolutional neural networks framework to perform joint embedding.

### 2.2 Graph Convolutional Networks

Graph convolutional networks (GCNs) [2, 12] have been widely used in classification and recommendation systems because they encode both the graph neighborhood and content information of nodes to be classified or recommended. The basic idea of GCNs is to learn the embeddings of nodes based on both their input features,  $\mathbf{X}$ , and graph structures. The graph structure is often denoted by an adjacency matrix,  $\mathbf{A}$ . A multi-layer GCN forward model could be expressed by:

$$\mathbf{H}^{(l+1)} = \sigma\left(\tilde{\mathbf{D}}^{-\frac{1}{2}}\tilde{\mathbf{A}}\tilde{\mathbf{D}}^{-\frac{1}{2}}\mathbf{H}^{(l)}\mathbf{W}^{(l)}\right), \quad (1)$$

where  $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}_N$ ,  $\mathbf{A}$  is the adjacency matrix for the input graph structure and  $\mathbf{I}_N$  is the identity matrix,  $\tilde{\mathbf{D}}_{ii} = \sum_j \tilde{\mathbf{A}}_{ij}$ ,  $\mathbf{W}^{(l)}$  is the weight matrix to be trained,  $\sigma(\cdot)$  is the activation function such as ReLU,  $\mathbf{H}^{(l)}$  is the hidden layer,  $l$  is the layer index and  $\mathbf{H}^{(0)} = \mathbf{X}$  is the input features of nodes.

Because the embeddings of papers and repositories generated by traditional GCNs for the two different platforms are not in the same space, it is necessary for us to jointly train the corresponding

neural networks subject to a constraint that aligns the projections of related papers and repositories onto the same (or neighboring) points in the embedding space. The above challenge leads to the following research problem formulation.

### 2.3 Problem Statement and Solution Idea

Suppose there are  $N_p$  papers in an academic search system, such as Microsoft Academic, and  $N_r$  repositories on GitHub. Some papers explicitly name their corresponding original repositories of source code on GitHub. We call these papers *bridge papers*. They can be used to connect the two platforms. For instance, paper  $P_1$  in Fig. 2 is a bridge paper that connects to its original repository  $R_1$  on GitHub. Each paper has an abstract, while each repository has tags and descriptions. Our main task is to learn high-quality embeddings of papers and repositories that can be used for recommendation across the two different platforms based on their content information and graph structures.

To this end, we construct two *undirected* context graphs, one for the paper platform and one for the repository platform, respectively. These graphs are introduced below:

- **Paper-citation graph:** We model the papers and their references as an undirected paper-citation graph, because only the connection between two papers is required in this work. We adopt an adjacency matrix with binary measurements to denote the graph structure. In this graph, each paper,  $p_i$ , is a node. Each node has an abstract as its input attributes.
- **Repository-repository context graph:** Since there is no direct citation relationship between repositories, it is more difficult to construct the repository-repository context graph. Motivated by the example in Section 2.1, we leverage tags and user starring to construct this graph. Specifically, as previously motivated in Fig. 2, we add a binary edge between two repositories if they are starred together by at least one user or share at least one term whose TF-IDF score is over a threshold (e.g., 0.3) in their description or tags. While binary edges ignore information available on the degree of overlap (in tags or starring users), we find that the simplification is adequate, especially that (at the end of the day) our embedding also considers full text of paper abstracts, repository descriptions, and tags.

After obtaining these two context graphs, we can combine the graph structures with the input content information of papers and repositories to generate high-quality embeddings. We can then compute similarity scores between neighboring paper and repository embeddings to recommend highly related repositories to a query paper. Namely, the top (*i.e.*, nearest) repository candidates will be recommended to the query papers. Below, we elaborate this general idea in more detail.

## 3 MODEL

In this section, we first introduce the overall architecture of the proposed *paper2repo* recommender system, then present techniques for model training.

### 3.1 *paper2repo* Architecture

The key contribution of this paper lies in developing a *joint embedding* that incorporates both text encoding and constrained GCN to generate embeddings that take into account similarity (between repositories and papers) in both text features and context graph structures.

Fig. 3 shows the overall architecture of the proposed *paper2repo* system. There are two main modules: text encoding and constrained GCN. The first module tries to develop a text encoding technique to encode the content information of papers and repositories into vectors as inputs of the constrained GCN model. The second module implements the GCN model with an added constraint. The constraint specifies that repositories explicitly mentioned in paper descriptions must be mapped to the same place as the referring paper, thereby linking the mappings of repositories and papers into a single consistent space. We detail these two modules in the following.

**Text encoding.** Fig. 4 shows the detailed framework of using text encoding to learn the features from both the tags and descriptions of repositories on GitHub. In general, the descriptions of repositories are limited and simple, so it is difficult to learn the overall features of repositories from them. Hence, we add the tags from repositories to enrich the descriptions, because tags often extract some key information about the topics of the repositories on GitHub. Therefore, we design the text encoding module to encode the features of both tags and descriptions. As illustrated in Fig. 4, there are four steps for description encoding and tag encoding of a repository:

- Step 1: convolution with multiple filters;
- Step 2: max-over-time pooling;
- Step 3: transformation with fully connected layers;
- Step 4: feature fusion.

