2019 IEEE International Conference on Data Mining (ICDM)

AMENDER: an Attentive and Aggregate Multi-layered Network for Dataset Recommendation

Yujun Chen*§¶, Yuanhong Wang*†, Yutao Zhang‡, Juhua Pu*§ ✉ and Xiangliang Zhang¶ ✉
*Engineering Research Center of Advanced Computer Application Technology, Ministry of Education,
Beihang University, Beijing, China
†X-Order Lab, Shanghai, China
‡Tsinghua University, Beijing, China
§Research Institute of Beihang University in Shenzhen, Shenzhen, China
¶King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
Email: {chenjohn, pujh, lucienwang}@buaa.edu.cn, zhangyutao@mail.tsinghua.edu.cn, xiangliang.zhang@kaust.edu.sa

***Abstract*—In this paper, we study the problem of recommending the appropriate datasets for authors, which is implemented to infer the proximity between authors and datasets by leveraging the information from a three-layered network, composed by authors, papers and datasets. To link author-dataset semantically by taking advantage of the rich content information of papers in the intermediate layer, we design an attentive and aggregate multi-layer network learning model. The aggregation is for integrating the intra-layer information of paper content and citations, while the attention is used for coordinating authors at the top-layer and datasets at the bottom-layer in the semantic space learned from papers in the intermediate layer. The experimental study demonstrates the superiority of our method compared with the solutions that extend existing models to our problem.**

***Index Terms*—Recommendation systems, Multi-layer network, Network embeddinng**

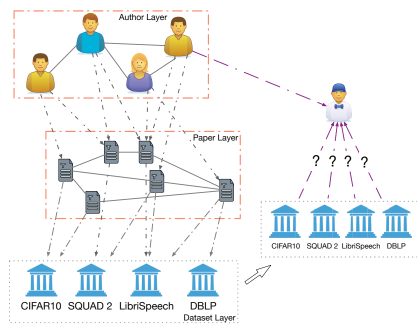

Fig. 1: Given the three-layer network of Author, Paper, and Dataset, we target on inferring the research interests of a user and recommend the appropriate datasets to be used.

I. Introduction

In modern research communities, researchers are living and breathing with data. State-of-the-art studies, such as BERT [1], RESNET [2] and Transformer-XL [3], made themselves persuasive by presenting their works' impact on benchmarking datasets. However, with the growing quantity and diversity of publicly available datasets, researchers spend much time on managing and exploring useful datasets [4], even on some well-organized and domain-specific dataset repositories. For example, when a researcher wants to discover a new field or a newly enrolled Ph.D. student starts a new project, most of the warm-up time is to find the most reliable and well-studied datasets in the field [5], [6]. Luckily, a few organizations and institutes, such as Google, Elsevier and Mendeley, have already offered their solutions: several datasets search engines are available for researchers all around the world [7]–[9]. However, these datasets search engines only provide general datasets searching service for researchers, none of them provide personalized recommendation for users.

In this paper, we aim at building a **customized dataset recommendation** model for researchers by mining their interactions and their semantics from publications. As presented in Figure 1, our problem is depicted as a three-layer network, including author, paper and dataset layers. The inter-layer links are formed via authorship between authors and papers, and by dataset citations between papers and datasets. At the paper layer, papers are linked by citations among them. At the researcher layer, people can be linked by their academic relations, e.g., co-authorship, co-affiliation interactions, follower/followee on ResearchGate or GoogleScholar. For a user on the right in Figure 1, we want to infer his preference to the existing datasets. Social relations have shown their usefulness on promoting recommendation systems [10]. We here face the challenge to model the interactions at two layers (researcher and paper), rather than the single layer social relation at the user side in standard recommendation systems. Moreover, we infer the relations by passing through the intermediate paper layer and ultimately learn the researcher and dataset proximity.

The special nature of our problem prevents the direct adoption of existing heterogeneous network models. Embedding approaches designed for heterogeneous networks [11], [12] or multi-layer network [13] do consider the node types and represent nodes by low-dimensional vectors. However, meta-path based methods [11] fuse the heterogeneous information but ignore the relationship between homogeneous nodes, like author-author and paper-paper relationship in our problem. Moreover, the learned embeddings may not be suitable for the recommendation tasks, considering the two significant chal-

✉ Corresponding Author

2374-8486/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDM.2019.00112

IEEE computer society

lenges in dataset recommendation, extreme *data sparsity* and *cold start* [14]. In the context of recommender systems, task-guided, path-augmented, and semantic-aware heterogeneous network embedding methods have been proposed in [14]–[17]. However, they are designed for bipartite networks (with two layers), while we learn from a three-layer network and fuse the content information of nodes and the intra-layer/inter-layer interactions.

