IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received November 27, 2020, accepted December 29, 2020, date of publication January 4, 2021, date of current version January 14, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3048932

Semantic-Aware Graph Convolutional Networks for Clinical Auxiliary Diagnosis and Treatment of Traditional Chinese Medicine

CHUNYANG RUAN[1], YINGPEI WU[2], YUN YANG[3], AND GUANGSHENG LUO[1], (Member, IEEE)

[1]Department of Data Science and Big Data Technology, School of Economics and Finance, Shanghai International Studies University, Shanghai 200083, China
[2]School of Software Engineering, Fudan University, Shanghai 200433, China
[3]Department of Oncology, Longhua Hospital Shanghai University of Traditional Chinese Medicine, Shanghai 200032, China

Corresponding author: Guangsheng Luo (luoguangsheng03@126.com)

This work was supported in part by the National Science Foundation of China under Grant 61672161, and in part by the Youth Foundation of Shanghai International Studies University under Grant 2020114096.

ABSTRACT Traditional Chinese Medicine (TCM) clinical informatization focuses on serving user-oriented health knowledge and facilitating online diagnosis. Regularities are hidden in clinical knowledge play a significant role in the improvement of the TCM informatization service. However, many regularities can hardly be discovered because of specific data-challenges in TCM prescriptions at present. Therefore, in this article, we propose an end-to-end model, called Semantic-aware Graph Convolutional Networks (SaGCN) model, to learn the latent regularities in three steps: (1) We first construct a heterogeneous graph based on prescriptions; (2) We stack Semantic-aware graph convolution to learn effective low-dimensional representations of nodes by meta-graphs and self-attention; (3) With the learned representations, we can detect regularities accurately by clustering and linked prediction. To the best of our knowledge, this is the first study to use metagraph and graph convolutional networks for modeling TCM clinical data and diagnosis prediction. Experimental results on three real datasets demonstrate SaGCN outperforms the state-of-the-art models for clinical auxiliary diagnosis and treatment.

INDEX TERMS Tranditional Chinese medicine, clinical knowledge discovery, metagraph, graph convolutional networks.

I. INTRODUCTION

Research on Traditional Chinese Medicine (TCM) clinical informatization has changed from providing literature resources to serving clinical auxiliary diagnosis and treatment. Promoting the research of TCM informatization and facilitating online diagnosis along with the development of data science are essential tasks of the clinical informatization of TCM. Prescriptions are important data accumulated over a long period of time in the clinical diagnosis and treatment of TCM. They contain considerable TCM knowledge and are the data basis of the TCM clinical informatization. In TCM, a clinical prescription is a group of herbs, symptoms, diseases, and other clinical entities, recording a personalized medical process for each patient. In clinical prescriptions, the essence of regularities is multiple relations among different data entities, such as symptoms, herbs, and diseases. Discovering regularities plays a significant role in improving TCM clinical diagnosis and treatment and the development of novel prescriptions [1].

Previous works proposed many machine learning-based methods that could discover regularities in TCM clinical prescriptions. They discover the latent relations among the herbs, symptoms, syndromes, and improve diagnosis to some extent [1]–[4]. However, the above methods failed to comprehensively explain how regularities are generated using multiple relations among different TCM entities or less consider domain knowledge of TCM well.

Actually, we should solve the below challenges to address shortcomings of the prior methods and support diagnosis decision-making. **(1)Random structure of data.** The real-world TCM clinical prescriptions are usually represented by natural languages and in free-text formats. To use entities in clinical prescription, we should first model them well from the text. **(2)Poor organization of data.** The prescriptions have their own way of organizing various clinical entities

The associate editor coordinating the review of this manuscript and approving it for publication was Giovanni Dimauro.

(e.g., herbs, symptoms, syndromes, and diseases), which are often put in a disordered way. For example, a herb in the front of the prescription may be correlated with the very last herb instead of its surroundings. Implementing feature engineering in a large number of semantic free-text prescriptions is a challenging and costly task. **(3)The sparsity of data.** The personalized views of TCM clinicians also influence datasets. For example, since TCM clinicians' prescribing habits often converge to their individual experiences, some herbs or symptoms would not be recorded in prescriptions. This may result in much difference and sparsity in recorded prescriptions of clinical data.

To tackle all the aforementioned challenges, we propose a graph convolutional networks based graph embedding model Semantic-aware Graph Convolutional Networks (SaGCN). In particular, we first construct a large graph from TCM clinical prescriptions, in which herbs, symptoms, syndromes, and diseases are defined as nodes. We then turn the massive free-text prescriptions analysis problem into a large graph analysis problem. To learn multiple relations in prescriptions, we define the TCM graph as a heterogeneous information network (HIN). HIN with various types of nodes and links has the superior ability in modeling heterogeneous data and learning the different semantics among nodes [5], and offers the advantage of straightforward handling of missing values. So, the TCM graph can be profoundly beneficial to better express the rich information of clinical entities. However, to analyze HIN is a complex problem. An effective approach to solve this problem is to utilize Graph Representation Learning (GRL) that uses deep learning and nonlinear dimensionality to encode graph structure into low-dimensional embeddings [6]. One of the main problems of GRL in HIN is semantic search while the central problem of GRL in the TCM graphs is how to incorporate TCM knowledge into the embedding model. Meta-graphs can capture various semantics among nodes on the graph [7]. We propose a meta-graph and attention mechanism-based approach to solve this problem. Next, we incorporate the semantics of the TCM graph into a graph convolutional networks (GCN) to obtain the node embedding of the TCM graph. After that, we optimize the overall model by using backpropagation and employ traditional machine learning algorithms to complete analysis for TCM clinical diagnoses supporting. We evaluate five state-of-the-art approaches and our proposed model *SaGCN* on three real-world TCM medical datasets for prediction and diagnostic tasks. The results show that our proposed model *SaGCN* outperforms other compared graph convolution-based models. Using 9000+ clinical lung tumor prescriptions, we also conduct two case studies, prescriptions prediction and disease prediction, to qualitatively reveal the robust power of *SaGCN* in capturing relation in TCM clinical data and reflect the clinical diagnostic patterns in TCM. To summarize, our main contributions are as follows:

- To the best of our knowledge, this is the first attempt to take advantage of HIN and GCN with self-attention for clinical auxiliary diagnosis and treatment task.
- We jointly model the clinical entities (herbs, symptoms, syndromes, and diseases) from clinical prescriptions as a large graph to provide effective and safe diagnosis prediction.
- We propose *SaGCN*, an accurate and robust learning model based on meta-graph and *semantic-aware convolution*-based GCN for TCM clinical prescriptions, which captures the multi-semantics and learns heterogeneous node embedding tailored for TCM diagnosis prediction tasks.
- We compare several state-of-the-art models on real TCM data qualitatively and quantitatively to demonstrate the effectiveness and robustness of *SaGCN*.

