Multi-layered HITS on Multi-sourced Networks

by

Haichao Yu

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved May 2018 by the
Graduate Supervisory Committee:

Hanghang Tong, Chair
Jingrui He
Yezhou Yang

ARIZONA STATE UNIVERSITY

August 2018

ABSTRACT

Network mining has been attracting a lot of research attention because of the prevalence of networks. As the world is becoming increasingly connected and correlated, networks arising from inter-dependent application domains are often collected from different sources, forming the so-called multi-sourced networks. Examples of such multi-sourced networks include critical infrastructure networks, multi-platform social networks, cross-domain collaboration networks, and many more. Compared with single-sourced network, multi-sourced networks bear more complex structures and therefore could potentially contain more valuable information.

This thesis proposes a multi-layered HITS (Hyperlink-Induced Topic Search) algorithm to perform the ranking task on multi-sourced networks. Specifically, each node in the network receives an authority score and a hub score for evaluating the value of the node itself and the value of its outgoing links respectively. Based on a recent multi-layered network model, which allows more flexible dependency structure across different sources (i.e., layers), the proposed algorithm leverages both within-layer smoothness and cross-layer consistency. This essentially allows nodes from different layers to be ranked accordingly. The multi-layered HITS is formulated as a regularized optimization problem with non-negative constraint and solved by an iterative update process. Extensive experimental evaluations demonstrate the effectiveness and explainability of the proposed algorithm.

TABLE OF CONTENTS

Page

LIST OF TABLES . . . iii

LIST OF FIGURES . . . iv

CHAPTER

1 INTRODUCTION . . . 1

2 PROBLEM DEFINITION . . . 5

3 MULTI-LAYERED HITS ALGORITHM . . . 8

3.1 Objective Function . . . 8

3.2 Iterative Algorithms . . . 9

3.2.1 Cross-Layer Ranking . . . 11

3.2.2 Cross-Layer Query . . . 11

3.3 Proof and Analysis . . . 12

4 EVALUATION . . . 16

4.1 Dataset & Preprocessing . . . 16

4.2 Experimental Design & Prototype . . . 18

4.3 Experimental Results & Analysis . . . 19

4.3.1 Effectiveness . . . 20

4.3.2 Explainability . . . 21

5 LITERATURE SURVEY . . . 24

5.1 Network of Networks . . . 24

5.2 Multi-layered Networks . . . 26

5.3 Knowledge Graph Based Mining . . . 28

6 CONCLUSION . . . 31

REFERENCES . . . 32

LIST OF TABLES

Table Page

2.1 Symbols 5

4.1 Amazon Co-purchasing Network Statistics 18

LIST OF FIGURES

Figure		Page
1.1	An Illustrative Multi-layered Networks	3
2.1	A Simple 4-layered Network for Biological Systems	7
4.1	Amazon Multi-layered Networks	17
4.2	Prototype Interface	19
4.3	Average Rating Regarding Relevance in Ranking Task	20
4.4	Average Rating Regarding Relevance in Query Task	20
4.5	Average Rating Regarding Explainability in Ranking Task	22
4.6	Average Rating Regarding Explainability in Query Task	22

Chapter 1

INTRODUCTION

In the era of big data, network is almost everywhere. As the world is becoming increasingly connected and correlated, the networks arising from inter-dependent application domains are often collected from different sources, forming the so-called multi-sourced networks. Classic examples of such kind of networks include multi-platform social networks, cross-domain collaboration networks, critical infrastructure networks, biological systems, etc. Compared with single-sourced network, multi-sourced networks bear more complex structures and therefore could potentially contain more valuable information. In recent years, how to perform data mining tasks on such multi-sourced networks to extract hidden information has become a popular research topic among the researchers, which has posed many fascinating research questions. One of the tasks we are going to explore in this thesis is ranking on multi-sourced networks.

Ranking is one of the most common but important tasks in the domain of data mining. Until now, there exist many effective and efficient ranking algorithms. The most basic one is eigenvector centrality [Newman (2008)], which ranks nodes by their influence in the network. Theoretically, a node is important if it is linked to by other important nodes. Another famous ranking algorithm is Google PageRank [Brin and Page (1998)], which is the algorithm behind many search engines. Google PageRank exploits the global structure of the link graph using a random walk and assigns an importance score to each node according to the number of links it receives from other nodes. Based on Google PageRank, Zhou et al. proposed a personalized Google PageRank algorithm [Zhou *et al.* (2004b)] based on the random walk with restart

(RWR). This algorithm allows data to be ranked with respect to the intrinsic manifold structure, which achieves better ranking performance. At the same time that Google PageRank was proposed, Jon Kleinberg developed the Hyperlink-Induced Topic Search (HITS) [Kleinberg (1999)], which is a link analysis algorithm that is used for ranking web pages. Each page will receive two scores: authority and hub. Authority measures the value of the page itself while hub measures the value of its links to other pages. These two scores present a mutually reinforcing relationship. Besides PageRank and HITS, there has been a number of extensions of those two algorithms proposed, such as [Amento *et al.* (2000); Bharat and Henzinger (1998); Chakrabarti *et al.* (1998); Chakrabarti (2001); Chakrabarti *et al.* (2001); Haveliwala (2002); Lempel and Moran (2000); Cai *et al.* (2004)]. All of these ranking algorithms work well in single-sourced networks, but few of work has been done to generalize them to the context of multi-sourced networks.

To tackle this challenge, one of the prior things to do is to build a mathematical model for the multi-sourced networks. In 2014, Ni *et al.* (2014) proposed a new data model called Network of Networks (NoN) for multi-sourced graph mining, where each node of the main network itself can be further represented as a domain-specific network. This model allows to compare and rank nodes beyond the atom level. In 2015, Chen *et al.* (2015) defined a new multi-layered network model called MuLaN, which is a further generalization of NoN model. Specifically, connections in the main network from NoN model are extended to inter-layer node-node dependencies in MuLaN. In this way, MuLaN provides more flexible dependency structure across different layers (domain-specific networks) than the NoN model.

In this paper, we propose multi-layered HITS, which is a generalization of regular HITS (Hyperlink-Induced Topic Search) on multi-layered networks. The proposed method mainly solve two problems: (1) it allows finding the top-ranked nodes in each layer according to their global popularity and importance. We name this setting as Cross-Layer Ranking. (2) it allows finding the most relevant nodes in each layer with respect to a query node. We name this setting as Cross-Layer Query.

Moreover, to further improve the performance and explainability of the multi-layered HITS algorithm, we combine the knowledge graphs (e.g., Google knowledge graph, Baidu knowledge graph, etc) as the knowledge layers into the multi-layered networks. And the original layers are considered as data layers. The multi-layered structure is shown in Figure 1.1. With the integration of data layers and knowledge layers, there is a more detailed description on how different entities are related with each other, which allows the multi-layered HITS algorithm to provide more accurate and reasonable ranking results.

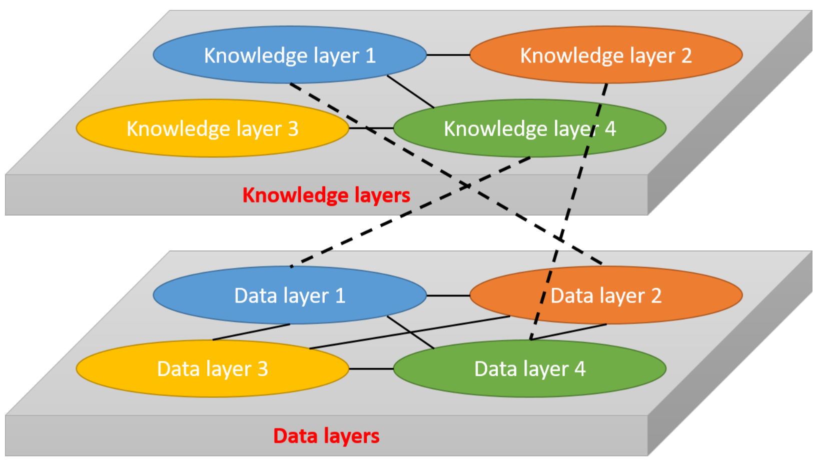

Figure 1.1: An Illustrative Multi-layered Networks

Generally speaking, the main contributions of this thesis can be summarized as the following 2 aspects:

- **Algorithms and Analysis**. We propose a new algorithm called multi-layered HITS, which performs the ranking task in multi-layered networks. And we analyze its convergence.

- **Empirical Evaluations**. We conduct comprehensive experiments on a real dataset to validate the effectiveness and explainability of the proposed algorithm.

The rest of the paper is organized as follows. Chapter 2 gives the problem definition of Cross-Layer Ranking and Cross-Layer Query. Chapter 3 proposes the multi-layered HITS algorithm with its analysis. Chapter 4 presents the experimental evaluations. Chapter 5 performs a literature survey regarding the related work. Chapter 6 conclude the paper.

Chapter 2

PROBLEM DEFINITION

In this chapter, we give the formal definitions of Cross-Layer Ranking problem and Cross-Layer Query Problem. Table 2.1 lists the main symbols used throughout this paper. We use bold capital letters to denote matrices (e.g., $\mathbf{G}$, $\mathbf{A}$, etc.), bold lower cases for vectors (e.g., $\mathbf{r}$) and calligraphic letters for sets (e.g., $\mathcal{A}$).

Symbol	Definition
$\mathbf{A}$, $\mathbf{B}$	adjacency matrices (bold upper case)
$\mathbf{a}$, $\mathbf{b}$	column vectors (bold lower case)
$\mathcal{A}$, $\mathcal{B}$	sets (calligraphic)
$\mathbf{A}(i,j)$	the element at i^{th} row j^{th} column in matrix $\mathbf{A}$
$\mathbf{a}(i)$	the element at i^{th} position in vector $\mathbf{a}$
$\mathbf{A}'$	the transpose of matrix $\mathbf{A}$
$\mathbf{G}$	layer-layer dependency matrix
$\mathcal{A}$	within-layer connectivity matrices of the network $\mathcal{A} = \{\mathbf{A}_1, ..., \mathbf{A}_g\}$
$\mathcal{D}$	cross-layer dependency matrices $\mathcal{D} = \{\mathbf{D}_{i,j}\ i,j = 1, ..., g\}$
θ, ϕ	one-to-one mapping function
$\mathbf{u}_i$	authority ranking vector for $\mathbf{A}_i$
$\mathbf{v}_i$	hub ranking vector for $\mathbf{A}_i$
g	the total number of layers
n_i	the number of nodes in graph $\mathbf{A}_i$ $(i = 1, ..., g)$
m_i	the number of edges in graph $\mathbf{A}_i$ $(i = 1, ..., g)$

Table 2.1: Symbols

At first, we give the definition of the multi-layered network model. In this paper, we use the MuLaN model proposed by Chen *et al.* (2015). For the full detail, we recommend readers to refer the original paper. Below is the definition for MuLaN:

Definition 1. A Multi-layered Network Model (MuLaN)

Given (1) a binary $g \times g$ abstract layer-layer dependency network $\mathbf{G}$, where $\mathbf{G}(i, j) = 1$ indicates layer j depends on layer i (or layer i support layer j), $\mathbf{G}(i, j) = 0$ means no direct dependency from layer i to layer j; (2) a set of within-layer adjacency matrices $\mathcal{A} = \{\mathbf{A}_1, ..., \mathbf{A}_g\}$ where $\mathbf{A}_i$ describes the connectivities/similarities between nodes within the layer i; (3) a set of inter-layer node-node dependency matrices $\mathcal{D}$, indexed by pair (i, j), $i, j \in [1, ..., g]$, such that for a pair (i, j), if $\mathbf{G}(i, j) = 1$, then $\mathbf{D}_{i,j}$ is an $n_i \times n_j$ dependency matrix; otherwise the corresponding dependency matrix $\mathbf{D}_{i,j}$ is absent; (4) θ is a one-to-one mapping function that maps each node in layer-layer dependency network $\mathbf{G}$ to the corresponding within-layer adjacency matrix $\mathbf{A}_i$ $(i = 1, ..., g)$; (5) ϕ is another one-to-one mapping function that maps each edge in $\mathbf{G}$ to the corresponding inter-layer node-node dependency matrix $\mathbf{D}_{i,j}$. We define a multi-layered network as a quintuple $\Gamma =< \mathbf{G}, \mathcal{A}, \mathcal{D}, \theta, \phi >$.

