Ranking on Network of Heterogeneous Information Networks

Zhe Xu*, Si Zhang*, Yinglong Xia†, Liang Xiong†, Hanghang Tong*
*University of Illinois at Urbana-Champaign, {zhexu3, sizhang2, htong}@illinois.edu
†Facebook, {yxia, lxiong}@fb.com

***Abstract*—Ranking on networks plays an important role in many high-impact applications, including recommender systems, social network analysis, bioinformatics and many more. In the age of big data, a recent trend is to address the *variety* aspect of network ranking. Among others, two representative lines of research include (1) heterogeneous information network with different types of nodes and edges, and (2) network of networks with edges at different resolutions. In this paper, we propose a new network model named Network of Heterogeneous Information Networks (NEOHIN for short) that is capable of simultaneously modeling both different types of nodes/edges, and different edge resolutions. We further propose two new ranking algorithms on NEOHIN based on the cross-domain consistency principle. Experiments on synthetic and real-world networks show that our proposed algorithms are (1) effective, which outperform other existing methods, and (2) efficient, without additional time cost per iteration to their counterparts.**

***Index Terms*—ranking; network of networks; graph mining**

I. INTRODUCTION

Ranking is a primary task in network analysis, which gives an order to nodes in the network w.r.t. the underlying network structure, preference, relevance, etc. Classic ranking algorithms, including PageRank [1], HITS [2] and their variants (e.g., personalized PageRank [3], random walk with restart [4], personalized propagation of neural predictions [5]), have shown a great success by exploiting the underlying network topology. They have been widely applied in numerous application areas, such as recommender systems [6], social networks [7], [8], network embedding [9], [10], bioinformatics [11], etc.

In the era of big data, networks arising from many important application domains are often accompanied with rich side information, beyond the underlying topology — a phenomenon that can be termed as the 'variety' aspect of the 4 Vs of big data. In response, more complex network models and sophisticated ranking algorithms have emerged. Among others, a remarkable line of research can be attributed to heterogeneous information network (HIN for short), which models nodes and edges of different types. Many ranking algorithms tailored for HIN have been proposed, including PathSim [12], Hetesim [13], SemRec [14], PReP [15], and so on. Most of the HIN-based ranking algorithms aim to utilize the semantic information among the interactions between multiple types of nodes and edges (e.g. mining meta path). Another line of research aims to accommodate edges at different resolutions, which models both the connections between different nodes and those between different networks. By co-learning the model with network-network interactions from the coarser resolution, noise and bias in each specific network can be alleviated or even eliminated. Examples include network of networks (NoN for short) [16], multi-layered networks [17], multimodal networks [18], multidimensional network [19], interdependent networks [20], and multiplex networks [21].

In the meanwhile, the multi-typed components and the multi-resolution characteristics of edges often co-exist with each other in many real-world networks. For example, heterogeneous scholarly networks often consist of different types of nodes (e.g., authors, papers, etc. in different research areas) and edges that can be viewed at multiple resolutions, such as edges modeling the relationships between multi-typed nodes and those between different research areas. Ranking problem with above multi-typed multi-resolution scenario has various real-world applications. For instance, for heterogeneous scholarly networks, interdisciplinary researchers often survey papers in a new domain (i.e. ranking target) while they are only expert in their established domain (i.e. preference input, represented as papers, keywords, and so on) and the ranking problem can be served as a cross-domain recommendation task. However, addressing the above task by modeling it into a single HIN can lead to the loss of domain-specific information (e.g. an interdisciplinary author may do great contribution in one domain but not be so famous in another one). Thus, it still remains a daunting task to advance network rankings by fully exploiting this rich information of networks and propose a well-tailored ranking model for it. Specifically, challenges can be characterized from two perspectives. First (network models), the existing network models can solely model either the multi-typed components at a single edge resolution (e.g., HIN) or multiple edge resolutions in homogeneous networks (e.g., NoN). How can we enjoy the best of both worlds? Second (ranking algorithms), given a network model that leverages both kinds of information, ranking algorithms can vary based on specific ranking tasks, such as ranking w.r.t. network structure or the relevance to certain query nodes.

In this paper, we aim to address these two challenges. We first propose a novel network model named Network of Heterogeneous Information Networks (NEOHIN for short). It contains a main network, whose nodes represent different domains and edges describe the interactions between different domains. Each node of the main network is further mapped to a HIN to model different types of relationships between

different types of nodes within a certain domain. In this way, the proposed NEOHIN is capable of simultaneously modeling both different types of nodes/edges, and different edge resolutions. Under the proposed network model, we formulate two ranking tasks based on the cross-domain consistency principle from the optimization perspectives and propose an efficient algorithm for each ranking task. The main contributions of this paper can be summarized as:

- **New network model.** In order to bridge the gap between real-world scenarios and existing methods. We propose a novel network model NEOHIN which can simultaneously model different types of nodes and edges in the network and view edges at different resolutions.
- **Algorithms.** We propose two ranking algorithms on the proposed model NEOHIN to address different ranking task scenarios named HITS-NEOHIN and PReP-NEOHIN. Our analyses show that both of our proposed algorithms provide theoretical guarantees on reaching local optimum without significant additional computational cost.
- **Experimental evaluations.** We conduct comprehensive evaluations on synthetic and real-world networks to demonstrate the proposed ranking algorithms (1) consistently outperform the existing algorithms on two ranking tasks, and (2) are efficient with a comparable computational time against their counterparts on existing network models.

II. PROBLEM DEFINITION

In this section, we introduce the formal definition of proposed NEOHIN ranking problem. Table I summarizes the main symbols and notations used throughout the paper. We use bold uppercase letters for matrices (e.g., $\mathbf{A}$), bold lowercase letters for vectors (e.g., $\mathbf{u}$) where $\mathbf{u}(x)$ is the x-th element of vector $\mathbf{u}$, and lowercase letters for scalars (e.g., c). We denote the transpose of a matrix/vector by the superscript $'$ (e.g., $\mathbf{A}'$ as the transpose of matrix $\mathbf{A}$).

Symbol	Definition
$\mathbf{G}$	adjacency matrix of main network
$\mathcal{V}$	set of nodes
$\mathcal{E}$	set of edges
g	number of domains in main network
i, j	indices of domains in main network
H_i	HIN of the i-th domain
$\mathbf{A}_i$	adjacency matrix of i-th domain
$\mathcal{I}_{ij}$	common nodes between H_i and H_j
$\mathbf{d}$	degree vector of main network
$\mathcal{S}$	set of node pairs
$\mathcal{T}$	set of candidate meta paths

TABLE I: Symbols and Notations

Our proposed model NEOHIN is built upon the heterogeneous information network, which is defined as follows [22].

Definition 1. ***Heterogeneous Information Network*** *is defined as a directed network $H = (\mathcal{V}, \mathcal{E})$. $\mathcal{V} = \cup_p \mathcal{V}_p$, $p \in \{1, 2, ..., P\}$ is the union of different types of nodes where $\mathcal{V}_p$ is the set of nodes of the p-th node type and P is the number of node types. Similarly, $\mathcal{E} = \cup_q \mathcal{E}_q$, $q \in \{1, 2, ..., Q\}$ denotes different types of edges where $\mathcal{E}_q$ is the set of edges of the q-th edge type and Q is the number of edge types.*

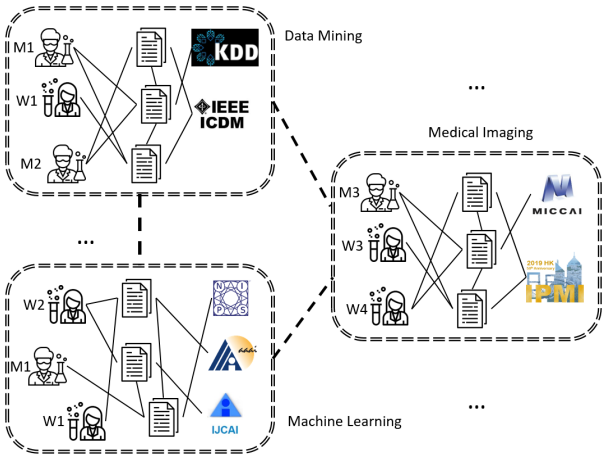


Fig. 1: An illustrative example of NEOHIN model. Each double dashed box represents a domain (i.e., a HIN) and nodes within each domain can represent authors, papers and conferences. Solid lines within each domain are the relationships between these nodes. Dashed lines between domains describe the domain-to-domain similarities which form the main network at a coarser resolution.