To measure the proximity of authors and datasets by their semantic representations, we take advantage of the rich content information of papers in the intermediate layer and develop an attentive and aggregate multi-layer network learning model for Dataset Recommendation, named *AMENDER*. The **aggregation** is for integrating the paper-layer structure and content information to produce paper representations. The **multi-head attention mechanism** is applied to model the interactions between papers and authors, and between papers and datasets, such that authors at the top-layer and datasets at the bottom-layer are represented in the semantic space learned from papers. To mitigate the impact of name ambiguity in scholarly data [18], we adopt the adversarial personalized ranking (APR) to stabilize our training. We demonstrate the effectiveness of our model on the collected scholar dataset, and also evaluate the most extendable models proposed for heterogeneous network recommendation [11], [12] and network representation learning methods [19], [20]. The extensive experiments show that our proposed AMENDER model has more accurate recommendation results than the other solutions that extend the existing models to our problem.

Here we briefly listed our contributions as follows:

- We study the author-dataset recommendation problem in big scholarly data, which can benefit the whole academic community.
- We design a model to measure the semantic proximity of authors and datasets by integrating the paper content information and the intra-layer/inter-layer interactions in a three-layer network.
- We demonstrate the effectiveness of our model on real-world scholar data.

II. Related Work

Prior works on similar recommendation problems adopt meta-path based framework and model the problem as a heterogeneous network problem [14], [16]. These methods generate a typical meta-path walk [11], like "author-paper-author", to obtain the pair-wise relationship between heterogeneous nodes in the network. A notable example of such approaches is TSR [14], a task guided, path-augmented network representation method which aims to build the paper and author representation in the same space using Bayesian Personalized Ranking (BPR) [21]. Similarly, HERec [12] uses the meta-path to represent the heterogeneous information network and makes the recommendation based on pair-wise ranking according to the meta-path. These methods are effective for learning heterogeneous relationships using pair-wise ranking loss. However, meta-path based methods ignore the information propagation among homogeneous nodes, which is extremely valuable for scholar data, such as author relations and papers citations.

Further, multi-layered network representation can also model the scholar network [13], but the matrix factorization will consume extremely high computational resources when the graph is large. General network representation methods can also deal with the representation learning problem [19], [20], but they map all nodes at the same representation space and ignore the difference between inter-layer and intra-layer relationships. Extending the above discussed solutions to our problem will fail to integrate certain types of information and deliver inappropriate results, as we will show in Section 4.

III. Methodology

A. Preliminaries

An *Author-Paper-Dataset (APD)* network $G = (V, E, X_p)$ is a special kind of heterogeneous information network [15], which is comprised of three types of nodes $V = (V_a, V_p, V_d)$ and four types of edges $E = (E_a, E_p, E_{ap}, E_{pd})$, with subscription a for authors, p for papers and d for datasets, as well as the additional features on paper vertices X_p. In this scenario, a multi-layer structure arises naturally from G that consists of three subgraphs $G_a = (V_a, E_a, \emptyset)$, $G_p = (V_p, E_p, X_p)$ and $G_d = (V_d, \emptyset, \emptyset)$ along with author-paper and paper-dataset affinity matrices E_{ap}, E_{pd}. Following the denotation, we can formalize the studied problem as follows.

Problem 1 (Dataset recommendation based on APD network): Given an APD network G and a set of author-dataset preference records $S = \{\langle v_a, v_d \rangle\} \subseteq V_a \bigotimes V_d$ where $\langle v_a, v_d \rangle$ means that the author v_a prefers the dataset v_d, our goal is to recommend the top-k relevant datasets for any $v_a \in V_a$.

To quantitatively measure the relevance of an author and a dataset and circumvent the data sparsity problem, we target on learning the representation for authors and datasets, denoted by $\mathbf{A} \in \mathbb{R}^{|V_a| \times d}$ and $\mathbf{D} \in \mathbb{R}^{|V_d| \times d}$, respectively, by fusing the graph structures and the paper content. The motivation is to characterize both authors and datasets in the same semantic space, where their proximity is measured regarding the latent research topics.

Unlike many previous models [12], [14] that only fuse the heterogeneous information by meta-path, our model will model both homogeneous and heterogeneous nodes interaction via two mechanisms: aggregation and attention. The overall framework is presented in Figure 2. We will first obtain paper representation by aggregating the neighborhood papers, which will then be passed to authors and datasets through attention mechanism to obtain author and dataset representations. Finally, we learn the proximity of between authors and datasets by pair-wise ranking loss with a specially designed training procedure.