II. RELATED WORK

A. TCM DIAGNOSIS PREDICTION

Minning over medical, health, or clinical data is considered the most challenging domain for data mining [8]. With the rapid development of machine learning, a large amount of work has been focused on finding out various kinds of hidden knowledge relations such as symptom and symptom, symptom and syndrome, and syndrome and disease for improving the quality of clinical diagnosis and healthcare via text mining [9]. Chen *et al.* [2] presented a HIN-based soft clustering approach to discover the categories of formulas. Li *et al.* [3] utilized seq2seq model with coverage mechanism to generate TCM prescription. Yao *et al.* [1] developed a novel topic model to detect the relation between herbs and symptoms and characterized the generative process of prescriptions. Although these models are effective in TCM exploration, they are limited to the traditional data mining methods or the characteristics of data. Compared with all the aforementioned NLP-based predictive methods, the proposed framework *SaGCN* has the following advantages: (1) It leverages the powerful data representation advantages to overcome the drawback of TCM clinical prescriptions; (2) It captures the semantics of TCM clinical prescriptions without loss of generality and simplicity, which takes a good predictive performance.

B. HIN AND METAGRAPH

HIN has attracted much attention in the past decade because of its capability of representing the rich type information, as well as the accompanying wide applications such as personalized recommendation [10], clustering [11], and outlier detection [12]. Exploring semantics is the foundation step of all HIN-based tasks [13]. Although meta paths have been shown to be useful in different applications, they can only express simple relations between source and target entities [14]. Previous works [10], [11], [15], [16] focus on using meta-path [17] to preserve the semantics in HIN. Recently many works have been adopting meta-graph to preserve the semantics in HIN, which measures semantics better than meta-path. For example, Huang *et al.* [14] proposed meta structure, a directed acyclic graph of entity types with edge types connecting in between, to measure the proximity

between entities. Fan *et al.* [18] presented a meta-graph based embedding model to depict the relatedness over files. Inspired by these works, we utilize meta-graph to incorporate more rich semantics into our GCN model.

C. GRAPH CONVOLUTIONAL NETWORKS

GCN is an extension of convolutional neural network for processing the graph data, which has received growing attentions recently. GCN has been successfully used in many tasks [19]–[23], such as neural recommendation [19], event detection [20], machine translation [21] and healthcare [22], [23]. Focusing on healthcare, most existing works aim to learn relation in biomedical networks for prediction tasks, such as medicine interaction prediction. In these tasks, the high-level graph representations help the final predictions. GCN has revolutionized the field of graph representation learning through effectively learned node embeddings, and achieved state-of-the-art results for many tasks [24]. Fout *et al.* [23] stacked multiple layers of convolution and learned effective latent representations that integrate information across the graph to predict protein interface. Sankar *et al.* [25] presented a novel spatial convolution operation to capture the key properties of local connectivity and translation invariance, using high order connection patterns and attention mechanism.

Inspired by previous works [23], [25], we employ heterogeneous graph representation learning based on meta-graph and GCN alongside with the joint learning framework to learn TCM clinical entities representations and their relations.

III. METHODOLOGY

A. PROBLEM FORMULATION

Definition 1 (TCM Clinical Prescriptions and Diagnosis Prediction): In TCM clinical settings, TCM doctors first record symptoms that they observe in their patients. Furthermore, they aim to determine the patient's syndrome according to the patient's symptoms. The doctors then prescribe herbs combination based on the patient's disease profile. These herbs, symptoms, syndromes, and diseases are captured and described as information entities such as "red ginseng" and "apricot kernel" in clinical prescriptions. Each full TCM clinical prescription of each patient can be represented as a word set of multivariate observations: $R^{(k)} = h^{(k)}, s^{(k)}, \hat{s}^{(k)}, d^{(k)}, k \in 1, 2, \cdots, K$ where $h^{(k)}$ is the herbs combination, $s^{(k)}$ is the group of symptom, $\hat{s}^{(k)}$ is a syndrome, $d^{(k)}$ is a disease and K is the total number of prescriptions. Prescribing herbs based on symptoms and discovering new herbs for disease(i.e. drug repositioning) are valuable diagnosis prediction. These prediction can be formulated as: $s \xrightarrow{f} d$, and $s \xrightarrow{f} h$ where f is a mapping function.

Definition 2 (TCM Graph): TCM prescriptions are modeled as a heterogeneous graph $G = (V, E)$ where V and E refer to the set of nodes and links respectively. Each node $v \in V$ is mapped to a specifc clinical entity O (e.g. herb, symptom, syndrome or disease) by an entity type mapping function $\phi : V \mapsto O$. And, each link $e = (v_i, v_j) \in E$ is mapped to type L by link type mapping function $\varphi : E \mapsto L$ where the two nodes v_i, v_j belong to two different types. Given TCM graph, its schema T_G is a directed graph defined over entity types O and link types L, i.e. $T_G = (O, L)$. The schema expresses all allowable link types between entity types [14]. Fig. 1 shows the abstract schema of the network illustrating node types and basic links.

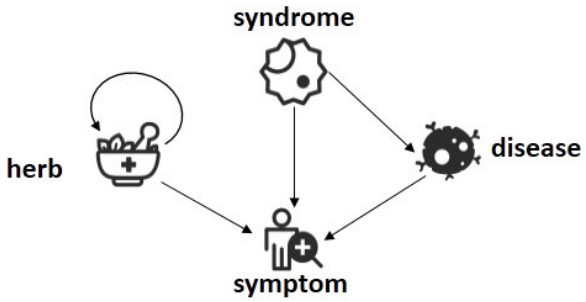


FIGURE 1. Structures of TCM graph. Different icons and line styles display different types of TCM nodes and links, respectively, which are closely correlated with one another.

Definition 3 (TCM Metagraph): A TCM metagraph is defined as $g = (\acute{V}, \acute{E}, v_s, v_t)$ defined on the given TCM graph schema $T_G = (O, L)$. g has only a single source node v_s and single target node v_t. $\acute{V}$ is a set of nodes and $\acute{E}$ is a set of links satisfying two constraints: (1) $\forall v_i \in \acute{V}, \exists \phi(v_i) \in O$; (2) $\forall (v_i, v_j) \in \acute{E}, \exists \varphi(\phi(v_i), \phi(v_j)) \in L$. Fig. 2(a) shows the abstract schema of the network illustrating node types and basic links.

Definition 4 (Instance of TCM Metagraph): Given TCM graph and metgraph, an instance of metagraph with target is a subgraph of G, denoted by $g_v = (\acute{V}_g, \acute{E}_g)$ such that there exists a mapping for g_v, $\psi : \acute{V}_g \mapsto \acute{V}$ satisfying two constraints: (1) $\forall v \in \acute{V}, \exists \phi(v) = \psi(v)$; (2) $\forall v, u \in \acute{V}, (v, u) \in (\notin)\acute{E}, \exists (\nexists)(\psi(v), \psi(u)) \in (\notin)\acute{E}$.