We first use Step 1 and Step 2 to encode the descriptions of a repository as shown in Fig. 4. Let  $\mathbf{x}_i \in \mathbb{R}^k$  denote the  $k$ -dimensional word vector of the  $i$ -th word in the description. The word vector comes from the pre-trained embeddings of Wiki words offered by the GloVe algorithm [15]. The length of each description is fixed to  $n$  (padded or cropped where necessary). So a description can be defined by

$$\mathbf{x}_{1:n} = \mathbf{x}_1 \oplus \mathbf{x}_2 \oplus \dots \oplus \mathbf{x}_n, \quad (2)$$

where  $\oplus$  is the concatenation operator that concatenates all the embeddings of words for a sentence. Then we apply filters  $\mathbf{w} \in \mathbb{R}^{hk}$  for the convolution operation in order to learn new features, where  $h$  is the window size of words for filters. Each filter computes the convolution to generate a new feature,  $c_i$ , from a possible window of words  $\mathbf{x}_{i:i+h-1}$ , which can be expressed by

$$c_i = f(\mathbf{w} \cdot \mathbf{x}_{i:i+h-1} + b), \quad (3)$$

where  $b$  is a bias parameter and  $f(\cdot)$  is a non-linear activation function such as ReLU. As the window slides from  $\mathbf{x}_{1:h}$  to  $\mathbf{x}_{n-h+1:n}$ , the filter yields a feature map

$$\mathbf{c} = [c_1, c_2, \dots, c_{n-h+1}]. \quad (4)$$

After obtaining the feature map, a max-over-time pooling operation is adopted to get the maximum value,  $\max\{\mathbf{c}\}$ , of each output feature map as illustrated in Step 2.

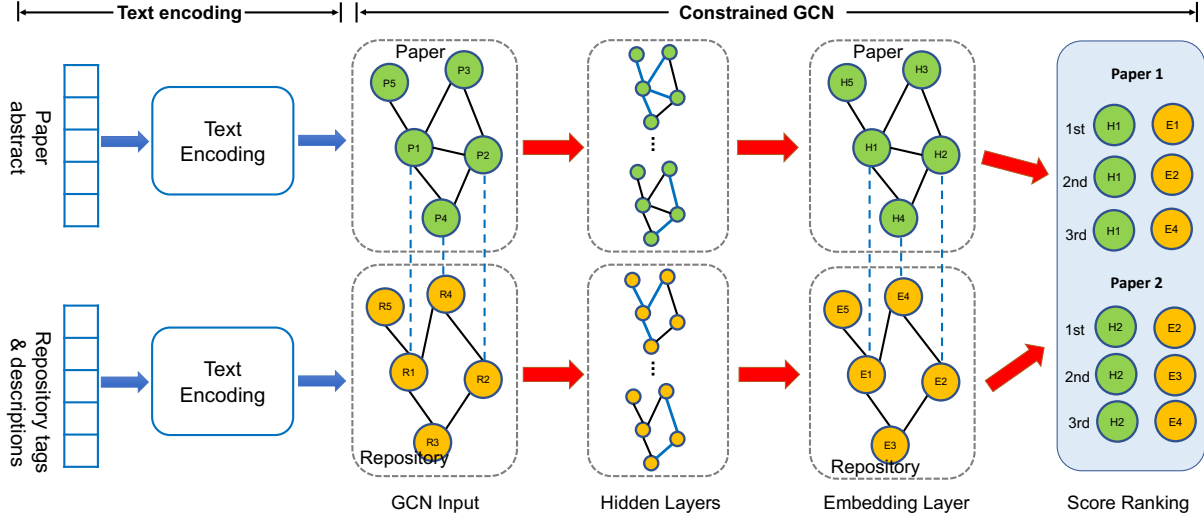


Figure 3: Overall architecture of the *paper2repo* system. There are two main modules: text encoding and constrained GCN model. The inputs of the *paper2repo* are the content information and graph structures of papers and repositories.

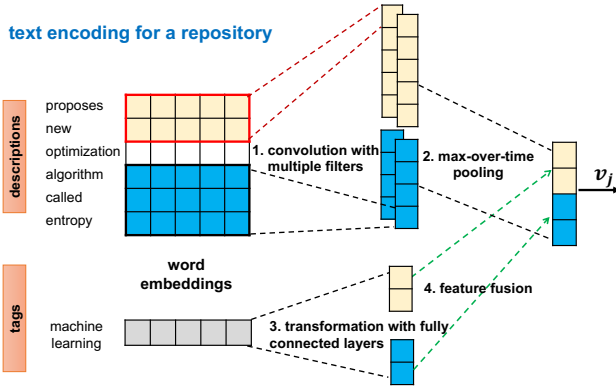


Figure 4: The framework of text encoding for tags and descriptions of repositories.

Next, we implement Step 3 to encode tags. Since there is no sequence for tags, we leverage fully connected layers to learn their features. For each tag, we first use the *fastText* trick [11] to get its word representation via merging the embeddings of each word. Here, *fastText* is a simple text encoding technique where word embeddings are averaged into a single vector. As illustrated in Fig. 4, the embedding of “machine learning” would be denoted by a new word vector by adding the word vectors of “machine” and “learning”. We then apply fully connected layers to produce the new features whose dimensions are aligned with the number of feature maps for description encoding. After that, feature fusion is adopted in Step 4 to add the new features of tags generated in Step 3 to the produced features in Step 2. Finally, the new features,  $v_j$ , integrated with description encoding and tag encoding are input of the constrained GCN model. In order to improve the stability

of model training, we adopt batch normalization to normalize the new features.

Similarly, for abstract encoding of a paper, we can propose the same methodology to learn the new features as the description encoding of a repository in Step 1 and Step 2. For brevity, we will not describe it in detail.

**Constrained GCN model.** Since papers and repositories are in two different platforms, the generated embeddings by the traditional GCN are not in the same space. Thus, we propose a constrained GCN model to constrain the embeddings to the same space. Specifically, it leverages the general GCN as the forward propagation model in Equation (1) and minimizes the distance between the embeddings of some bridge papers and their original repositories as a constraint. We use the cosine similarity to measure the distance. In order to compute the cosine similarity distance for the embeddings, we normalize their embeddings as  $\mathbf{p}_i$  and  $\mathbf{r}_j$ , respectively. Let  $\mathbf{p}'_i$  and  $\mathbf{r}'_j$  denote the normalized embeddings of the  $i$ -th bridge paper and the corresponding original repository, respectively. Then, for each pair of bridge paper and repository, the corresponding constraint could be expressed as:

$$1 - \mathbf{p}'_i{}^\top \mathbf{r}'_j \leq \epsilon, \quad (5)$$

where  $\epsilon$  is a small error term, such as 0.001.