There are two key concepts in HIN, including the meta path [12] and the path count. A meta path is the concatenation of node types and edge types to encode a specific semantic meaning, such as [author] $\xrightarrow{\text{writes}}$ [paper] $\xrightarrow{\text{cited\_by}}$ [paper] $\xrightarrow{\text{written\_by}}$ [author], which describes the citation relationship between papers from two authors. Path count is the number of concrete path instances between a starting node and an ending node given a meta path. For our example with aforementioned meta path, the path count represents the number of citations from author A's papers to author B's papers.

HIN is powerful to describe different types of relationships between different types of nodes. Nonetheless, the representation power of the classic HIN is restricted to a single domain (e.g., papers, authors and conferences in the data mining domain). In some applications, we are often faced with multiple inter-correlated domains. For example, when mining on the scholar data (e.g., DBLP), there exist multiple research domains (see Figure 1), each of which is represented as a HIN to describe the relationships between authors, papers and conferences in that domain. At a different edge resolution, research domains can interact with others if they are closely related, e.g., the data mining domain with a strong connection to machine learning domain. In order to model such a collection of inter-correlated HINs, we generalize an existing network of networks model [16] to a network of HINs as follows.

Definition 2. ***Network of HINs (NEOHIN)*** *is composed of*

a $g \times g$ main network $\mathbf{G}$, a set of domain-specific HINs $\mathcal{H} = \{H_1, ..., H_g\}$ and an one-to-one mapping function ψ which maps each node in the main network $\mathbf{G}$ to a domain specific HIN. In general, a NEOHIN *is defined as $W =< \mathbf{G}, \mathcal{H}, \psi >$.*

Generally speaking, the ranking problem on the NEOHIN can be formally defined as follows.

Problem 1. NEOHIN *Ranking*

Given: *(1) a* NEOHIN *$W =< \mathbf{G}, \mathcal{H}, \psi >$, (2) target nodes for ranking, (3) query node(s) of interests (optional), and (4) meta path(s) of interests (optional).*

Find: *the rankings among target nodes w.r.t. query node(s).*

Depending on (1) the specific way to leverage the input NEOHIN, and (2) the specific query node(s) as well as target nodes, Problem 1 embraces several ranking scenarios. In the next two sections, we will present two new ranking algorithms on NEOHIN in these two ranking scenarios. The key idea of our proposed algorithms is the cross-domain consistency principle that the influence (e.g. ranking score) of a common element (e.g. node/edge/meta path) in i-th domain is similar to that in j-th domain if these two domains themselves are similar to each other (i.e. with a large $\mathbf{G}(i,j)$). For instance, in Figure 1, if a researcher appears in two similar domains (e.g., data mining and machine learning), the ranking scores of this researcher in these two domains should be similar with each other.

III. PROPOSED ALGORITHM #1: HITS-NEOHIN

In this ranking scenario, we are interested in ranking all the nodes in terms of their hub and authority scores by exploiting the connectivity structure of the input NEOHIN (referred to as Scenario #1). Here, the query node(s) of interests could be absent if we want to calculate the global hub/authority scores which are independent on any specific query node(s).

A. Preliminaries: HITS for a Single HIN

HITS is a classic network ranking algorithm that computes a hub score and an authority score for each node. From the optimization perspective [23], HITS can be viewed as a rank-1 non-negative factorization of the input adjacency matrix by minimizing the following cost function.

$$\begin{aligned} J_i(\mathbf{u}_i, \mathbf{v}_i) = & \frac{c}{2}||\mathbf{A}_i - \mathbf{u}_i\mathbf{v}_i'||_F^2 \\ & + (1-c)(||\mathbf{u}_i - \mathbf{e}_{ui}||_2^2 + ||\mathbf{v}_i - \mathbf{e}_{vi}||_2^2) \\ & s.t.\ \forall x,\ \mathbf{u}_i(x) \geq 0,\ \mathbf{v}_i(x) \geq 0 \end{aligned} \tag{1}$$

where i is the index of a domain, $\mathbf{A}_i$ is the adjacency matrix of the given i-th HIN with $P_i \times P_i$ blocks, P_i is the number of node types, $0 < c < 1$ is a regularization parameter, $\mathbf{u}_i$ and $\mathbf{v}_i$ are the hub vector and the authority vector respectively. $\mathbf{e}_{ui}$ and $\mathbf{e}_{vi}$ are the preference vectors given a priori. For example, if users are interested with the k-th node, then the k-th element of $\mathbf{e}_{ui}$ and $\mathbf{e}_{vi}$ are set to 1 and other elements are set to be 0. When the query node is absent, $\mathbf{e}_{ui}$ and $\mathbf{e}_{vi}$ could be set as the uniform vectors.

B. Optimization Formulation

In order to generalize the classic HITS algorithm to the NEOHIN model, we propose to obtain the hub vectors $\mathbf{u}_i$ and the authority vectors $\mathbf{v}_i$ $(i = 1, ..., g)$ that minimize the following cost function:

$$\begin{aligned} J(\mathbf{u}, \mathbf{v}) = & \sum_{i=1}^{g} J_i(\mathbf{u}_i, \mathbf{v}_i) \\ & + a\sum_{i=1}^{g}\sum_{j=1}^{g} ||\frac{\mathbf{u}_i(\mathcal{I}_{ij})}{\sqrt{\mathbf{d}(i)}} - \frac{\mathbf{u}_j(\mathcal{I}_{ij})}{\sqrt{\mathbf{d}(j)}}||_2^2 \mathbf{G}(i,j) \\ & + a\sum_{i=1}^{g}\sum_{j=1}^{g} ||\frac{\mathbf{v}_i(\mathcal{I}_{ij})}{\sqrt{\mathbf{d}(i)}} - \frac{\mathbf{v}_j(\mathcal{I}_{ij})}{\sqrt{\mathbf{d}(j)}}||_2^2 \mathbf{G}(i,j) \\ & s.t.\ \forall x,\ \mathbf{u}_i(x) \geq 0,\ \mathbf{v}_i(x) \geq 0. \end{aligned} \tag{2}$$

where the first term $J_i(\mathbf{u}_i, \mathbf{v}_i)$ comes from Eq. (1), $\mathbf{d}$ is the degree vector of the main network, $\mathcal{I}_{ij}$ is the set of common nodes between the i-th domain and j-th domain, and $\mathbf{G}$ is the adjacency matrix of the main network. The key idea of our formulation lies in the second and the third terms, which instantiate the cross-domain consistency principle outlined in Section 2. That is, the hub/authority scores of a common node x ($x \in \mathcal{I}_{ij}$) across two domains should be close to each other (i.e., small $(\frac{\mathbf{u}_i(x)}{\sqrt{\mathbf{d}(i)}} - \frac{\mathbf{u}_j(x)}{\sqrt{\mathbf{d}(j)}})^2$ and small $(\frac{\mathbf{v}_i(x)}{\sqrt{\mathbf{d}(i)}} - \frac{\mathbf{v}_j(x)}{\sqrt{\mathbf{d}(j)}})^2$) if these two domains $(i,\ j)$ are similar with each other (i.e., large $\mathbf{G}(i,j)$).

