7-2020

Semi-supervised co-clustering on attributed heterogeneous information networks

Yugang JI

Chuan SHI

Yuan FANG
Singapore Management University, yfang@smu.edu.sg

Xiangnan KONG

Mingyang YIN

Semi-supervised Co-Clustering on Attributed Heterogeneous Information Networks

Yugang Ji[a], Chuan Shi[a,*], Yuan Fang[b], Xiangnan Kong[c], Mingyang Yin[d]

[a] *Beijing University of Posts and Telecommunications, Beijing, China*
[b] *Singapore Management University, Singapore*
[c] *Worcester Polytechnic Institute, USA*
[d] *Alibaba Group, Hangzhou, China*

Abstract

Node clustering on heterogeneous information networks (HINs) plays an important role in many real-world applications. While previous research mainly clusters same-type nodes independently via exploiting structural similarity search, they ignore the correlations of different-type nodes. In this paper, we focus on the problem of *co-clustering* heterogeneous nodes where the goal is to mine the latent relevance of heterogeneous nodes and simultaneously partition them into the corresponding type-aware clusters. This problem is challenging in two aspects. First, the similarity or relevance of nodes is not only associated with multiple meta-path-based structures but also related to numerical and categorical attributes. Second, clusters and similarity/relevance searches usually promote each other.

To address this problem, we first design a learnable overall relevance measure that integrates the structural and attributed relevance by employing meta-paths and attribute projection. We then propose a novel approach, called SCCAIN, to co-cluster heterogeneous nodes based on constrained orthogonal non-negative matrix tri-factorization. Furthermore, an end-to-end framework is developed to jointly optimize the relevance measures and co-clustering. Extensive experi-

*This is to indicate the corresponding author.
Email addresses: jiyugang@bupt.edu.cn (Yugang Ji), shichuan@bupt.edu.cn (Chuan Shi), yfang@smu.edu.sg (Yuan Fang), xkong@wpi.edu (Xiangnan Kong), hengyang.ymy@alibaba-inc.com (Mingyang Yin)

ments on real-world datasets not only demonstrate that SCCAIN consistently outperforms state-of-the-art methods but also validate the effectiveness of integrating attributed and structural information for co-clustering.

Keywords: co-clustering, heterogeneous information network, meta-paths, matrix tri-factorization, semi-supervised learning

1. Introduction

In recent years, heterogeneous information networks (HINs), consisting of various nodes and multiple relations among these nodes, have been proposed to model the complex real-world data [1, 2]. Figure 1 illustrates a toy HIN with different types of nodes such as authors, conferences and papers. Compared with traditional homogeneous networks where both nodes and edges belong to a single type, HINs are able to effectively fuse more structural information and carry richer semantics. Given the advantage of HINs in modeling real-world data, many innovative data mining tasks have been performed on HINs.

For clustering on HINs, it is vital to do similarity search among nodes connected by various paths. As defined in [3] to describe the order of types within paths, meta-paths have been widely adopted to extract structural semantics of heterogeneous connections between nodes on HINs. Taking Figure 1 as an example, given the authors (A), papers (P) and conferences (C), we can utilize the meta-path A-P-A to capture cooperation of authors (e.g., A_1 and A_2), while adopting the meta-path A-P-C to describe the relation of submission (e.g., A_1 and C_1). The earlier works [3, 4, 5] are to evaluate the similarity/relevance of nodes connected by a single meta-path. Recently, considering there are multiple meta-paths between nodes (e.g., authors can be connected by meta-paths A-P-A and A-P-C-P-A), it becomes popular to compose multiple meta-paths together and automatically learn the importance of these meta-paths via semi-supervised manners [6, 7, 8]. These methods mainly take advantage of the structural information on HINs, while ignoring attributes of nodes which can contribute significantly to the relevance or similarity between nodes.

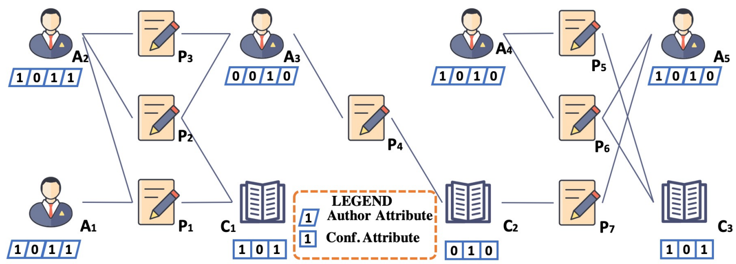


Figure 1: A toy example of attributed heterogeneous information network. A, P and C respectively denote authors, papers and conferences. The attributes of authors are their interest in research areas including network embedding, anomaly detection, NMF, and co-clustering, while the attributes of conferences are topics such as clustering, topic modeling, and recommender systems.

Additional attributed information has been proved beneficial for dealing with many data mining tasks. For example, in Figure 1, authors form two clusters $\{A_1, A_2, A_3\}$ and $\{A_4, A_5\}$ based on the structural information alone. With additional attributes on authors, more precise, fine-grained clustering is possible, such as separating A_3 from $\{A_1, A_2\}$. Although some studies have explored attributes on homogeneous networks [9, 10, 11], earlier clustering methods on HINs [6] fail to leverage node attributes.

To integrate attributes into HINs, a naïve idea is to represent each attribute as nodes of a new type [12]. However, only categorical attributes, such as cities and keywords can be integrated in this way, leaving out ordinal attributes, such as age and number of co-authors. To describe both ordinal and categorical attributes in HINs, a better way is to consider attributes of a node as a vector, where each dimension denotes one attribute [13]. While recent works focus on similarity search on same-type nodes, they fail to jointly analyze the clusters of different-type nodes in HINs. Actually, the different-type clusters are usually associated with each other because of the latent relevance/similarity between different-type nodes. For instance, the clusters of authors help to guide conference clustering and vice versa.

In order to make full use of the relevance for clustering different-type nodes and mining latent correlation of clusters at the same time, co-clustering has been a good choice [14, 15, 16, 17]. Different from traditional clustering methods, co-clustering leverages the duality between features and samples to achieve the simultaneous clustering of features and samples. Moreover, co-clustering methods have the ability to derive latent correspondence between clusters of different node types, making the resulting clusters more interpretable. Furthermore, must-link or cannot-link pair-wise constraints [18, 15, 19], which are to limit whether nodes should be assigned to the same clusters, have been utilized to guide co-clustering. Among these methods, non-negative matrix factorization (NMF) [20, 14, 15, 21] has been commonly adopted. Given a similarity matrix of different-type nodes, NMF would like to factorize the matrix into several latent non-negative factors and generate both clusters of heterogeneous nodes by considering the row and column factors as the distribution of clusters. However, when dealing with attributed HINs, since there are multiple relevance measures based on structures and attributes, these previous models are unable to perform co-clustering on HINs.

Challenges and Insights. In view of the shortcomings in existing method, we focus on the problem of co-clustering nodes of different types on attributed HINs. For instance, as shown in Figure 1, given the links among authors, papers and conferences as well as several node attributes, our goal is to co-cluster authors and conferences at the same time by analyzing both structural and attributed relevance between authors and conferences. We identify two major challenges here, and highlight the corresponding insights.

First, it is difficult to make full use of both attributed and structural information for relevance search on heterogeneous nodes. On the one hand, structural relevance alone can be captured by multiple meta-paths while previous co-clustering methods [15, 16] just focus on the connection of nodes. On the other hand, the attributed relevance of different-type nodes cannot be calculated directly because attributes of heterogeneous nodes contain quite different meanings while current integration [13] just deals with same-typed nodes. Notice that both structures and attributes contribute to the overall relevance.

Second, it is desirable to jointly optimize both co-clustering and relevance measure with pair-wise constraints. On HINs, there could be must-link and cannot-link pairwise constraints of nodes to help co-clustering and relevance measure. For example, on bibliographic graphs, we have constraints on author-conference, author-author, and conference-conference pairs. Moreover, the results of co-clustering promote relevance measure, and vice versa. How to design the unified framework to make full use of abundant semantics and constraints? It needs to be delicately designed.