Figure 2.1 shows a typical example of multi-layered networks. The abstract layer-layer dependency network $\mathbf{G}$ in this example is a line graph. There are 4 within-layer adjacency matrices in $\mathcal{A}$: the chemical network ($\mathbf{A}_1$), the drug network ($\mathbf{A}_2$), the disease network ($\mathbf{A}_3$) and the protein-protein interaction (PPI) network ($\mathbf{A}_4$). Across these layers, there are 3 non-empty dependencies in $\mathcal{D}$: the chemical-drug dependency ($\mathbf{D}_{1,2}$), the drug-disease dependency ($\mathbf{D}_{2,3}$) and the disease-protein dependency ($\mathbf{D}_{3,4}$).

Based on the MuLaN model, we define the problems of Cross-Layer Ranking and Cross-Layer Query as below:

Problem 1: Cross-Layer Ranking

Given: (1) an multi-layered network $\Gamma =< \mathbf{G}, \mathcal{A}, \mathcal{D}, \theta, \phi >$;

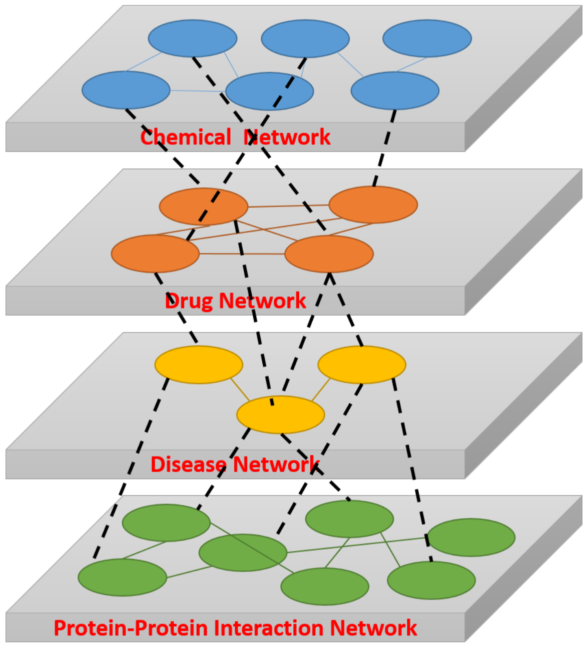

Figure 2.1: A Simple 4-layered Network for Biological Systems

Find: the authority ranking vectors $\mathbf{u}_i$ and the hub ranking vectors $\mathbf{v}_i$ for the nodes in each layer $\mathbf{A}_i$.

Problem 2: Cross-Layer Query

Given: (1) an multi-layered network $\Gamma =< \mathbf{G}, \mathcal{A}, \mathcal{D}, \theta, \phi >$, (2) a query node, and (3) an integer K;

Find: the top-K most relevant nodes (ranked by authority score and ranked by hub score) from each layer $\mathbf{A}_i$ with respect to the query node.

Chapter 3

MULTI-LAYERED HITS ALGORITHM

In this chapter, we propose the multi-layered HITS algorithm to solve the Cross-Layer Ranking problem and Cross-Layer Query problem. We first formulate it as a regularized optimization problem, then present an iterative algorithm to solve it, followed by some theoretical analysis.

3.1 Objective Function

To formulate the ranking task on multi-layered networks, there are two types of constraints we need to take care of. The first constraint is within-layer smoothness, which requires similar nodes on the same layer to have similar ranking scores. This is easy to handle because we can directly apply the regular HITS algorithm within the layer. Refer to Cai and Chakravarthy (2014), regular HITS algorithm is essentially a non-negative matrix factorization problem, which makes it easy to obtain its objective function accordingly. The second constraint is cross-layer consistency that similar nodes across different layers should share similar ranking scores. This is an over-arching principle to perform mining and learning with multi-layered networks. Inspired by Sindhwani and Melville (2008), where the authors think the sentiment-polarity scores between connected word and document should be similar, we assume the nodes who have cross-layer connections should share similar authority scores and hub scores. This gives us a metric for the cross-layer consistency.

According to above analysis, we present the following regularized objective function $J(\mathbf{u}_1, ..., \mathbf{u}_g; \mathbf{v}_1, ..., \mathbf{v}_g)$. The optimal authority ranking vectors $\mathbf{u}_i$ and hub ranking vectors $\mathbf{v}_i$ are the solution for minimizing the objective function.

$$\begin{aligned}\underset{\mathbf{u}_i\geq 0,\mathbf{v}_i\geq 0}{J} =& \underbrace{\sum_{i=1}^{n}\frac{1}{2}||\frac{1}{e_i}\mathbf{A}_i - \mathbf{u}_i\mathbf{v}_i'||_F^2}_{\text{within-layer smoothness}} \\ &+ \mu \underbrace{\sum_{i,j:\mathbf{G}(i,j)=1}\ \sum_{x,y:\mathbf{D_{i,j}}(x,y)=1}[(\mathbf{u}_i(x)-\mathbf{u}_j(y))^2+(\mathbf{v}_i(x)-\mathbf{v}_j(y))^2]}_{\text{cross-layer consistency}}\end{aligned} \tag{3.1}$$

where $||\cdot||_F$ denotes the Frobenius norm, e_i is the total number of links in layer i, i.e., $e_i = \sum_{x,y}\mathbf{A}_i(x,y)$, $\mu > 0$ is a regularization parameter.

Vectorizing the objective function, we have

$$\begin{aligned}\underset{\mathbf{u}_i\geq 0,\mathbf{v}_i\geq 0}{J} =& \underbrace{\sum_{i=1}^{n}\frac{1}{2}||\frac{1}{e_i}\mathbf{A}_i - \mathbf{u}_i\mathbf{v}_i'||_F^2}_{\text{within-layer smoothness}} \\ &+ \mu \underbrace{\sum_{i,j:\mathbf{G}(i,j)=1}\left(\begin{bmatrix}\mathbf{u}_i' & \mathbf{u}_j'\end{bmatrix}\mathbf{L}_{i,j}\begin{bmatrix}\mathbf{u}_i\\ \mathbf{u}_j\end{bmatrix} + \begin{bmatrix}\mathbf{v}_i' & \mathbf{v}_j'\end{bmatrix}\mathbf{L}_{i,j}\begin{bmatrix}\mathbf{v}_i\\ \mathbf{v}_j\end{bmatrix}\right)}_{\text{cross-layer consistency}}\end{aligned} \tag{3.2}$$

where $\mathbf{L}_{i,j} = \begin{bmatrix}\mathbf{T}_{i,j} & 0\\ 0 & \mathbf{T}_{i,j}'\end{bmatrix} - \begin{bmatrix}0 & \mathbf{D}_{i,j}\\ \mathbf{D}_{i,j}' & 0\end{bmatrix} = \begin{bmatrix}\mathbf{T}_{i,j} & -\mathbf{D}_{i,j}\\ -\mathbf{D}_{i,j}' & \mathbf{T}_{i,j}'\end{bmatrix}$ where $\mathbf{T}_{i,j}$ is the diagonal degree matrix associated with $\mathbf{D}_{i,j}$, i.e., $\mathbf{T}_{i,j}(r,r) = \sum_s \mathbf{D}_{i,j}(r,s)$.

3.2 Iterative Algorithms

The optimization problem defined in equation (3.2) is quadratic, so we start by finding the derivatives of J with respect to $\mathbf{u}_i$ and $\mathbf{v}_i'$. Only consider terms related to $\mathbf{u}_i$, we have

$$J(\mathbf{u}_i) = \frac{1}{2}||\frac{1}{e_i}\mathbf{A}_i - \mathbf{u}_i\mathbf{v}_i'||_F^2 + \mu\sum_{j:\mathbf{G}(i,j)=1}(\mathbf{u}_i'\mathbf{T}_{i,j}\mathbf{u}_i - \mathbf{u}_j'\mathbf{D}_{i,j}'\mathbf{u}_i - \mathbf{u}_i'\mathbf{D}_{i,j}\mathbf{u}_j + \mathbf{u}_j'\mathbf{T}_{i,j}'\mathbf{u}_j) \tag{3.3}$$

Also, since we know $||\mathbf{X}||_F^2 = Tr(\mathbf{X}\mathbf{X}')$, then

$$\begin{aligned}\frac{1}{2}||\frac{1}{e_i}\mathbf{A}_i - \mathbf{u}_i\mathbf{v}_i'||_F^2 &= \frac{1}{2}Tr[(\frac{1}{e_i}\mathbf{A}_i - \mathbf{u}_i\mathbf{v}_i')(\frac{1}{e_i}\mathbf{A}_i' - \mathbf{v}_i\mathbf{u}_i')] \\ &= \frac{1}{2}Tr(\frac{1}{e_i^2}\mathbf{A}_i\mathbf{A}_i' - \frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i\mathbf{u}_i' - \mathbf{u}_i\mathbf{v}_i' \cdot \frac{1}{e_i}\mathbf{A}_i' + \mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i\mathbf{u}_i') \\ &= \frac{1}{2}Tr(\frac{1}{e_i^2}\mathbf{A}_i\mathbf{A}_i') - Tr(\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i\mathbf{u}_i') + \frac{1}{2}Tr(\mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i\mathbf{u}_i') \\ &= \frac{1}{2}Tr(\frac{1}{e_i^2}\mathbf{A}_i\mathbf{A}_i') - \mathbf{u}_i' \cdot \frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i + \frac{1}{2}\mathbf{v}_i'\mathbf{v}_i\mathbf{u}_i'\mathbf{u}_i\end{aligned} \tag{3.4}$$

Drop constant terms, we have

$$J(\mathbf{u}_i) = \frac{1}{2}\mathbf{v}_i'\mathbf{v}_i\mathbf{u}_i'\mathbf{u}_i - \mathbf{u}_i' \cdot \frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i + \mu \sum_{j:\mathbf{G}(i,j)=1} (\mathbf{u}_i'\mathbf{T}_{i,j}\mathbf{u}_i - \mathbf{u}_j'\mathbf{D}_{i,j}'\mathbf{u}_i - \mathbf{u}_i'\mathbf{D}_{i,j}\mathbf{u}_j) \tag{3.5}$$

Compute the derivative of J with respect to $\mathbf{u}_i$

$$\frac{\partial J}{\partial \mathbf{u}_i} = (\mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i - \frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i) + 2\mu \sum_{j:\mathbf{G}(i,j)=1} (\mathbf{T}_{i,j}\mathbf{u}_i - \mathbf{D}_{i,j}\mathbf{u}_j) \tag{3.6}$$