In this paper, we adopt weighted approximate-rank pairwise (WARP) loss [24] to train the *paper2repo* model in order to recommend the target repositories to a query paper at the top of the ranked candidates. Let  $\mathcal{L}$  be the WARP loss for the labeled pairs of papers and repositories during training, and  $m$  be the number of pairs for bridge papers and their original repositories in the training data. Then, the constrained GCN model is defined as:

$$\begin{aligned} \min \quad & \mathcal{L} \\ \text{subject to} \quad & \sum_{i=1}^m (1 - \mathbf{p}'_i{}^\top \mathbf{r}'_j) \leq \epsilon. \end{aligned} \quad (6)$$

In the above (6), the WARP loss function  $\mathcal{L}$  is defined by:

$$\mathcal{L} = \sum_{n_k} L(\text{rank}^\Delta(\mathbf{p}^\top \mathbf{r}_+)) \frac{|\Delta - \mathbf{p}^\top \mathbf{r}_+ + \mathbf{p}^\top \mathbf{r}_{n_k}|_+}{\text{rank}^\Delta(\mathbf{p}^\top \mathbf{r}_+)}, \quad (7)$$

where  $(0 < \Delta < 1)$  is the margin hyper-parameter;  $n_k$  denotes the number of negative examples in a batch during model training;  $\mathbf{r}_+$  and  $\mathbf{r}_{n_k}$  denote the representations of positive and negative repositories, respectively;  $|t|_+$  is the positive part of  $t$ . In addition,  $L(\cdot)$  and  $\text{rank}^\Delta(\mathbf{p}^\top \mathbf{r}_+)$  will be introduced below.

First,  $L(\cdot)$  in (7) is a transformation function which transforms the rank into a loss, defined as

$$L(K) = \sum_{j=1}^K \frac{1}{j}, \quad (8)$$

where  $K$  denotes the position of a positive example in the ranked list. In addition,  $\text{rank}^\Delta(\mathbf{p}^\top \mathbf{r}_+)$  is the margin-penalized rank of positive examples, defined as:

$$\text{rank}^\Delta(\mathbf{p}^\top \mathbf{r}_+) = \sum_{n_k} I(\Delta - \mathbf{p}^\top \mathbf{r}_+ + \mathbf{p}^\top \mathbf{r}_{n_k} > 0), \quad (9)$$

where  $I(\cdot)$  is the indicator function that outputs 1 or 0.

After defining the WARP loss, the next step is to solve the above non-linear optimization model in (6). One possible solution is to transform it into a dual problem. Based on the method of Lagrange multipliers, we can rewrite the above constrained model (6) as:

$$\min \quad \mathcal{L} + \lambda \sum_{i=1}^m (1 - \mathbf{p}_i^\top \mathbf{r}'_i). \quad (10)$$

It is very hard to obtain a closed-form solution that minimizes the objective function (10) due to the nonlinear nature of the optimization problem. Instead, we feed this objective function as a new loss function to the neural network. Convergence of training of this neural network to a solution that meets both the original loss function,  $\mathcal{L}$  (left term in objective function 10), and the joint embedding constraint (right term in objective function 10), requires that the two terms have comparable weights. Otherwise, learning gradients will favor optimizing the larger term and ignore the smaller. Unfortunately, the loss function  $\mathcal{L}$  dynamically drops during training, making it difficult to choose the hyper-parameter,  $\lambda$ , to create gradients that properly trade off the loss function,  $\mathcal{L}$ , and the constraint error. To address this issue, we replace  $\lambda$  in the second term with the varying WARP loss,  $\mathcal{L}$ , and normalize the constraint error instead of using the total error. Accordingly, our model can be formulated as:

$$\min \quad (1 + C_e) \mathcal{L}, \quad (11)$$

where  $C_e$  is the average constraint error, defined as:

$$C_e = \frac{1}{2m} \sum_{i=1}^m (1 - \mathbf{p}_i^\top \mathbf{r}'_i), \quad (12)$$

Note that,  $C_e \in [0, 1]$ . This is because the cosign similarity for any two vectors,  $\mathbf{p}_i^\top \mathbf{r}'_i$ , ranges between 1 and -1. Thus, each term in the summation above ranges between 0 and 2. The entire summation ranges between 0 and  $2m$ , and the normalization leads to a value of  $C_e$  between 0 and 1. Using the mean constraint error,  $C_e$ , helps keep the scale of the two terms in the objective function (11) comparable. While hypothetically, we can even remove the constant, 1, from

**Table 1: Main notations in our model.**

Notations	Descriptions
$\mathbf{p}_i$	$i$ -th paper among all the papers
$\mathbf{r}_j$	$j$ -th repositories among all the repositories
$\mathbf{p}'_i$	$i$ -th bridge paper
$\mathbf{r}'_i$	bridge (original) repository of $i$ -th bridge paper
$\mathcal{L}$	WARP loss
$C_e$	Mean of the constraint error
$\Delta$	Margin hyper-parameter for WARP loss
$T$	Number of top related repositories for a paper

objective function (11), we find that keeping it improves numeric stability and convergence.

In the new formulation, we no longer need to dynamically adjust the hyper-parameter,  $\lambda$ , for model training. The constrained optimization problem in (6) can be solved by minimizing the new loss function (11) by feeding it to the graph neural network.