Inspired by the above, our goal is to obtain more accurate co-clusters by integrating both relevance measure and constrained non-negative matrix factorization. In this paper, we propose a novel **S**emi-supervised **C**o-**C**lustering framework on **A**ttributed heterogeneous **I**nformation **N**etwork (SCCAIN). In this framework, we introduce a new relevance measure which takes into account both structural and attributed relevance of different-type nodes by utilizing multiple meta-paths and projection matrices. Moreover, we present the constrained negative matrix tri-factorization (ONMTF) to co-cluster nodes with constraints. To integrate both relevance measure and co-clustering, we set the results of co-clustering as the sharing factors and propose to design a unified semi-supervised learning framework to jointly optimize the co-clustering and relevance measure using the given constraints.

Contributions. In summary, the contributions of this paper are as follows.

- To the best of our knowledge, this is the first to study the problem of co-clustering nodes on attributed HINs. By designing the unified framework SCCAIN, we can effectively co-cluster nodes of different types at the same time.
- Our work overcomes the mentioned challenges including the integration of structural and attributed information and the suitable sharing factors of co-clustering and relevance measure. We propose a novel relevance measure that utilizes multiple meta-paths with learnable weights for the structural

relevance and a parameterized attributed relevance measure for attributes in different spaces. Furthermore, we alternately optimize overall relevance measure and co-clustering by setting the relevance matrix as sharing factors.

- We perform extensive experiments on three real-world datasets, including Aminer, DBLP and a subset of Alibaba user activity dataset. We compare SCCAIN against the various state of the arts and the experimental results discriminate that our model outperforms the baselines.

2. Related work

Since our work is to address the problem of co-clustering in attribute HINs, here we briefly introduce the most related work about similarity measure, clustering on HINs and co-clustering.

As a fundamental task in data mining, clustering is to simultaneously group similar nodes and separate dissimilar nodes[12, 22, 23]. Traditional clustering methods, such as K-means, adopt cosine function or Euclidean distance to evaluate the feature-based similarity of nodes. Recently, with HINs being more and more popular to model more complex data, similarity measure on HINs attaches much attention of researchers. Sun et al. [3] firstly propose the meta-path to capture semantics of different-typed nodes and put forward the PathSim to evaluate the similarity of same-typed nodes in HINs. To measure the relevance of different-typed nodes, Ni et al. [4] and Chuan et al. [5] respectively propose the asymmetric PCRW and symmetric HeteSim based on random walk. Taking the influence of different meta-paths into consideration, Luo et al. [7] propose to combining similarity measures of multiple meta-paths. Wang et al. [8] design the weighted PathSim upon meta structure in a semi-supervised manner. However, these methods just pay attention to link-based similarity but ignore the attributes of nodes.

For clustering on HINs, previous works mainly make use of the heterogeneous relations. Sun et al. [24, 25] utilize the ranking information of nodes of other types to cluster the nodes of the pointed type. Deng et al. [26, 27]

propagate topic distribution among different-type nodes (i.e., papers, authors and conferences) to cluster papers. However, these methods fail to make full use of structural semantics in HINs because of ignoring the composited relations (i.e., meta-paths) existing in such networks. Since different links between nodes represent different semantic, Sun et al. [6] integrate meta-path selection for clustering with user guide information. Li et al. [28] design a unsupervised non-negative matrix factorization to cluster nodes by integrating node similarity of different meta-paths with weights. Since nodes in HINs contain their own attributes to describe themselves, more and more research attempts to do clustering in attributed networks. Sun et al. [29] put forward a probabilistic model by utilizing both different-typed links and attributes of the pointed nodes. In [12], the authors reconstruct attributed networks where not only the entities but also the attributes are formed as nodes. Both Hsu et al. [9] and Perozzi et al. [30] assume attributes of nodes as vectors, and weight the links with the attribute-based similarity between nodes. Recently, Li et al. [13] propose a semi-supervised clustering method in attributed networks. Zhao et al. [31] integrate structural and attributed information by respectively constructing the graphs of links and attributes. Although take multiple nodes into consideration, these methods can only cluster nodes of the same type.

Taking the heterogeneity of nodes into consideration, there have been some researches on simultaneously generating clusters for different-typed nodes, namely, co-clustering [19, 32]. Dhillon et al. [33] consider documents as a bipartite spectral graph, and then co-cluster documents and words in terms of finding minimum cut vertex partitions in such a graph. Since matrix factorization performs well in detecting latent factors of different-type nodes (e.g., users and items) [34], some researches attempt to integrate non-negative matrix factorization for co-clustering [35, 20, 14, 36]. Nie et al. [16] attempt to learn a bipartite graph with exactly k connected components with some constraints. Zhang et al. [37] propose to co-cluster different-typed nodes by factorize meta-path based similarity matrices at the same time. These co-clustering methods are often applied for networks neglecting node attributes but fail to deal with attributed

HINs. Yao et al [38] propose to co-cluster multi-view data by measuring the relevance of nodes in views like the texture view and a color view, while ignoring the structural information. Recently, some works [10, 11, 39] attempt to aggregate attributed information to reconstruct homogeneous/heterogeneous graph embedding, which has attracted much attention. However, they focus on learning a general graph embedding rather than aiming at co-clustering.

Thus, we consider this work is valuable and meaningful.

3. Preliminaries and problem definition

Here we introduce the relevant concepts and the problem of semi-supervised co-clustering on attributed HINs. Main notations are summarized in Table 1.

Definition 1. ***Heterogeneous Information Networks (HINs)****: A HIN is denoted as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{T}, \mathcal{R}\}$. where $\mathcal{V}$ is the set of nodes, $\mathcal{E}$ is the set of links, $\mathcal{T}$ is the set of node types and $\mathcal{R}$ is the set of types of relations or links. There are two mapping functions on HINs, one of which is node type mapping $\phi : \mathcal{V} \to \mathcal{T}$ to obtain the type of a node, and the other is link type mapping $\psi : \mathcal{E} \to \mathcal{R}$ to obtain the type of a link. Notice that $|\mathcal{T}| + |\mathcal{R}| > 2$.*

Definition 2. ***Attributed HINs****: An attributed HIN is a special type of HINs in the form of $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{F}\}$. Compared with traditional HINs, there are abundant attribute information on attributed HINs, namely, $\mathcal{F} = \{\boldsymbol{f}_v\}$ where $\boldsymbol{f}_v$ is the attribute vector of node v. Notice that, the attribute vectors of heterogeneous nodes may be of different sizes and meanings.*

Taking Figure 1 as an example, there are three types of nodes (i.e., $\mathcal{T} = \{A,\ P,\ C\}$), two types of links (i.e., $\mathcal{R} = \{$ *"write"*, *"submit"*$\}$). Moreover, both authors and conferences contain several attributes in the form of vectors. Since the attributes of authors and conferences denote different meanings, here we respectively utilize parallelograms and squares to distinguish them.

Definition 3. ***Meta-path****: A meta-path $\mathcal{P} : \mathcal{T}_1 \xrightarrow{\mathcal{R}_1} \mathcal{T}_2 \xrightarrow{\mathcal{R}_2} \cdots \xrightarrow{\mathcal{R}_l} \mathcal{T}_{l+1}$ represents the connection from the source node of type $\mathcal{T}_1$ to the target node of*

Table 1: Notations

Symbols	Descriptions
$\boldsymbol{V}_s, \boldsymbol{V}_t$	source nodes, target nodes
$\mathcal{P}$	the given meta-path
$\boldsymbol{\Lambda}$	the weights of meta-path based relevance
$\boldsymbol{f}_{s,i}, \boldsymbol{f}_{t,j}$	the attributes of node $v_{s,i}$ and $v_{t,j}$
D_t, D_s	the dimensions of $\boldsymbol{f}_{t,i}$ and $\boldsymbol{f}_{s,j}$
$\boldsymbol{A}$	the relevance of attribute spaces
$\boldsymbol{M}_{ss}, \boldsymbol{C}_{ss}$	the must/cannot-link constraints of source nodes
$\boldsymbol{M}_{tt}, \boldsymbol{C}_{tt}$	the must/cannot-link constraints of target nodes
$\boldsymbol{M}_{st}, \boldsymbol{C}_{st}$	the must/cannot-link constraints between source and target nodes
K_s, K_t	the number of source and target clusters
α	the weight of structure and attributes
$\boldsymbol{X}_L$	the link-based relevance $\in \mathbb{R}_+^{\lvert\boldsymbol{V}_s\rvert\times\lvert\boldsymbol{V}_t\rvert}$
$\boldsymbol{X}_A$	the attribute-based relevance $\in \mathbb{R}_+^{\lvert\boldsymbol{V}_s\rvert\times\lvert\boldsymbol{V}_t\rvert}$
$\boldsymbol{X}$	the overall relevance $\in \mathbb{R}_+^{\lvert\boldsymbol{V}_s\rvert\times\lvert\boldsymbol{V}_t\rvert}$
$\boldsymbol{S}$, $\boldsymbol{T}$	cluster distribution of $\boldsymbol{V}_s$/ $\boldsymbol{V}_t$
$\boldsymbol{W}$	the matrix to absorb scales of $\boldsymbol{S}$ and $\boldsymbol{T}$

type $\mathcal{T}_{l+1}$ based on the composite relation $\mathcal{R} = \mathcal{R}_1 \circ \mathcal{R}_2 \circ \cdots \circ \mathcal{R}_l$.