B. SEMANTIC-AWARE

TCM clinical prescriptions contain rich TCM domain knowledge. For example, the jun (emperor) herbs treat the main cause or primary symptoms of a disease, and the zuo (assistant) herbs are used to improve the effects of jun and chen, and to counteract the toxic or side effects of these herbs [1]. In Fig. 2(b), a disease D_1 connects two herbs H_1 and H_4 via the same meta-graph Ms_3. H_1 and H_4 may play a different role for the disease. How to distinguish the semantics of two nodes in this meta-graph comprehensively? In this article, we model this issue as a weight measure. We employ Point-wise Mutual Information (PMI) and degree to develop a structure-aware approach.

To utilize the global characteristics of the relation between two nodes, we use PMI with a constant size sliding window on all documents in the prescriptions to gather co-occurrence statistics of nodes. The weight of the link between node i and node j can be defined as follows:

$$\rho_{ij} = \log \frac{\sharp Count * Count(i, j)}{Count(i) * Count(j)} \tag{1}$$

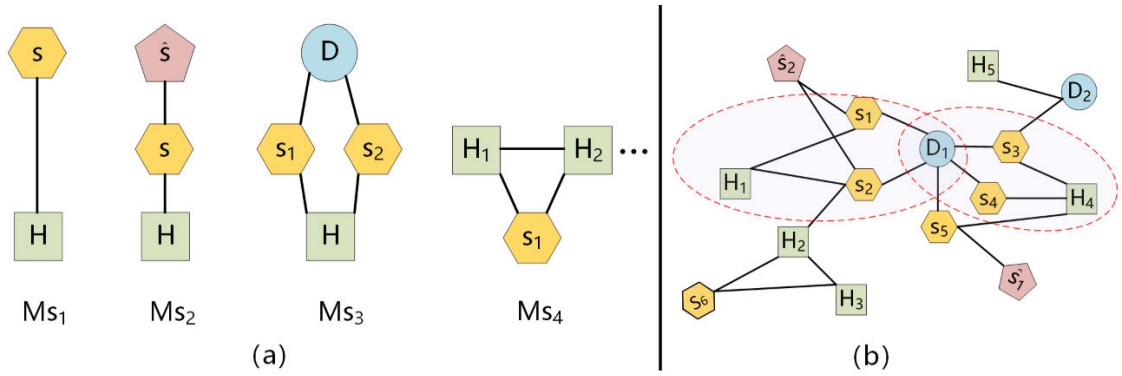


FIGURE 2. (a) Sample metagraph in TCM graph. Node Types: Herb (H), Symptom (S), Syndrome ($\hat{S}$) and Disease (D); (b) Example subgraph with instances of Ms_3 for target disease D.

where $Count(i)$ and $Count(j)$ are the number of sliding windows in all prescriptions that contain entity i and j respectively, $Count(i, j)$ is the number of sliding windows that contain both entity i and entity j, and $\sharp Count$ is the total number of sliding windows in all prescriptions. A high ρ value implies a high semantic correlation of node in the TCM graph, while a low ρ value indicates little or no semantic correlation in the graph. Connecting to many nodes, however, a node (e.g., zuo herb) just plays a minor role in the TCM graph. The degree of nodes can well reflect the structures of graph [26]. So, to solve this problem, we define a degree based approach to optimize (1) as follows:

$$\overline{\rho}_{ij} = \frac{\overline{\rho}_{ij}}{\sum_{k=1}^{|V|} \overline{\rho}_{ik} \sum_{k=1}^{|V|} \overline{\rho}_{jk}} \tag{2}$$

where $|V|$ is number of nodes in TCM graph, $\overline{\rho}_{ik}$ and $\overline{\rho}_{jk}$ are the the values of PMI between node k and i, j respectively. Next, we explore the semantic characteristics of a meta-graph's instance. Different walk paths of a meta-graph encode different semantics for TCM knowledge. In Fig. 2(b), a disease S_2 connects two herbs H_1 and H_2 via the same meta-graph Ms_1, the H_1 also connects with S_1 so that it may be a zuo (assistant) herb. So, the two herb nodes H_1 and H_2 have different relevance with symptom node S_2. Formally, because the attention mechanism aims to focus on the most pertinent information, and if an instance of meta-graph m_s with source nodes v_s and target node v_t is given, we define the self-attention-based approach to preserve different semantics between walk paths in m_s as follows:

$$\begin{aligned} a_l &= softmax(s_l^{(k)}(v_s, v_t) \cdot w_l \cdot \frac{1}{L} \sum_{i=1}^{L} s_i^{(k)}(v_s, v_t)), \\ s_l^{(k)}(v_s, v_t) &= s(v_s, v_t|[1:|l|]) = \frac{\rho_{sm}}{\sharp \rho_s} \times s(v_m, v_t|[2:|l|]) \end{aligned} \tag{3}$$

where w_l is a parameter mapping between the context semantics of all walk path L and each semantics $s_l^{(k)}(v_s, v_t)$ in m_s, which is learned as part of the training process. $|l|$ is a length of a walk path of l, L is number of walk path in m_s. ρ_{sm} is the weight of link between node v_s and node v_m, and $\sharp \rho_s$ is the sum of weights among v_s and its neighboring nodes. The value of $\rho_{sm}/\sharp \rho_s$ is the transition probability between node v_s and node v_m.

C. SEMANTIC MATRIX

A semantic matrix $\mathbf{S}^{(k)}$ is a similarity matrix to encode the relevance of nodes in each unique semantic role over all instances of Ms_k in the graph G. $\mathbf{S}_{ij}^{(k)}$ is the transition probability between souce node v_i and target node v_j in an instance of the Ms_k. Formally, $\mathbf{s}_{ij}^{(k)}$ can be formulated as an iterative function:

$$s_{ij}^{(k)} = s^{(k)}(v_i, v_j|g_{v_i \to v_j}) \tag{4}$$

where g is an instance of Ms_k. $s^{(k)}(v_i, v_j|g_{v_i \to v_j})$ can be computed by the aggregation of all value of $s_l^{(k)}(v_s, v_t)$ in (3). For complex meta-graphs, computing $\mathbf{S}_{ij}^{(k)}$ is very complicated because of the various ways to pass through the meta-graph [27]. For Ms_1 in Fig. 2(a), there is only one path to pass through Ms_1, and the transition probability between source node and target node can be calculated by Eq.3. For Ms_3 in Fig. 2(a), however, there are two ways to pass through the meta-graph, which are $H - S_1 - D$ and $H - S_2 - D$. Note that S_1 and S_2 represent the different entity type symptoms in the TCM graph. In Ms_3, the path $H - S_1 - D$ means that herb can cure a symptom of a disease, so that herb and disease have some similarities. Similarly, in the path $H - S_2 - D$, herb and disease have some similarities as well. Therefore, we should define the logic of similarity and semantics when there are multiple ways passing from the source node H to the target node D in the meta-graph. Inspired by [27], we propose a approach to obtain matrix $\mathbf{S}^{(k)}$. Algorithm(1) depicts the example of the similarity-based semantic matrix operations for Ms_3 where $\odot$ is the Hadamard product, and the elements of matrix T is the transition probability between source node and target node in the current path, which is calculated by Eq.4, and $\mathbf{W}$ is the weight matrix consisting of the elements a_l in Eq.3. Note that this algorithm is not limited to Ms_3. Any meta-graph with complex paths can be computed by Hadamard product and multiplication on the corresponding matrixes. We then can get different semantics between source node and target node by computing semantic matrix for all