### 3.2 Model Training

We summarize the main notations used in our model in Table 1. The output of our trained network is the ranked inner products of the embeddings for pairs of papers and repositories. The closer the embeddings of a paper and a repository, the higher their inner product, and the higher the ranking. Given that output, one can identify the top ranked recommended repositories for a given paper (or the top ranked papers for a given repository). Training computes the weights and biases of each convolutional neural network layer, the weights and biases of each fully connected layer (for tag encoding), and the weights and biases of filters for the convolution operation. Besides, we need to tune the number of filters used for text encoding, the filter window  $h$ , the number of fully connected layers for tag encoding of repositories, the number of hidden layers of neural networks, and the dimension of output embeddings of each node as well as the margin hyper-parameter for WARP loss. We set the output dimension of representations (embeddings) of each paper and repository equal in order to compute their similarity scores. We then select positive and negative samples as follows:

- **Positive samples:** We use the bridge papers and the corresponding repositories as labeled training data. Let  $\mathbf{p}_i$  be the  $i$ -th paper and  $\mathbf{r}_j$  be its highly related repositories. Such pairs,  $(\mathbf{p}_i, \mathbf{r}_j)$ , constitute positive samples. We further hypothesize that if users, who star repository A, also star repository B more often than C, then (on average) B is more related than C to repository A. Thus, besides bridge repositories, we collect the top  $T$  related repositories, ranked by the frequency they are starred by users who star the original bridge repository. In order to get more training data, we also sample some one-hop neighbors of bridge papers combined with the corresponding bridge repositories to be positive examples (at most  $T$ ). Different values of  $T$  represent how liberally one interprets “related” (and will result in different recommender performance). We can think of this method as a form of distant supervision; an imperfect heuristic to label more samples than what can be inferred from explicit bridge references. It is important to note that we *do not* use the

**Algorithm 1:** Constrained GCN algorithm

---

**input** : Features of papers and repos  $X_p, X_r$ , adjacency matrix  $A_p, A_r$  for paper-citation graph and repository-repository graph

**output**: Embeddings of paper  $p_i$  and repository  $r_j$

```

1 /* For paper embeddings */
2  $\tilde{A} = A_p + I_N, \tilde{D} = \sum_j \tilde{A}_{ij}, H^{(0)} = X_p$ ;
3  $p_i \leftarrow \text{ReLU}(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)})$ ;
4  $p_i \leftarrow p_i / \|p_i\|_2$ ;
5 /* For repo embeddings */
6  $\tilde{A} = A_r + I_N, \tilde{D} = \sum_j \tilde{A}_{ij}, H^{(0)} = X_r$ ;
7  $r_j \leftarrow \text{ReLU}(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^{(l)} W^{(l)})$ ;
8  $r_j \leftarrow r_j / \|r_j\|_2$ ;
9 /* Constraint for paper  $p'_i$  and repository  $r'_i$  */
10  $1 - p_i^\top r'_i \leq \epsilon$ ;
11 /* Compute similarity score */
12  $y_{score} \leftarrow p_i^\top r_j$ ;

```

---

same heuristic for recommender *evaluation*. Rather, as we describe in a later section, we use Mechanical Turk workers to evaluate resulting recommendation accuracy, allowing us to understand the efficacy of our distant supervision framework.

- **Negative samples:** We also introduce negative examples, referring to repositories that are not highly related to a query paper. In this work, we randomly sample  $n_k$  negative examples of repositories across the entire graph to train the model. We expect the similarity scores of positive examples to be larger than those of negative examples.

We briefly summarize the proposed constrained GCN algorithm in **Algorithm 1**. In this algorithm, the first  $l$  layers learn the hidden vectors (embeddings) of each paper and repository using the GCN algorithm in (1). Then we use the normalized embeddings of paper  $p_i$ , and repository  $r_j$  to compute their similarity scores. Additionally, we try to minimize the distance between the embeddings of bridge papers and their original repositories as an added constraint.

## 4 EXPERIMENTS

We carry out experiments to evaluate the performance of the proposed *paper2repo* on the real-world data sets collected from GitHub and Microsoft Academic. Further, we show that the proposed model performs well on larger data sets without any hyper-parameter tuning. In addition, we conduct ablation experiments to explore how design choices impact performance. Finally, we conduct a case study to evaluate the effectiveness of the proposed method.

### 4.1 Datasets

We collected a repository data set and a research paper data set from GitHub and Microsoft Academic, respectively. Microsoft Academic provides an API for users to query the detailed entities of academic papers, such as paper title, abstract, and venue names. For a proof of concept, we query the top 20 venue names and 8 journals of computer science, such as KDD, WWW, ACL, ICML, NIPS, CVPR and TKDE, to retrieve the entities of 59,404 raw papers from year

2010 to 2018. After that, we query for the titles of these papers to collect some original open source repositories through the GitHub API. We obtain about 2,427 original repositories corresponding to the named papers. We define bridge papers as those for which we found a matching repository (which we call the bridge repository). In addition, we collect about 8,000 popular repositories from users who star the bridge repositories. After data cleaning and preprocessing, we have 32,029 research papers and 7,571 repositories, including 2,107 bridge repositories. In our experiments, we evaluate the performance of the proposed model on both small and full data sets as illustrated in Table 2.

**Table 2: Information summary about paper and repositories data sets.**

Data	# papers	# repositories	# bridge papers
Small	11,272	7,516	1,386
Large	32,029	7,571	2,107

**Testing and ground truth estimation.** To evaluate the accuracy of our recommender system, its results need to be compared to ground truth (that is not part of training data). For that purpose, we asked human graders on Amazon Mechanical Turk (AMT) to evaluate produced recommendations. Specifically, for each pair to be evaluated, we provided the graders with (i) the paper title, abstract, and (ii) the description, README (if available) and URL the corresponding repository on GitHub. The graders were asked to grade similarity on a three point scale: Score “2” meant that the paper and repository were highly related; score “1” meant that they were somewhat related; and score “0” meant that the pair was not related. Three graders were used per pair. After labelling, we considered those pairs that received a score of 1 or 2 from all graders to be correct matches. Pairs that received a score of 1 or 0 from all graders (and received at least one zero) were considered to be incorrect matches. Finally, pairs where graders disagreed, receiving scores 0 and 2 together (about 7% result), were graded by an additional grader. An average was then computed. If the average score was less than 1, the pair was considered unrelated (an incorrect match). Otherwise, it was considered a correct match.