C. Optimization Solution and Analysis

We propose an iterative algorithm to find $\mathbf{u}_i$ and $\mathbf{v}_i$ $(i = 1, ..., g)$ by minimizing Eq. (2). Since these two groups of variables ($\mathbf{u}_i$ and $\mathbf{v}_i$) are symmetric, we only present the details about the update of $\mathbf{u}_i$ due to the space limitation and directly present the update of $\mathbf{v}_i$. To simplify the description, we introduce the following auxiliary notations.[1]

Define matrix $\mathbf{C}$ as a $g \times g$ block matrix where the (i,j)-th block is computed by $\mathbf{G}(i,j)\mathbf{C}_{ij}$. $\mathbf{C}_{ij}$ is an $n_i \times n_j$ matrix where n_i, n_j are the number of nodes in i-th domain and j-th domain respectively. If the x-th node in $\mathbf{A}_i$ and the y-th node in $\mathbf{A}_j$ represent the same node (i.e., a common node), $\mathbf{C}_{ij}(x,y) = 1$. Since $\mathbf{C}$ might be singular, we further define $\mathbf{L} = \mathbf{C} + \mathbf{D}_P$, where $\mathbf{D}_P = \text{diag}(\mathbf{d}(1)\mathbf{I}_{n_1}, ..., \mathbf{d}(g)\mathbf{I}_{n_g}) - \mathbf{D}_C$. $\mathbf{D}_C$ and $\mathbf{D}_L$ are the degree matrices of $\mathbf{C}$ and $\mathbf{L}$, respectively. By defining $\tilde{\mathbf{L}} = \mathbf{D}_L^{-\frac{1}{2}}\mathbf{L}\mathbf{D}_L^{-\frac{1}{2}}$, $\mathbf{u} = (\mathbf{u}_1', ..., \mathbf{u}_g')'$, $\mathbf{v} = (\mathbf{v}_1', ..., \mathbf{v}_g')'$, $\mathbf{e} = (\mathbf{e}_{u1}', ..., \mathbf{e}_{ug}')'$, and aggregated diagonal block matrix $\mathbf{A} = \text{diag}(\mathbf{A}_1, ..., \mathbf{A}_g)$, we can simplify Eq. (2) w.r.t. the hub vector $\mathbf{u}$ as below,

$$\begin{aligned} J(\mathbf{u}) = & \frac{c}{2}(\mathbf{v}'\mathbf{v}\mathbf{u}'\mathbf{u} - 2\mathbf{u}'\mathbf{A}\mathbf{v}) \\ & + (1-c)||\mathbf{u} - \mathbf{e}||_2^2 + 2a\mathbf{u}'(\mathbf{I} - \tilde{\mathbf{L}})\mathbf{u}. \end{aligned}$$

[1]They were first introduced in [16] to generalize PageRank to network of networks.

Algorithm 1 Optimization algorithm of HITS-NEOHIN

Input: (1) the adjacency matrix of main network $\mathbf{G}$; (2) a set of adjacency matrices of domain networks $\{\mathbf{A}_i\}$; (3) a set of common node indicating matrix $\{\mathbf{C}_{ij}\}$.

Output: (1) The hub and authority ranking vectors of each domain $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$.

1: Pre-processing: generate the aggregated diagonal block matrix $\mathbf{A}$ and auxiliary matrix $\tilde{\mathbf{L}}$ as introduced in Sec. III-C.
2: Initialization: (1) set $\mathbf{u}_i(x) = 1$ and $\mathbf{v}_i(x) = 1$ if the x-th node in the i-th domain is of interest. Otherwise, set them as uniform vectors; (2) concatenate $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ as $\mathbf{u}$ and $\mathbf{v}$, respectively.
3: **while** not converged **do**
4: Update $\mathbf{u}$ based on Eq. (3)
5: Update $\mathbf{v}$ based on Eq. (4)
6: Normalize vectors $\mathbf{u}$ and $\mathbf{v}$ s.t. $\sum_{x=1}^{n_i} \mathbf{u}_i(x) = 1$ *and* $\sum_{x=1}^{n_i} \mathbf{v}_i(x) = 1, \forall i = 1, ..., g$
7: **end while**
8: **return** two groups of ranking vectors $\{\mathbf{u}_i\}$ and $\{\mathbf{v}_i\}$ by deconcatenation of vectors $\mathbf{u}$ and $\mathbf{v}$.

With the non-negativity constraint on $\mathbf{u}$, we propose the following multiplicative update rule to solve for $\mathbf{u}$:

$$\mathbf{u}(x) \leftarrow \mathbf{u}(x)\sqrt{\frac{[c\mathbf{A}\mathbf{v} + 4a\tilde{\mathbf{L}}\mathbf{u} + 2(1-c)\mathbf{e}](x)}{[c\mathbf{u}\mathbf{v}'\mathbf{v} + 4a\mathbf{u} + 2(1-c)\mathbf{u}](x)}} \tag{3}$$

After each iteration, we normalize vector $\mathbf{u}$ according to node types (i.e. $\sum_{x=1}^{n_i} \mathbf{u}_i(x) = 1, i = 1, ..., g$). For group of variables $\{\mathbf{v}_i\}$, it has symmetric updating method as variables $\{\mathbf{u}_i\}$ as follows,

$$\mathbf{v}(x) \leftarrow \mathbf{v}(x)\sqrt{\frac{[c\mathbf{A}\mathbf{u} + 4a\tilde{\mathbf{L}}\mathbf{v} + 2(1-c)\mathbf{e}](x)}{[c\mathbf{v}\mathbf{u}'\mathbf{u} + 4a\mathbf{v} + 2(1-c)\mathbf{v}](x)}} \tag{4}$$

The whole algorithm is provided in Alg. 1. Next, we provide the theoretical analysis of the updating rule of $\mathbf{u}$. We first analyze the convergence by proving its monotonicity in Lemma 1. Then we prove in Lemma 2 that the fixed-point solution of Eq. (3) satisfies the KKT conditions. Finally in Lemma 3 we show that our algorithm is efficient.

Lemma 1. *Under Eq.* (3)*, the objective function in Eq.* (2) *is monotonically non-increasing.*

Proof. Now we have formulas:

$$J(\mathbf{u}) = (\frac{c}{2}\mathbf{v}'\mathbf{v} + 1 - c + 2a)\underbrace{\mathbf{u}'\mathbf{u}}_{Q_1} - c\underbrace{\mathbf{u}'\mathbf{A}\mathbf{v}}_{Q_2} - 2(1-c)(\underbrace{\mathbf{u}'\mathbf{e}}_{Q_3}) - 2a(\underbrace{\mathbf{u}'\tilde{\mathbf{L}}\mathbf{u}}_{Q_4})$$

Based on the auxiliary function method [24], [25], using the following inequality:

$$z \geq 1 + \log z$$

, we get the auxiliary function:

$$H(\mathbf{u}, \tilde{\mathbf{u}}) = (\frac{c}{2}\mathbf{v}'\mathbf{v} + 1 - c + 2a)Q_1' - cQ_2' - 2(1-c)Q_3' - 2aQ_4',$$

$$where \begin{cases} Q_1' &= Q_1, \\ Q_2' &= \sum_x [\mathbf{A}\mathbf{v}](x)\tilde{\mathbf{u}}(x)(1 + \log\frac{\mathbf{u}(x)}{\tilde{\mathbf{u}}(x)}), \\ Q_3' &= \sum_x \mathbf{e}(x)\tilde{\mathbf{u}}(x)(1 + \log\frac{\mathbf{u}(x)}{\tilde{\mathbf{u}}(x)}), \\ Q_4' &= \sum_{x,y} \tilde{\mathbf{L}}(x,y)\tilde{\mathbf{u}}(x)\tilde{\mathbf{u}}(y)(1 + \log\frac{\mathbf{u}(x)\mathbf{u}(y)}{\tilde{\mathbf{u}}(x)\tilde{\mathbf{u}}(y)}). \end{cases}$$

Thus $H(\mathbf{u}, \tilde{\mathbf{u}}) \geq J(\mathbf{u})$ and $H(\mathbf{u}, \mathbf{u}) = J(\mathbf{u})$. By setting the gradient of $H(\mathbf{u}, \tilde{\mathbf{u}})$ to zero, we get the updating function as follows:

$$\mathbf{u}^2(x) = \tilde{\mathbf{u}}^2(x)\frac{[c\mathbf{A}\mathbf{v} + 4a\tilde{\mathbf{L}}\mathbf{u} + 2(1-c)\mathbf{e}](x)}{[c\mathbf{u}\mathbf{v}'\mathbf{v} + 4a\mathbf{u} + 2(1-c)\mathbf{u}](x)},$$

which is same as Eq. (3).