Similarly, we can compute the derivative of J with respect to $\mathbf{v}_i'$

$$\frac{\partial J}{\partial \mathbf{v}_i'} = (\mathbf{u}_i'\mathbf{u}_i\mathbf{v}_i' - \mathbf{u}_i' \cdot \frac{1}{e_i}\mathbf{A}_i) + 2\mu \sum_{j:\mathbf{G}(i,j)=1} (\mathbf{v}_i'\mathbf{T}_{i,j}' - \mathbf{v}_j'\mathbf{D}_{i,j}') \tag{3.7}$$

According to the fix-point solution with the non-negativity constraint on the authority ranking vector $\mathbf{u}_i$ and the hub ranking vector $\mathbf{v}_i$, we have the following multiplicative updating rule

$$\begin{aligned}\mathbf{u}_i(x) \leftarrow & \mathbf{u}_i(x)\sqrt{\frac{[\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i + 2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{D}_{i,j}\mathbf{u}_j](x)}{[\mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i + 2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{T}_{i,j}\mathbf{u}_i](x)}} \\ \mathbf{v}_i'(x) \leftarrow & \mathbf{v}_i'(x)\sqrt{\frac{[\mathbf{u}_i' \cdot \frac{1}{e_i}\mathbf{A}_i + 2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{v}_j'\mathbf{D}_{i,j}'](x)}{[\mathbf{u}_i'\mathbf{u}_i\mathbf{v}_i' + 2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{v}_i'\mathbf{T}_{i,j}'](x)}}\end{aligned} \tag{3.8}$$

With the multiplicative updating rules, it is easy to summarize the algorithms for Cross-Layer Ranking and Cross-Layer Query.

3.2.1 Cross-Layer Ranking

To solve the problem of Cross-Layer Ranking, we can directly apply the iterative algorithm to the multi-layered network model. The algorithm is summarized in Algorithm 1.

Algorithm 1: Cross-Layer Ranking

Input : (1) an multi-layered network $\Gamma =< \mathbf{G}, \mathcal{A}, \mathcal{D}, \theta, \phi >$;

Output: the authority ranking vectors $\mathbf{u}_i$ and the hub ranking vectors $\mathbf{v}_i$ for nodes in each layer $\mathbf{A}_i$;

for $i \leftarrow 1$ **to** g **do**

  Initialize $\mathbf{u}_i$, $\mathbf{v}_i$ as $n_i \times 1$ non-negative random vectors;

end

while *not converge* **do**

  $\mathbf{u}_i(x) \leftarrow \mathbf{u}_i(x)\sqrt{\frac{[\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{D}_{i,j}\mathbf{u}_j](x)}{[\mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{T}_{i,j}\mathbf{u}_i](x)}}$;

  $\mathbf{v}_i'(x) \leftarrow \mathbf{v}_i'(x)\sqrt{\frac{[\mathbf{u}_i'\cdot\frac{1}{e_i}\mathbf{A}_i+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{v}_j'\mathbf{D}_{i,j}'](x)}{[\mathbf{u}_i'\mathbf{u}_i\mathbf{v}_i'+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{v}_i'\mathbf{T}_{i,j}'](x)}}$;

end

return *the authority ranking vectors* $\mathbf{u}_i$ *and the hub ranking vectors* $\mathbf{v}_i$*;*

3.2.2 Cross-Layer Query

To solve the problem of Cross-Layer Query, we need to modify the Cross-Layer Ranking algorithm with respect to the query node. One solution is to extract a subgraph with respect to the query node using breadth-first search (BFS) or random walk with restart (RWR). Then we still use the MuLaN to model the extracted subgraph. In this way, we can obtain a smaller multi-layered network which centers on the query node. Applying the iterative algorithm to this new multi-layered networks, the ranking vectors with respect to the query node will be returned. Then the only thing we need to do is to take the top-K ranked results from the ranking vectors. This algorithm is summarized in Algorithm 2.

Algorithm 2: Cross-Layer Query

Input : (1) an multi-layered network $\Gamma =< \mathbf{G}, \mathcal{A}, \mathcal{D}, \theta, \phi >$, (2) a query node, and (3) an integer K;

Output: the top-K most relevant nodes (ranked by authority score and ranked by hub score) from each layer $\mathbf{A}_i$ with respect to the query node.

Extract a subgraph from the query node using BFS or RWR;

Form the subgraph as a smaller multi-layered network;

Use the extracted multi-layered network as the new input;

for $i \leftarrow 1$ **to** $\hat{g}$ **do**

  Initialize $\mathbf{u}_i$, $\mathbf{v}_i$ as $\hat{n}_i \times 1$ non-negative random vectors;

end

while *not converge* **do**

  $\mathbf{u}_i(x) \leftarrow \mathbf{u}_i(x)\sqrt{\frac{[\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{D}_{i,j}\mathbf{u}_j](x)}{[\mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{T}_{i,j}\mathbf{u}_i](x)}}$;

  $\mathbf{v}_i'(x) \leftarrow \mathbf{v}_i'(x)\sqrt{\frac{[\mathbf{u}_i'\cdot\frac{1}{e_i}\mathbf{A}_i+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{v}_j'\mathbf{D}_{i,j}'](x)}{[\mathbf{u}_i'\mathbf{u}_i\mathbf{v}_i'+2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{v}_i'\mathbf{T}_{i,j}'](x)}}$;

end

return *the top-K ranked nodes from each layer $\mathbf{A}_i$;*

3.3 Proof and Analysis

Here, we analyze the proposed multi-layered HITS algorithm in terms of its effectiveness by proving that the proposed multi-layered HITS algorithm indeed finds a local optimum solution to Equation 3.2. At first, we give the following theorem, which says that the fixed point solution of Equation 3.8 satisfies the KKT condition.

Theorem 3.1. The fixed point solution of Equation 3.8 satisfies the KKT condition.

Proof. The Lagrangian function of Equation 3.5 can be written as

$$\begin{aligned} L(\mathbf{u}_i) = & \frac{1}{2}\mathbf{v}_i'\mathbf{v}_i\mathbf{u}_i'\mathbf{u}_i - \mathbf{u}_i' \cdot \frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i \\ & + \mu \sum_{j:\mathbf{G}(i,j)=1} (\mathbf{u}_i'\mathbf{T}_{i,j}\mathbf{u}_i - \mathbf{u}_j'\mathbf{D}_{i,j}'\mathbf{u}_i - \mathbf{u}_i'\mathbf{D}_{i,j}\mathbf{u}_j) - \boldsymbol{\alpha}\mathbf{u}_i \end{aligned} \tag{3.9}$$

where $\boldsymbol{\alpha}$ is the Lagrange multiplier. Setting the derivative of $L(\mathbf{u}_i)$ to 0, we get

$$(\mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i - \frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i) + 2\mu \sum_{j:\mathbf{G}(i,j)=1} (\mathbf{T}_{i,j}\mathbf{u}_i - \mathbf{D}_{i,j}\mathbf{u}_j) = \boldsymbol{\alpha} \tag{3.10}$$

By the KKT complementary slackness condition, we have

$$[(\mathbf{u}_i\mathbf{v}_i'\mathbf{v}_i + 2\mu \sum_{j:\mathbf{G}(i,j)=1} \mathbf{T}_{i,j}\mathbf{u}_i) - (\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i + 2\mu \sum_{j:\mathbf{G}(i,j)=1} \mathbf{D}_{i,j}\mathbf{u}_j)](x)\mathbf{u}_i(x) = 0 \tag{3.11}$$

Similarly,

$$[(\mathbf{u}_i'\mathbf{u}_i\mathbf{v}_i' + 2\mu \sum_{j:\mathbf{G}(i,j)=1} \mathbf{v}_i'\mathbf{T}_{i,j}') - (\mathbf{u}_i' \cdot \frac{1}{e_i}\mathbf{A}_i + 2\mu \sum_{j:\mathbf{G}(i,j)=1} \mathbf{v}_j'\mathbf{D}_{i,j}')](x)\mathbf{v}_i(x) = 0 \tag{3.12}$$

Therefore, we can see that the fixed point solution of Equation 3.8 satisfies the above equations. □

The convergence of the proposed multi-layered HITS algorithm is given by the following lemma.

Lemma 3.2. Under the updating rule in Equation 3.8, the objective function in Equation 3.2 decreases monotonically.

Proof. From Equation 3.5, we have

$$J(\mathbf{u}_i) = \underbrace{-\mathbf{u}_i' \cdot \frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i}_{Q_1} + \underbrace{\frac{1}{2}\mathbf{v}_i'\mathbf{v}_i\mathbf{u}_i'\mathbf{u}_i}_{Q_2} + \mu \sum_{j:\mathbf{G}(i,j)=1} (\underbrace{\mathbf{u}_i'\mathbf{T}_{i,j}\mathbf{u}_i}_{Q_3} \underbrace{-\mathbf{u}_j'\mathbf{D}_{i,j}'\mathbf{u}_i}_{Q_4} \underbrace{-\mathbf{u}_i'\mathbf{D}_{i,j}\mathbf{u}_j}_{Q_5}) \tag{3.13}$$

Following the auxiliary function approach in Lee and Seung (2001), the auxiliary function $H(\mathbf{u}_i, \tilde{\mathbf{u}}_i)$ of $J(\mathbf{u}_i)$ must satisfy

$$H(\mathbf{u}_i, \tilde{\mathbf{u}}_i) \geq J(\mathbf{u}_i),\ H(\mathbf{u}_i, \mathbf{u}_i) = J(\mathbf{u}_i) \tag{3.14}$$

Define

$$\mathbf{u}_i^{(t+1)} = arg \min_{\mathbf{u}_i} H(\mathbf{u}_i, \mathbf{u}_i^{(t)}) \tag{3.15}$$

by this construction, we have

$$J(\mathbf{u}_i^{(t+1)}) \leq H(\mathbf{u}_i^{(t+1)}, \mathbf{u}_i^{(t)}) \leq H(\mathbf{u}_i^{(t)}, \mathbf{u}_i^{(t)}) = J(\mathbf{u}_i^{(t)}) \tag{3.16}$$

which proves that $J(\mathbf{u}_i^{(t)})$ decreases monotonically.

Next, we prove that (1) we can find such an auxiliary function which satisfies the constraints and (2) the updating rule in Equation 3.8 leads to the global minimum solution to the auxiliary function.