It remains to describe how we partition data into the training set and test set. We already described how positive and negative *training* pairs were selected. It is tempting to try and choose *test* papers and repositories at random. However, due to the large number of papers and repositories available on the respective platforms, this random sampling leads to a predominance of unrelated pairs, unless the test set is very large (which would be very expensive to label). Clearly, restricting testing to predominantly unrelated papers and repositories would artificially reduce the recommender’s ability to find matches. In other words, the test set has to include related pairs. Hence, we first selected 580 bridge papers and their one-hop neighbors *not present in the training set*. We then included their repositories as well as the two hop neighborhood of those repositories according to the context graph. Thus, for each paper in the test set, we included some repositories that are likely related to different degrees, and hundreds of repositories that are likely not related (i.e., repositories related to other papers). This allowed



for a much richer diversity in the test set and Mechanical Turk outputs. It is important to underscore that the criteria above were used merely for paper and repository selection for inclusion in the test set. The *labeling* of the degree of match for pairs returned by the recommender was done strictly by human graders and not by machine heuristics.

## 4.2 Experimental Settings

**Baseline methods.** We compare the proposed *paper2repo* with the algorithms below:

- NCSR [23]: This is a cross-domain recommendation framework that combines deep fully connected layers with graph Laplacian to recommend items from information domain to social domains.
- KGCN [21]: This method leverages knowledge graph and graph convolutional neural networks to recommend interested items to users.
- CDL [20]: This is a hybrid recommendation algorithm, which jointly performs deep representation learning for the content information and collaborative filtering for the ratings matrix.
- NCF [7]: This is a neural collaborative filtering model for recommendation, which combines the matrix factorization (MF) and MLP to learn the user-item relationship.
- LINE [18]: This is a graph embedding algorithm that uses BFS to learn the embedding of nodes in the graph with unsupervised learning. In order to better perform LINE, we construct the entire graph of papers and repositories via the bridge papers and their original repositories.
- MF [19]: This method is widely used for traditional recommender systems due to its good performance for dense data. It is a supervised algorithm, so we use 50% positive examples as training data and the remaining data as testing data.
- BPR [17]: This is an optimized MF model with Bayesian analysis for implicit recommendation. It is a supervised model and we use the same method as MF to do experiments.

**Evaluation measures.** In order to evaluate the performance of the proposed *paper2repo* recommender system, this paper adopts three commonly used information retrieval measures: HR@K (hit ratio), MAP@K (mean average precision), MRR@K (mean reciprocal rank). In general, HR is used to measure the accuracy about the number of correct repositories recommended by the *paper2repo* system, while MAP and MRR measure the accuracy of the rank of recommended items.

**Model architecture and parameters.** In order to compare the performance of our proposed *paper2repo* to the other recommender algorithms, we tune the hyper-parameters during model training. We use 90% of *training data* to train the model and the remaining training data as validation data. After extensive experiments (not shown), we obtain the best parameters for our model below: we set the number of layers to 2 for the graph convolutional neural networks and the size of each layer to 256. The number of fully connected layers is 2 for tags encoding of repositories. The length of each abstract from papers, and description of repositories is fixed to 200 and 50, respectively. For paper abstract encoding, the filter window ( $h$ ) is 2, 3, 5, 7 with 64 feature maps each. For repository descriptions encoding, its filter windows ( $h$ ) is 2, 4 with 64 and 32

feature maps, respectively. The pre-trained embeddings of words produced by GloVe are used with size of 200. We set the learning rate to 0.0005. For training, we set the number of positive repositories,  $T$ , to 6 given a paper. In addition, we randomly sample 44 negative samples to train the model. We set the margin hyper-parameter,  $\Delta$ , to 0.5 in this experiment. For testing, we sample 50 examples include  $T$  positive examples and the remaining  $50 - T$  negative example to measure the three metrics above. Our evaluation results are obtained by averaging 3 runs per point.

## 4.3 Performance Comparison

To understand dependence on scale, we start with a subset of 11, 272 papers and 7, 516 repositories, including 1, 386 pairs of bridge papers and original repositories, to conduct experiments (later we shall use the full data set). We compare the performance of the proposed *paper2repo* with the competing methods. Fig. 5 illustrates the comparison results of different methods as  $K$  increases with three measures: HR@K, MRR@K and MAP@K. We can observe that the proposed *paper2repo* outperforms the other recommendation algorithms for all the three metrics. This is because our *paper2repo* encodes both the contents and graph structures of papers and repositories to learn their embeddings. In addition, *paper2repo* performs better than the cross-domain NCSR algorithm for two reasons. First of all, NCSR leverages spectral embeddings method without using node features in social domain, while our model encodes both content information and graph structures. Secondly, the items in our data sets have more attributes (keywords and tags) than that (20 attributes) in the NCSR paper, making it hard to train the NCSR model. In Fig. 5(c), we also find that the MAP of *paper2repo* slightly drops. This reason is that more positive repositories are recommended but not at the top of ranking list, leading to lower MAP. Moreover, we can observe that CDL and NCF do not perform well in our case. The main reason is that the number of positive examples for model training is very small such that they suffer from data sparsity as the traditional MF. Besides, KGCN does not work very well because it only adopts one hot embedding without taking node features into account in the paper domain.

## 4.4 Performance on Larger Data Sets

We conduct another experiment to evaluate the performance of the *paper2repo* on larger data sets. We do not otherwise perform any data-set-specific tuning. We change the number of papers from 11, 272 to 32, 029, and the number of pairs of bridge papers and their original repositories from 1, 386 to 2, 107. Fig. 6 illustrates the comparison results for two different sizes of data sets. From this figure, it can be observed that the proposed *paper2repo* performs better on larger data sets than that on small data sets in terms of MAP and MRR, and their hit ratios are comparable. This is because our model can learn more information from different nodes in a large-scale graph.