B. Aggregation at Paper-layer

The intermediate layer including paper content and citation relations can provide us abundant semantic information for our tasks. To build the semantic space where authors and dataset will be mapped, we first aggregate the paper content and paper

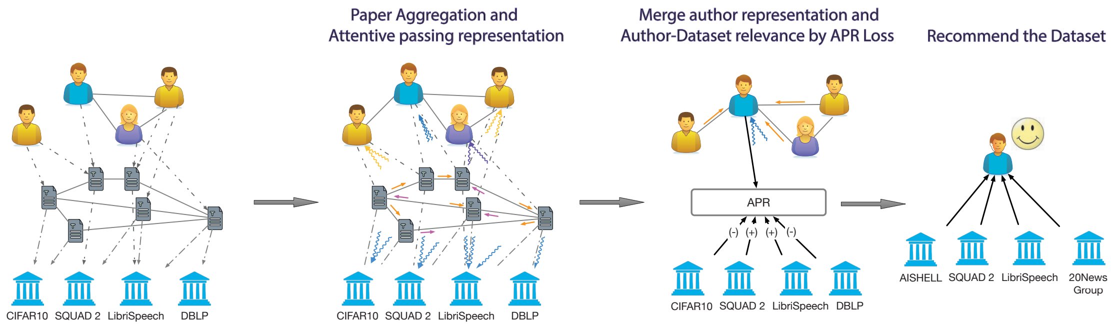

Fig. 2: The overview figure of our method

citations to learn a latent space that captures the research topics of the papers.

Formally, given the citation network $G_p = (V_p, E_p, X_p)$, we learn the paper representation $\mathbf{P} \in \mathbb{R}^{|V_p| \times d}$, where d is the dimension of representation vector, by the recent inductive attributed graph embedding model GraphSAGE [22]. For each paper $v_p \in V_p$, we obtain its representation by aggregating its immediate neighborhood paper information K times. At k-th aggregation,

$$\mathbf{P}^k_{v_p} = g^k \left(\text{AGGREGATE}_k(\{\mathbf{P}^{k-1}_{v_p}\} \cup \mathbf{P}^{k-1}_{\mathcal{N}_{pp}(v_p)}) \right), \tag{1}$$

where $\mathbf{P}^{k-1}_{\mathcal{N}_{pp}(v_p)} = \{\mathbf{P}^{k-1}_{v} \mid v \in \mathcal{N}_{pp}(v_p)\}$, $\mathbf{P}^{k-1}_{v_p}$ is the (k-1)-th aggregation of paper v_p, $\mathcal{N}_{pp}(v_p)$ is the immediate neighborhood of v_p, and g^k is a non-linear function. Paper representation $\mathbf{P}$ is the output of the final aggregation, as $\mathbf{P} = \mathbf{P}^K$. And the initial representations $\mathbf{P}^0$ are generated from the paper's content features X_p.

Our results are not sensitive to the selection of aggregators. Here we adopt the simplest mean aggregation operator,

$$\text{AGGREGATE}_k(\cdot) = \sigma_1(\mathbf{W}^k_{pp} \cdot \text{MEAN}(\cdot)), \tag{2}$$

where σ_1 is the ReLU activation function. Lastly, the normalization is introduced to constrain the representation, which leads to the final result,

$$\mathbf{P}^k_{v_p} = \frac{g^k \left(\text{AGGREGATE}_k(\{\mathbf{P}^{k-1}_{v_p}\} \cup \mathbf{P}^{k-1}_{\mathcal{N}_{pp}(v_p)}) \right)}{||g^k \left(\text{AGGREGATE}_k(\{\mathbf{P}^{k-1}_{v_p}\} \cup \mathbf{P}^{k-1}_{\mathcal{N}_{pp}(v_p)}) \right) ||_2}. \tag{3}$$

C. Learning Author and Dataest Representation

The paper representation $\mathbf{P}$ obtained in the previous section, along with author relationships and author-paper affinity (or only dataset-paper affinity), is then utilized to learn the author representation (or dataset representations).