In Figure 1, the source nodes and target nodes of meta-path A-P-C are authors and conferences. Furthermore, different meta-paths will capture different semantics, which is helpful for clustering. For instance, A_4 and C_3 can be connected by A-P-C or A-P-A-P-C, the first meta-path represents the "publication" while the second path is to capture the relevance of authors and conferences through co-authors.

The problem of semi-supervised co-clustering on attributed HINs: Given an attributed HIN formed as $\mathcal{G} = \{\mathcal{V}, \mathcal{E}, \mathcal{F}\}$, some meta-paths connecting source nodes $\boldsymbol{V}_s$ and target nodes $\boldsymbol{V}_t$, and some must-link ($\boldsymbol{M}_{ss}, \boldsymbol{M}_{st}$ and $\boldsymbol{M}_{tt}$)/cannot-link ($\boldsymbol{C}_{ss}, \boldsymbol{C}_{st}$ and $\boldsymbol{C}_{tt}$) constraints between nodes, the goal is to simultaneously generate clusters of $\boldsymbol{V}_s$ and $\boldsymbol{V}_t$ (i.e., $\boldsymbol{S}$ and $\boldsymbol{T}$) with the overall relevance matrix $\boldsymbol{X}$ considering both structural and attributed information. Notice that $\boldsymbol{V}_s$ and $\boldsymbol{V}_t$ respectively denote the row instances and column instances of $\boldsymbol{X}$. In addition, the subscripts of $\boldsymbol{M}/\boldsymbol{C}$ denote the types of constraints. For instance, $\boldsymbol{M}_{st}$ are the must-link constraints between $\boldsymbol{V}_s$ and $\boldsymbol{V}_t$.

This problem is meaningful and promising in real-world applications. On the one hand, we focus on making full use of heterogeneous structures and attributes rather than the single connection of graphs. On the other hand, the latent relevance of different-typed nodes can be detected for further recommender systems.

4. The proposed model

In this section, we propose our method ***S****emi-supervised* ***C****o-****C****lustering in* ***A****ttributed heterogeneous* ***I****nformation* ***N****etwork* (SCCAIN). We will introduce the overall framework as well as zoom into each component in the following sections.

4.1. The overall framework

We outline the overall framework of SCCAIN in Figure 2. In this framework, we respectively design the structural relevance measure based on the importance of meta-paths $\boldsymbol{\Lambda}$ and the attributed relevance measure based on the latent parameter $\boldsymbol{A}$. Considering both kinds of relevance, we compose them into an overall relevance measure, as introduced in Section 4.2. Secondly, we design an ONMTF based semi-supervised co-clustering model in Section 4.3 which factorizes the relevance matrix into two cluster distributions, $\boldsymbol{S}$ and $\boldsymbol{T}$, and an auxiliary matrix $\boldsymbol{W}$. Furthermore, since the performance of relevance measure and co-clustering can mutually influence each other, we integrate the two parts into a joint framework and optimize them in Section 4.4 to learn the final clusters of heterogeneous nodes.

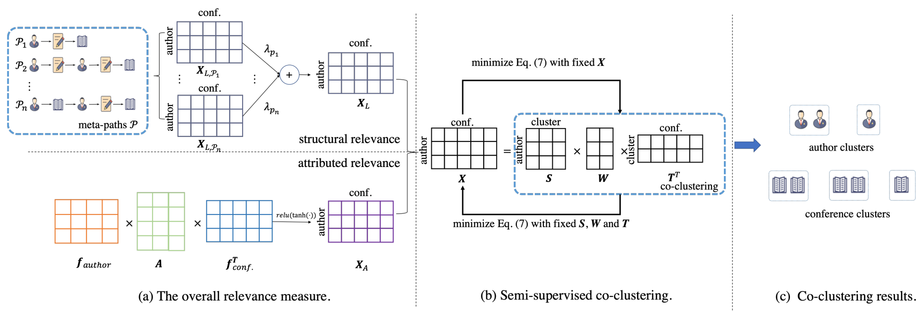


Figure 2: The overall framework of SCCAIN. (a) is to measure the overall relevance based on both attributes and structures, (b) is to co-cluster nodes of different types and (c) is the results of co-clustering.

4.2. Relevance measure on attributed HINs

4.2.1. Structural relevance

We adopt HeteSim [5] to measure the relevance between i^{th} type-s nodes and j^{th} type-t nodes, denoted as

$$\begin{aligned} &HS(v_{s,i}, v_{t,j}|\mathcal{R}_1 \circ \cdots \circ \mathcal{R}_l) \\ &= \frac{\sum_{i'=1}^{O(v_{s,i}|\mathcal{R}_1)} \sum_{j'=1}^{I(v_{t,j}|\mathcal{R}_1)} HS(v_{s,i'}, v_{t,j'}|\mathcal{R}_2 \circ \cdots \circ \mathcal{R}_{l-1})}{|O(v_{s,i}|\mathcal{R}_1)||I(v_{t,j}|\mathcal{R}_l)|}, \end{aligned} \tag{1}$$

where $v_{s,i}$ and $v_{t,j}$ respectively denotes the i^{th} source node and the j^{th} target node, $HS(v_{s,i}, v_{t,j}|\mathcal{R}_1 \circ \cdots \circ \mathcal{R}_l)$ is the HeteSim[5] value between $v_{s,i}$ and $v_{t,j}$ on the meta-path $\mathcal{R}_1 \circ \cdots \circ \mathcal{R}_l$, $O(v_{s,i}|\mathcal{R}_1)$ is the out-neighbors of $v_{s,i}$ based on relation $\mathcal{R}_1$, $I(v_{t,j}|\mathcal{R}_l)$ is the input-neighbors of $v_{t,j}$ based on relation $\mathcal{R}_l$. $HS(v_{s,i'}, v_{t,j'}|\mathcal{R}_{\frac{l+1}{2}}) = 1$ if $v_{s,i'} = v_{t,j'}$, or else 0. Different from traditional PathSim [3] which only calculates the similarity of homogeneous nodes or PCRW [4] in which the relevance is not symmetric, HeteSim can measure the relevance of different-type nodes.

Considering that there are several meta-paths and each meta-path indicates one form of structural relevance, as is shown in Figure 2, we assign a meta-path importance weight $\lambda_{\mathcal{P}}$ to the specific relevance $HS(v_{s,i}, v_{t,j}|\mathcal{P})$, and then calculate the structural relevance, namely,

$$\boldsymbol{X}_L(v_{s,i}, v_{t,j}|\boldsymbol{\Lambda}) = \sum_{\mathcal{P}} \lambda_{\mathcal{P}} \cdot HS(v_{s,i}, v_{t,j}|\mathcal{P}), \tag{2}$$

where $\mathcal{P}$ denotes a specific meta-path, $HS(v_{s,i}, v_{t,i}|\mathcal{P})$ denotes the structural relevance on $\mathcal{P}$, $\lambda_{\mathcal{P}} \in \boldsymbol{\Lambda}$ is the weight of the corresponding relevance and $\sum_{\mathcal{P}} \lambda_{\mathcal{P}} = 1$.