## 4.5 Ablation Studies

**Effect of Pre-trained Embeddings.** We first explore how the settings of pre-trained embeddings impact the performance of the proposed *paper2repo*. We compare three different cases: (i) fixed pre-trained embeddings; (ii) pre-trained embeddings are used as



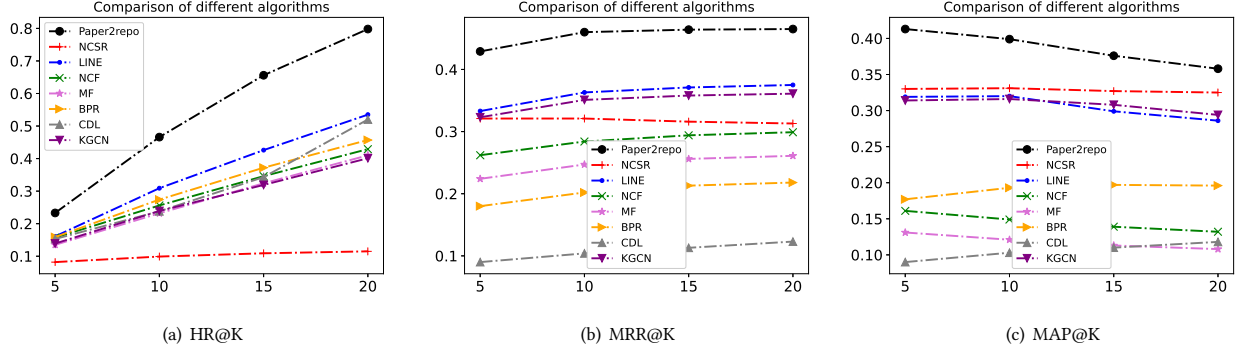
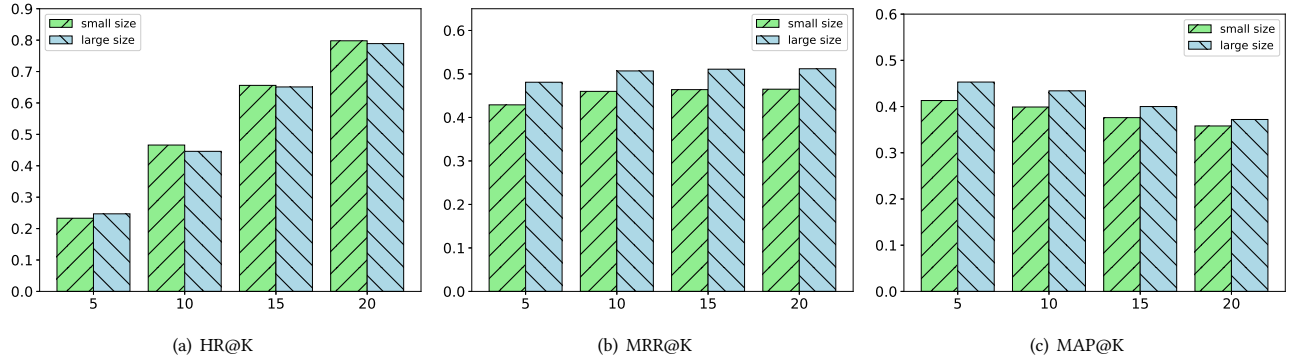
Figure 5: Performance comparison of different methods as the number of candidates  $K$  increases.

Figure 6: Performance comparison of different sizes of data sets using same parameters.

initialization (not fixed); (iii) fixed and not fixed embeddings are concatenated together. Table 3 illustrates the comparison results for these three cases when  $K = 10$ . We can see that when the pre-trained embeddings are fixed, it performs better than the other two cases. The main reason is that the pre-trained embeddings are produced from the large Wikipedia corpus by GloVe. The performance of concatenated embeddings seems not very good, because it is more difficult to train the model with more complicated networks.

Table 3: Performance comparison for different settings of pre-trained embeddings when  $K = 10$ ,  $T = 6$ .

Pre-trained embeddings	MAP	MRR	HR
fixed	<b>0.399</b>	<b>0.460</b>	<b>0.466</b>
not fixed	0.330	0.364	0.367
fixed & not fixed	0.332	0.374	0.374

**Effect of top  $T$  positive repositories.** Next, we explore how the number of positive repositories in training,  $T$ , influences the performance of *paper2repo*. In our experiment, the number of positive repositories,  $T$ , varies from 3 to 7 with step 1. From Table 4, it can be seen that the performance of *paper2repo* gradually boosts as  $T$  increases from 3 to 6. When  $T = 6$  and 7, their performance are close to each other. Thus, we can find that *paper2repo* has a better performance when the number of positive repositories in training is large in some degree. In this paper, we use  $T = 6$  to do experiments

Table 4: Performance comparison for different  $T$  when  $K = 10$ .

Top $T$	3	4	5	6	7
MAP	0.269	0.323	0.350	<b>0.399</b>	0.388
MRR	0.291	0.36	0.390	<b>0.460</b>	0.438
HR	0.409	0.414	0.461	<b>0.466</b>	0.432

because most of papers in the testing data labelled by graders have at most 6 positive examples.

**Effect of the Margin Hyper-parameter.** We also study the impact of the margin hyper-parameter on the performance of the *paper2repo*. We change the margin parameter,  $\Delta$ , from 0.1 to 0.7 with step 0.1 while keeping the other hyper-parameters unchanged. Table 5 illustrates the evaluation results for different margin parameters when  $K = 10$ . We can observe that the performance of *paper2repo* gradually increases as  $\Delta$  rises from 0.1 to 0.4, and then gradually drops as  $\Delta$  increases from 0.4 to 0.7. This is to say, when  $\Delta = 0.4$ , it performs best among them. The main reason is that as margin  $\Delta$  is too small, it is hard to separate the positive and negative examples. At the same time, a larger  $\Delta$  may result in higher loss during training.