1) Attention from Papers: To learn author and dataset representation in the same semantic space with $\mathbf{P}$, we leverage the multi-head attention network [23] to link all the related papers to authors and datasets. The graph attention network component is a natural fit to learn inter-layer information propagation for it only computes the subgraph information, such as only author and his authored papers, rather than all instances in the training set as original attention structure [24]. Formally, for an author $v_a \in V_a$ (or a dataset $v_d \in V_d$), we will obtain its representation $\mathbf{A}^{(p)}_{v_a} \in \mathbb{R}^d$ (or $\mathbf{D}_{v_d}$) by passing the representation of all related papers $\mathcal{N}_{ap}(v_a) = \{v_p \mid v_p \in V_p, (v_a, v_p) \in E_{ap}\}$ (or $\mathcal{N}_{dp}(v_d)$) using attention

$$\mathbf{A}^{(p)}_{v_a} = \text{ATTN}(\mathbf{P}, \mathcal{N}_{ap}(v_a)), \tag{4}$$

$$\mathbf{D}^{(p)}_{v_d} = \text{ATTN}(\mathbf{P}, \mathcal{N}_{dp}(v_d)). \tag{5}$$

For brevity, we will only illustrate the learning of $\mathbf{A}^{(p)}$ here, and building $\mathbf{D}^{(p)}$, i.e. $\mathbf{D}$, will follow the same process.

The attention mechanism will pass papers' representation to authors by first compute the attention coefficient between each related paper representation $\mathbf{P}_{v_{p|a}}, v_{p|a} \in \mathcal{N}_{ap}(v_a)$ and latest author's representation $\mathbf{A}^{(p)}_{v_a}$,

$$\mathbf{e}_{v_a, v_{p|a}} = f(\mathbf{W}_{ap} \cdot \mathbf{A}^{(p)}_{v_a}, \mathbf{W}_{ap} \cdot \varphi(\mathbf{P}_{v_{p|a}})), \tag{6}$$

where f is a 2-layer feedforward neural network parameterized by the non-linear LeakyReLU activation function, and φ is a linear mapping function. The $\mathbf{e}_{v_a, v_{p|a}}$ measures the importance of paper $v_{p|a}$ to author v_a. To make coefficients comparable, we adopt softmax function for normalizing the attention coefficient,

$$\alpha_{v_a, v_{p|a}} = \frac{\exp(\mathbf{e}_{v_a, v_{p|a}})}{\sum_{v_p \in \mathcal{N}_{ap}(v_a)} \exp(\mathbf{e}_{v_a, v_p})}. \tag{7}$$

The normalized attention coefficients are used to compute a combination of all related paper representations with ReLU activation σ_2 as the final output for each author,

$$\text{ATTN}(\mathbf{P}, \mathcal{N}_{ap}(v_a)) = \sigma_2(\sum_{v_{p|a} \in \mathcal{N}_{ap}(v_a)} \alpha_{v_a, v_{p|a}} \mathbf{W}^{attn}_{ap} P_{v_{p|a}}). \tag{8}$$

In addition, the multi-head strategy is applied to sample variant information from related papers and to make the learning more stable as proved in [24]. Specifically, we use M independent attention mechanism $\{\text{ATTN}^m \mid i = 1, \cdots, M\}$ in our model. The final author representation is the multiplication between concatenated attention results and a parameter matrix $\mathbf{W}^{proj}_{ap} \in \mathbb{R}^{Md \times d}$:

$$\text{ATTN}(\mathbf{P}, \mathcal{N}_{ap}(v_a)) = \Big\Vert_{m=1}^{M} \text{ATTN}^m(\mathbf{P}, \mathcal{N}_{ap}(v_a)) \cdot \mathbf{W}^{proj}_{ap}, \tag{9}$$

where $\|$ represents the concatenation.

2) Extra Aggregation for Author Representation: In the above process, author representations $\mathbf{A}^{(p)}$ are obtained by only using the paper representations. To integrate the author relationship that may contain extra information at the author-layer, aggregation is used again to refine the author representations, analogous to paper aggregation in Section III-B, with initial $\mathbf{A}^{(a)0} = \mathbf{A}^{(p)}$. For those authors have not published a paper, the mean vector of their neighborhood representation will be adopted for prediction. Author representation from papers $\mathbf{A}^{(p)}$, and from author neighborhood $\mathbf{A}^{(a)}$ are then combined for the final $\mathbf{A} = (\mathbf{A}^{(p)} + \mathbf{A}^{(a)})/2$.