□

Lemma 2. *At convergence, the fixed-point solution of Eq.* (3) *satisfies the KKT conditions.*

Proof. The Lagrangian function of Eq. (2) is:

$$L(\mathbf{u}) = \frac{c}{2}\mathbf{v}'\mathbf{v}\mathbf{u}'\mathbf{u} - c\mathbf{u}'\mathbf{A}\mathbf{v} + (1-c)||\mathbf{u} - \mathbf{e}||_2^2 + 2a\mathbf{u}'(\mathbf{I} - \tilde{\mathbf{L}})\mathbf{u} - \boldsymbol{\alpha}'\mathbf{u} \tag{5}$$

with the $\boldsymbol{\alpha}$ as Lagrange multiplier and setting the derivative of $L(\mathbf{u})$ w.r.t. $\mathbf{u}$ to 0, we get:

$$c\mathbf{u}\mathbf{v}'\mathbf{v} - c\mathbf{A}\mathbf{v} + 2(1-c)(\mathbf{u} - \mathbf{e}) + 4a(\mathbf{I} - \tilde{\mathbf{L}})\mathbf{u} = \boldsymbol{\alpha} \tag{6}$$

by KKT complementary slackness condition, we have:

$$[c\mathbf{u}\mathbf{v}'\mathbf{v} + 2(1-c)\mathbf{u} + 4a\mathbf{u} - (4a\tilde{\mathbf{L}}\mathbf{u} + c\mathbf{A}\mathbf{v} + 2(1-c)\mathbf{e})](x)\mathbf{u}(x) = 0 \tag{7}$$

Therefore Eq. (3) satisfies KKT conditions. □

Lemma 3. *The gap of the time complexity between the fixed-point solution of HITS-*NEOHIN *(Eq.* (3)*) and HITS is close.*

Proof. The difference of the objective functions between our HITS-NEOHIN algorithm and regular HITS algorithm lies at the second and third terms in Eq.(2), and they lead to the $4a\tilde{\mathbf{L}}\mathbf{u}$ and $4a\mathbf{u}$ terms in Eq.(3). For the update solution of regular HITS, its time complexity is dominant by term $c\mathbf{A}\mathbf{v}$, which is $O(\sum_i |\mathcal{E}_i|)$ due to the sparsity of $\mathbf{A}$. For the time complexity of $4a\tilde{\mathbf{L}}\mathbf{u}$, it is $O(\sum_{i,j} |\mathcal{I}_{ij}| + \sum_i |\mathcal{V}_i|)$, due to the sparsity of $\tilde{\mathbf{L}}$. In the real-world setting the number of nodes is much less than the number of edges, so put them together the time complexity increases minorly. □

IV. Proposed Algorithm #2: PReP-NeoHIN

In this scenario, we focus on ranking all the nodes of the same type w.r.t. a query node, by exploring certain type(s) of meta path of the input NeoHIN (referred to as Scenario #2) from a given meta path candidate pool. For instance in Figure 1, we might be interested in ranking all the author nodes w.r.t. a given researcher by exploring the following meta path t: [author] $\xrightarrow{\text{writes}}$ [paper] $\xrightarrow{\text{cited\_by}}$ [paper] $\xrightarrow{\text{written\_by}}$ [author].

A. Preliminaries: PReP for a Single HIN

PReP is one of the most recent ranking algorithm in a single HIN [15]. It regards the number of meta path t between a node pair (y, z) in the HIN as being generated from an exponential distribution: $pc_{(y,z),t} \sim \text{Exp}(\lambda)$. By designing the parameter λ and imposing prior distributions on parameters, it minimizes the negative log likelihood of $pc_{i,(y,z),t}$ between every pair of nodes (y, z) as follows.

$$\begin{aligned} L_{O,i}(\eta_i, \rho_i, \phi_i, \theta_i) &= -\log(p(pc_i, \eta_i, \rho_i, \phi_i, \theta_i|\alpha_i, \beta_i)) \\ &= -\Big\{ \sum_{y \in V_i} \log(\Gamma(\rho_{iy}; (\alpha_{iy}, 1))) \\ &+ \sum_{(y,z) \in S_i} \log(\text{Dir}_{|\mathcal{M}|}(\phi_{i,(y,z),m}; \beta_i)) \\ &+ \sum_{(y,z) \in S_i} \sum_{t=1}^{|\mathcal{T}|} \log(\text{Exp}(pc_{i,(y,z),t}; \lambda_i)) \Big\} \end{aligned} \tag{8}$$

where i is the index of a domain, $\lambda_i = \frac{\eta_{it}}{\rho_{iy}\rho_{iz}\sum_{m=1}^{|\mathcal{M}|}\phi_{i,(y,z),m}\theta_{imt}}$ is the designed parameter for the i-th domain, t is the index of a meta path, $\mathcal{T}$ is the set of candidate meta paths, $\mathcal{M}$ is the set of generating patterns. The model has the following parameters: $\rho_i = \{\rho_{iy}\}$ where ρ_{iy} describes the visibility of node y (i.e., number of paths connecting to node y); $\eta_i = \{\eta_{it}\}$ where η_{it} describes the selectivity of meta path t (i.e., significance of a path t for calculating relevance); $\phi_i = \{\phi_{i,(y,z),m}\}$ where $\phi_{i,(y,z),m}$ is the projection probability from a pair of nodes (y, z) to a generating pattern m such that $\sum_m \phi_{i,(y,z),m} = 1$; $\theta_i = \{\theta_{imt}\}$ where θ_{imt} is the projection probability from a generating pattern m to a meta path t s.t. $\sum_t \theta_{imt} = 1$. The Gamma prior distribution $\Gamma(\cdot)$ on parameter ρ_{iy} is to prevent the trivial re-scaling of parameters ρ_{iy}, ρ_{iz} and η_{it}. The Dirichlet prior $\text{Dir}(\cdot)$ is designed for sparse distribution over $|\mathcal{M}|$ generating patterns. α_{iy} and β_i are parameters for these two distributions. Once the parameters η_{it}, ρ_{iy}, $\phi_{i,(y,z),m}$ and θ_{imt} are learnt by minimizing $L_{O,i}$, given query node z, PReP [15] ranks node y of the same type based on the relevance measure as follows,

$$\begin{aligned} \mathbf{R}_i(y, z) = &\sum_{t=1}^{|\mathcal{T}|} \frac{\eta_{it} pc_{i,(y,z),t}}{\rho_{iy}\rho_{iz}\sum_{m=1}^{|\mathcal{M}|}\phi_{i,(y,z),m}\theta_{imt}} \\ &+ (1 - \beta_i) \sum_{m=1}^{|\mathcal{M}|} \log \phi_{i,(y,z),m} \end{aligned} \tag{9}$$