4.2.2. Attributed relevance

Given the features of i^{th} source node $\boldsymbol{f}_{s,i}$ and j^{th} target node $\boldsymbol{f}_{t,j}$, it is impossible to directly measure the relevance of $\boldsymbol{f}_{s,i}$ and $\boldsymbol{f}_{t,j}$. By mapping attributes in different spaces into the same space, here we calculate the attributed relevance measure of $v_{s,i}$ and $v_{t,j}$ as follows.

$$\boldsymbol{X}_A(v_{s,i}, v_{t,j}) = \sigma(\boldsymbol{f}_{s,i}\boldsymbol{A}\boldsymbol{f}_{t,j}^T + b), \tag{3}$$

where $\boldsymbol{X}_A$ is the attributed relevance matrix, $\boldsymbol{A} \in \mathbb{R}^{D_s \times D_t}$ is the relevance parameters of attribute vectors of different spaces, $\sigma(\cdot)$ is the activation function and we adopt $relu$ to keep attributed relevances positive.

4.2.3. Overall relevance

Taking both structural and attributed information into consideration, SC-CAIN evaluates the overall relevance of two nodes based on their structural relevance and attributed relevance. By setting a balance parameter $\alpha \in [0, 1]$, the overall relevance of $v_{s,i}$ and $v_{t,j}$ is defined as

$$\boldsymbol{X}(v_{s,i}, v_{t,j}) = \alpha \boldsymbol{X}_A(v_{s,i}, v_{t,j}) + (1-\alpha)\boldsymbol{X}_L(v_{s,i}, v_{t,j}|\boldsymbol{\Lambda}). \tag{4}$$

To learn the parameters more effectively, we utilize additional constraints to guide the optimization. The corresponding loss function with constraints is denoted as

$$L_1 = -\frac{1}{m}[\sum MC_{i,j} \log(\boldsymbol{X}_{i,j}) + (1 - MC_{i,j})\log(1 - \boldsymbol{X}_{i,j})], \tag{5}$$

where m is the number of labels, $MC_{i,j}$ is the constraints of different-type nodes, according to the given must-link set $\boldsymbol{M}_{s,t}$ and cannot-link set $\boldsymbol{C}_{s,t}$ and

. $MC_{i,j} = 1$ if $\boldsymbol{M}_{st,i,j} = 1$ while $MC_{i,j} = 0$ if $\boldsymbol{C}_{st,i,j} = 1$, $\boldsymbol{X}_{i,j} = \boldsymbol{X}(v_{s,i}, v_{t,j})$. These constraints, together with constraints on same-type nodes, can also be utilized to guide co-clustering, as we will discuss next.

4.3. Semi-supervised co-clustering

In this section, we design the semi-supervised non-negative matrix tri-factorization with orthogonal limitation to simultaneously cluster nodes of different types.

$$
\begin{aligned}
L_2 =&||\boldsymbol{X} - \boldsymbol{S}\boldsymbol{W}\boldsymbol{T}^T||^2 - \sum_{i,j}(\boldsymbol{M}_{ss,i,j} - \boldsymbol{C}_{ss,i,j})\boldsymbol{S}_i\boldsymbol{S}_j^T \\
&- \sum_{i,j}(\boldsymbol{M}_{tt,i,j} - \boldsymbol{C}_{tt,i,j})\boldsymbol{T}_i\boldsymbol{T}_j^T \\
s.t. \quad & \boldsymbol{S} \geq 0, \boldsymbol{W} \geq 0, \boldsymbol{T} \geq 0, \boldsymbol{S}^T\boldsymbol{S} = \boldsymbol{I}, \boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{I},
\end{aligned} \tag{6}
$$

where $\boldsymbol{X} \in \mathbb{R}_+^{|\boldsymbol{V}_s|\times|\boldsymbol{V}_t|}$ is the relevance matrix, $\boldsymbol{S} \in \mathbb{R}_+^{|\boldsymbol{V}_s|\times K_s}$ and $\boldsymbol{T} \in \mathbb{R}_+^{|\boldsymbol{V}_t|\times K_t}$ are the distributions of node clusters, K_s and K_t are the number of clusters of source/target nodes, $\boldsymbol{W} \in \mathbb{R}_+^{K_s\times K_t}$ is an extra factor to absorb $\boldsymbol{S}$ and $\boldsymbol{T}$, $\boldsymbol{M}_{ss} \in \mathbb{R}^{|\boldsymbol{V}_s|\times|\boldsymbol{V}_s|}$ and $\boldsymbol{C}_{ss} \in \mathbb{R}^{|\boldsymbol{V}_s|\times|\boldsymbol{V}_s|}$ are the must-link and cannot-link pair-wise constraints of $\boldsymbol{V}_s$, while $\boldsymbol{M}_{tt} \in \mathbb{R}^{|\boldsymbol{V}_t|\times|\boldsymbol{V}_t|}$ and $\boldsymbol{C}_{tt} \in \mathbb{R}^{|\boldsymbol{V}_t|\times|\boldsymbol{V}_t|}$ are the must-link and cannot-link pair-wise constraints of $\boldsymbol{V}_t$.

4.4. Joint optimization

Given the different-type nodes, $\boldsymbol{V}_s$ and $\boldsymbol{V}_t$, our goal is to simultaneously cluster $\boldsymbol{V}_s$ and $\boldsymbol{V}_t$ by utilizing the structural information and attributed information, as well as some constraints including must-link/cannot-link pairs. To optimize both the co-clustering and relevance measure together in this model, we design a joint model to learn the corresponding parameters including the weights of meta-paths Λ and the clustering distributions $\boldsymbol{S}$ and $\boldsymbol{T}$. Specifically, we consider the relevance matrix $\boldsymbol{X}$ as a variable $\boldsymbol{X}(\Theta)$ related to parameters $\Theta = \{\boldsymbol{\Lambda}, \boldsymbol{A}\}$, and the loss function is denoted as

$$
L = L_1(\Theta) + L_2(\Theta) + \gamma(||\Theta||^2). \tag{7}
$$

In SCCAIN, we learn the parameters Θ and $(\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{T})$ with an iterative update approach, and each iteration is made up of two steps.

Update $\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{T}$ with fixed Θ Given Θ, the main goal in this step is to select the solution $(\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{T})$ of the semi-supervised co-clustering model. With the fixed $\boldsymbol{X}$, L can be written as:

$$\begin{aligned} L_{cocl} &= Tr(\boldsymbol{X}^T\boldsymbol{X} - 2\boldsymbol{X}^T\boldsymbol{S}\boldsymbol{W}\boldsymbol{T}^T + \boldsymbol{S}\boldsymbol{W}\boldsymbol{T}^T\boldsymbol{T}\boldsymbol{W}^T\boldsymbol{S}^T) \\ &\quad + Tr((\boldsymbol{C}_{ss} - \boldsymbol{M}_{ss})\boldsymbol{S}\boldsymbol{S}^T + (\boldsymbol{C}_{tt} - \boldsymbol{M}_{tt})\boldsymbol{T}\boldsymbol{T}^T) \\ &\quad s.t. \quad \boldsymbol{S} \geq 0, \boldsymbol{W} \geq 0, \boldsymbol{T} \geq 0, \boldsymbol{S}^T\boldsymbol{S} = \boldsymbol{I}, \boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{I}. \end{aligned} \tag{8}$$

There are three parameters with some constraints in this function. We respectively fix two of the parameters to optimize the other one.