Algorithm 1 Metagraph Based Computing Semantic for Ms_3

Input: TCM graph $G = (V, E)$
Output: $\mathbf{T}_{HD}$
$\mathbf{T}_{HD_1}$ matrix
computation:$\mathbf{T}_{HD_1} \leftarrow \mathbf{T}_{HS_1} \times \mathbf{T}_{S_1D} \times \mathbf{W}_1$;
$\mathbf{T}_{HD_2}$ matrix
computation:$\mathbf{T}_{HD_2} \leftarrow \mathbf{T}_{HS_2} \times \mathbf{T}_{S_2D} \times \mathbf{W}_2$;
$\mathbf{T}_{HD}$ matrix computation: $\mathbf{T}_{HD} \leftarrow \mathbf{T}_{HD_1} \times \mathbf{T}_{HD_2}$;

meta-graphs denoted by $\{Ms_1, Ms_2, Ms_3, \cdots, Ms_k\}$ in TCM graph.

D. SEMANTIC-BASED CONVOLUTION

In this section, we consider learning representation for a specific node type via spatial convolution operation, which preserves the spatial locality and precise semantics.

Specifically, we provide a meta-graph Ms with target node $v_i \in V$ as input. Convolution, intrinsically, is an aggregation operation between local inputs and filters [28]. In this article, filters should be in a position to aggregate local inputs with diverse topological structures and semantics. A semantic filter for Ms is defined by a weight matrix $W^{(t)}$ for target node v_i and a weight matrix $W^{(N)}$ for the sources of target node in Ms, and each weight in $W^{(N)}$ differentiates the semantic roles of source nodes in the receptive field. Our objective is to design a semantic convolution kernel that can be applied to heterogeneous graphs with spatial locality and rich semantics. To summarize, we would like to learn a mapping function at each node in the graph, which has the form:

$$y_i = \sigma_W(x_i, \{x_{s_1}, x_{s_2}, \cdots, x_{s_k}\}) \tag{5}$$

where $\{x_{s_1}, x_{s_2}, \cdots, x_{s_k}\}$ are the source nodes of node v_i that defines the receptive field of the convolution; σ is a non-linear activation function, and W is the filter as learned parameters of the function. For the target node v_i, its spatial locality and semantics are captured by relation structure-aware. Accordingly, we define semantic convolution as follows:

$$y_i = \sigma(W^{(t)}x_i + \frac{1}{|N_k|}\sum_{v_j \in N_k} W^{(N)}S_{ij}x_j + b) \tag{6}$$

where N_k is the set of source nodes of node v_i, and b is is a vector of biases.

E. COMBINING MULTI-SEMANTIC

Given multiple semantics with certain characteristic meta-graphs, a good proximity measure must account for different semantics. Since different semantics can vary in their importance for the graph representation, we face the challenge of appropriately weighting the extracted proximity at the end of each semantics for effective feature propagation. To tackle this challenge, we leverage attention mechanism to help learn stable and robust node embedding of graph. Inspired by the recent progress of the self-attention for machine translation [29], we propose a semantic-attention model to weight the importance of different semantics for each node dynamically.

$$z_i = \sum_{k=1}^{K} a_{(k,i)} \cdot y_i^{(k)} \tag{7}$$

where z_i is defined as the weighted summation of every semantic vectors $y_i^{(k)}$ for node v_i, $k = 1, 2, \cdots, K$, corresponding to the semantic index for each node. For each semantic vector $y_i^{(k)}$ of node v_i, we compute a positive weight $a_{(k,i)}$ which can be interpreted as the probability that $y_i^{(k)}$ is assigned by node v_i. We define the weight of semantics k for the node v_i using a softmax function as follows:

$$\begin{aligned} a_{(k,i)} &= \frac{\exp(\lambda_{(k,i)})}{\sum_{k=1}^{K}\exp(\lambda_{(k,i)})}, \\ \lambda_{(k,i)} &= {y_i^{(k)}}^{\top} \cdot H \cdot y_{(s,i)}, \\ y_{(s,i)} &= \frac{1}{K}\sum_{k=1}^{K} y_i^{(k)} \end{aligned} \tag{8}$$

where $y_{(s,i)}$ is the average of different semantic vectors, which can capture the global context of the semantic information. H is a vector mapping between the global context embedding $y_{(s,i)}$ and each semantic vector $y_i^{(k)}$, which is learned as part of the training process. By introducing an attentive vector $y_i^{(k)}$, we compute the relevance of each semantic vector to the node v_i. If $y_i^{(k)}$ and $y_{(s,i)}$ have a large dot product, this node believes that semantic k is an informative semantics. For example, the weight of semantic k for this node will be largely based on the definition.

Once we obtain the weighted node vector representation z_i, an objective function is used to learn a low-dimensional node embedding. We use negative samples based hinge loss to minimize the reconstruction error:

$$J(\theta) = \sum_{j=1}^{J} \max(0, 1 - z_i x_i + z_i n_j) \tag{9}$$

where n_j is the negative node, which has no connection with the target node in any meta-graph. J is the times of node sampling, and θ is the set of parameters to be solved. In addition, we adopt the asynchronous stochastic gradient descent (ASGD) and the backpropagation algorithm to optimize the objective function. We also use dropout to prevent over-fitting. A multi-semantic convolutional layer is illustrated in Fig. 3.