B. Optimization Formulation

In order to generalize the PReP algorithm into the NeoHIN model, we propose to minimize the following cost function:

$$L = \sum_{i=1}^{g} L_{O,i}(\eta_i, \rho_i, \phi_i, \theta_i) + \gamma \sum_{i=1}^{g} L_{C,i}(\eta_i)$$

where γ is the regularization parameter. The first term $L_{O,i}$ is same as we defined in Eq. (8) representing the negative log likelihood in each specific domain. The second term $L_{C,i}$ is the cross-domain objective function which aims to encode the cross-domain consistency principle and can be written as follows.

$$L_{C,i}(\eta_i) = \sum_{j=1}^{g} \sum_{t=1}^{|\mathcal{T}|} \left(\frac{\eta_{it}}{\sqrt{\mathbf{d}(i)}} - \frac{\eta_{jt}}{\sqrt{\mathbf{d}(j)}}\right)^2 \mathbf{G}(i, j) \tag{10}$$

where $\mathbf{d}(i), \mathbf{d}(j)$ and $\mathbf{G}(i, j)$ have the same meaning as in Eq. (2). The intuition of $L_{C,i}$ is that if a given meta path t appears in both the i-th domain and the j-th domain, its contributions η_{it} and η_{jt} (i.e., its significance for calculating relevance) in these two domains should be similar (i.e., small $(\frac{\eta_{it}}{\sqrt{\mathbf{d}(i)}} - \frac{\eta_{jt}}{\sqrt{\mathbf{d}(j)}})^2$) if i-th domain and j-th domain are similar with each other (i.e., with a large $\mathbf{G}(i, j)$). Thus, if there exists noise in the data from a specific domain, which misleads the learning of $\{\eta_{it}\}$, the constraints from the similar domain via $\{\eta_{jt}\}$ could alleviate or even eliminate the influence of noise data. On the other hand, compared with methods mixing all the domains together and viewing it as a single HIN, our method can keep the domain specific characteristics. For example, if we aim to evaluate the relevance between two authors, for authors in relatively specific domains like medical imaging, attending the same conference (i.e. meta path [author] $\xrightarrow{\text{writes}}$ [paper] $\xrightarrow{\text{accepted\_by}}$ [conference] $\xrightarrow{\text{accepts}}$ [paper] $\xrightarrow{\text{written\_by}}$ [author]) can describe great similarity between authors compared with general domains like machine learning.

C. Optimization Solution and Analysis

We propose an alternating optimization strategy to optimize different groups of variables η_{it}, ρ_{iy}, $\phi_{i,(y,z),m}$ and θ_{imt}. Since the variables ρ_{iy}, $\phi_{i,(y,z),m}$ and θ_{imt} only appear in the cost function $L_{O,i}$, we can apply the same algorithm to infer them as PReP [15] and we omit the details due to the space limitations. To optimize η_{it}, we first compute the gradient as follows,

$$\begin{aligned} \frac{\partial L}{\partial \eta_{it}} = &-\frac{|\mathcal{S}|}{\eta_{it}} + \sum_j 2\gamma \left(\frac{\eta_{it}}{\sqrt{\mathbf{d}(i)}} - \frac{\eta_{jt}}{\sqrt{\mathbf{d}(j)}}\right) \frac{\mathbf{G}(i, j)}{\sqrt{\mathbf{d}(i)}} \\ &+ \sum_{(y,z) \in S_i} \frac{pc_{i,(y,z),t}}{\rho_{iy}\rho_{iz}\sum_m \phi_{i,(y,z),m}\theta_{imt}}. \end{aligned} \tag{11}$$

Algorithm 2 Optimization algorithm of PReP-NEOHIN

Input: (1) a set of adjacency matrices between all types of nodes in each domain $\{\mathbf{A}_i^{type1,type2}\}$; (2) candidate meta paths $\mathcal{T}$; (3) number of generating patterns $|\mathcal{M}|$; (4) a set of target node pairs $\{y, z\}$ to measure their relevance.

Output: (1) the relevance score $\mathbf{R}_i(y, z)$ between target node pairs (y, z) in the domain i.

Pre-processing: generate the path count $pc_{i,(y,z),t}$ between every pair of target node pairs (y, z) in domain i following meta path t by the multiplication of adjacency matrices between different types of nodes.
while not converged **do**
  for the i-th domain , $1 \leq i \leq g$ **do**
    for meta path t, $1 \leq t \leq |\mathcal{T}|$ **do**
      Update η_{it} based on Eq. (12).
    end for
    Update ρ_i, ϕ_i, θ_i followed by method in [15] with $pc_{i,(y,z),t}$ and $|\mathcal{M}|$.
  end for
end while
return relevance score $\mathbf{R}_i(y, z)$ between target node pairs (y, z) in the domain i by Eq. (9).

where $\mathcal{S}$ is the set of node pairs. By setting Eq. (11) to be 0, we adopt a following solution to update the η_{it}.

$$\eta_{it} = \frac{-B\mathbf{d}(i) + \sqrt{B^2\mathbf{d}(i)^2 + 8\gamma\sum_j \mathbf{G}(i,j)|\mathcal{S}|\mathbf{d}(i)}}{4\gamma\sum_j \mathbf{G}(i,j)} \quad (12)$$

where

$$B = \sum_{(y,z)\in S_i} \frac{pc_{i,(y,z),t}}{\rho_{iy}\rho_{iz}\sum_m \phi_{i,(y,z),m}\theta_{imt}} - \sum_j \frac{2\gamma\mathbf{G}(i,j)\eta_{jt}}{\sqrt{\mathbf{d}(j)\mathbf{d}(i)}}$$

We have the following lemma to support that Eq. (12) is a local optimal solution for updating the parameter η_{it}.

Lemma 4. *Eq.* (12) *is a local optimal solution for updating the parameter* η_{it}.

Proof. By setting Eq. (11) to be 0, we get a quadratic equation w.r.t. variable η_{it} as follows,

$$-|\mathcal{S}| + \frac{2\gamma\eta_{it}^2}{\mathbf{d}(i)}\sum_j \mathbf{G}(i,j) + B\eta_{it} = 0, \quad (13)$$

where B has same definition as in Eq.(12). Apparently, Eq. (13) has a root on the positive half axis where the derivative of Eq. (13) is positive around that root. Thus, $L(\cdot)$ is convex around the positive root, which gives the minimum of $L(\cdot)$ (i.e. Eq.(12)) as a local optimal solution for the parameter η_{it}. □

After finish inferring all the parameters, PReP-NEOHIN ranks the same-type nodes by computing the relevance scores (i.e., Eq. (9)) with the learnt parameters. We present the whole algorithm in Alg. 2

Remarks. From the generative model perspective, we have the lemma as follows.

Lemma 5. *The effect of* $L_{C,i}$ *in Eq.* (10) *is equivalent to imposing a prior distribution* $f(\eta_{it})$ *on* η_{it}. *Specifically,* $f(\eta_{it}) \propto \prod_{j=1}^{g}\prod_{t=1}^{|\mathcal{T}|}\sqrt{\frac{\mathbf{G}(i,j)}{\pi\mathbf{d}(i)}}\exp\left(-\frac{(\eta_{it}-\sqrt{\frac{\mathbf{d}_i}{\mathbf{d}_j}}\eta_{jt})^2}{\frac{\mathbf{d}_i}{\mathbf{G}(i,j)}}\right)$.

Proof. We have

$$f(\eta_{it}) = \kappa\prod_{j=1}^{g}\prod_{t=1}^{|\mathcal{T}|}\sqrt{\frac{\mathbf{G}(i,j)}{\pi\mathbf{d}(i)}}\exp\left(-\frac{(\eta_{it}-\sqrt{\frac{\mathbf{d}_i}{\mathbf{d}_j}}\eta_{jt})^2}{\frac{\mathbf{d}_i}{\mathbf{G}(i,j)}}\right)$$

where κ is a normalization term to ensure the integral of $f(\eta_{it})$ to be 1. By setting the negative log likelihood of $f(\eta_{it})$ as loss function, we have the following equation.

$$-\log f(\eta_{it}) = \sum_{j=1}^{g}\sum_{t=1}^{|\mathcal{T}|}(\frac{\eta_{it}}{\sqrt{\mathbf{d}(i)}} - \frac{\eta_{jt}}{\sqrt{\mathbf{d}(j)}})^2\mathbf{G}(i,j) + const$$

By ignoring the constant term in the loss function, the above equation is equivalent to Eq. (10). □

In addition, in Lemma 6 we show that our algorithm is efficient as well.