$$\boldsymbol{S}_{i,k} \longleftarrow \boldsymbol{S}_{i,k}\sqrt{\frac{[\boldsymbol{X}\boldsymbol{T}\boldsymbol{W}^T + \boldsymbol{M}_{ss}\boldsymbol{S}]_{i,k}}{[\boldsymbol{S}\boldsymbol{S}^T(\boldsymbol{X}\boldsymbol{T}\boldsymbol{W}^T - \boldsymbol{C}_{ss}\boldsymbol{S} + \boldsymbol{M}_{ss}\boldsymbol{S}) + \boldsymbol{C}_{ss}\boldsymbol{S}_{i,k}}}. \tag{9}$$

$$\boldsymbol{T}_{i,k} \longleftarrow \boldsymbol{T}_{i,k}\sqrt{\frac{[\boldsymbol{X}^T\boldsymbol{S}\boldsymbol{W} + \boldsymbol{M}_{tt}\boldsymbol{T}]_{i,k}}{[\boldsymbol{T}\boldsymbol{T}^T(\boldsymbol{X}^T\boldsymbol{S}\boldsymbol{W} - \boldsymbol{C}_{tt}\boldsymbol{T} + \boldsymbol{M}_{tt}\boldsymbol{T}) + \boldsymbol{C}_{tt}\boldsymbol{T}]_{i,k}}}. \tag{10}$$

$$\boldsymbol{W}_{i,k} \longleftarrow \boldsymbol{W}_{i,k}\sqrt{\frac{[\boldsymbol{S}^T\boldsymbol{X}\boldsymbol{T}]_{i,k}}{[\boldsymbol{S}^T\boldsymbol{S}\boldsymbol{W}\boldsymbol{T}^T\boldsymbol{T}]_{i,k}}}. \tag{11}$$

To obtain the accurate $\boldsymbol{S}$, $\boldsymbol{W}$ and $\boldsymbol{T}$, we iteratively update these three parameters until they are stable. With the update process done, we fix $\boldsymbol{S}$, $\boldsymbol{W}$, $\boldsymbol{T}$ to optimize the relevance measure.

Update Θ with fixed $\boldsymbol{S}, \boldsymbol{W}, \boldsymbol{T}$ For the fixed $\boldsymbol{S}$, $\boldsymbol{W}$, $\boldsymbol{T}$, L is a function of $\Theta = \{\boldsymbol{\Lambda}, \boldsymbol{A}\}$. The global loss function can be rewritten as follows.

$$L_{rele} = ||\boldsymbol{X}(\Theta) - \boldsymbol{S}\boldsymbol{W}\boldsymbol{T}^T||^2 + \gamma(||\Theta||^2), \tag{12}$$

where $\boldsymbol{X}(\Theta)$ is calculated by Eq. (4) with the parameters Θ, and $\boldsymbol{S}\boldsymbol{W}\boldsymbol{T}^T$ is the fixed value. In addition, considering $\boldsymbol{\Lambda} \geq 0$, we update $\lambda_{\mathcal{P}}$ by $max(0, \lambda_{\mathcal{P}})$.

Finally, we respectively get the clusters of source nodes $\boldsymbol{G}_s$ and the clusters of target nodes $\boldsymbol{G}_t$ from the optimized $\boldsymbol{S}$ and $\boldsymbol{T}$. Specifically, $\boldsymbol{G}_{s,i} = \arg\max_{k_s} \boldsymbol{S}_{i,k_s}$, and $\boldsymbol{G}_{t,j} = \arg\max_{k_t} \boldsymbol{T}_{j,k_t}$.

5. Experiments

In this section, we evaluate the empirical performance of our method on three public real-world attributed HINs, including two bibliographic networks (Aminer and DBLP) and a user-item Alibaba recommendation network. More Specifically, we study the effectiveness and efficiency of our method.

5.1. Datasets and metrics

The statistics of our three public datasets, namely, Aminer, DBLP and an Alibaba recommendation dataset, are summarized in Table 2. The details of these datasets are described as follows.

1) Aminer dataset: This is a public benchmark dataset[1], which consists of three types of nodes, authors (A), papers (P) and conferences (C), and the corresponding relations including "publish" (P-C), "participation" (A-C) and "write" (A-P). There are five main research areas, including *Data Mining*, *Medical Informatics*, *Theory*, *Visualization* and *Database*, and each node is assigned to a specific area. We focus on clustering authors and conferences at the same time. The structural relevance is calculated based on three meta-paths including A-P-C, A-P-A-P-C and A-P-C-P-A-P-C. The attributes of both authors and conferences are the related abstracts of papers and we utilize doc2vec [40] to model texts as dense vectors.

2) DBLP dataset: This is a public sub-network[2] involving major conferences in four research areas: *Database*, *Data Mining*, *Artificial Intelligence* and *Information Retrieval*. There are four types of nodes, authors (A), papers(P), conference (C) and topics (T). In this network, we focus on co-clustering authors and conferences, too. The structural relevance is calculated based on three meta-paths, A-P-C, A-P-T-P-C and A-P-A-P-C. Here we set the number of papers written by authors at the 20 conferences as the attributes of authors, and

[1] Available at `https://www.aminer.cn/topic_paper_author`

[2] Available at `http://shichuan.org/HIN_dataset.html`

Table 2: Description of datasets.

Dataset	#nodes	#types	G_s	G_t	$\mathcal{P}$	D_s	D_t
Aminer	2,593	3	5	5	3	64	64
DBLP	14,495	4	4	4	3	20	20
Alibaba	5,415	2	9	65	3	5	22

set the number of links through the meta-path C-P-A-P-C as the attributes of conferences.

3) Alibaba dataset: This is a subset of user activity dataset[3] which captures user actions in one week on a public Alibaba platform. It consists of two types of nodes, users (U) and items (I), as well as multiple types of relations, “click”, “cart”, “favorite” and “buy” between users and items. The structural relevance is calculated based on three meta-paths, U $\xrightarrow{click}$ I, U $\xrightarrow{cart}$ I and U $\xrightarrow{favorite}$ I. We consider the “buy” action as the constraint information. In this dataset, users contain 5 numerical attributes including the total number of “click”, the rate of “cart” and so on, while items contain 22 numerical attributes including the number of transactions, the rate of “cart” and so on. Users are of 9 age groups, and items are of 65 main categories. In this network, we focus on co-clustering users and items by utilizing the above relations and attributes.

5.2. Baselines and experimental settings

We first compare our SCCAIN[4] with the state-of-the-art methods including three co-clustering methods and two graph embedding methods. And then, we analyze the contributions of integrating attributes and structures by comparing SCCAIN with its modified versions, **SCCAIN(L)** and **SCCAIN(A)**, where the former focuses on structures while the latter focuses on attributes. The details of baselines are listed as follow.

- **DNMTF** [14]: This is a matrix tri-factorization method that optimizes both

[3]Available at `https://tianchi.aliyun.com/dataset/dataDetail?dataId=9716`
[4]The source is available at https://github.com/yuduo93/SCCAIN

matrix factorization and graph dual regularization when co-clustering. For a fair comparison, both the k-nearest neighbor of nodes and the pair-wise constraints are set as dual regularization. Here we utilize the links within attributed HINs as the input matrix.

- **CPSSCC** [15]: This is a semi-supervised co-clustering method which utilizes both row constraint projections and column constraint projection to guide cluster source nodes and target nodes in a low-dimensional space.

- **ONMTF(HS)** [20]: This is a non-negative matrix tri-factorization. Different from traditional ONMTF proposed in [20], we utilize the average relevance of multple meta-paths as the similarity matrix and integrate pair-wise constraints during training to discuss the effectiveless of learning relevance measure.

- **GCN(K)** [10]: This is a popular attributed graph embedding learning method which aggregates graph information to reconstruct node embedding. We generate the basic embedding of nodes by using the supervised information and adopt the K-means method to cluster nodes of each type respectively.

- **H2V(K)** [41]:This is a heterogeneous graph embedding model with K-means. H2V samples neighbors according to heterogeneous edges and learns the embedding of both nodes and edges. In this model, we integrate the pair-wise constraints into the graph for a fair comparison.

For SCCAIN, we set the learning rate as 0.001, the max-iterations as 200 and γ as 0.01. We utilize Adam to minimize the loss of L_1. Since the must-link and cannot-link pairs are provided as supervision information, α are tuned by cross-validation. For Aminer, DBLP and Alibaba datasets, We respectively generate fixed number of must-link and cannot-link neighbors as the total constraints, and then sample 2.5%, 5%, 7.5% and 10% of the constraints for learning. The baselines and our SCCAIN all run ten times and we report the mean value as the performance. Both the Normalized Mutual Information (NMI)[42] and Purity[43] are adopted as the metrics. $NMI \in [0, 1]$ and $Purity \in [0, 1]$, and

Table 3: The NMI and Purity value on the three datasets of different scales. For DBLP and Aminer, the source nodes and target nodes respectively denote authors and conferences. For Alibaba, the source nodes and target nodes respectively denote users and items. The best method is bolded.