F. DIAGNOSIS PREDICTION

We obtain node embedding (e.g., herb, symptom, and disease) using model *SaGCN*. These node embeddings can be further extended to diagnosis tasks according to specific data and problems. For example, given a set of herbs, the incompatible herb pair can be summarized in a latent embedding space by similarity calculation:

$$f(h1, h2) = \sqrt{\sum_{i=1}^{d}(h1_i - h2_i)^2} \tag{10}$$

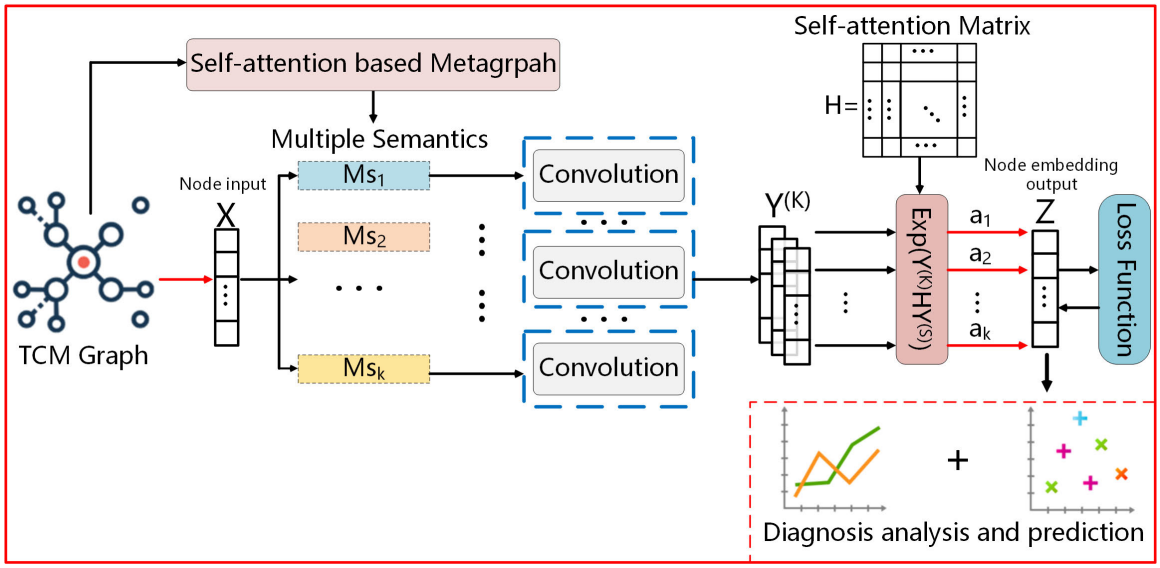


FIGURE 3. An overview of the proposed SaGCN model.

where $h1, h2$ are the different herbs, and d is the dimensionality of herb in embedding space. To optimize the aforementioned model, we set the goal is to minimize J as a function of θ. In addition, we use the cross-entropy between the ground truth visit $\bar{f}$ and the predicted visit f to calculate the loss for each prescription from all the herb-pair as follows:

$$L(h1, h2, \cdots, hm; f_1, f_2, \cdots, f_m) = -\frac{1}{m-1}\sum_{i=1}^{m-1}(\bar{f}_i \log(f_i) + (1-\bar{f}_i)\log(1-f_i)) \quad (11)$$

IV. EXPERIMENTS

In this section, we evaluate *SaGCN* on three real-world TCM datasets. First, it is evaluated under node clustering and link prediction tasks. Then its TCM diagnosis prediction performance is compared to various state-of-the-art TCM predictive models. Finally, it is qualitatively evaluated through a case study.

A. DATASET AND CONFIGURATIONS

We use three TCM datasets in our evaluation.

- **TCMRel [30]:** This is a candidate relation graph composed of four node types: herbs (H), formula (F), syndrome ($\hat{S}$) and disease (D), connected by five link types: F-D, F-$\hat{S}$, H-D, H-$\hat{S}$, D-$\hat{S}$. We use a subset with H, D, $\hat{S}$, and their correlations.
- **CMD**[1]**:** We use chinese medical information to create a graph with four node types: herbs (H), symptom (S), syndrome ($\hat{S}$), disease (D) linked by six fundamental types: H-D, H-S, H-$\hat{S}$, D-S, D-$\hat{S}$, S-$\hat{S}$.
- **CLLT:** We construct the clinical graph with three node types: herb (H), symptom (S), disease D) and their correlations from 7,000 clinical prescriptions for lung tumor.

[1]http://cmekg.pcl.ac.cn/

To validate the predictive performance of the proposed *SaGCN*, we compare it with the following state-of-the-art approaches:

- **ASPEm** [15] is a node embedding learning for HIN, which observes multiple aspects existing in HIN and extends the skip-gram model to obtain the graph representation.
- **GWCN** [31] is a recently proposed spectral GCN, which leverages wavelet transform to implement efficient convolution on graph data.
- **MotifCNN** [25] is a novel spatial convolution operation to model the key properties of heterogeneous local connectivity and translation invariance, using high-order connection patterns.
- **HeteroMed** [5] is capable of capturing informative relation for the diagnosis goal and uses the best relation sampling strategy for learning clinical event representations for EHR data.
- **TM2P** [1] is a novel prescription topic model incorporated TCM knowledge to discover regularities on the herbs composition and corresponding symptoms.

B. SIMILAR DISCOVERY AND VISUALIZATION

We conduct a series of comparative experiments of node clustering to simulate the clinical analysis and detect the combination of herbs curing a syndrome. Getting the node embedding, we select a set of syndrome ($\hat{S}$) nodes, which are assigned more labels from a TCM doctor selected set and we use their representations as feature vectors to learn and test a clustering algorithm Density-Based Spatial Clustering of Applications with Noise (DBSCAN). We use *accuracy* and *normalized mutual information (NMI)* as metrics for evaluation. *NMI* is often employed to determine the gap between the results of division and the true partition.

In fact, this clustering step can select groups without label according to specific data. For clinical diagnostic prediction

TABLE 1. Performance Evaluation of Similar Discovery.

Dataset	Model	Metric					
		Accuracy@herb	NMI@herb	Accuracy@symptom	NMI@symptom	Accuracy@disease	NMI@disease
TCMRel	ASPEm	0.7012	0.4811	0.7123	0.5312	0.6879	0.5034
	GWCN	0.6431	0.4580	0.6620	0.4651	0.5431	0.4678
	MotifCNN	0.7310	0.5132	0.7520	0.5531	0.6977	0.5178
	TM2P	0.5010	0.3210	0.5390	0.4160	0.4960	0.2901
	HeteroMed	0.7296	0.5015	0.7351	0.5394	0.6901	0.5095
	SaGCN	**0.7452**	**0.5270**	**0.7564**	**0.5582**	**0.7025**	**0.5306**
CMD	ASPEm	0.7231	0.4897	0.6890	0.4791	0.7115	0.5097
	GWCN	0.6701	0.4618	0.5907	0.4983	0.6702	0.4134
	MotifCNN	0.7353	0.5221	0.7208	0.5302	0.7256	0.5458
	TM2P	0.5720	0.3790	0.5640	0.3710	0.5271	0.4021
	HeteroMed	0.7300	0.4981	0.7115	0.4804	0.7193	0.5058
	SaGCN	**0.7587**	**0.5412**	**0.7491**	**0.5533**	**0.732**	**0.5563**
CLLT	ASPEm	0.6900	0.4617	0.6865	0.4601	0.7840	0.5131
	GWCN	0.7012	0.5019	0.6123	0.4765	0.7650	0.4872
	MotifCNN	0.7141	0.5208	0.7085	0.5012	0.7901	0.5233
	TM2P	0.5230	0.3410	0.5180	0.3351	0.5334	0.3908
	HeteroMed	0.6756	0.4850	0.6952	0.4634	0.7853	0.5121
	SaGCN	**0.7218**	**0.5395**	**0.7161**	**0.5142**	**0.7943**	**0.5304**

task, we obtain groups as candidate sets in order to reduce the number of node relation to be predicted.