Lemma 6. *The gap of the time complexity between the inference algorithms of PReP-*NEOHIN *and PReP is close.*

Proof. The difference between inference algorithms of PReP-NEOHIN and PReP lies at the update of η_{it}. They both need to compute the first term of B in Eq. (11) whose time complexity is $O(|\mathcal{S}||\mathcal{M}|)$. The difference lies at the computing of $\frac{2\gamma\eta_{it}^2}{\mathbf{d}(i)}\sum_j \mathbf{G}(i,j)$ and $\sum_j \frac{2\gamma\mathbf{G}(i,j)\eta_{jt}}{\sqrt{\mathbf{d}(j)\mathbf{d}(i)}}$ whose time complexity are $O(g)$, which is minor compared with $O(|\mathcal{S}||\mathcal{M}|)$. □

V. EXPERIMENT

In this section, we conduct various experiments aiming to answer the following questions:

- If the proposed model and algorithms are effective for various scenarios?
- If the proposed algorithms converge stably and quickly?
- If the proposed algorithms are as efficient as their counterparts?

A. Effectiveness on synthetic dataset

We design a comparative experiment to illustrate the effectiveness of our proposed NEOHIN model. We first construct two synthetic networks with $2,000$ nodes following the same power-law distribution using the famous Barabási-Albert method [26]. Each of them would be described as a domain network. The domain-domain similarity is set as 1 since they follow the almost same distribution. We assign node type to each node based on its degree. Specifically, we assign nodes with degree larger than $\frac{2*\max degree}{3}$ as Type I, assign nodes with degree lower than $\frac{\max degree}{3}$ as Type III, and assign the others as Type II. In total, there are $3,587$ Type I nodes, 343 Type II nodes, and 70 Type III nodes.

Algorithm	Accuracy					
	K=5	K=10	K=15	K=20	K=25	K=30
CrossRank	**0.045**	**0.090**	0.135	0.180	0.202	0.225
HITS-NoN	0.034	0.056	0.112	0.124	0.135	0.180
HITS-NEOHIN	**0.045**	**0.090**	**0.146**	**0.191**	**0.213**	**0.236**

TABLE II: Results of cross-domain link prediction on synthetic dataset.

Algorithm	ROC-AUC	AUPRC
PReP	0.553	0.307
PReP-NEOHIN	**0.566**	**0.404**

TABLE III: Results of meta path-based link prediction on synthetic dataset.

In the following parts, we would present two comparative experiments for the proposed HITS-NEOHIN, PReP-NEOHIN and their counterparts.

1) Task 1: Cross-domain Link Prediction: We assume that there exists cross-domain links between nodes from different domain-specific networks with the same degree. We connect node pairs with same degree greedily and name those node pairs as target node pairs. We set half of the target node pairs as common nodes between domains and the other half of them as the ground truths to be predicted. To predict the aforementioned target cross-domain links, we set the query vector using the adjacent nodes of the source node from the 1-st domain, and rank all the nodes in the 2-nd domain. If the ground truth (i.e. target node of the cross-domain link in the 2-nd domain) is in the top-*K* ranking results, we view the prediction is accurate (i.e., a 'hit'). In this scenario we focus on comparing our HITS-NEOHIN with following two network of networks-based algorithms:

- CrossRank [16] which formulates the ranking problem on NoN based on the cross-domain consistency principle,
- HITS-NoN which is a natural extension of the HITS algorithm [23] to the NoN setting by encoding cross-domain consistence.

From the results presented in Tab. II we observe that our proposed HITS-NEOHIN consistently outperforms other two comparison methods with the increment of K which is because (1) HITS-NEOHIN makes better use of the node type information, and (2) HITS-NEOHIN evaluates the importance of nodes from hub and authority perspectives simultaneously.

2) Task 2: Meta Path-based Link Prediction: We follow the same synthetic dataset setting as we illustrated in Sec. V-A and have three node types. Here we aim to predict if two Type II nodes are connected or not based on the following meta paths:

- [Type II] $\rightarrow$ [Type I] $\rightarrow$ [Type II],
- [Type II] $\rightarrow$ [Type III] $\rightarrow$ [Type II].

We compare our PReP-NEOHIN with its counterpart PReP [15] and adopt the receiver operating characteristic curve (ROC-AUC) and area under precision-recall curve (AUPRC) as evaluation metrics.

From Tab. III we can observe that under the NEOHIN framework, our PReP-NEOHIN algorithm outperforms its counterpart PReP. Although these two synthetic domain-specific networks are generated based on the same settings and method, it is inevitable to import noise when predicting the linking status between same-type nodes (i.e. Type II in our experiment) based on given meta paths. Co-learning on multiple domain-specific networks are effective to alleviate the bias in any of them.

B. Effectiveness on Aminer Dataset

In this section we evaluate our algorithms on the scholar dataset `Aminer` [27]. We extract a subset of data that consists of five research domains, including data mining, machine learning, database, information retrieval and bioinformatics. The domain-domain similarity is based on the citation proportion between two domains. For example, among all references of data mining papers, 50% of them are data base papers and for references of data base papers, 40% of them are data mining papers. Then, the similarity between these two domains are $0.4 \times 0.5 = 0.2$. We consider four node types in the network, including author, paper, keyword and conference. In total, there are 27,665 authors, 19,206 papers, 12,478 keywords and 29 conferences. We design two different tasks to show the effectiveness of our proposed algorithms HITS-NEOHIN and PReP-NEOHIN, respectively.

1) Task 1: Cross-Domain Citation Prediction: For the task of cross-domain citation prediction, we partition the dataset into two periods based on the publication years, including 2005-2010 (P1) and 2010-2015 (P2). We focus on the data mining papers which were published in P2 and cited databases paper(s) published in P1. We set prior vectors based on the keywords extracted from abstracts. If the ground truth is in the top-*K* ranking results, we view the prediction is accurate. We compare the proposed HITS-NEOHIN with the following five baseline methods, including

- PageRank [3] which measures the importance of nodes w.r.t. the network topology,
- HITS [2] which measures the importance of nodes in the network from two perspectives: authority and hub,
- CrossRank [16] as introduced in Sec. V-A,
- HITS-NoN as introduced in Sec. V-A,
- HITS-HIN which is a natural extension of HITS to HIN by normalizing the ranking vectors $\mathbf{u}$ and $\mathbf{v}$ based on node types in each iteration.

We summarize the experiment results of our proposed HITS-NEOHIN compared with all the baselines in Table IV. We have the following observations. First, the proposed HITS-NEOHIN achieves an up to 10% prediction improvement compared to the classic ranking algorithms PageRank and HITS. In the meanwhile, our method outperforms the state-of-the-art ranking algorithm CrossRank by a large amount. Second, we can observe that our method also outperforms the baseline methods HITS-NoN and HITS-HIN, which further demonstrate simultaneously incorporating the various rich side information (i.e., heterogeneity and different edge resolutions) leads to more accurate rankings. In addition, it is unsurprising that the ranking accuracy increases with the increment of K.

Algorithm	Accuracy				
	K=100	K=200	K=300	K=400	K=500
PageRank	0.016	0.092	0.131	0.150	0.198
CrossRank	0.063	0.120	0.162	0.223	0.258
HITS	0.042	0.087	0.128	0.154	0.172
HITS-NoN	0.064	0.130	0.172	0.233	0.273
HITS-HIN	0.016	0.082	0.126	0.143	0.167
HITS-NEOHIN	**0.109**	**0.160**	**0.203**	**0.246**	**0.291**

TABLE IV: Results of cross-domain citation prediction.