D. Training the Model

1) Author and Dataset Proximity: Given the representation $\mathbf{A}$ and $\mathbf{D}$ as well as the preference records S, where $S = \{\langle v_a, v_d, \rangle \mid \exists v_p \in V_p, (v_a, v_p) \in E_{ap}, (v_p, v_d) \in E_{pd}\}$ in our model. We use the triplet personalized ranking loss function, derived from Bayesian Personalized Ranking (BPR) [21], to build the main loss for AMENDER. Denoting the negative samples by $S' = \{\langle v_a, v'_d \rangle \mid v_a \in V_a, v'_d \in V_d, \langle v_a, v'_d \rangle \notin S\}$ we have the BPR loss

$$\mathcal{L}_{ad} = \sum_{v_a \in V_a} \sum_{\substack{\langle v_a, v_d \rangle \in S \\ \langle v_a, v'_d \rangle \in S'}} \ell(v_a, v_d, v'_d) \tag{10}$$

where

$$\ell(v_a, v_d, v'_d) = \left(1 - \sigma_3 \left(h(\mathbf{A}_{v_a}, \mathbf{D}_{v_d}) - h(\mathbf{A}_{v_a}, \mathbf{D}_{v'_d})\right)\right), \tag{11}$$

where the h is a distance metric (e.g., cosine distance in our experiments) and σ_3 is a non-linear mapping function to normalize the representation distance (e.g., sigmoid function).

2) Smooth Loss of Author Representation: To stabilize the author representation, we push $\mathbf{A}^{(p)}$ and $\mathbf{A}^{(a)}$ close by introducing a smooth loss,

$$\mathcal{L}_{aa} = \sum_{v_a \in V_a} d\left(\mathbf{A}^{(p)}_{v_a}, \mathbf{A}^{(a)}_{v_a}\right), \tag{12}$$

where d is the pair-wise cosine similarity.

To the end, our objective function is defined as the combination of $\mathcal{L}_{aa}$ and $\mathcal{L}_{ad}$:

$$\mathcal{L} = \mathcal{L}_{aa} + \mathcal{L}_{ad}. \tag{13}$$

3) Adversarial Personalized Ranking (APR): The traditional BRP method is vulnerable to adversarial noises [25]. When adding small noises to the recommendation items, the final representation might change severely. Recall the name ambiguity problem for the citation network, even the most advanced scholarship data sources will suffer from recognizing wrong authors [18]. Here we do not aim to propose a specific solution for this problem, but utilizing the adversarial personalized ranking loss to mitigate the impact of such a problem. As proposed in [25], we add the adversarial noises Δ to the author representations $\mathbf{A}$ only and build the loss:

$$\mathcal{L}_{APR} = \mathcal{L} + \lambda(\mathcal{L}_{aa,\Delta_{adv}} + \mathcal{L}_{ad,\Delta_{adv}}), \tag{14}$$

where

$$\Delta_{adv} = \underset{\Delta, ||\Delta|| \leq \epsilon}{\operatorname{argmax}} \ \mathcal{L}_{aa,\Delta} + \mathcal{L}_{ad,\Delta}, \tag{15}$$

$\mathcal{L}_{\cdot,\Delta}$ is the loss with adversarial noises added to author's representation and λ is he adversarial regularization parameter.

IV. Experiments

A. Datasets

Our experiment is conducted on the records collected from a dataset search engine called Delve[1], covering academic papers and datasets mostly from artificial intelligence domains like machine learning, data mining and computer vision, etc. In our experiment, the network we used contains 73,855 authors, 79,775 papers and 6,863 datasets in total. Here we treat the citation network as paper relationships, co-author network as the author relationships (they may have co-authored on other papers out of this collection). In total, there are 191,096 author-paper edges and 15,572 paper-dataset edges. The network is sparse, as only around 10K authors have used one or more dataset, and the others do not have direct links to any datasets we collected.

B. Compared Methods

We extend 11 related state-of-the-art methods that span three categories to our problem for comparison:

1) **Network Representations**: Network representation methods learn the representations for author, paper and datasets for general purposes. We select homogeneous model Node2Vec [19], distributional representation model Graph2Gaus [20], heterogeneous model Metapath2Vec [11] and multi-layered network model MANE [13]. Note that MANE is highly memory demand, and we can only learn two layers, author and datasets, on our 128 Gigabytes machine.
2) **Ranking based Recommendations**: Ranking based recommendation methods use the ranking loss to learn the relevance between authors and datasets. Here we select a baseline Matrix Factorization method, content-based model BPR [21] and WARP [26], and heterogeneous information network based model HERec [12].
3) **Social Recommendation**: Social recommendation methods leverage author relationships for promoting recommendation performance. We select two state-of-the-art social recommendation models TSR [14] and CSE [27].
4) **Proposed Method**: One may be interested in how important the role of the intermediate layer of papers is playing in our model, because a simple operation can form an author-dataset bipartite network by linking an author and a dataset if the dataset is used by author. We evaluate our proposed model on this setting by filtering out the paper information, named AMENDER$_{\cancel{paper}}$, and compare it with the original model AMENDER.