Table 1 shows the performance of all the approaches on all the three real-world TCM datasets. We can observe that our proposed approache *SaGCN* achieves the best performance compared with all the baselines in terms of the values of all the measures. On all datasets, the overall performance of traditional topic model based on approache TM2P is worse than that of the deep learning based approaches, and SaGCN obtains the highest score among all baselines with respect to *accuracy* and *NMI*. For deep learning baselines, ASPEm, MotifCNN and HeteroMed achieved higher performance than GWCN on datasets TCMRel and CMD because GWCN can not capture rich semantic proximity in large graph with various nodes. But on dataset CLLT with less node types, GWCN performanced better compared with ASPEm and HeteroMed because its spectral convolution can capture the structure of graph effectively. As for the representation learning model for heterogeneous graph, the performance of MotifCNN is close to that of the winner *NMI* on all datasets, while ASPEm and HeteroMed perform similarly. To intuitively illustrate the significance of node clustering for clinical diagnostic prediction, we give an illustrative visualization of the herbs clustering. We use clustering to predicte relation among herbs based on learned node embeding. We randomly selected herbs, and use k-means to obtain the herb-clusters. We used *FVIC* [32] to test results. In addition, we asked two TCM doctors to verify the clusters, they confirmed that the herbs in the same cluster had the same function or belonged to the frequently occurring herb-pairs, which basically conformed to the rule of clinical medication. Fig 4 shows the visualization results. We can see that our model SaGCN distinguishes different herbs much better than another node embedding models, which also shows the power of SaGCN on the task of unsupervised learning.

C. RELEVENT SYMPTOM-HERB RELATION DISCOVERY

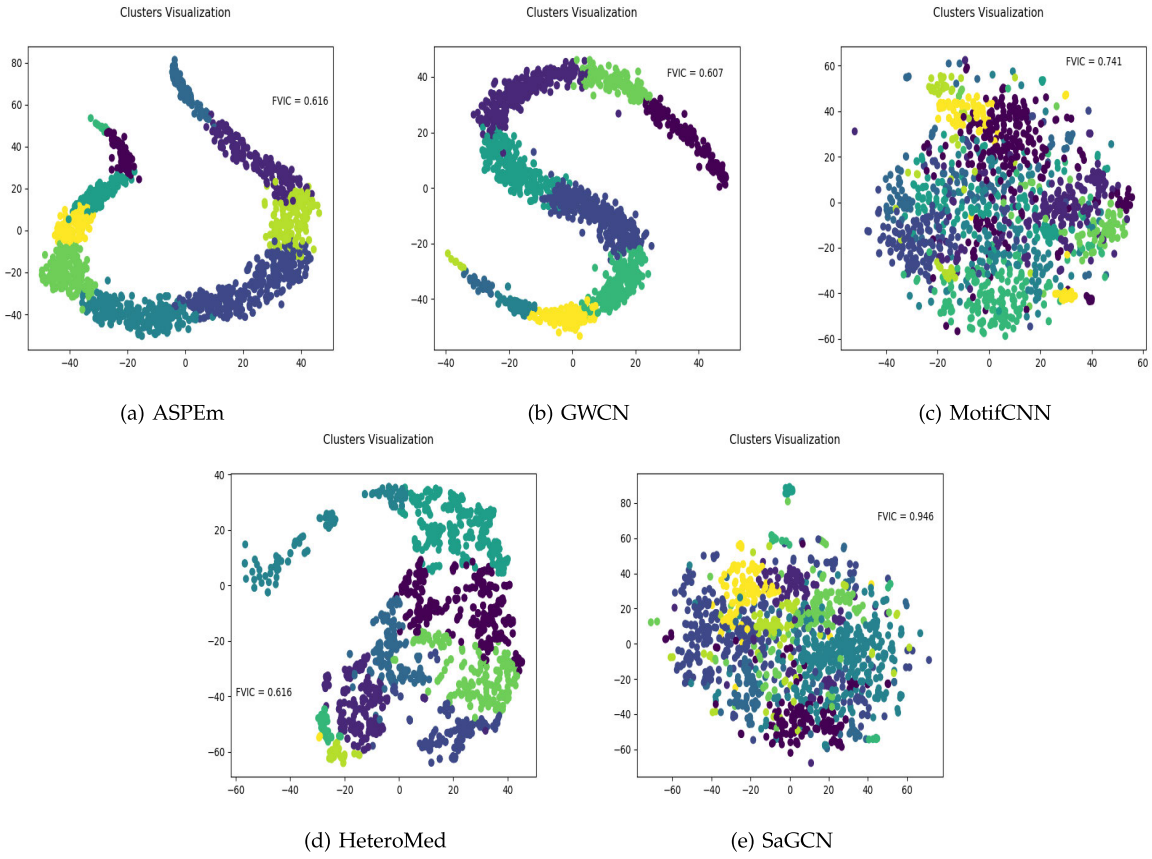
In this article, we model the relation discovery problem as a link prediction that aims to rank node pairs in terms of their relevancy, which may lead to a potential linkage between them. Using the symptom and herb distributions of different syndromes of diseases from candidate sets, we can further derive symptom-herb relation. Specifically, given a TCM graph, we first generate a subgraph by selecting a herb-symptom class and randomly remove a certain fraction (30% in our experiments) of links of the selected class as missing links. Since the logistic regression converges faster, we use it to predict the missing edges between herb–symptoms pairs as testing instances. We then use *Mean Average Precision (MAP)* as metric for evaluation.

From the main results on link prediction presented in Table 2, we have observed consistent with the clustering tasks that all the graph embedding methods perform better performance than the topic model based TM2P without considering type information. For graph embedding methods, SaGCN outperforms all other models. MotifCNN and GWCN achieve better results than ASPEm and HeteroMed. GWCN has superior performance than ASPEm and HeteroMed, which preserves the important structural information in the graph with fewer node types. Overall, we can see that SaGCN outperforms all the other methods in the task of link prediction.

In addition, We show the herbs with the highest relation with two test symptoms respectively in Table 3. Each rank is the conditional probability of herb given a test symptom. We can see that our model not only discovered herbs for highly frequent symptoms like ''bitter figwort'' for symptom ''sore throat'' but also found important infrequent herbs for this symptoms like ''oroxylum indicum''. In fact, ''oroxylum indicum'' can relieve asthma which has been verified in TCM literature ''Compendium of Materia Medica''.