Algorithm	ROC-AUC	AUPRC
PathCount	0.414	0.464
PathSim	0.491	0.513
JoinSim	0.574	0.579
PReP	0.542	0.524
PReP-NEOHIN	**0.584**	**0.607**

TABLE V: Results of similar author identification.

Yet, our proposed algorithm can still consistently outperform other baseline methods.

2) Task 2: Similar Author Identification: For the task of identifying similar authors, we select the top 20 researchers whose research interests lie in both database and data mining based on the citations of their publications. The ground truth provides the information whether each two of these distinguished researchers are similar or not, based on the their publications. We evaluate the performance of the algorithms based on the ground truth under the relevance measure computed by Eq. (9). We compare PReP-NEOHIN with four meta path-based baselines, including

- PathCount [12] which simply use path counts as the relevance measure,
- PathSim [12] which makes a further step on PathCount by normalizing the relevance with the sum of importance of the node pair,
- JoinSim [28] which normalizes the relevance by the square root of the product of importance of the node pair,
- PReP [15] which is introduced in Sec. IV-A.

In all these methods, we use the following meta paths:

- [author] $\xrightarrow{\text{writes}}$ [paper] $\xrightarrow{\text{written\_by}}$ [author]
- [author] $\xrightarrow{\text{writes}}$ [paper] $\xrightarrow{\text{accepted\_by}}$ [conference] $\xrightarrow{\text{accepts}}$ [paper] $\xrightarrow{\text{written\_by}}$ [author]
- [author] $\xrightarrow{\text{writes}}$ [paper] $\xrightarrow{\text{contains}}$ [keyword] $\xrightarrow{\text{is\_in}}$ [paper] $\xrightarrow{\text{written\_by}}$ [author]

We use ROC-AUC and AUPRC as evaluation metrics which are same as we illustrated in Sec. V-A2. The results are summarized in Table V. We can observe that PReP-NEOHIN consistently outperforms all the baseline methods, which further indicates that different edge resolutions (i.e., the network of networks structure) can boost the learning of model in every domain-specific network, alleviate the impact of noise effectively, and improve the ranking performance.

C. Effectivenss on DisGeNET Dataset — a Case Study

In this section, we conduct a case study on the DisGeNET dataset, a widely-used real-world dataset, to show that our algorithm can provide meaningful ranking results. DisGeNet is composed by human disease-gene associated network [29] and gene-gene interaction network [30]. Since gene-gene connections can often vary in different human tissues and a pair of gene-disease can be closely related in some human tissues but not all, it is of a great importance to consider such tissue-specific connection patterns in the ranking problem. Correspondingly, we construct our proposed NEOHIN model as follows. First, we extract the human tissues (i.e., domains) from the cardiovascular system, nervous system, and musculoskeletal system. Then, for each tissue, we construct the domain-specific heterogeneous network with the nodes as genes and disease, as well as the edges among nodes. At a coarse resolution, the edges between domains represent the interactions between different tissues. In this NEOHIN model, there are 6,880 diseases and 1,348 genes in total.

Based on this constructed network, we aim to discover the common yet critical diseases in certain human systems (e.g., skeletal muscle). We present the top-10 critical diseases in each domain in Table VI. Due to the settings of our dataset, there exists some overlaps between the concepts of diseases, but surprisingly we find that those diseases with general definitions are ranked high in our results such as 'heart disease', 'peripheral nervous system disease', and 'brain disease', which verifies the correctness of our algorithm. After removing those general diseases, the ranking of specific diseases could provide additional information about the importance and commonality of the diseases by leveraging the structures of tissue-specific heterogeneous networks as well as the information of tissues.

D. Convergence Results

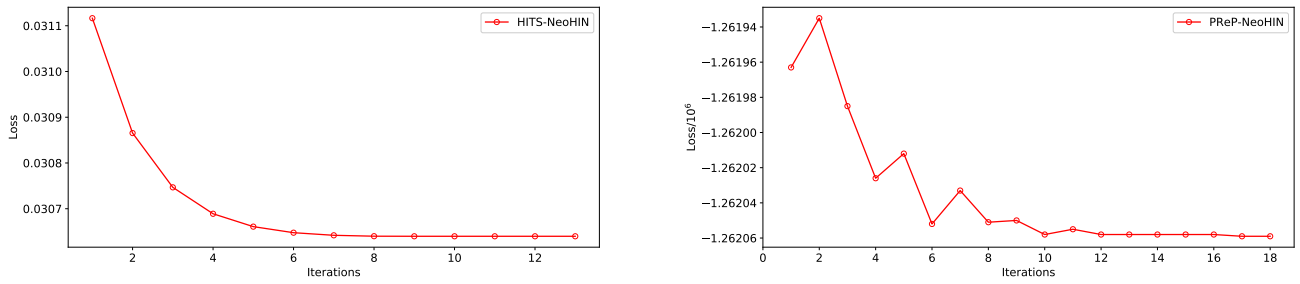
For two of our proposed algorithms, despite the local optimum guarantees provided in Lemma 1, 2, and 4, the updating methods of them are composed by alternative updating of multiple parameters. Aim to answer that whether the algorithms converge quickly and stably, we empirically study the convergence of them on the `AMiner` dataset [27] in two tasks illustrated in Sec.V-B, respectively. We show the convergence results in Figure 2 where we observe that both of our proposed algorithms converge very quickly (within 20 iterations), and stably (with few fluctuation).

E. Efficiency Results

For the optimization of our proposed algorithms, Sec.V-D shows that proposed algorithms can converge within few iterations. In this section, we further evaluate the efficiency of our proposed algorithms HITS-NEOHIN in Sec. V-B1 and PReP-NEOHIN in Sec. V-B2 in terms of the running time per iteration. We summarize the efficiency results in Table VII. As we can see, in both tasks, our proposed algorithms consistently have a comparable computation time per iteration against their counterparts without NEOHIN framework, which verifies the proved Lemma 3 and 6 that our proposed NEOHIN only adds no additional time cost .

Rank	Cardiovascular system	Nervous system	Musculoskeletal system
1	heart disease	peripheral nervous system disease	mitochondrial encephalomyopathy
2	heart valve disease	amyotrophic lateral sclerosis	MELAS syndrome
3	cardiovascular system disease	hepatic encephalopathy	myopathy
4	pulmonary hypertension	brain disease	rheumatoid
5	hypertension	epilepsy syndrome	mitochondrial myopathy
6	congenital heart disease	nervous system disease	arthritis
7	coronary artery disease	Parkinson's disease	muscular atrophy
8	congestive heart failure	Alzheimer's disease	osteoporosis
9	myocardial infarction	movement disease	gout
10	vascular disease	migraine	congenital diaphragmatic hernia

TABLE VI: Ranking results of diseases in cardiovascular system, nervous system, and musculoskeletal system domains.



(a) HITS-NEOHIN (b) PReP-NEOHIN

Fig. 2: Convergence of the proposed algorithms.

Scenarios	Algorithm	Seconds/Iteration
Task 1	HITS	0.026
	HITS-NoN	0.027
	HITS-HIN	0.028
	HITS-NEOHIN	0.027
Task 2	PReP	5.586
	PReP-NEOHIN	5.602

TABLE VII: Computational time of different algorithms.

VI. Related Work

We briefly review the related works on network ranking and network of networks.