First, we show that the following function is one of the auxiliary function of Equation 3.13:

$$H(\mathbf{u}_i, \tilde{\mathbf{u}}_i) = (Q'_1 + Q'_2) + \mu \sum_{j:\mathbf{G}(i,j)=1} (Q'_3 + Q'_4 + Q'_5) \tag{3.17}$$

where

$$Q'_1 = -\sum_{x=1}^{n_i} [\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i](x)\tilde{\mathbf{u}}_i(x)(1 + log\frac{\mathbf{u}_i(x)}{\tilde{\mathbf{u}}_i(x)}) \tag{3.18}$$

$$Q'_2 = \frac{1}{2}\mathbf{v}'_i\mathbf{v}_i \sum_{x=1}^{n_i} \mathbf{u}_i^2(x) \tag{3.19}$$

$$Q'_3 = \sum_{x=1}^{n_i} \frac{[\mathbf{T}_{i,j}\tilde{\mathbf{u}}_i](x)\mathbf{u}_i^2(x)}{\tilde{\mathbf{u}}_i(x)} \tag{3.20}$$

$$Q'_4 = -\sum_{x=1}^{n_i} [\mathbf{D}_{i,j}\mathbf{u}_j](x)\tilde{\mathbf{u}}_i(x)(1 + log\frac{\mathbf{u}_i(x)}{\tilde{\mathbf{u}}_i(x)}) \tag{3.21}$$

$$Q'_5 = -\sum_{x=1}^{n_i} [\mathbf{D}_{i,j}\mathbf{u}_j](x)\tilde{\mathbf{u}}_i(x)(1 + log\frac{\mathbf{u}_i(x)}{\tilde{\mathbf{u}}_i(x)}) \tag{3.22}$$

Using the inequality $z \geq 1 + logz$, we have

$$Q_1' \geq -\sum_{x=1}^{n_i} [\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i](x)\mathbf{u}_i(x) = Q_1 \tag{3.23}$$

$$Q_4' \geq -\sum_{x=1}^{n_i} [\mathbf{D}_{i,j}\mathbf{u}_j](x)\mathbf{u}_i(x) = Q_4 \tag{3.24}$$

$$Q_5' \geq -\sum_{x=1}^{n_i} [\mathbf{D}_{i,j}\mathbf{u}_j](x)\mathbf{u}_i(x) = Q_5 \tag{3.25}$$

For Q_2', we have

$$Q_2' = Q_2 \tag{3.26}$$

For Q_3', by using the following inequality in Ding *et al.* (2006)

$$\sum_{i=1}^{n}\sum_{p=1}^{k} \frac{[\mathbf{A}\mathbf{S}^*\mathbf{B}]\mathbf{S}^2(i,p)}{\mathbf{S}^*(i,p)} \geq Tr(\mathbf{S}'\mathbf{A}\mathbf{S}\mathbf{B}) \tag{3.27}$$

where $\mathbf{A} \in \mathbb{R}_+^{n\times n}$, $\mathbf{B} \in \mathbb{R}_+^{k\times k}$, $\mathbf{S} \in \mathbb{R}_+^{n\times k}$, $\mathbf{S}^* \in \mathbb{R}_+^{n\times n}$, and $\mathbf{A}$, $\mathbf{B}$ are symmetric, we have

$$Q_3' >= Tr(\mathbf{u}_i'\mathbf{T}_{i,j}\mathbf{u}_i) = \mathbf{u}_i'\mathbf{T}_{i,j}\mathbf{u}_i = Q_3 \tag{3.28}$$

Putting the above inequalities together, we have $H(\mathbf{u}_i, \tilde{\mathbf{u}}_i) \geq J(\mathbf{u}_i)$.

Then we find the global minimum solution to $H(\mathbf{u}_i, \tilde{\mathbf{u}}_i)$. The gradient of $H(\mathbf{u}_i, \tilde{\mathbf{u}}_i)$ is computed as

$$\begin{aligned}\frac{\partial H(\mathbf{u}_i, \tilde{\mathbf{u}}_i)}{\partial \mathbf{u}_i(x)} &= -\frac{[\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i](x)\tilde{\mathbf{u}}_i(x)}{\mathbf{u}_i(x)} + \mathbf{v}_i'\mathbf{v}_i\mathbf{u}_i(x) \\ &+ 2\mu \sum_{j:\mathbf{G}(i,j)=1} \frac{[\mathbf{T}_{i,j}\tilde{\mathbf{u}}_i](x)\mathbf{u}_i(x)}{\tilde{\mathbf{u}}_i(x)} - 2\mu \sum_{j:\mathbf{G}(i,j)=1} \frac{[\mathbf{D}_{i,j}\mathbf{u}_j](x)\tilde{\mathbf{u}}_i(x)}{\mathbf{u}_i(x)}\end{aligned} \tag{3.29}$$

Setting the gradient to zero, we have

$$\mathbf{u}_i^2(x) = \tilde{\mathbf{u}}_i^2(x)\frac{[\frac{1}{e_i}\mathbf{A}_i\mathbf{v}_i + 2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{D}_{i,j}\mathbf{u}_j](x)}{[\tilde{\mathbf{u}}_i\mathbf{v}_i'\mathbf{v}_i + 2\mu\sum_{j:\mathbf{G}(i,j)=1}\mathbf{T}_{i,j}\tilde{\mathbf{u}}_i](x)} \tag{3.30}$$

Recall that we have set $\mathbf{u}_i^{(t+1)} = \mathbf{u}_i$ and $\mathbf{u}_i^{(t)} = \tilde{\mathbf{u}}_i$. The above equation proves that the updating rule for $\mathbf{u}$ in 3.8 decreases monotonically. By reversing the roles of $\mathbf{u}$ and $\mathbf{v}$, it is simple to show that the updating rule for $\mathbf{v}$ decreases monotonically, too. □

Chapter 4

EVALUATION

In this chapter, we evaluate the proposed multi-layered HITS algorithm on a real dataset from the following two aspects:

- **Effectiveness**. How effective are the multi-layered HITS algorithm on ranking the nodes within different layers?
- **Explainability**. Is the ranking result of multi-layered HITS algorithm explainable?

4.1 Dataset & Preprocessing

We use the real dataset of Amazon product co-purchasing network[1] [Yang and Leskovec (2015)] for our experiments. In this network, each node represents a product and each edge represents an undirected co-purchasing relationship. Here, if a product i is frequently purchased with product j, i and j are defined as being co-purchased.

From the Amazon product metadata[2] [Leskovec *et al.* (2007)], we further extract product group information. Each product belongs to a group, which can be Book, DVD, Music, or Video. This allows us to divides the products into 4 groups. Utilizing the MuLaN model, we can build a 4-layered products co-purchasing network.

Besides the group information, Amazon product metadata also contains data about customer reviews. This could be regarded as a kind of knowledge for improving the ranking performance of the algorithm. Specifically, we add a customer

[1]Please refer to `http://snap.stanford.edu/data/com-Amazon.html`

[2]Please refer to `http://snap.stanford.edu/data/amazon-meta.html`

layer to the multi-layered products co-purchasing network, where each edge between customer layer and product layer represents a review. Since each review has a rating (1 stands for most negative, 5 stands for most positive) and we do not consider any edge attributes in our algorithm, we want to consider only one polarity for consistency. Therefore, we only keep the reviews whose ratings are greater than or equal to 4. Also, to avoid introducing too much noisy information, we only consider valuable reviews which have at least 20 votes and more than 80% helpfulness.

After preprocessing, we build a 5-layered Amazon network model. In this multi-layered network model, both within-layer connectivities and the cross-layer dependencies are binary and undirected. The structure is shown in Figure 4.1. The statistics is shown in Table 4.1.

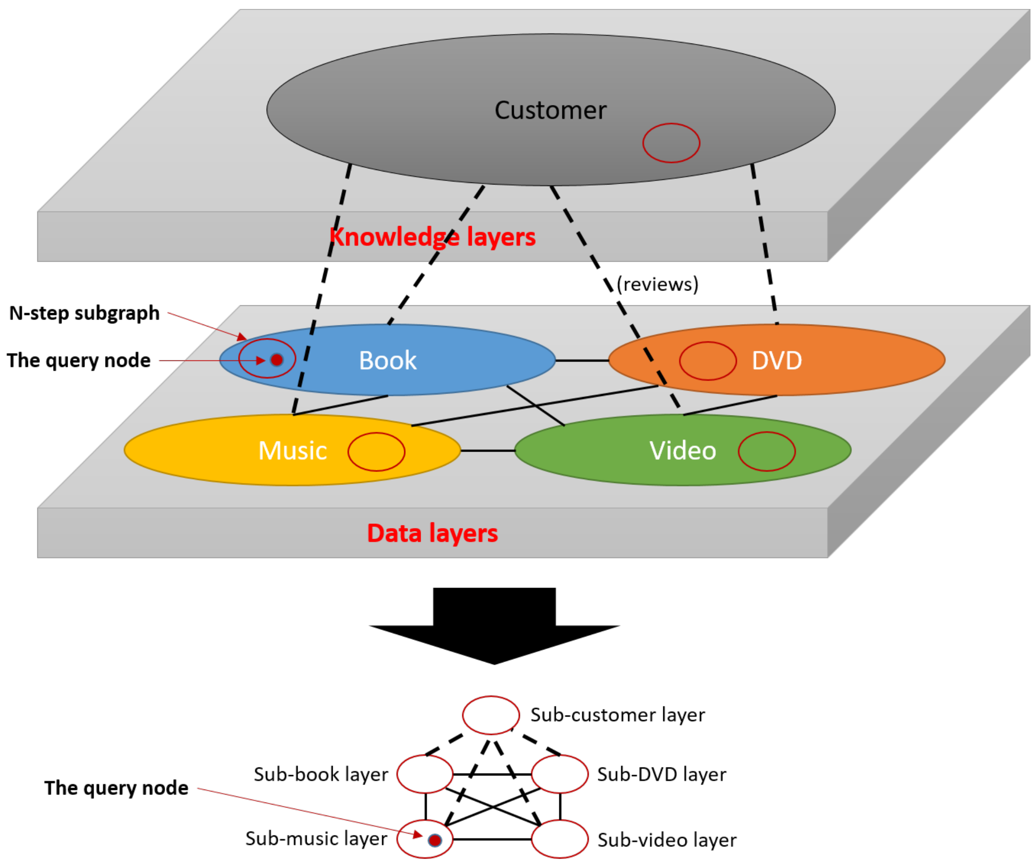

Figure 4.1: Amazon Multi-layered Networks

Dataset	# of Layer	# of Nodes	# of Links	# of CrossLinks
AMAZON	$4+1$	399452	1088423	203510

Table 4.1: Amazon Co-purchasing Network Statistics

4.2 Experimental Design & Prototype

Based on the processed Amazon co-purchasing network, we compare the performance of the following methods on both ranking and query task: (1) Regular HITS algorithm on a single layer (e.g., Book layer); (2) Regular HITS on flattened network; (3) Multi-layered HITS algorithm ($\mu = 0.1$) on 4-layered co-purchasing networks (customer layer is NOT included); (4) Multi-layered HITS algorithm ($\mu = 0.1$) on 5-layered co-purchasing networks (customer layer is included).

All of the algorithms above return ranking vectors (authority and hub) for different layers. Here, we only take the top-5 ranked products from each layer and invite 10 graduate students to evaluate their relevance and explainability. Specifically, the top-5 ranked Books are evaluated by the relevance (for ranking, relevance stands for global importance/popularity; for query, relevance stands for similarity w.r.t the query product); the top-5 ranked DVDs, Musics, and Videos are evaluated by the explainability, that is, the helpfulness of the top-5 ranked products in terms of interpreting the top-5 ranked books. Here, the intuition for explainability is that we assume two relevant ranking results can somehow explain each other.

For the convenience of performing human rating, we also implement a prototype [3] which allows users to run real-time experiments and evaluate the experimental results (users can submit their ratings online). This prototype is built with a tech stack of React.js, Flask, and MongoDB. The interface is shown in Figure 4.2. For the query task, we select 6 query products for user evaluation.