Dataset	Rate	Metric	Target Clustering						Source Clustering					
			DNMTF	CPSSCC	GCN(K)	H2V(K)	ONMTF(HS)	SCCAIN	DNMTF	CPSSCC	GCN(K)	H2V(K)	ONMTF(HS)	SCCAIN
DBLP	2.5%	NMI	0.5048	0.2491	0.2219	0.2232	0.2336	**0.6374**	0.2187	0.0743	0.0220	0.0168	0.0755	**0.2594**
	5%		0.6264	0.2925	0.3141	0.3011	0.3214	**0.6353**	0.5123	0.1737	0.3641	0.3819	0.2030	**0.5282**
	7.5%		**0.7013**	0.2951	0.3816	0.3418	0.3816	0.6794	0.4833	0.3187	0.4898	0.6259	0.8410	**0.8834**
	10%		**0.7424**	0.3683	0.4365	0.4177	0.4365	0.6265	0.6386	0.3655	0.7477	0.7731	0.9399	**0.9666**
	2.5%	Purity	0.5971	0.4012	0.4730	0.4839	0.3500	**0.7013**	0.4437	0.3194	0.3480	0.3439	0.3614	**0.4565**
	5%		0.6688	0.4128	0.5232	0.5102	0.4510	**0.7002**	**0.6842**	0.3526	0.4543	0.6522	0.4666	0.6290
	7.5%		0.7105	0.4505	0.5555	0.5672	0.5325	**0.7328**	0.6722	0.5154	0.6345	0.7481	0.9502	**0.9551**
	10%		**0.7536**	0.4989	0.6150	0.6003	0.6000	0.6992	0.8070	0.5689	0.7099	0.7969	0.9840	**0.9911**
Aminer	2.5%	NMI	0.3038	0.2474	0.4287	0.4528	0.3595	**0.7688**	0.1062	0.0379	0.0205	0.0378	0.0386	**0.6285**
	5%		0.6285	0.3209	0.4027	0.3020	0.4635	**0.7416**	0.1670	0.0607	0.0243	0.0465	0.0669	**0.6130**
	7.5%		0.7492	0.3609	0.3952	0.2867	0.5593	**0.7504**	0.1879	0.1396	0.0285	0.0451	0.1674	**0.6952**
	10%		0.7514	0.3957	0.3059	0.5025	0.5987	**0.7681**	0.2120	0.1992	0.0303	0.0466	0.2278	**0.7002**
	2.5%	Purity	0.4089	0.4022	0.5909	0.5292	0.4545	**0.7676**	0.2929	0.2797	0.2653	0.2785	0.4730	**0.7052**
	5%		0.6818	0.4523	0.5458	0.4101	0.5909	**0.7218**	0.3520	0.2886	0.2633	0.2777	0.3298	**0.6943**
	7.5%		0.7625	0.5027	0.5001	0.4023	0.6364	**0.7473**	0.3831	0.3423	0.2730	0.2785	0.4189	**0.7620**
	10%		**0.7727**	0.5909	0.4545	0.5455	0.6455	0.7510	0.4352	0.4002	0.2999	0.2843	0.5010	**0.7466**
Alibaba	2.5%	NMI	0.0812	0.0755	0.1421	0.0488	0.2225	**0.2828**	0.0294	0.0334	0.0153	0.0635	0.0521	**0.2271**
	5%		0.0876	0.0724	0.1707	0.3929	0.4156	**0.4543**	0.0231	0.0333	0.0695	0.3528	0.3983	**0.5233**
	7.5%		0.3181	0.0661	0.3641	**0.5879**	0.5507	0.5699	0.0227	0.0347	0.2065	0.4464	0.6479	**0.7708**
	10%		**0.6246**	0.0658	0.5113	0.6198	0.6001	0.6126	0.0482	0.0445	0.3458	0.6822	**0.8306**	0.8216
	2.5%	Purity	0.3918	0.3900	0.4298	0.4057	0.5576	**0.5672**	0.3737	0.3731	0.3695	0.4069	0.3936	**0.4665**
	5%		0.4033	0.3846	0.5142	0.6896	0.7764	**0.7987**	0.3719	0.3737	0.3984	0.5310	0.6920	**0.7269**
	7.5%		0.5443	0.3918	0.6462	0.7498	0.9216	**0.9295**	0.3725	0.3719	0.5467	0.5546	0.8433	**0.9403**
	10%		0.7743	0.3876	0.7257	0.8348	0.9699	**0.9735**	0.3773	0.3779	0.6010	0.6828	**0.9689**	0.9632

the larger NMI or $Purity$ value indicate the better performance.

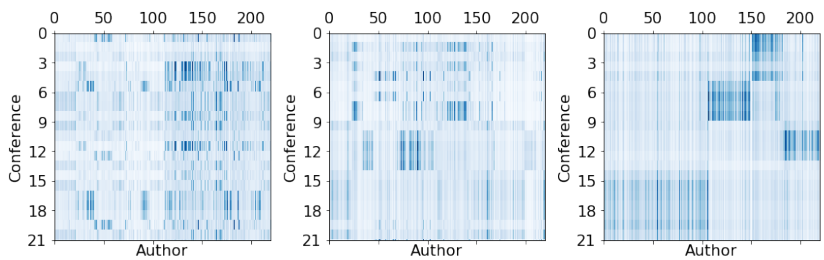
5.3. Effectiveness analysis

In this section, we firstly compare the performance of our method on clustering of both source nodes and target nodes. The experimental results of co-clustering on three datasets are reported in Table 3. And then, we show the co-clustering visualization in Figure 3 to describe the performance of detecting latent correlations between heterogeneous clusters.

5.3.1. The performance of co-clustering

By comparing the co-clustering performance of these baselines and our model, we summary two main observations and list them as follows

- SCCAIN generally achieves the best performance for all three datasets. Compared with DNMTF, CPSSCC, and ONMTF(HS), the main improvement is



(a) DNMTF (b) ONMTF(HS) (c) SCCAIN

Figure 3: The relevance of co-clustering on Aminer dataset with 10% constraints. Both authors and conferences are assigned into five groups. The deeper colored blocks denotes the higher relevance of clusters.

from the learnable overall relevance. Compared with GCN(K) and H2V(K), our method is a unified model considering both heterogeneity and attributes within attributed HINs. In addition, DNMTF and CPSSCC perform worse on Alibaba dataset becasue of the sparsity of this network. Moreover, some methods only work effectively in clustering one type of nodes, e.g., CPSSCC.

- With little supervision, SCCAIN performs much better than most baselines on both clustering of source nodes and clustering of target nodes. This phenomenon indicates the advantages of SCCAIN on graphs without too many constraints. However, the performance of some other models, e.g., GCN(K) and DNMTF, depends on the scale of supervised information heavily and increases more slowly than SCCAIN.
- Compared with ONMTF(HS) which adopts static relevance, our SCCAIN performs quite better with little supervision because of the adaptive overall relevance measure.

5.3.2. Co-clustering visualization

As is mentioned above, our model is to simultaneously cluster different-type nodes. In this section, we analyze the relevance of these clusters by setting the co-clustering on Aminer dataset as an example. In Figure 3, we rearrange relevance matrices of DNMTF, ONMTF(HS) (which perform better than other

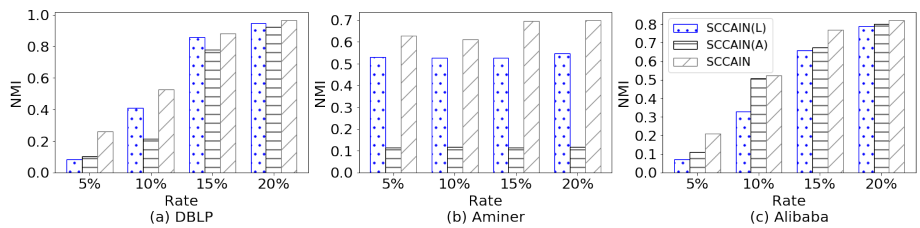


Figure 4: The NMI performance of SCCAIN(A), SCCAIN(L) and SCCAIN on co-clustering.

baselines in this dataset) and our SSCAIN based on authors' clusters and conferences' clusters, and then show the corresponding matrix visualization. By comparing the blocks in Figure 3a, Figure 3b and Figure 3c, we can easily find that our SCCAIN have a better ability to detect the relevance of different-type clusters because of obvious blocks. It will be helpful to utilize such information for recommendation systems and some other valuable tasks.