**Effect of Number of Bridge Papers.** Finally, we explore how the number of bridge papers and repositories affect the recommendation performance of *paper2repo*. Table 6 illustrates the comparison results under different ratios of 1, 386 bridge papers in the small

**Table 5: Performance comparison for different margin parameters when  $K = 10$ .**

$\Delta$	0.1	0.2	0.3	0.4	0.5	0.6	0.7
MAP	0.394	0.392	0.394	<b>0.399</b>	0.349	0.307	0.274
MRR	0.436	0.456	0.440	<b>0.460</b>	0.400	0.345	0.312
HR	0.446	<b>0.500</b>	0.466	0.466	0.412	0.381	0.310

data set. We can observe from it that the evaluation metrics, HR, MAP and HRR, gradually rise as the number of bridge papers increases. This is because more bridge papers and repositories can make the embeddings of the similar nodes closer to each other in the graph.

**Table 6: Performance comparison for different number of bridge papers when  $K = 10$ .**

Ratio	0.4	0.6	0.8	1.0
MAP	0.298	0.314	0.341	<b>0.399</b>
MRR	0.325	0.364	0.391	<b>0.460</b>
HR	0.339	0.369	0.379	<b>0.466</b>

#### 4.6 Discussion of *pape2repo*

We further discuss the advantages and disadvantages of *pape2repo*. According to our experiments, we discover that the recommended repositories are relevant to a query paper when there exists substantial user co-starring between bridge repositories and other repositories or when there are multiple overlapped tags between them. The main reason is that two repositories starred by many of the same users or that have multiple overlapped keywords are very likely to involve similar research topics. However, when only few users star both repositories, the recommendation performance is not very good. For example, when two repositories are only starred by a couple of users together, it is hard to judge whether these two repositories are similar or not. As a result, the recommended repositories seem not to be very relevant to the query papers. The need to find a sufficient number of bridge repositories is another limitation. Extensions of this joint embedding framework to domains with no natural analogue to bridge papers/repositories can be a good topic for future investigation.

Cold start is one of the most important research topics in recommender systems. As we know, when users release new source code, their repositories have very little user starring in the beginning. Thus, in practice, cold start is an issue similar to lack of sufficient co-starring, discussed above. Our *pape2repo* is able to partially deal with cold start because, as mentioned in Section 2.3, we construct the repository-repository graph using tags as well. Even if a repository has very few stars, we can still use its tags to construct the repository graph. Therefore, we can still recommend some repositories to the query papers, although the quality of recommendation will be impacted.

The accuracy of our evaluation results is impacted by the accuracy of estimating ground truth. Due to budget constraints, we used three Mechanical Turk graders per item, when they coincided. Grading reliability can be improved by increasing the graders and performing grader training.

Other limitations include the general applicability of repository recommendations. While software libraries are becoming an increasingly important research enabler, some fields (including theory) are less likely to benefit from this system. The system is also less relevant to research enabled by artifacts not in the public domain. One may also argue that authors will tend to cite relevant repositories as a best practice. This is true, but our recommender system can also uncover *subsequently created* repositories that impact a given paper, such as novel implementations of relevant libraries, or algorithm implementations on new hardware platforms and accelerators. This is the same reason why one might not exclusively rely on citations in a paper to find other relevant papers, especially if the paper in question is a few years old.

## 5 RELATED WORK

This section reviews related work on recommender systems that use deep learning and graph neural networks, especially for cross-domain recommendations.

### 5.1 Cross-domain Recommendation

In order to deal with data sparsity and cold start issues in recommender systems, recent research proposed cross-domain recommendations [3, 9, 10, 23] to enrich user preferences with auxiliary information from other platforms. Cross-domain recommendations get rich information on items that users prefer buying based on their history on multiple platforms. For example, some users like to purchase furniture on Walmart and buy clothes on Amazon. So we can recommend furniture to the users on Amazon next time, which can improve the diversity of recommendations. Motivated by this observation, a multi-view deep learning model was proposed by Elkahky *et al.* [3] to jointly learn the users' features from their past preferred items on different domains. Following this work, Lian *et al.* [13] further developed a CCCFNet model that integrates collaborative filtering and content-based filtering in a unified framework to address the data sparsity problem. However, these algorithms mainly adopt the idea of transfer learning to boost users' preferences via jointly learning users' items on multiple platforms or domains. In addition, Wang *et al.* [23] proposed a user-item cross-domain framework, NSCR, to recommend items from information domains to users on social media. However, a limitation is that NSCR is a supervised learning algorithm. In contrast, we propose a item-item cross-domain recommendation framework with a distant-supervised learning model (without human labelling for training).

### 5.2 Recommendation with Graph Neural Networks

In recent years, graph neural networks have been widely proposed in recommender systems [1, 5, 14, 22, 27, 28] because they can encode both node information and graph structure. For instance, Fan *et al.* [5] developed a new framework, called GraphRec, to jointly capture interactions and opinions in the user-item graph. In addition, Shaohua *et al.* [4] proposed a meta-path guided heterogeneous Graph Neural Network for intent recommendation on an e-commerce platform. However, these efforts apply graph neural networks to recommender systems on a single platform.

Different from prior work, we propose a novel item-item cross-domain recommendation framework that automatically recommends related repositories on GitHub to a query paper in an academic search system, helping users find their repositories of interest quickly by a *joint embeddings across platforms*.

## 6 CONCLUSIONS AND FUTURE WORK

This paper developed an item-item cross-platform recommender system, *paper2repo* that can automatically recommend repositories on GitHub matching a specified paper in the academic search system. We proposed a joint model that incorporates a text encoding technique into a constrained GCN formulation to generate joint embeddings of papers and repositories from the two different platforms. Specifically, the text encoding technique was leveraged to learn sequence information from paper abstracts and descriptions/tags of repositories. In order to map the representations of papers and repositories onto the same space, we adopted a constrained GCN model that forces the embeddings of bridge papers and their corresponding repositories to be equal as a constraint. Finally, we conducted experiments to evaluate the performance of *paper2repo* on real-world data sets. Our evaluation results demonstrated that the proposed *paper2repo* systems can achieve a higher recommendation accuracy than prior methods. In the future, we can extend our method to consider other entities (such as venues and authors), in addition to papers, to construct the knowledge graph. We can improve text embedding (e.g., by use of phrase embedding instead of word embedding). We can also investigate generalizations of our joint embedding, especially to domains with no analogue to bridge papers to serve as mapping constraints.