For fair comparison, we set representation dimension $d = 64$ for all methods. In the proposed method, the adversarial regularization parameter λ is set as 0.2, and the number of multi-head attention M is set as 3 for the best performance. Paper content vectors are pre-trained using all abstracts via doc2vec [28]. Other methods adopt the default parameters.

[1] http://adatahub.com

TABLE I: Recommendation accuracy to existing authors (the best results in **bold**, while the second best with a $*$)

Methods	Metrics	Training Percentage		
		50%	80%	95%
Node2Vec	AUC	0.844	0.890	0.901
	AP	0.802	0.859	0.870
Graph2Gauss	AUC	0.891*	0.918*	0.932*
	AP	0.884*	0.908*	0.925*
MANE	AUC	0.863	0.878	0.885
	AP	0.855	0.868	0.872
MetaPath2Vec	AUC	0.838	0.864	0.854
	AP	0.860	0.886	0.884
MF	AUC	0.714	0.750	0.768
	AP	0.774	0.816	0.836
BPR	AUC	0.821	0.862	0.881
	AP	0.796	0.851	0.888
WARP	AUC	0.828	0.878	0.881
	AP	0.798	0.869	0.875
HERec	AUC	0.866	0.917	0.922
	AP	0.826	0.891	0.900
TSR	AUC	0.824	0.846	0.887
	AP	0.818	0.839	0.868
CSE	AUC	0.826	0.885	0.901
	AP	0.794	0.875	0.891
AMENDER$_{\cancel{paper}}$	AUC	0.881	0.892	0.910
	AP	0.866	0.878	0.900
AMENDER	AUC	**0.933**	**0.945**	**0.955**
	AP	**0.925**	**0.940**	**0.951**

TABLE II: Recall@k for cold start users (the best results in **bold**, while the second best with a $*$)

Methods	Recall				Precision	
	@3	@10	@50	@100	@3	@10
Node2Vec	0.235	0.305	0.364	0.387	0.194	0.095
Graph2Gauss	0.202	0.306	0.487	0.577	0.196	0.094
MANE	0.241	0.300	0.432	0.521	0.206	0.087
MetaPath2Vec	0.584	0.678	0.753	0.761	0.277	0.143
MF	0.480	0.739	0.795	0.809	0.215	0.121
BPR	0.615	0.721	0.826	0.867	0.394*	0.163
WARP	0.621	0.747	0.839	0.868	0.348	0.150
HERec	0.626	0.672*	0.860	0.896*	0.338	0.148
TSR	0.637*	0.725	0.803	0.855	0.382	0.166
CSE	0.605	0.735	0.821	0.854	0.290	0.146
AMENDER$_{\cancel{paper}}$	0.587	0.738	0.856	0.889	0.375	0.155
AMENDER	**0.738**	**0.835**	**0.914**	**0.940**	**0.427**	**0.171**

C. *Recommendation to Existing Authors*

We first evaluate our model's effectiveness by recommending datasets to existing authors. Here we formalize the problem as pair-wise ranking orders between correlated and uncorrelated author-dataset pairs. We create three test sets that contains 50%, 20% and 5% randomly selected correlated author-dataset pairs and equal number of randomly selected uncorrelated pairs. By convention, we report both the AUC (Area Under ROC Curve) score and AP (Average Precision) score to evaluate how highly ground-truth correlated datasets have been ranked over other datasets.

The performance of all methods is presented in Table I. We can observe that AMENDER consistently outperforms all others at all the ratio of training and testing set. Specifically, AMENDER can obtain about 0.93 AUC with only 50% training samples, which is higher than many methods using 95% for training. AMENDER$_{\cancel{paper}}$ performs also worse than AMENDER,which indicates that our model can better capture author-dataset correlation with the supplement of papers.

Further, it's interesting that Graph2Gauss obtains the second best performance (better than recommendation methods), probably because the objective of network representation methods is pair-wise distance, which totally fits the task.

D. *Cold Start Recommendations*

A more realistic but challenging setting is to infer non-existing user's preference for datasets. For example, young students have no published papers but only linked with their advisors. We randomly selected ten percent of authors, and removed all their used datasets and their published papers from the scholarly network. When recommending datasets to one of these new users, only the relations to authors in training set are used. For baselines that cannot deal with the cold start problem, we use the mean vector of all neighborhood authors representation in the author graph as the target representation. Here we use the Recall@k and Precision@k for evaluating the recommendation correctness.