5.4. Model analysis

In this section, we analyze the characteristics of SCCAIN, including the advantages of integrating attributes and structures for co-clustering, the convergences, and the parameter sensitivity.

5.4.1. Ablation study

Here we compare SCCAIN with SCCAIN(L) and SCCAIN(A). As shown in Figure 4, compared with both SCCAIN(A) and SCCAIN(L), our model performs better on both three datasets. On Aminer dataset and DBLP dataset, SCCAIN(L) achieves better performance than SCCAIN(A), However, SCCAIN(A) is better on Alibaba dataset. These phenomenons authenticate the effectiveness of integrating both attributed and structural information for co-clustering. Moreover, although containing similar structures, SCCAIN is better than ONMTF(HS) because of the auto-learning meta-path weights.

5.4.2. Convergence study and parameter analysis

To analyze the convergence of SCCAIN, we increase the max iteration from 0 to 200 and showcase the NMI value of SCCAIN with different supervised

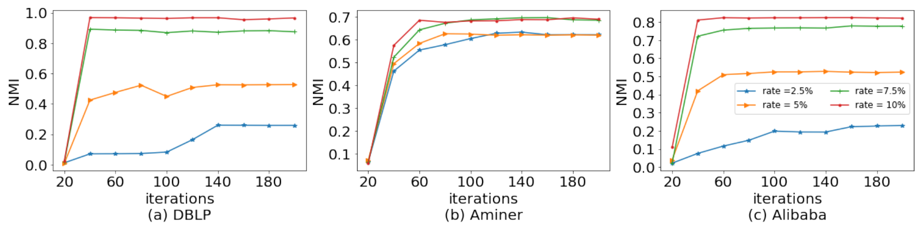


Figure 5: The NMI value of SCCAIN on Aminer dataset with different iterations.

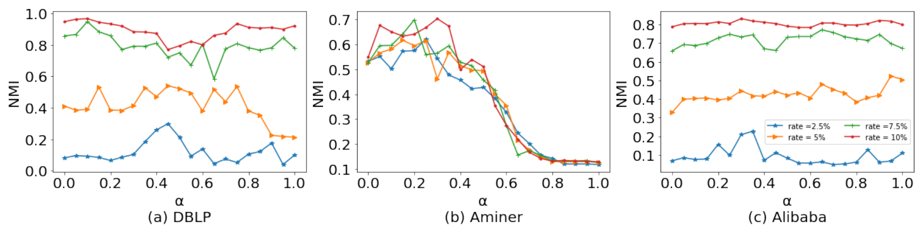


Figure 6: The NMI value of SCCAIN on Aminer dataset with different α.

information. As shown in Figure 5, we can easily observe that SCCAIN can more quickly converge with the supervised information increasing on all three datasets. This phenomenon proves the effectiveness of supervised information as well as the optimization framework.

To analyze the influence of the balance parameter α, we adjust it from 0 to 1 and report the NMI value of SCCAIN in Figure 6. By comparing the performance on each dataset, we can find that a suitable α will improve the NMI value of clustering. By comparing the trend of performances on different datasets, we can observe that α is quite more sensitive on Aminer dataset. This phenomenon is reasonable. Specifically, the attributes of authors and conferences on Aminer dataset are the average of associated abstract vectors. It could be difficult to distinguish authors or conferences based on too much attributed relevance (namely, $\alpha \geq 0.5$). On the one hand, papers of a conference often belong to multiple domains, so that the average of their attribute vectors could become similar for many different conferences. On the other hand, there could be some noisy information in modeling the representation of abstract texts since the corpus is not very large. On DBLP and Alibaba dataset, the stable performance indicates that the relevance of node attributes can be more easily learnt to help co-clustering.

6. Conclusion

In this paper, we address the problem of co-clustering on attributed heterogeneous information networks by making full use of both attributed and structural information. We design a joint model called SCCAIN to integrate the semi-supervised ONMTF and the overall relevance measure for co-clustering different-type nodes. We test our model on three public real-world datasets and the experimental results demonstrate the effectiveness of SCCAIN comparing with representative methods. Furthermore, we analyze the key factors of SCCAIN including the co-clustering visualization, the integration of attributes and structures, the convergence and the balance parameters, to showcase the effectiveness of our solutions.

Co-clustering different-typed nodes on attributed HINs still remains an open problem on evolving networks. It is worth of considering the dynamics of edges and attributes and modeling the evolving of relevance by splitting attributed HINs into several snapshots of HINs. More future work can be done along this line.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China (No. U1936220, 61532006, 61772082), National Research Foundation, Prime Minister's Office, Singapore under its Campus for Research Excellence and Technological Enterprise (CREATE) programme, the Ministry of Education, Singapore under its Academic Research Funding Tier 1 (18-C220-SMU-006), and Alibaba Group under its Alibaba Innovative Research (AIR) programme.

[1] Y. Sun, J. Han, Mining heterogeneous information networks: a structural analysis approach, ACM SIGKDD Explorations Newsletter 14 (2012) 20–28.

[2] L. Liu, S. Wang, Meta-path-based outlier detection in heterogeneous information network, Frontiers Comput. Sci. 14 (2020) 388–403.

[3] Y. Sun, J. Han, X. Yan, P. S. Yu, T. Wu, Pathsim: Meta path-based top-k similarity search in heterogeneous information networks, Proceedings of the VLDB Endowment 4 (2011) 992–1003.

[4] N. Lao, W. W. Cohen, Fast query execution for retrieval models based on path-constrained random walks, in: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2010), Washington, DC, USA, 2010, pp. 881–888. URL: https://doi.org/10.1145/1835804.1835916.

[5] C. Shi, X. Kong, Y. Huang, P. S. Yu, B. Wu, Hetesim: A general framework for relevance measure in heterogeneous networks, IEEE Transactions on Knowledge and Data Engineering 26 (2014) 2479–2492.

[6] Y. Sun, B. Norick, J. Han, X. Yan, P. S. Yu, X. Yu, Pathselclus: Integrating meta-path selection with user-guided object clustering in heterogeneous information networks, ACM Transactions on Knowledge Discovery from Data 7 (2013) 11:1–11:23.

[7] C. Luo, W. Pang, Z. Wang, Semi-supervised clustering on heterogeneous information networks, in: Proceedings of the 18th Pacific-Asia Conference (PAKDD 2014), Tainan, Taiwan, 2014, pp. 548–559. URL: https://doi.org/10.1007/978-3-319-06605-9_45.

[8] C. Wang, Y. Song, H. Li, M. Zhang, J. Han, Knowsim: A document similarity measure on structured heterogeneous information networks, in: Proceedings of International Conference on Data Mining (ICDM 2015),

Atlantic City, NJ, USA, 2015, pp. 1015–1020. URL: https://doi.org/10.1109/ICDM.2015.131.

[9] C. Hsu, Y. Lai, W. Chen, M. Feng, S. Lin, Unsupervised ranking using graph structures and node attributes, in: Proceedings of the 10th ACM International Conference on Web Search and Data Mining (WSDM 2017), Cambridge, United Kingdom, 2017, pp. 771–779. URL: https://doi.org/10.1145/3018661.3018668.

[10] T. N. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proceedings of the 5th International Conference on Learning Representations (ICLR 2017), Toulon, France, 2017. URL: https://openreview.net/forum?id=SJU4ayYgl.

[11] W. L. Hamilton, Z. Ying, J. Leskovec, Inductive representation learning on large graphs, in: Proceedings of Annual Conference on Neural Information Processing Systems (NIPS 2017) Long Beach, CA, USA, 2017, pp. 1024–1034. URL: http://papers.nips.cc/paper/6703-inductive-representation-learning-on-large-graphs.

[12] M. P. Boobalan, D. Lopez, Spatio-temporal graph clustering algorithm based on attribute and structural similarity, International Journal of Knowledge-based and Intelligent Engineering Systems 20 (2016) 149–160.