[3] Please refer to `http://thesis.haichaoy.com/`

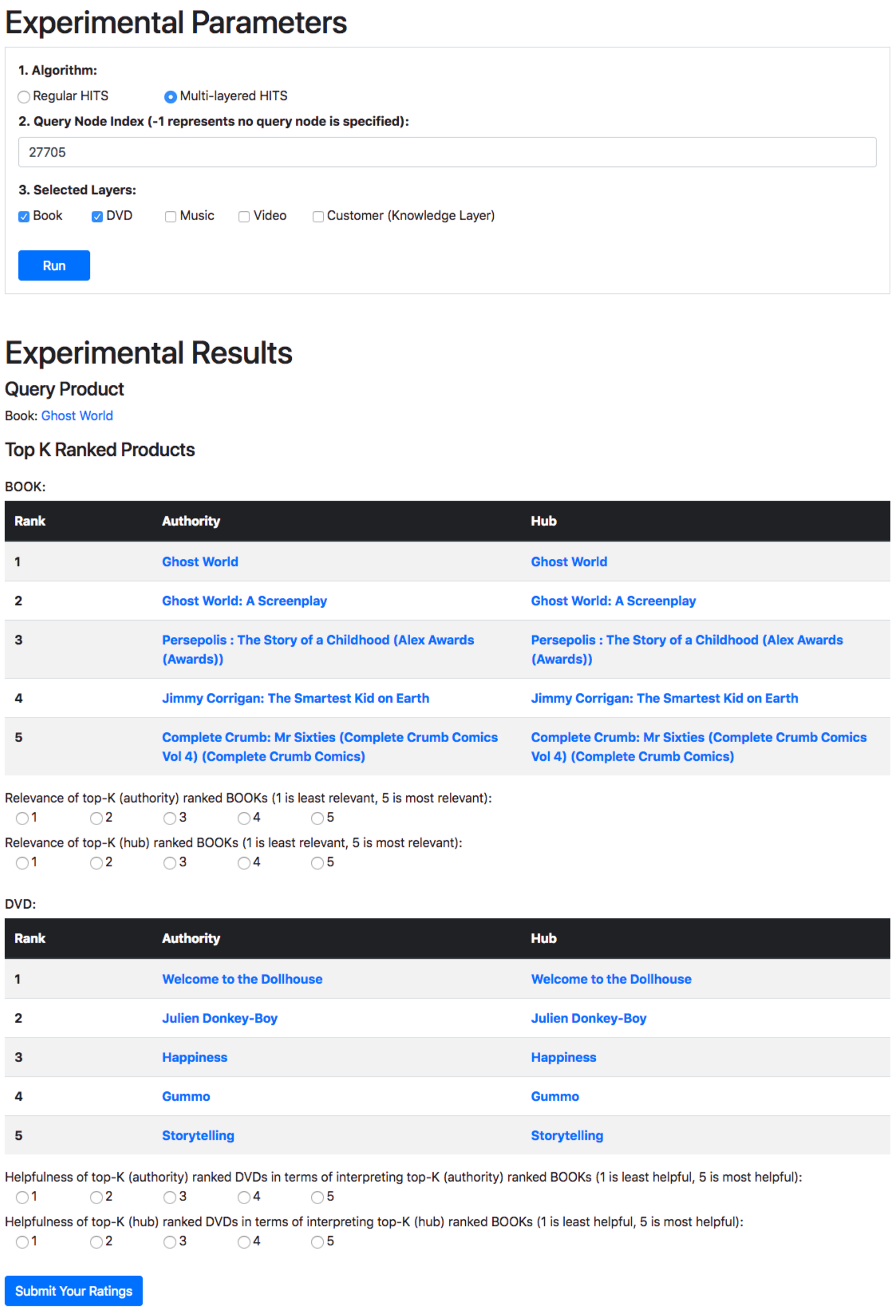

Figure 4.2: Prototype Interface

4.3 Experimental Results & Analysis

After collecting the ratings from 10 graduate students, we compute the average ratings for ranking task and query task regarding the relevance and explainability.

4.3.1 Effectiveness

Generally, a ranking algorithm is considered more effective if it gives more relevant ranking results. As we mentioned in the previous section, Book layer is taken to evaluate the effectiveness of the algorithm. The experimental results for ranking and query tasks are shown in Figure 4.3 and Figure 4.4.

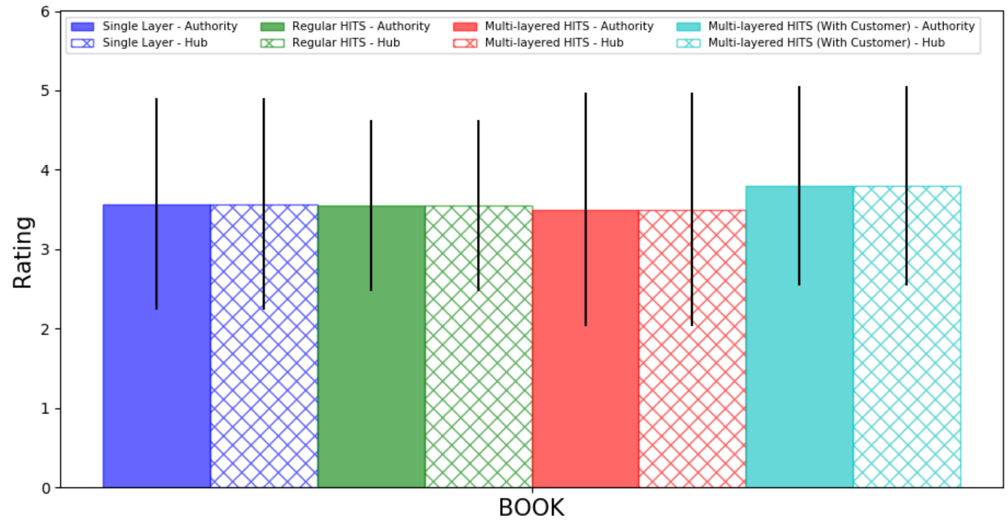

Figure 4.3: Average Rating Regarding Relevance in Ranking Task

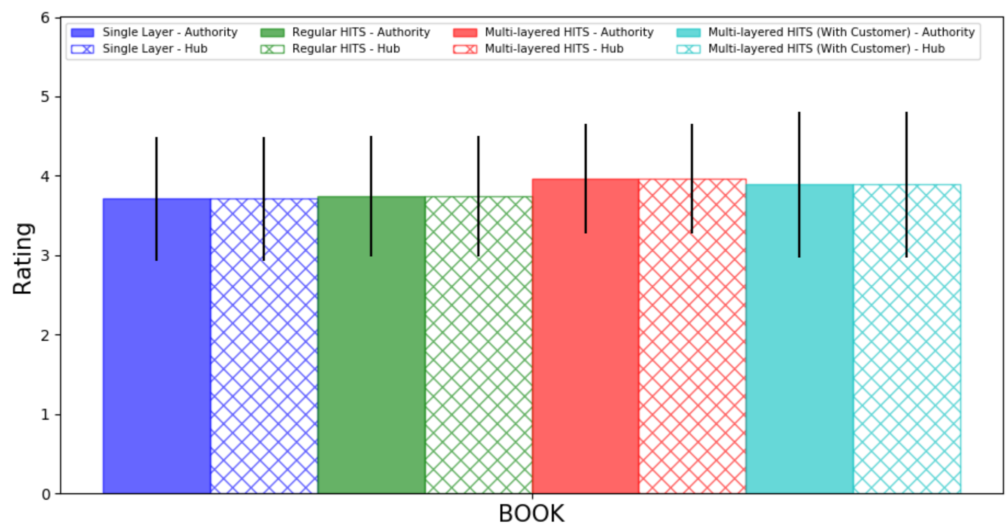

Figure 4.4: Average Rating Regarding Relevance in Query Task

For the ranking task, the regular HITS on single layer network and regular HITS on flattened network achieve similar relevance ratings because they give similar ranking results. These two methods are considered as the baseline methods. Without considering the customer (knowledge) layer, the proposed multi-layered HITS algorithm performs slightly worse than the baseline method. The reason is that the most important/popular products across different layers are not necessarily relevant to each other. For example, the most popular books are not necessarily relevant to the most popular DVDs. However, the multi-layered HITS algorithm performs much better if the customer (knowledge) layer is considered. This is because the reviews we consider are positive and helpful, which helps the algorithm to find more important/popular books through the cross-layer dependency.

For the query task, the multi-layered HITS algorithm (without customer) achieves the best performance. Here, the top-5 ranked products from each layer are more relevant to each other. Yet, it seems that the positive customer reviews do not have much contribution in this circumstance. Instead, they may introduce some noisy information to weaken the performance of the algorithm.

Also, compared with the rating results in query task, the rating results in ranking task have much higher variance. The is probably caused by three reasons: (1) There are more experiments and ratings for query task (6 query nodes); (2) Compared with evaluating similarity, evaluating global importance/similarity is vaguer because of a lack of strict criterion; (3) Co-purchasing relationship does not reflect global importance/popularity precisely.

4.3.2 Explainability

To evaluate the explainability, we ask the users to give ratings to the top-5 ranked products in DVD, Music, and Video layer by the helpfulness in terms of interpreting the top-5 ranked books. The experimental results for ranking and query tasks are shown in Figure 4.5 and Figure 4.6.