D. CASE STUDY FOR DIAGNOSIS PREDICTION

We choose clinical prescriptions from test datasets based on the consideration of demonstrating the model effect on more challenging cases: there are complex herbs composition and corresponding symptoms. To examine the potential of this direction, we use 75% prescriptions as training data and the



(a) ASPEm (b) GWCN (c) MotifCNN

(d) HeteroMed (e) SaGCN

FIGURE 4. Comparison of herb groups visualization using unsupervised node embeddings on TCMRel.

TABLE 2. Performance Evaluation of herb-symptom link prediction.

Dataset	Model	Metric		
		MAP@3	MAP@5	MAP@10
TCMRel	ASPEm	0.7642	0.7830	0.7940
	GWCN	0.8050	0.8242	0.8501
	MotifCNN	0.8076	0.8309	0.8542
	TM2P	0.3101	0.3261	0.3304
	HeteroMed	0.7825	0.8023	0.8140
	SaGCN	**0.8213**	**0.8507**	**0.8745**
CMD	ASPEm	0.8091	0.8213	0.8279
	GWCN	0.8143	0.8350	0.8315
	MotifCNN	0.8176	0.8472	0.8342
	TM2P	0.3157	0.3210	0.3467
	HeteroMed	0.8132	0.8401	0.8312
	SaGCN	**0.8239**	**0.8501**	**0.8398**
CLLT	ASPEm	0.8176	0.8203	0.8219
	GWCN	0.8459	0.8432	0.8354
	MotifCNN	0.8452	0.8518	0.8406
	TM2P	0.3190	0.3341	0.3619
	HeteroMed	0.8234	0.8601	0.8517
	SaGCN	**0.8524**	**0.8603**	**0.8577**

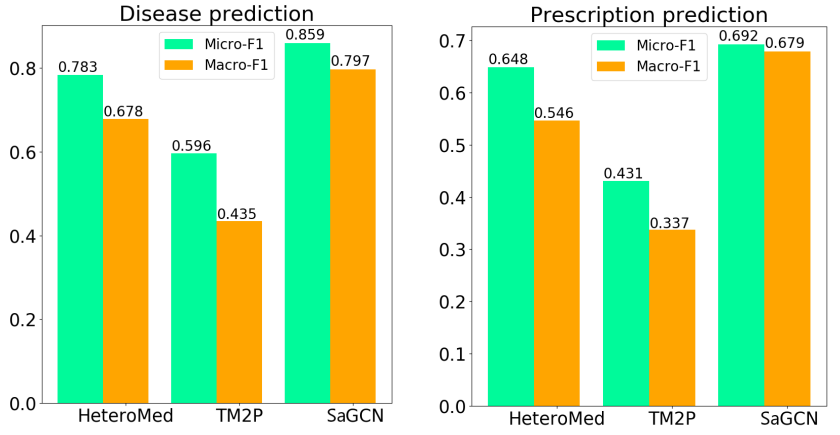
remaining 25% as test data. We compare the results of disease and treatment prediction to the state-of-the-art predictive models. We employ three standard metrics commonly used to evaluate this prediction task: F_1 score.

We compute the group of symptoms given a disease via soft clustering. We then use the top N symptoms with the largest similarity as the recommended symptoms. Fig. 5(a) gives the performance of each model. Our model SaGCN achieves

TABLE 3. The example of Top-5 herbs given a symptom.

Symptom	Herb	Rank
sore throat	bitter figwort	0.1823
	oroxylum indicum	0.1024
	caulis trachelospermi	0.0594
	folium isatidis	0.0568
	lonicera confusa	0.0546
anorexia	hippophae rhamnoides	0.4398
	poria cocos	0.3474
	ginseng	0.0712
	dioscorea opposita	0.0712
	ziziphus jujubae	0.0700

a significant increase generally. Numerically, our model SaGCN achieves macro-F_1 and micro-of F_1 of respectively, which is much higher than that of HeteroMed and TM2P. Because they require the number of contexts, HeteroMed and TM2P are limited to the missing value. Our model SaGCN is not hampered by this restriction. In addition to supporting treatment, our model can also recommend prescriptions as references for doctors. We obtain a strong relation among herbs and symptoms for a specific disease using link prediction. From Fig. 5(b), we can see that our model SaGCN achieves a higher macro-F_1 than HeteroMed and TM2P. This suggests that SaGCN can recover the herbs that the doctor actually prescribed to a disease, while also predict many herbs that were not prescribed before. HeteroMed achieves a relatively low F_1, and TM2P performances much worse.



(a) Performance of disease prediction (b) Performance of prescription prediction

FIGURE 5. Comparison among our model SaGCN, HeteroMed and TM2P in disease and treatment prediction.

TABLE 4. An example of the comparison between a diagnosis prediction from our model MsGCN and that from TCM doctor.

Symptoms	cough(咳嗽), stringy pulse(弦脉), yellow greasy coating(舌苔黄腻), sticky phlegmlight red tongue(痰黏舌红), hard to expectorate phlegm(痰多难咳出), fever(发热).	
	Disease	**Herbs**
Prediction from SaGCN	**phlegm-dampness due to spleen deficiency (脾虚痰湿)**	**houttuynia cordata(鱼腥草), scutellaria baicalensis(黄芩), hedyotis diffusa(白花蛇舌草), salvia chinensis(华鼠尾草), artemisia annua(青蒿素), bupleurum chinense(北柴胡),** asparagus cochinchinensis(天门冬), glehnia littoralis(北沙参), **dioscorea opposita(山药), fructus ligustri lucidi(女贞子), ophiopogon japonicus(麦冬), herba epimedii(淫羊藿), pericarpium citrus reticulata(陈皮), cuscuta chinensis(菟丝子), pinellia ternata(半夏),** caulis bambusae in taeniam(竹茹), et corneum gigeriae galli(鸡内金), poria cocos(茯苓), gynostemma pentaphyllum(绞股蓝), **selaginella doederleinii(深绿卷柏), platycodon grandiflorum(桔梗),** massa medicata fermentata(神曲).
Prediction from doctor	**phlegm-dampness due to spleen deficiency (脾虚痰湿)**	**houttuynia cordata(鱼腥草), scutellaria baicalensis(黄芩), hedyotis diffusa(白花蛇舌草),** aster tataricus(紫菀), **salvia chinensis(华鼠尾草), artemisia annua(青蒿素), platycodon grandiflorum(桔梗), selaginella doederleinii(深绿卷柏),** herba inulae(金沸草), **ophiopogon japonicus(麦冬), pericarpium citrus reticulata(陈皮), pinellia ternata(半夏), herba epimedii(淫羊藿), dioscorea opposita(山药),** astragalus membranaceus(黄芪), **fructus ligustri lucidi(女贞子), cuscuta chinensis(菟丝子), bupleurum chinense(北柴胡).**
Precision=0.682	Recall=0.833	F1=0.75

To examine this in detail, we compare the results predicted by using our model SaGCN with the herbs prescribed by the doctors diagnosing the disease, and the comparing results are shown in Table 4. Given test symptoms, we can see that of the eighteen prescription herbs prescribed by doctors, and our model prescribed fifteen identical prescription herbs. Our model also recommended seven herbs not prescribed by the doctor. A doctor verified that these herbs are all known to be associated with spleen deficiency. For example, ''caulis bambusae in taeniam'' can replace ''aster tataricus'' to cure the symptom ''cough''. At the bottom of the table, we give the results of the quantitative validation for match degree between the herbs given by the doctor and that generated by SaGCN.