Table II shows the performance of all models at different setting of k. We find that network representation methods cannot perform well when addressing such a recommendation problem. The only effective network embedding method is Metapath2Vec which is good at learning heterogeneous representations. Recommendation based methods (e.g., HERec) perform well in this case, showing the effectiveness of incorporation extra layer for recommendation. AMENDER again has the best performance, and obtain about 10% relative recall gain when only a few candidate datasets are given.

E. *Topics of Datasets and Authors from their Representation*

To interpret the learned dataset and author representation, we visualize the obtained $\mathbf{A}$ and $\mathbf{D}$ in 2-dim space by t-SNE [29]. Due to the space limit, we randomly selected 100 authors (triangles) and 100 datasets (squares) and show them in Figure 3. Also, we apply k-means to cluster the representations and show a word-cloud for interpreting each cluster[2]. Each cluster has different keywords. For example, the orange cluster is about *system performance*, while red cluster is about *data mining*. It is interesting that *system performance* and *data mining* have close representations. To investigate further, we zoom in the overlapping area, and found that most of the authors and datasets in that area are related to *database performance and theory*. This is reasonable because the *database* serves as infrastructure for *data mining*. Meanwhile, the efficiency and scalability of database is a heated topic for *system* field. For example, professor Xiaodong Zhang, from Ohio state university, has research interests mainly focuse on "data management in scalable system". Also, Michaela Götz, a former Ph.D. student in Cornell University, has published her most influential paper in SIGMOD, a top conference in database theories. These author names can be found in the overlapping area. Datasets in this area are also interesting to discover. For example, both *TREC-1* and *TREC-3* (very closely located) are from the Text REtrievel Conference (TREC). They contain billions of webpages and often serve as datasets

[2]Each cluster is interpreted by the paper contents published by the group of authors, and the papers citing the datasets.

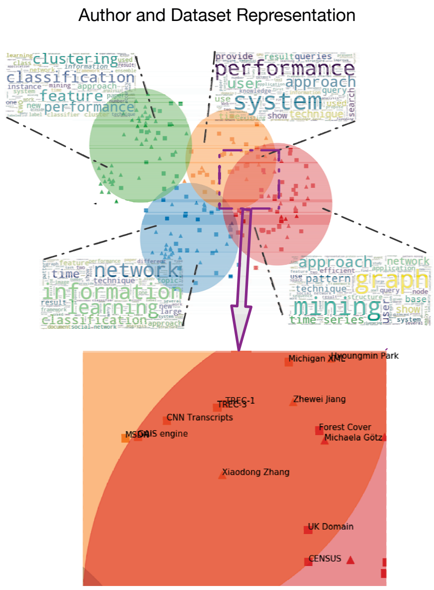

Fig. 3: Visualize author (triangle) and dataset (square) representation. Clusters are interpreted by the word-cloud of research topics.

for query performance evaluation and text mining. Also, the CNN Transcripts contain a large corpus of CNN transcripts. The dataset is constantly used for text classification or the performance benchmarks for *information retrieval*. According to the interesting visualization results, we confirm that the learned representations are meaningful and well capture the relations between authors and datasets.

V. Conclusion

In this paper, we studied the problem to recommend appropriate datasets for authors. We designed AMENDER to tackle the relevance between authors and datasets. AMENDER can better leverage both intra-layer and inter-layer information, and can better model the author-dataset relationships by utilizing the adversarial ranking loss, comparing to the solutions that extend existing heterogenous network embedding models and social recommendation models. We plan to deploy our model and benefit researchers in AI domain in the future.

Acknowledgement

The authors would like to thank the anonymous reviewers for their helpful comments. This work is supported by the King Abdullah University of Science and Technology (KAUST), Saudi Arabia, National Key Research and Development Program of China (2017YFB1002000), Science Technology and Innovation Commission of Shenzhen Municipality (JCYJ20180307123659504), and the State Key Laboratory of Software Development Environment in Beihang University.