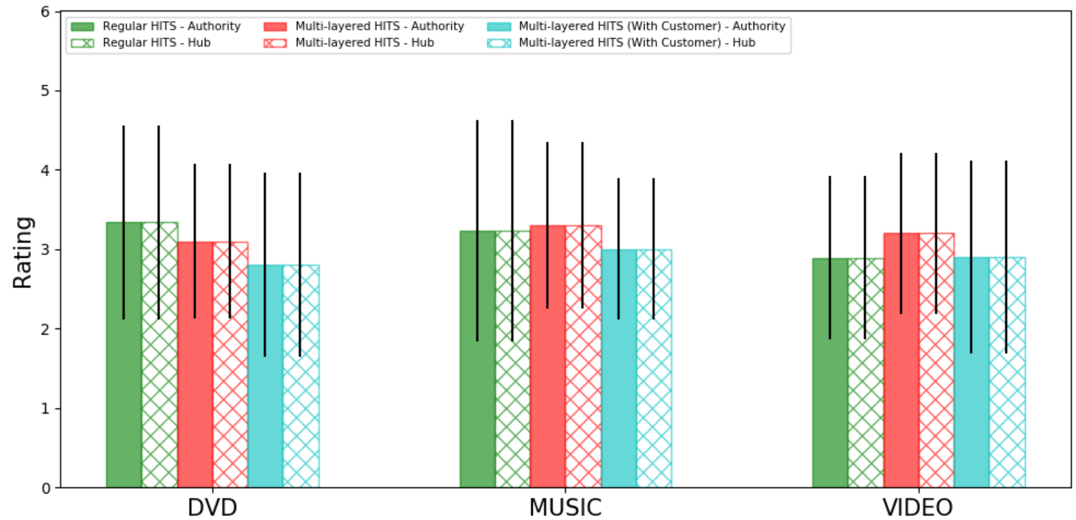

Figure 4.5: Average Rating Regarding Explainability in Ranking Task

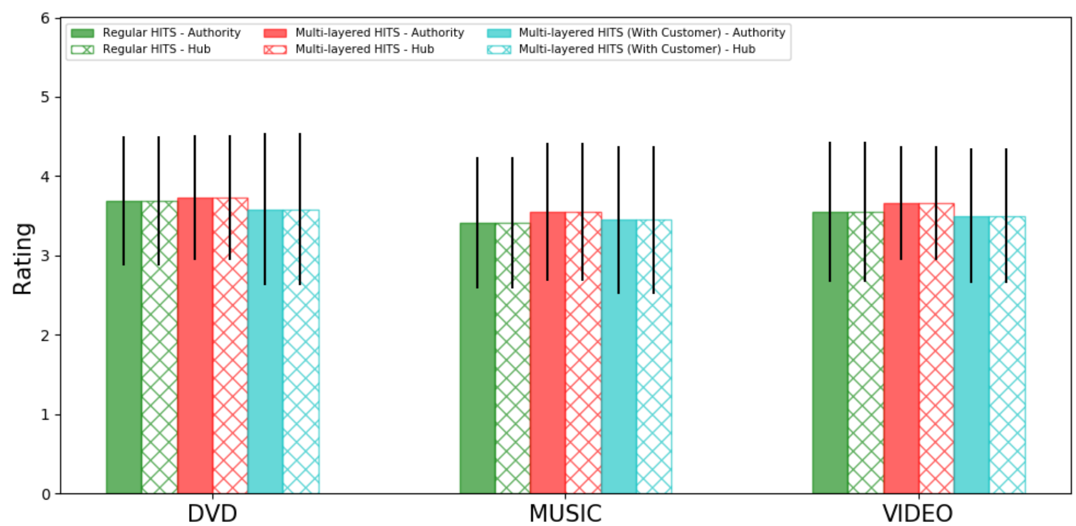

Figure 4.6: Average Rating Regarding Explainability in Query Task

In the ranking task, the multi-layered HITS (without customer) receives the highest explainability ratings in all layers except the DVD layer. The possible reason might be that there are dense connections between Book layer and DVD layer so that even if we do not consider cross-layer consistency HITS algorithm still can give a quite explainable ranking result for DVD layer. However, for layers who do not have such dense connections with Book layer (e.g., Music layer and Video layer), multi-layered HITS performs slightly better because of cross-layer consistency.

In the query task, multi-layered HITS (without customer) receives the highest explainability rating in all of the 3 layers. This essentially demonstrates that multi-layered HITS algorithm is more explainable than regular HITS algorithm. However, the knowledge layer (customer reviews) does not help improve the explainability of the algorithm.

Chapter 5

LITERATURE SURVEY

5.1 Network of Networks

Network of Networks (NoN) is a new data model for multi-sourced graph mining, where each node of the main network itself can be further represented as another domain-specific network. It has been demonstrated that this NoN model is effective and efficient in both ranking task and clustering task.

In ranking task, Ni *et al.* (2014) assumes the ranking scores for the same node should be consistent across highly similar domain-specific networks. Based on this assumption, they generalized the random walk with restart (also called manifold ranking, personalized PageRank, which is a popular ranking method on homogeneous network) [Zhou *et al.* (2004b,a); He *et al.* (2004)] to the NoN model, and developed a new ranking algorithm (CrossRank & CrossQuery) for the multi-sourced networks. This new method has shown its effectiveness and efficiency on the co-authorship networks [Tang *et al.* (2008)] and the protein-protein interaction (PPI) networks [Hamosh *et al.* (2005); Magger *et al.* (2012); Lage *et al.* (2008)].

In clustering task, Ni *et al.* (2015) assumes the clustering results (which are often encoded as a cluster-membership vector for each node) for the same domain node should be consistent across two domain-specific networks if these two domain networks likely belong to the same cluster (i.e., main network clustering). In this way, the clustering task is performed in two different levels - one on node/data level, the other on network level. Based on this assumption, the authors proposed a new clustering algorithm (NoNClus) for NoN model, which is shown to be effective according to the extensive experimental evaluation.

Both NoN ranking method [Ni *et al.* (2014)] and NoN clustering method [Ni *et al.* (2015)] made an assumption, that the same domain node across similar domain-specific networks should share consistent properties (i.e., ranking scores, clustering results). This assumption is so-called cross-network consistency, which is an over-arching principle to perform mining and learning with the NoN model. From the transfer-learning perspective, this principle enables us to transfer the knowledge across different domain networks through the main network.

Besides NoN model, there are many other advanced network models for ranking and clustering tasks, but NoN model has many advantages compared with them. (1) For ranking tasks, there are various advanced network models, such as multi-relational network [Ng *et al.* (2011)], heterogeneous information network [Sun *et al.* (2009)], and hypergraph [Zhou *et al.* (2007)]. Multi-relational network incorporates different types relations into objects, while heterogeneous information network contains different types of objects. A novel method called Multi-Relational Influence Propagation (MRIP) [Yang *et al.* (2012)] combined them together. A tensor-based co-ranking framework [Ng *et al.* (2011)] is proposed for multi-relational network to determine the importance of objects and relations simultaneously. In heterogeneous information network, ranking-based clustering [Sun *et al.* (2009)] and ranking-based classification [Ji *et al.* (2011)] are developed, where ranking and clustering/classification can be mutually enhanced. In a hypergraph [Zhou *et al.* (2007)], an edge is a subset of vertices, but the set of vertices do not form any network topology. Compared with these network model, the main advantage of NoN model is its generality and flexibility. Also, the NoN model enables to compare the nodes in a broader context and rank them at a finer granularity. (2) For clustering tasks, the most common network model is multi-view networks, where views can be networks [Zhou and Burges (2007);

Kumar *et al.* (2011); Kumar and Daumé (2011)] or data-feature matrices [Akata *et al.* (2011); Liu *et al.* (2013); Wang *et al.* (2013b)]. Ensemble clustering [Strehl and Ghosh (2002); Fern and Brodley (2004); Greene and Cunningham (2009)] is based on multi-view clustering, where a consensus clustering is obtained by applying the same clustering algorithm on different views or applying different clustering algorithms on the same view. Recently, some work extends the traditional multi-view clustering to allow incomplete views. Also, a multi-domain graph clustering is proposed to allow flexible network sizes and the number of clusters. The commonality of these clustering models is that they assume there is a common clustering structure shared across all views/domains, while this assumption does not hold in many real-life applications. Compared with these models, the biggest advantage of NoN model is that it allows multiple underlying clustering structures across different networks, which is more flexible and robust.

Based on NoN model, there are many exciting research achievements have been made. A new multiple network clustering method (ComClus) [Ni *et al.* (2016a)] has been proposed to simultaneously group networks and detect common clusters, and enhance clustering accuracy by group-wise consensus. Also, many works such as [Ni *et al.* (2016b, 2017); Chen and Xu (2017)] applied NoN model to bioinformatics domain and outperformed traditional methods. In community detection, the new methods [Li *et al.* (2016); Kim *et al.* (2016)] on heterogeneous social network and multi-layer graph achieved great performance.

5.2 Multi-layered Networks

Multi-layered network has become a hot research topic in recent years. Kivelä *et al.* (2014) provides a comprehensive summary about different types of multi-layered networks, including multi-modal networks [Heath and Sioson (2009)], multi-dimensional networks [Berlingerio *et al.* (2011)], multiplex networks [Battiston *et al.* (2014)], inter-dependent networks [Buldyrev *et al.* (2009)], etc. Compared with single-layered networks, multi-layered networks are more vulnerable as even a small disturbance on supporting layer/network might cause a ripple effect to all the dependent layers, leading to a catastrophic/cascading failure [Buldyrev *et al.* (2009); Vespignani (2010)]. Therefore, network robustness [Gao *et al.* (2011)] becomes one of the most studied problems in inter-dependent networks. Yet, most of the previous research work focuses on the two-layered networks [Buldyrev *et al.* (2009); Gao *et al.* (2012); Parshani *et al.* (2010); Shao *et al.* (2011)]. Few of work has been done in a more broader context.

In 2015, Chen *et al.* (2015) proposed a new multi-layered network model called MuLaN, which admits an arbitrary number of layers. This MuLaN model is a further generalization of Network of Networks (NoN). Specifically, this model extends the main node connections in NoN to inter-layer node-node dependency matrices. In this way, the multi-layered network model has more flexible dependency structure among different layers (domain networks). In this MuLaN model, the authors defined a new optimal connectivity control problem (Opera), and developed a near-optimal control algorithm to solve this problem (find a set of nodes which have the maximum overall impact on the whole network).

Compared with single-layered network, the unique topological characteristic of multi-layered networks lies in its cross-layer dependency. In the MuLaN model, the cross-layer dependency is assumed to be prior knowledge, which, however, is difficult to obtain in real-life applications. To address this problem of inferring missing cross-layer dependencies on multi-layered network, Chen et al. proposed a new algorithm called FASCINATE [Chen *et al.* (2016)] and proved its effectiveness and efficiency. Also, Chen *et al.* (2017) re-explained this problem from a collaborative filtering perspective.

5.3 Knowledge Graph Based Mining

In the past decades, a lot of effort has been made to reduce the labeling work done by humans for supervised learning or to improve the performance of unsupervised learning with only minimal supervision. For example, semi-supervised learning [Chapelle *et al.* (2009)] is proposed to use only a part of labeled data to achieve the same performance as the fully supervised learning. Transfer learning [Pan and Yang (2010)] uses the labeled data from source domain to help the learning task in target domain. The intuition is that the labeled data in source domain is easier to acquire than in target domain. Yet, the human labeling work is still necessary, especially in those very specific domains where human experts are needed. The labeling cost is huge, and how to reduce this cost has become a very important research question.

In the era of big data, with the spring up of general-purpose knowledge bases (or knowledge graphs), e.g., Cyc project [Lenat and Guha (1989)], Wikipedia, Freebase [Bollacker *et al.* (2008)], KnowItAll [Etzioni *et al.* (2004)], TextRunner [Banko *et al.* (2007)], ReVerb [Fader *et al.* (2011)], Ollie [Schmitz *et al.* (2012)], WikiTaxonomy [Ponzetto and Strube (2007)], Probase [Wu *et al.* (2012)], DBpedia [Auer *et al.* (2007)], YAGO [Suchanek *et al.* (2007)], NELL [Mitchell *et al.* (2015)] and Knowledge Vault [Dong *et al.* (2014)], how to apply knowledge graph to machine learning and data mining to reduce the labeling cost and improve the learning performance has attracted many research attentions. Generally, there are two ways for doing that: (1) incorporate domain knowledge into machine learning; (2) incorporate world knowledge into machine learning.

The idea of incorporate domain knowledge into machine learning algorithms has been studied carefully in natural language processing community. A constraint conditional model (CCM) [Chang *et al.* (2012)] has been proposed to inject high-level knowledge as a soft constraint into linear models. Posterior Regularization (PR) [Ganchev *et al.* (2010)] works on incorporating indirect supervision via constraints on posterior distributions of probabilistic models with latent variables. The different between CCM and PR is that CCMs allows the use of hard constraints, while PR uses expectation constraints. Samdani *et al.* (2012) proposed a unified Expectation-Maximization algorithm to combine CCR and PR together. To gain better performance, many other models have been studied to incorporate domain knowledge, such as Markov Logic Network (MLN) [Richardson and Domingos (2006)], a combination of Bayesian network model [Dechter and Mateescu (2004)], etc. Also, transfer learning [Pan and Yang (2010)] is a direction on leveraging domain knowledge for better machine learning performance.