A. Ranking

Ranking is a classic topic of network analysis. PageRank [1] and HITS [2] lay a solid foundation of random walk-based algorithms, followed by works like personalized PageRank [3], SimRank [31], fast random walk with restart [4] and so on. After that, with the development of heterogeneous information networks, more algorithms tailored for multi-type nodes and edges are proposed. PathSim [12] introduces several important ideas and concepts like meta path, net schema, and path count to measure the similarity between same-type nodes. Hetesim [13] models the similarity between nodes of different types based on relevance path with a similar definition as meta path. PReP [15] offers a new generative model prospective for the meta path-based similarity measure. Besides, real-world application based on HIN ranking including recommender system [32], drug discovery [33], and many more.

B. Network of Networks

Network of networks (NoN) model [16] is another line of research to mine the compatible and complementary information within data. It is introduced to analyze network at a finer granularity with global view across different domains. [34] addresses the clustering problem for networks collected from multiple domains by a two-step framework. MuLaN [35] [17] organizes the data into a new model named multi-layered networks, which consider cross domain node-node dependency with a cross-layer consistency constraint. Additionally, excluding aforementioned multi-domain networks, [36] offers a comprehensive survey about multi-layered network (also referred as network of networks), including multimodal networks [18], and multiplex networks [21], etc.

VII. Conclusions

As the cornerstone of network analysis, ranking plays an important role in a variety of applications. In this paper, we study the ranking problem in a collection of inter-connected heterogeneous information networks. In particular, we propose a new network model named NEOHIN that can simultaneously capture the network heterogeneity and different edge resolutions. We propose two new ranking algorithms in different ranking scenarios based on the cross-domain consistency principle, followed by some theoretical analyses. We conduct extensive experiments on synthetic and real-world data to demonstrate the efficacy of the proposed algorithms. Future works include generalizing the current model and algorithms to dynamic networks.

VIII. Acknowledgment

This work is supported by National Science Foundation under grant No. 1947135, and 2003924 by the NSF Program on Fairness in AI in collaboration with Amazon under award No. 1939725, and Department of Homeland Security under Grant Award Number 17STQAC00001-03-03. The content of the information in this document does not necessarily reflect the position or the policy of the Government or Amazon, and no official endorsement should be inferred. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation here on.

References

[1] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[2] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[3] G. Jeh and J. Widom, "Scaling personalized web search," in *Proceedings of the 12th international conference on World Wide Web*. Acm, 2003, pp. 271–279.

[4] H. Tong, C. Faloutsos, and J.-Y. Pan, "Fast random walk with restart and its applications," in *Sixth International Conference on Data Mining (ICDM'06)*. IEEE, 2006, pp. 613–622.

[5] J. Klicpera, A. Bojchevski, and S. Günnemann, "Predict then propagate: Graph neural networks meet personalized pagerank," in *International Conference on Learning Representations*, 2018.

[6] M. Gori, A. Pucci, V. Roma, and I. Siena, "Itemrank: A random-walk based scoring algorithm for recommender engines." in *IJCAI*, vol. 7, 2007, pp. 2766–2771.

[7] J. Weng, E.-P. Lim, J. Jiang, and Q. He, "Twitterrank: finding topic-sensitive influential twitterers," in *Proceedings of the third ACM international conference on Web search and data mining*. ACM, 2010, pp. 261–270.

[8] Z. Gyongyi, H. Garcia-Molina, and J. Pedersen, "Combating web spam with trustrank," in *Proceedings of the 30th international conference on very large data bases (VLDB)*, 2004.

[9] B. Perozzi, R. Al-Rfou, and S. Skiena, "Deepwalk: Online learning of social representations," in *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014, pp. 701–710.

[10] A. Grover and J. Leskovec, "node2vec: Scalable feature learning for networks," in *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 2016, pp. 855–864.

[11] Y. Xu, X. Zhou, and W. Zhang, "Microrna prediction with a novel ranking algorithm based on random walks," *Bioinformatics*, vol. 24, no. 13, pp. i50–i58, 2008.

[12] Y. Sun, J. Han, X. Yan, P. S. Yu, and T. Wu, "Pathsim: Meta path-based top-k similarity search in heterogeneous information networks," *Proceedings of the VLDB Endowment*, vol. 4, no. 11, pp. 992–1003, 2011.

[13] C. Shi, X. Kong, Y. Huang, S. Y. Philip, and B. Wu, "Hetesim: A general framework for relevance measure in heterogeneous networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 10, pp. 2479–2492, 2014.

[14] 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 *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 453–462.

[15] Y. Shi, P.-W. Chan, H. Zhuang, H. Gui, and J. Han, "Prep: Path-based relevance from a probabilistic perspective in heterogeneous information networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017, pp. 425–434.

[16] J. Ni, 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*. ACM, 2014, pp. 1356–1365.

[17] C. Chen, H. Tong, L. Xie, L. Ying, and Q. He, "Fascinate: Fast cross-layer dependency inference on multi-layered networks," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016, pp. 765–774.

[18] L. S. Heath and A. A. Sioson, "Multimodal networks: Structure and operations," *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, vol. 6, no. 2, pp. 321–332, 2009.

[19] M. Berlingerio, M. Coscia, F. Giannotti, A. Monreale, and D. Pedreschi, "Foundations of multidimensional network analysis," in *2011 international conference on advances in social networks analysis and mining*. IEEE, 2011, pp. 485–489.

[20] S. V. Buldyrev, R. Parshani, G. Paul, H. E. Stanley, and S. Havlin, "Catastrophic cascade of failures in interdependent networks," *Nature*, vol. 464, no. 7291, pp. 1025–1028, 2010.

[21] F. Battiston, V. Nicosia, and V. Latora, "Structural measures for multiplex networks," *Physical Review E*, vol. 89, no. 3, p. 032804, 2014.

[22] Y. Sun and J. Han, "Mining heterogeneous information networks: a structural analysis approach," *Acm Sigkdd Explorations Newsletter*, vol. 14, no. 2, pp. 20–28, 2013.

[23] Y. Cai and S. Chakravarthy, "Hits vs. non-negative matrix factorization," 2014.

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

[25] C. H. Ding, T. Li, and M. I. Jordan, "Convex and semi-nonnegative matrix factorizations," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 1, pp. 45–55, 2010.

[26] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *science*, vol. 286, no. 5439, pp. 509–512, 1999.

[27] J. Tang, 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*. ACM, 2008, pp. 990–998.

[28] Y. Xiong, Y. Zhu, and S. Y. Philip, "Top-k similarity join in heterogeneous information networks," *IEEE Transactions on Knowledge and Data Engineering*, vol. 27, no. 6, pp. 1710–1723, 2015.

[29] J. Piñero, À. Bravo, N. Queralt-Rosinach, A. Gutiérrez-Sacristán, J. Deu-Pons, E. Centeno, J. García-García, F. Sanz, and L. I. Furlong, "Disgenet: a comprehensive platform integrating information on human disease-associated genes and variants," *Nucleic acids research*, p. gkw943, 2016.

[30] B. Wang, A. Pourshafeie, M. Zitnik, J. Zhu, C. D. Bustamante, S. Batzoglou, and J. Leskovec, "Network enhancement as a general method to denoise weighted biological networks," *Nature communications*, vol. 9, no. 1, p. 3108, 2018.

[31] G. Jeh and J. Widom, "Simrank: a measure of structural-context similarity," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2002, pp. 538–543.

[32] X. Yu, X. Ren, Y. Sun, Q. Gu, B. Sturt, U. Khandelwal, B. Norick, and J. Han, "Personalized entity recommendation: A heterogeneous information network approach," in *Proceedings of the 7th ACM international conference on Web search and data mining*, 2014, pp. 283–292.

[33] B. Chen, Y. Ding, and D. J. Wild, "Assessing drug target association using semantic linked data," *PLoS Comput Biol*, vol. 8, no. 7, p. e1002574, 2012.

[34] J. Ni, 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*. ACM, 2015, pp. 835–844.

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

[36] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, "Multilayer networks," *Journal of complex networks*, vol. 2, no. 3, pp. 203–271, 2014.