References

[1] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2018.
[2] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *CVPR 16’*, pp. 770–778, 2016.
[3] Z. Dai, Z. Yang, Y. Yang, J. G. Carbonell, Q. V. Le, and R. Salakhutdinov, “Transformer-xl: Attentive language models beyond a fixed-length context,” *arXiv preprint arXiv: 1901.02860*, vol. abs/1901.02860, 2019.
[4] M. D. Wilkinson, M. Dumontier, Aalbersberg *et al.*, “The fair guiding principles for scientific data management and stewardship,” *Scientific data*, vol. 3, 2016.
[5] A. Halevy, F. Korn, N. F. Noy, C. Olston, N. Polyzotis, S. Roy, and S. E. Whang, “Goods: Organizing google’s datasets,” *SIGMOD*, 2016.
[6] A. Chapman, E. Simperl, L. Koesten, G. Konstantinidis, L.-D. Ibáñez-Gonzalez, E. Kacprzak, and P. Groth, “Dataset search: a survey,” *arXiv preprint arXiv:1901.00735*, 2019.
[7] Mendeley, https://data.mendeley.com/, 2019, (Accessed on 02/21/2019).
[8] Elsevier, “Elsevier datasearch,” https://datasearch.elsevier.com/, 2019, (Accessed on 02/21/2019).
[9] Google, “Google dataset search,” https://toolbox.google.com/datasetsearch, 2019, (Accessed on 02/21/2019).
[10] J. Tang, X. Hu, and H. Liu, “Social recommendation: a review,” *Social Network Analysis and Mining*, vol. 3, no. 4, pp. 1113–1133, 2013.
[11] Y. Dong, N. V. Chawla, and A. Swami, “metapath2vec: Scalable representation learning for heterogeneous networks,” in *SIGKDD 17’*. ACM, 2017, pp. 135–144.
[12] C. Shi, B. Hu, W. X. Zhao, and S. Y. Philip, “Heterogeneous information network embedding for recommendation,” *IEEE TKDE*, vol. 31, no. 2, pp. 357–370, 2019.
[13] J. Li, C. Chen, H. Tong, and H. Liu, “Multi-layered network embedding,” in *SDM 18’*. SIAM, 2018, pp. 684–692.
[14] C. Zhang, L. Yu, X. Zhang, and N. V. Chawla, “Task-guided and semantic-aware ranking for academic author-paper correlation inference.” in *IJCAI 18’*, 2018, pp. 3641–3647.
[15] C. Shi, Z. Zhang, P. Luo, P. S. Yu, Y. Yue, and B. Wu, “Semantic path based personalized recommendation on weighted heterogeneous information networks,” in *CIKM 15’*. ACM, 2015, pp. 453–462.
[16] T. Chen and Y. Sun, “Task-guided and path-augmented heterogeneous network embedding for author identification,” in *WSDM 17’*. ACM, 2017, pp. 295–304.
[17] C. Zhang, C. Huang, L. Yu, X. Zhang, and N. V. Chawla, “Camel: Content-aware and meta-path augmented metric learning for author identification,” in *WWW 18’*. International World Wide Web Conferences Steering Committee, 2018, pp. 709–718.
[18] Y. Zhang, F. Zhang, P. Yao, and J. Tang, “Name disambiguation in aminer: Clustering, maintenance, and human in the loop.” in *KDD 18’*. ACM, 2018, pp. 1002–1011.
[19] A. Grover and J. Leskovec, “node2vec: Scalable feature learning for networks,” in *KDD 16’*. ACM, 2016, pp. 855–864.
[20] A. Bojchevski and S. Günnemann, “Deep gaussian embedding of graphs: Unsupervised inductive learning via ranking,” *arXiv preprint arXiv:1707.03815*, 2017.
[21] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme, “Bpr: Bayesian personalized ranking from implicit feedback,” in *UAI 19’*. AUAI Press, 2009, pp. 452–461.
[22] W. Hamilton, Z. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” in *NIPS 17’*, 2017, pp. 1024–1034.
[23] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” *ICLR 18’*, 2017.
[24] D. Bahdanau, K. Cho, and Y. Bengio, “Neural machine translation by jointly learning to align and translate,” *arXiv preprint arXiv:1409.0473*, 2014.
[25] X. He, Z. He, X. Du, and T.-S. Chua, “Adversarial personalized ranking for recommendation,” in *SIGIR 18’*. ACM, 2018, pp. 355–364.
[26] J. Weston, H. Yee, and R. J. Weiss, “Learning to rank recommendations with the k-order statistic loss,” ser. RecSys ’13. New York, NY, USA: ACM, 2013, pp. 245–248. [Online]. Available: http://doi.acm.org/10.1145/2507157.2507210
[27] C.-M. Chen, C.-J. Wang, M.-F. Tsai, and Y.-H. Yang, “Collaborative similarity embedding for recommender systems,” *arXiv preprint arXiv:1902.06188*, 2019.
[28] Q. Le and T. Mikolov, “Distributed representations of sentences and documents,” in *International Conference on Machine Learning*, 2014, pp. 1188–1196.
[29] L. v. d. Maaten and G. Hinton, “Visualizing data using t-sne,” *JMLR*, vol. 9, no. Nov, pp. 2579–2605, 2008.