Instead of using domain knowledge, incorporating world knowledge into machine learning algorithm is a better and more challenging way to improve performance. Most of existing work consider world knowledge as a source of features, and use them in tasks like text classification [Gabrilovich and Markovitch (2005, 2006, 2007, 2009); Wang *et al.* (2014a)], clustering [Hotho *et al.* (2003); Hu *et al.* (2008, 2009a,b); Fodeh *et al.* (2011); Song *et al.* (2011, 2015)], information retrieval [Egozi *et al.* (2011); Hua *et al.* (2013); Song *et al.* (2014); Wang *et al.* (2016b)], mining knowledge from text for information retrieval [Wang *et al.* (2013a)], etc. However, the knowledge in the knowledge bases indeed has annotations of types, categories, etc, and the above approaches just ignored them. Some work utilizes world knowledge as distant supervision [Mintz *et al.* (2009); Wang *et al.* (2014b); Xu *et al.* (2014)] for entity and relation extraction and embedding. This is a direct use of the facts in world knowledge bases, where the entities in the knowledge bases are matched in the context regardless of the ambiguity.

In 2015, [Wang *et al.* (2015, 2016a)] proposed a new learning framework with world knowledge as indirect supervision. This work specifies the world knowledge to domains by resolving the ambiguity of the entities and their types, then represents the data with world knowledge as a heterogeneous information network [Han *et al.* (2010)]. Based on this network, a new clustering algorithm is proposed, which significantly outperform the state-of-the-art clustering algorithms as well as clustering algorithms enhanced with world knowledge features. Compared with previous approaches, the main advantage of indirect supervision is that it can extend the knowledge about entities and relations to more generic text analytics problems.

Chapter 6

CONCLUSION

In this thesis, our goal is to apply ranking algorithms to multi-sourced networks. By generalizing regular HITS to multi-layered networks, we propose a new algorithm called multi-layered HITS to solve the problems of Cross-Layer Ranking and Cross-Layer Query with the leverage of both within-layer smoothness and cross-layer consistency. Compared to other network models, the biggest advantage of the multi-layered network model is that it allows more flexible layer-layer dependency structure across different layers. This essentially allows the multi-layered HITS algorithm to potentially mine more valuable information from the networks.

We give detailed theoretical analysis for the proposed algorithm regarding its effectiveness. Also, we try to integrate the knowledge graph into the multi-layered network to further improve the ranking and query performance, but unfortunately the experimental result does not reach our expectation. This will be one of our future tasks to study. How to include valuable knowledge without introducing too much noisy information is the key to solve the problem. Furthermore, We conduct comprehensive experiments to evaluate the proposed algorithm based on a real dataset (Amazon co-purchasing network) for both ranking and query tasks. Compared with the baseline method regular HITS algorithm, multi-layered HITS algorithm achieves relatively higher human evaluation scores regarding both relevance and explainability. This essentially proves that multi-layered HITS algorithm gives more accurate and reasonable ranking result by considering the cross-layer dependency structure.

REFERENCES

Akata, Z., C. Thurau and C. Bauckhage, "Non-negative matrix factorization in multimodality data for segmentation and label prediction", in "16th Computer vision winter workshop", (2011).

Amento, B., L. Terveen and W. Hill, "Does authority mean quality? predicting expert quality ratings of web documents", in "Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval", pp. 296–303 (ACM, 2000).

Auer, S., C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak and Z. Ives, "Dbpedia: A nucleus for a web of open data", The semantic web pp. 722–735 (2007).

Banko, M., M. J. Cafarella, S. Soderland, M. Broadhead and O. Etzioni, "Open information extraction from the web.", in "IJCAI", vol. 7, pp. 2670–2676 (2007).

Battiston, F., V. Nicosia and V. Latora, "Structural measures for multiplex networks", Physical Review E **89**, 3, 032804 (2014).

Berlingerio, M., M. Coscia, F. Giannotti, A. Monreale and D. Pedreschi, "Foundations of multidimensional network analysis", in "Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on", pp. 485–489 (IEEE, 2011).

Bharat, K. and M. R. Henzinger, "Improved algorithms for topic distillation in a hyperlinked environment", in "Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval", pp. 104–111 (ACM, 1998).

Bollacker, K., C. Evans, P. Paritosh, T. Sturge and J. Taylor, "Freebase: a collaboratively created graph database for structuring human knowledge", in "Proceedings of the 2008 ACM SIGMOD international conference on Management of data", pp. 1247–1250 (AcM, 2008).

Brin, S. and L. Page, "The anatomy of a large-scale hypertextual web search engine", Computer networks and ISDN systems **30**, 1-7, 107–117 (1998).

Buldyrev, S. V., R. Parshani, G. Paul, H. E. Stanley and S. Havlin, "Catastrophic cascade of failures in interdependent networks", arXiv preprint arXiv:0907.1182 (2009).

Cai, D., X. He, J.-R. Wen and W.-Y. Ma, "Block-level link analysis", in "Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval", pp. 440–447 (ACM, 2004).

Cai, Y. and S. Chakravarthy, "Hits vs. non-negative matrix factorization", (2014).

Chakrabarti, S., "Integrating the document object model with hyperlinks for enhanced topic distillation and information extraction", in "Proceedings of the 10th international conference on World Wide Web", pp. 211–220 (ACM, 2001).

Chakrabarti, S., B. Dom, P. Raghavan, S. Rajagopalan, D. Gibson and J. Kleinberg, "Automatic resource compilation by analyzing hyperlink structure and associated text", Computer networks and ISDN systems **30**, 1-7, 65–74 (1998).

Chakrabarti, S., M. Joshi and V. Tawde, "Enhanced topic distillation using text, markup tags, and hyperlinks", in "Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval", pp. 208–216 (ACM, 2001).

Chang, M.-W., L. Ratinov and D. Roth, "Structured learning with constrained conditional models", Machine learning **88**, 3, 399–431 (2012).

Chapelle, O., B. Scholkopf and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]", IEEE Transactions on Neural Networks **20**, 3, 542–542 (2009).

Chen, C., J. He, N. Bliss and H. Tong, "On the connectivity of multi-layered networks: Models, measures and optimal control", in "Data Mining (ICDM), 2015 IEEE International Conference on", pp. 715–720 (IEEE, 2015).

Chen, C., H. Tong, L. Xie, L. Ying and Q. He, "Fascinate: Fast cross-layer dependency inference on multi-layered networks.", in "KDD", pp. 765–774 (2016).

Chen, C., H. Tong, L. Xie, L. Ying and Q. He, "Cross-dependency inference in multi-layered networks: A collaborative filtering perspective", ACM Transactions on Knowledge Discovery from Data (TKDD) **11**, 4, 42 (2017).

Chen, Y. and R. Xu, "Context-sensitive network-based disease genetics prediction and its implications in drug discovery", Bioinformatics **33**, 7, 1031–1039 (2017).

Dechter, R. and R. Mateescu, "Mixtures of deterministic-probabilistic networks and their and/or search space", in "Proceedings of the 20th conference on Uncertainty in artificial intelligence", pp. 120–129 (AUAI Press, 2004).

Ding, C., T. Li, W. Peng and H. Park, "Orthogonal nonnegative matrix t-factorizations for clustering", in "Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining", pp. 126–135 (ACM, 2006).

Dong, X., E. Gabrilovich, G. Heitz, W. Horn, N. Lao, K. Murphy, T. Strohmann, S. Sun and W. Zhang, "Knowledge vault: A web-scale approach to probabilistic knowledge fusion", in "Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining", pp. 601–610 (ACM, 2014).

Egozi, O., S. Markovitch and E. Gabrilovich, "Concept-based information retrieval using explicit semantic analysis", ACM Transactions on Information Systems (TOIS) **29**, 2, 8 (2011).

Etzioni, O., M. Cafarella, D. Downey, S. Kok, A.-M. Popescu, T. Shaked, S. Soderland, D. S. Weld and A. Yates, "Web-scale information extraction in knowitall:(preliminary results)", in "Proceedings of the 13th international conference on World Wide Web", pp. 100–110 (ACM, 2004).

Fader, A., S. Soderland and O. Etzioni, "Identifying relations for open information extraction", in "Proceedings of the Conference on Empirical Methods in Natural Language Processing", pp. 1535–1545 (Association for Computational Linguistics, 2011).

Fern, X. Z. and C. E. Brodley, "Solving cluster ensemble problems by bipartite graph partitioning", in "Proceedings of the twenty-first international conference on Machine learning", p. 36 (ACM, 2004).

Fodeh, S., B. Punch and P.-N. Tan, "On ontology-driven document clustering using core semantic features", Knowledge and information systems **28**, 2, 395–421 (2011).

Gabrilovich, E. and S. Markovitch, "Feature generation for text categorization using world knowledge", in "IJCAI", vol. 5, pp. 1048–1053 (2005).

Gabrilovich, E. and S. Markovitch, "Overcoming the brittleness bottleneck using wikipedia: Enhancing text categorization with encyclopedic knowledge", in "AAAI", vol. 6, pp. 1301–1306 (2006).

Gabrilovich, E. and S. Markovitch, "Computing semantic relatedness using wikipedia-based explicit semantic analysis.", in "IJcAI", vol. 7, pp. 1606–1611 (2007).

Gabrilovich, E. and S. Markovitch, "Wikipedia-based semantic interpretation for natural language processing", Journal of Artificial Intelligence Research **34**, 443–498 (2009).

Ganchev, K., J. Gillenwater, B. Taskar *et al.*, "Posterior regularization for structured latent variable models", Journal of Machine Learning Research **11**, Jul, 2001–2049 (2010).

Gao, J., S. V. Buldyrev, S. Havlin and H. E. Stanley, "Robustness of a network of networks", Physical Review Letters **107**, 19, 195701 (2011).

Gao, J., S. V. Buldyrev, H. E. Stanley and S. Havlin, "Networks formed from interdependent networks", Nature physics **8**, 1, 40 (2012).

Greene, D. and P. Cunningham, "A matrix factorization approach for integrating multiple data views", Machine Learning and Knowledge Discovery in Databases pp. 423–438 (2009).

Hamosh, A., A. F. Scott, J. S. Amberger, C. A. Bocchini and V. A. McKusick, "Online mendelian inheritance in man (omim), a knowledgebase of human genes and genetic disorders", Nucleic acids research **33**, suppl_1, D514–D517 (2005).

Han, J., Y. Sun, X. Yan and P. S. Yu, “Mining knowledge from databases: an information network analysis approach”, in “Proceedings of the 2010 ACM SIGMOD International Conference on Management of data”, pp. 1251–1252 (ACM, 2010).

Haveliwala, T. H., “Topic-sensitive pagerank”, in “Proceedings of the 11th international conference on World Wide Web”, pp. 517–526 (ACM, 2002).

He, J., M. Li, H.-J. Zhang, H. Tong and C. Zhang, “Manifold-ranking based image retrieval”, in “Proceedings of the 12th annual ACM international conference on Multimedia”, pp. 9–16 (ACM, 2004).

Heath, L. S. and A. A. Sioson, “Multimodal networks: Structure and operations”, IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) **6**, 2, 321–332 (2009).

Hotho, A., S. Staab and G. Stumme, “Ontologies improve text document clustering”, in “Data Mining, 2003. ICDM 2003. Third IEEE International Conference on”, pp. 541–544 (IEEE, 2003).

Hu, J., L. Fang, Y. Cao, H.-J. Zeng, H. Li, Q. Yang and Z. Chen, “Enhancing text clustering by leveraging wikipedia semantics”, in “Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval”, pp. 179–186 (ACM, 2008).