[13] X. Li, Y. Wu, M. Ester, B. Kao, X. Wang, Y. Zheng, Semi-supervised clustering in attributed heterogeneous information networks, in: Proceedings of the 26th International Conference on World Wide Web (WWW 2017), Perth, Australia, 2017, pp. 1621–1629. URL: https://doi.org/10.1145/3038912.3052576.

[14] F. Shang, L. Jiao, F. Wang, Graph dual regularization non-negative matrix factorization for co-clustering, Pattern Recognition 45 (2012) 2237–2250.

[15] S. Huang, H. Wang, T. Li, Y. Yang, T. Li, Constraint co-projections for

semi-supervised co-clustering, IEEE transactions on cybernetics 46 (2016) 3047–3058.

[16] F. Nie, X. Wang, C. Deng, H. Huang, Learning A structured optimal bipartite graph for co-clustering, in: Proceedings of Annual Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 2017, pp. 4129–4138. URL: `http://papers.nips.cc/paper/7001-learning-a-structured-optimal-bipartite-graph\-for-co-clustering`.

[17] B. Altinel, M. C. Ganiz, Semantic text classification: A survey of past and recent advances, Information Processing & Management 54 (2018) 1129–1153.

[18] X. Shi, W. Fan, P. S. Yu, Efficient semi-supervised spectral co-clustering with constraints, in: Proceedings of the 10th IEEE International Conference on Data Mining (ICDM 2010), Sydney, Australia, 2010, pp. 1043–1048. URL: `https://doi.org/10.1109/ICDM.2010.64`.

[19] S. Huang, Z. Xu, I. W. Tsang, Z. Kang, Auto-weighted multi-view co-clustering with bipartite graphs, Information Sciences 512 (2020) 18–30.

[20] C. H. Q. Ding, T. Li, W. Peng, H. Park, Orthogonal nonnegative matrix t-factorizations for clustering, in: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2006), Philadelphia, PA, USA, 2006, pp. 126–135. URL: `https://doi.org/10.1145/1150402.1150420`.

[21] H. Ma, W. Zhao, Q. Tan, Z. Shi, Orthogonal nonnegative matrix tri-factorization for semi-supervised document co-clustering, in: Proceedings of the 14th Pacific-Asia Conference (PAKDD 2010), Hyderabad, India, 2010, pp. 189–200. URL: `https://doi.org/10.1007/978-3-642-13672-6_19`.

[22] D. Bo, X. Wang, C. Shi, M. Zhu, E. Lu, P. Cui, Structural deep clustering network, in: Proceedings of The Web Conference 2020, 2020, pp. 1400–1410.

[23] S. Fan, X. Wang, C. Shi, E. Lu, K. Lin, B. Wang, One2multi graph autoencoder for multi-view graph clustering, in: Proceedings of The Web Conference 2020, 2020, pp. 3070–3076.

[24] Y. Sun, Y. Yu, J. Han, Ranking-based clustering of heterogeneous information networks with star network schema, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2009), Paris, France, 2009, pp. 797–806. URL: https://doi.org/10.1145/1557019.1557107.

[25] Y. Sun, J. Han, P. Zhao, Z. Yin, H. Cheng, T. Wu, Rankclus: integrating clustering with ranking for heterogeneous information network analysis, in: Proceeding of the 12th International Conference on Extending Database Technology (EDBT 2009), Saint Petersburg, Russia, 2009, pp. 565–576. URL: https://doi.org/10.1145/1516360.1516426.

[26] H. Deng, J. Han, B. Zhao, Y. Yu, C. X. Lin, Probabilistic topic models with biased propagation on heterogeneous information networks, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2011), San Diego, CA, USA, 2011, pp. 1271–1279. URL: https://doi.org/10.1145/2020408.2020600.

[27] H. Deng, B. Zhao, J. Han, Collective topic modeling for heterogeneous networks, in: Proceeding of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2011), Beijing, China, 2011, pp. 1109–1110. URL: https://doi.org/10.1145/2009916.2010073.

[28] X. Li, B. Kao, Z. Ren, D. Yin, Spectral clustering in heterogeneous information networks, in: Proceedings of the 33th AAAI Conference on Artificial

Intelligence (AAAI 2019), Honolulu, USA, 2019, AAAI Press, 2019, pp. 4221–4228. URL: https://doi.org/10.1609/aaai.v33i01.33014221.

[29] Y. Sun, C. C. Aggarwal, J. Han, Relation strength-aware clustering of heterogeneous information networks with incomplete attributes, Proc. VLDB Endow. 5 (2012) 394–405.

[30] B. Perozzi, L. Akoglu, P. I. Sánchez, E. Müller, Focused clustering and outlier detection in large attributed graphs, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (SIGKDD 2014), New York, NY, USA, 2014, pp. 1346–1355. URL: https://doi.org/10.1145/2623330.2623682.

[31] J. Zhao, D. Xiao, L. Hu, C. Shi, Coupled semi-supervised clustering: Exploring attribute correlations in heterogeneous information networks, in: Proceedings of Web and Big Data - Third International Joint Conference (APWeb-WAIM 2019), Chengdu, China, August 1-3, 2019, volume 11641, 2019, pp. 95–109. URL: https://doi.org/10.1007/978-3-030-26072-9_7.

[32] S. Huang, Z. Xu, J. Lv, Adaptive local structure learning for document co-clustering, Knowledge-Based Systems 148 (2018) 74–84.

[33] I. S. Dhillon, Co-clustering documents and words using bipartite spectral graph partitioning, in: Proceedings of the 7th ACM SIGKDD international conference on Knowledge discovery and data mining (SIGKDD 2001), San Francisco, CA, USA, 2001, pp. 269–274. URL: http://portal.acm.org/citation.cfm?id=502512.502550.

[34] D. Vo, E. Bagheri, Feature-enriched matrix factorization for relation extraction, Information Processing & Management 56 (2019) 424–444.

[35] J. Yoo, S. Choi, Orthogonal nonnegative matrix tri-factorization for co-clustering: Multiplicative updates on stiefel manifolds, Information Processing & Management 46 (2010) 559–570.

[36] J. J. Whang, I. S. Dhillon, Non-exhaustive, overlapping co-clustering, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM 2017), Singapore, 2017, pp. 2367–2370. URL: https://doi.org/10.1145/3132847.3133078.

[37] X. Zhang, H. Li, W. Liang, J. Luo, Multi-type co-clustering of general heterogeneous information networks via nonnegative matrix tri-factorization, in: IEEE 16th International Conference on Data Mining, ICDM 2016, December 12-15, 2016, Barcelona, Spain, 2016, pp. 1353–1358. URL: https://doi.org/10.1109/ICDM.2016.0185. doi:10.1109/ICDM.2016.0185.

[38] S. Yao, G. Yu, J. Wang, C. Domeniconi, X. Zhang, Multi-view multiple clustering, in: Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI 2019), Macao, China, 2019, pp. 4121–4127. URL: https://doi.org/10.24963/ijcai.2019/572.

[39] V. W. Zheng, M. Sha, Y. Li, H. Yang, Y. Fang, Z. Zhang, K. Tan, K. C. Chang, Heterogeneous embedding propagation for large-scale e-commerce user alignment, in: Proceedings of IEEE International Conference on Data Mining (ICDM 2018), Singapore, 2018, pp. 1434–1439. URL: https://doi.org/10.1109/ICDM.2018.00198.

[40] Q. V. Le, T. Mikolov, Distributed representations of sentences and documents, in: Proceedings of the 31th International Conference on Machine Learning (ICML 2014), Beijing, China, 2014, pp. 1188–1196. URL: http://proceedings.mlr.press/v32/le14.html.

[41] T. Fu, W. Lee, Z. Lei, Hin2vec: Explore meta-paths in heterogeneous information networks for representation learning, in: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (CIKM 2017), Singapore, 2017, pp. 1797–1806. URL: https://doi.org/10.1145/3132847.3132953.

[42] A. Strehl, J. Ghosh, Cluster ensembles — A knowledge reuse framework

for combining multiple partitions, Journal of machine learning research 3 (2002) 583–617.

[43] Y. Zhao, G. Karypis, Empirical and theoretical comparisons of selected criterion functions for document clustering, Machine Learning 55 (2004) 311–331.