## ACKNOWLEDGMENTS

Research reported in this paper was sponsored in part by DARPA award W911NF-17-C-0099, DTRA award HDTRA1-18-1-0026, and the Army Research Laboratory under Cooperative Agreements W911NF-09-2-0053 and W911NF-17-2-0196.

## REFERENCES

- [1] Chong Chen, Min Zhang, Chenyang Wang, Weizhi Ma, Minming Li, Yiqun Liu, and Shaoping Ma. 2019. An Efficient Adaptive Transfer Neural Network for Social-aware Recommendation. In *Proceedings of SIGIR*.
- [2] Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems*. 3844–3852.
- [3] Ali Mamdouh Elkahky, Yang Song, and Xiaodong He. 2015. A multi-view deep learning approach for cross domain user modeling in recommendation systems. In *Proceedings of the 24th International Conference on World Wide Web*. 278–288.
- [4] Shaohua Fan, Junxiong Zhu, Xiaotian Han, Chuan Shi, Linmei Hu, Biyu Ma, and Yongliang Li. 2019. Metapath-guided Heterogeneous Graph Neural Network for Intent Recommendation. In *Proceedings of KDD*.
- [5] Wenqi Fan, Yao Ma, Qing Li, Yuan He, Eric Zhao, Jiliang Tang, and Dawei Yin. 2019. Graph Neural Networks for Social Recommendation. In *The World Wide Web Conference*. 417–426.
- [6] William L Hamilton, Rex Ying, and Jure Leskovec. 2017. Representation learning on graphs: Methods and applications. *arXiv preprint arXiv:1709.05584* (2017).
- [7] Xiangnan He, Lizi Liao, Hanwang Zhang, Lijiang Nie, Xia Hu, and Tat-Seng Chua. 2017. Neural collaborative filtering. In *Proceedings of the WWW*. 173–182.
- [8] Jyun-Yu Jiang, Pu-Jen Cheng, and Wei Wang. 2017. Open Source Repository Recommendation in Social Coding. In *Proceedings of the 40th International ACM SIGIR*. 1173–1176.
- [9] Meng Jiang, Peng Cui, Xumin Chen, Fei Wang, Wenwu Zhu, and Shiqiang Yang. 2015. Social recommendation with cross-domain transferable knowledge. *IEEE Transactions on Knowledge and Data Engineering* 27, 11 (2015), 3084–3097.
- [10] Meng Jiang, Peng Cui, Nicholas Jing Yuan, Xing Xie, and Shiqiang Yang. 2016. Little is much: Bridging cross-platform behaviors through overlapped crowds. In *Thirtieth AAAI Conference on Artificial Intelligence*.
- [11] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759* (2016).
- [12] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
- [13] Jianxun Lian, Fuzheng Zhang, Xing Xie, and Guangzhong Sun. 2017. CCCFNet: a content-boosted collaborative filtering neural network for cross domain recommender systems. In *Proceedings of the 26th International Conference on World Wide Web Companion*. 817–818.
- [14] Namyoung Park, Andrey Kan, Xin Luna Dong, Tong Zhao, and Christos Faloutsos. 2019. Estimating Node Importance in Knowledge Graphs Using Graph Neural Networks. In *Proceedings of KDD*.
- [15] Jeffrey Pennington, Richard Socher, and Christopher Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 EMNLP*. 1532–1543.
- [16] Juan Ramos et al. 2003. Using tf-idf to determine word relevance in document queries. In *Proceedings of the first instructional conference on machine learning*, Vol. 242. 133–142.
- [17] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner, and Lars Schmidt-Thieme. 2009. BPR: Bayesian personalized ranking from implicit feedback. In *Proceedings of the twenty-fifth conference on uncertainty in artificial intelligence*. 452–461.
- [18] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. Line: Large-scale information network embedding. In *Proceedings of the 24th International Conference on World Wide Web*. 1067–1077.
- [19] Aaron Van den Oord, Sander Dieleman, and Benjamin Schrauwen. 2013. Deep content-based music recommendation. In *Advances in neural information processing systems*. 2643–2651.
- [20] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. 2015. Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD*. 1235–1244.
- [21] Hongwei Wang, Miao Zhao, Xing Xie, Wenjie Li, and Minyi Guo. 2019. Knowledge graph convolutional networks for recommender systems. In *The World Wide Web Conference*. ACM, 3307–3313.
- [22] Xiang Wang, Xiangnan He, Yixin Cao, Meng Liu, and Tat-Seng Chua. 2019. KGAT: Knowledge Graph Attention Network for Recommendation. In *Proceedings of KDD*.
- [23] Xiang Wang, Xiangnan He, Lijiang Nie, and Tat-Seng Chua. 2017. Item silk road: Recommending items from information domains to social users. In *Proceedings of the 40th International ACM SIGIR*. 185–194.
- [24] Jason Weston, Samy Bengio, and Nicolas Usunier. 2011. Wsabie: Scaling up to large vocabulary image annotation. In *IJCAI*, Vol. 11. 2764–2770.
- [25] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2017. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 1492–1500.
- [26] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826* (2018).
- [27] Rex Ying, Ruining He, Kaifeng Chen, Pong Eksombatchai, William L Hamilton, and Jure Leskovec. 2018. Graph Convolutional Neural Networks for Web-Scale Recommender Systems. *arXiv preprint arXiv:1806.01973* (2018).
- [28] Jiaxuan You, Yichen Wang, Aditya Pal, Pong Eksombatchai, Chuck Rosenberg, and Jure Leskovec. 2019. Hierarchical Temporal Convolutional Networks for Dynamic Recommender Systems. In *The World Wide Web Conference*. ACM, 2236–2246.
- [29] Fuzhen Zhuang, Yingmin Zhou, Fuzheng Zhang, Xiang Ao, Xing Xie, and Qing He. 2018. Cross-domain novelty seeking trait mining for sequential recommendation. *arXiv preprint arXiv:1803.01542* (2018).