Hu, X., N. Sun, C. Zhang and T.-S. Chua, “Exploiting internal and external semantics for the clustering of short texts using world knowledge”, in “Proceedings of the 18th ACM conference on Information and knowledge management”, pp. 919–928 (ACM, 2009a).

Hu, X., X. Zhang, C. Lu, E. K. Park and X. Zhou, “Exploiting wikipedia as external knowledge for document clustering”, in “Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining”, pp. 389–396 (ACM, 2009b).

Hua, W., Y. Song, H. Wang and X. Zhou, “Identifying users’ topical tasks in web search”, in “Proceedings of the sixth ACM international conference on Web search and data mining”, pp. 93–102 (ACM, 2013).

Ji, M., J. Han and M. Danilevsky, “Ranking-based classification of heterogeneous information networks”, in “Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining”, pp. 1298–1306 (ACM, 2011).

Kim, J., J.-g. Lee and S. Lim, “Differential flattening: A novel framework for community detection in multi-layer graphs”, ACM Transactions on Intelligent Systems and Technology (TIST) **8**, 2, 27 (2016).

Kivelä, M., A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno and M. A. Porter, “Multilayer networks”, Journal of complex networks **2**, 3, 203–271 (2014).

Kleinberg, J. M., “Authoritative sources in a hyperlinked environment”, Journal of the ACM (JACM) **46**, 5, 604–632 (1999).

Kumar, A. and H. Daumé, "A co-training approach for multi-view spectral clustering", in "Proceedings of the 28th International Conference on Machine Learning (ICML-11)", pp. 393–400 (2011).

Kumar, A., P. Rai and H. Daume, "Co-regularized multi-view spectral clustering", in "Advances in neural information processing systems", pp. 1413–1421 (2011).

Lage, K., N. T. Hansen, E. O. Karlberg, A. C. Eklund, F. S. Roque, P. K. Donahoe, Z. Szallasi, T. S. Jensen and S. Brunak, "A large-scale analysis of tissue-specific pathology and gene expression of human disease genes and complexes", Proceedings of the National Academy of Sciences **105**, 52, 20870–20875 (2008).

Lee, D. D. and H. S. Seung, "Algorithms for non-negative matrix factorization", in "Advances in neural information processing systems", pp. 556–562 (2001).

Lempel, R. and S. Moran, "The stochastic approach for link-structure analysis (salsa) and the tkc effect1", Computer Networks **33**, 1-6, 387–401 (2000).

Lenat, D. B. and R. V. Guha, *Building large knowledge-based systems; representation and inference in the Cyc project* (Addison-Wesley Longman Publishing Co., Inc., 1989).

Leskovec, J., L. A. Adamic and B. A. Huberman, "The dynamics of viral marketing", ACM Transactions on the Web (TWEB) **1**, 1, 5 (2007).

Li, Z., Z. Pan, Y. Zhang, G. Li and G. Hu, "Efficient community detection in heterogeneous social networks", Mathematical Problems in Engineering **2016** (2016).

Liu, J., C. Wang, J. Gao and J. Han, "Multi-view clustering via joint nonnegative matrix factorization", in "Proceedings of the 2013 SIAM International Conference on Data Mining", pp. 252–260 (SIAM, 2013).

Magger, O., Y. Y. Waldman, E. Ruppin and R. Sharan, "Enhancing the prioritization of disease-causing genes through tissue specific protein interaction networks", PLoS computational biology **8**, 9, e1002690 (2012).

Mintz, M., S. Bills, R. Snow and D. Jurafsky, "Distant supervision for relation extraction without labeled data", in "Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 2-Volume 2", pp. 1003–1011 (Association for Computational Linguistics, 2009).

Mitchell, T. M., W. W. Cohen, E. R. Hruschka Jr, P. P. Talukdar, J. Betteridge, A. Carlson, B. D. Mishra, M. Gardner, B. Kisiel, J. Krishnamurthy *et al.*, "Never ending learning.", in "AAAI", pp. 2302–2310 (2015).

Newman, M. E., "The mathematics of networks", The new palgrave encyclopedia of economics **2**, 2008, 1–12 (2008).

Ng, M. K.-P., X. Li and Y. Ye, “Multirank: co-ranking for objects and relations in multi-relational data”, in “Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining”, pp. 1217–1225 (ACM, 2011).

Ni, J., W. Cheng, W. Fan and X. Zhang, “Self-grouping multi-network clustering”, in “Data Mining (ICDM), 2016 IEEE 16th International Conference on”, pp. 1119–1124 (IEEE, 2016a).

Ni, J., H. Fei, W. Fan and X. Zhang, “Cross-network clustering and cluster ranking for medical diagnosis”, in “Data Engineering (ICDE), 2017 IEEE 33rd International Conference on”, pp. 163–166 (IEEE, 2017).

Ni, J., M. Koyuturk, H. Tong, J. Haines, R. Xu and X. Zhang, “Disease gene prioritization by integrating tissue-specific molecular networks using a robust multi-network model”, BMC bioinformatics **17**, 1, 453 (2016b).

Ni, J., H. Tong, W. Fan and X. Zhang, “Inside the atoms: ranking on a network of networks”, in “Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining”, pp. 1356–1365 (ACM, 2014).

Ni, J., H. Tong, W. Fan and X. Zhang, “Flexible and robust multi-network clustering”, in “Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining”, pp. 835–844 (ACM, 2015).

Pan, S. J. and Q. Yang, “A survey on transfer learning”, IEEE Transactions on knowledge and data engineering **22**, 10, 1345–1359 (2010).

Parshani, R., S. V. Buldyrev and S. Havlin, “Interdependent networks: reducing the coupling strength leads to a change from a first to second order percolation transition”, Physical review letters **105**, 4, 048701 (2010).

Ponzetto, S. P. and M. Strube, “Deriving a large scale taxonomy from wikipedia”, in “AAAI”, vol. 7, pp. 1440–1445 (2007).

Richardson, M. and P. Domingos, “Markov logic networks”, Machine learning **62**, 1, 107–136 (2006).

Samdani, R., M.-W. Chang and D. Roth, “Unified expectation maximization”, in “Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies”, pp. 688–698 (Association for Computational Linguistics, 2012).

Schmitz, M., R. Bart, S. Soderland, O. Etzioni *et al.*, “Open language learning for information extraction”, in “Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning”, pp. 523–534 (Association for Computational Linguistics, 2012).

Shao, J., S. V. Buldyrev, S. Havlin and H. E. Stanley, “Cascade of failures in coupled network systems with multiple support-dependence relations”, Physical Review E **83**, 3, 036116 (2011).

Sindhwani, V. and P. Melville, “Document-word co-regularization for semi-supervised sentiment analysis”, in “Data Mining, 2008. ICDM’08. Eighth IEEE International Conference on”, pp. 1025–1030 (IEEE, 2008).

Song, Y., H. Wang, W. Chen and S. Wang, “Transfer understanding from head queries to tail queries”, in “Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management”, pp. 1299–1308 (ACM, 2014).

Song, Y., H. Wang, Z. Wang, H. Li and W. Chen, “Short text conceptualization using a probabilistic knowledgebase”, in “Proceedings of the Twenty-Second international joint conference on Artificial Intelligence-Volume Volume Three”, pp. 2330–2336 (AAAI Press, 2011).

Song, Y., S. Wang and H. Wang, “Open domain short text conceptualization: A generative+ descriptive modeling approach.”, in “IJCAI”, pp. 3820–3826 (2015).

Strehl, A. and J. Ghosh, “Cluster ensembles-a knowledge reuse framework for combining partitionings”, in “Aaai/iaai”, pp. 93–99 (2002).

Suchanek, F. M., G. Kasneci and G. Weikum, “Yago: a core of semantic knowledge”, in “Proceedings of the 16th international conference on World Wide Web”, pp. 697–706 (ACM, 2007).

Sun, Y., J. Han, P. Zhao, Z. Yin, H. Cheng and T. Wu, “Rankclus: integrating clustering with ranking for heterogeneous information network analysis”, in “Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology”, pp. 565–576 (ACM, 2009).

Tang, J., J. Zhang, L. Yao, J. Li, L. Zhang and Z. Su, “Arnetminer: extraction and mining of academic social networks”, in “Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining”, pp. 990–998 (ACM, 2008).

Vespignani, A., “Complex networks: The fragility of interdependency”, Nature **464**, 7291, 984–985 (2010).

Wang, C., N. Duan, M. Zhou and M. Zhang, “Paraphrasing adaptation for web search ranking.”, in “ACL (2)”, pp. 41–46 (2013a).

Wang, C., Y. Song, A. El-Kishky, D. Roth, M. Zhang and J. Han, “Incorporating world knowledge to document clustering via heterogeneous information networks”, in “Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining”, pp. 1215–1224 (ACM, 2015).

Wang, C., Y. Song, D. Roth, M. Zhang and J. Han, “World knowledge as indirect supervision for document clustering”, ACM Transactions on Knowledge Discovery from Data (TKDD) **11**, 2, 13 (2016a).

Wang, C., Y. Sun, Y. Song, J. Han, Y. Song, L. Wang and M. Zhang, "Relsim: relation similarity search in schema-rich heterogeneous information networks", in "Proceedings of the 2016 SIAM International Conference on Data Mining", pp. 621–629 (SIAM, 2016b).

Wang, F., Z. Wang, Z. Li and J.-R. Wen, "Concept-based short text classification and ranking", in "Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management", pp. 1069–1078 (ACM, 2014a).

Wang, H., F. Nie and H. Huang, "Multi-view clustering and feature learning via structured sparsity", in "International Conference on Machine Learning", pp. 352–360 (2013b).

Wang, Z., J. Zhang, J. Feng and Z. Chen, "Knowledge graph and text jointly embedding.", in "EMNLP", vol. 14, pp. 1591–1601 (2014b).

Wu, W., H. Li, H. Wang and K. Q. Zhu, "Probase: A probabilistic taxonomy for text understanding", in "Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data", pp. 481–492 (ACM, 2012).

Xu, C., Y. Bai, J. Bian, B. Gao, G. Wang, X. Liu and T.-Y. Liu, "Rc-net: A general framework for incorporating knowledge into word representations", in "Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management", pp. 1219–1228 (ACM, 2014).

Yang, J. and J. Leskovec, "Defining and evaluating network communities based on ground-truth", Knowledge and Information Systems **42**, 1, 181–213 (2015).

Yang, Y., N. Chawla, Y. Sun and J. Hani, "Predicting links in multi-relational and heterogeneous networks", in "Data Mining (ICDM), 2012 IEEE 12th International Conference on", pp. 755–764 (IEEE, 2012).

Zhou, D., O. Bousquet, T. N. Lal, J. Weston and B. Schölkopf, "Learning with local and global consistency", in "Advances in neural information processing systems", pp. 321–328 (2004a).

Zhou, D. and C. J. Burges, "Spectral clustering and transductive learning with multiple views", in "Proceedings of the 24th international conference on Machine learning", pp. 1159–1166 (ACM, 2007).

Zhou, D., J. Huang and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding", in "Advances in neural information processing systems", pp. 1601–1608 (2007).

Zhou, D., J. Weston, A. Gretton, O. Bousquet and B. Schölkopf, "Ranking on data manifolds", in "Advances in neural information processing systems", pp. 169–176 (2004b).