E. COMPUTATIONAL EFFICIENCY

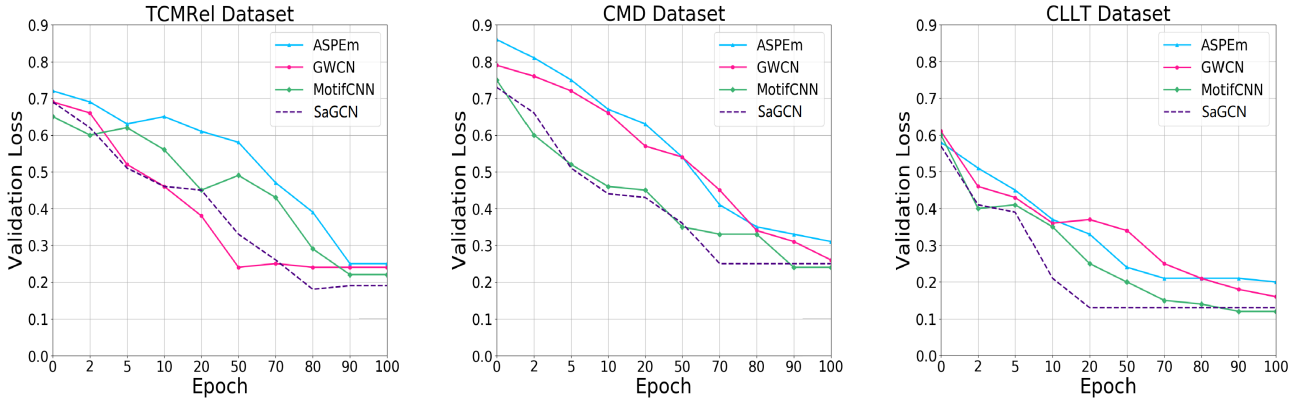
We report running times on an Inter(R) Core(TM) i7-7700HQ CPU @2.80GHz with 8 cores and 64GB memory.

1) STABILITY

We compare the convergence rates of different graph representation models by depicting the validation set loss in Fig 6. Overall, SaGCN achieves faster convergence and lower error in comparison to other models on all datasets. We think the reason is that the SaGCN rationally combines more semantics from all types of entities, which helps it to achieve steady performance.

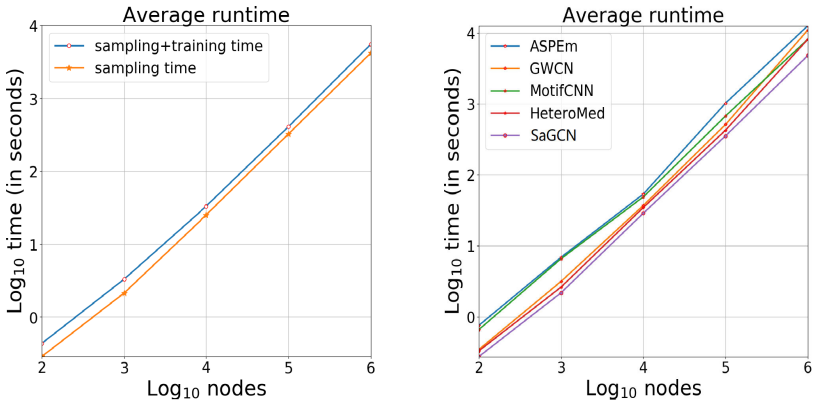
2) SCALABILITY

In order to illustrate its scalability, we apply SaGCN to learn node representation on TCM datasets. We compute the average runtime with increasing sizes from 100 to 1,000,000 nodes and average degree of 10. In Fig 6(a) we empirically observe that SaGCN scales linearly with increase in number of nodes generating representations for one million nodes in less than three hours. In order to speed up training the deep model, we use GCN with negative sampling. The sampling procedure comprises of preprocessing for computing transition probabilities for our semantic. The optimization phase is made efficient using negative sampling. In addition, we recorded the average runtime of each heterogeneous graph representation model along with the increasing nodes on TCMRel. Fig 6(b) shows that SaGCN achieved lower average runtime in comparison to other heterogeneous graph



(a) Loss (b) Loss (c) Loss

FIGURE 6. Comparison of validation loss w.r.t. epochs for all graph representation model.



(a) Scalability test (b) Scalability comparison

FIGURE 7. (a) Scalability test of MsGCN on TCM graph with an average degree of 10 and (b) scalability comparison of heterogeneous graph representation model.

representation models since it leverages relevant semantics and negative sampling simultaneously.

V. CONCLUSION

In this article, we study an approach to discover regularities in prescriptions, and we propose a model of clinical entities of prescriptions as a heterogeneous TCM graph to address shortcomings of previous methods pursuing the same goals. Using meta-graph and self-attention, our proposed model *SaGCN* is capable of capturing semantics for HIN. *SaGCN* effectively fuses semantics from multiple meta-graphs to learn clinical entities embedding through novel GCN. Experimental results show that *SaGCN* can achieve significantly better performance in diagnosis task and prove the effectiveness and robustness of *SaGCN*. The model is helpful for clinical research and practice. Future work includes incorporating more diverse types of clinical information such as herbal dosage, and more domain knowledge such as syndrome category as prior knowledge into our model. Evaluating herb roles inferred by our model is another interesting task we are going to investigate.

REFERENCES

[1] L. Yao, Y. Zhang, B. Wei, W. Zhang, and Z. Jin, "A topic modeling approach for traditional Chinese medicine prescriptions," *IEEE Trans. Knowl. Data Eng.*, vol. 30, no. 6, pp. 1007–1021, Jun. 2018.

[2] X. Chen, C. Ruan, Y. Zhang, and H. Chen, "Heterogeneous information network based clustering for categorizations of traditional Chinese medicine formula," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 839–846.

[3] W. Li, Z. Yang, and X. Sun, "Exploration on generating traditional chinese medicine prescription from symptoms with an end-to-end method," 2018, *arXiv:1801.09030*. [Online]. Available: http://arxiv.org/abs/1801.09030

[4] K. Yang, R. Zhang, L. He, Y. Li, W. Liu, C. Yu, Y. Zhang, X. Li, Y. Liu, and W. Xu, "Multistage analysis method for detection of effective herb prescription from clinical data," *Frontiers Med.*, vol. 2, no. 7, pp. 1–12, 2017.

[5] A. Hosseini, T. Chen, W. Wu, Y. Sun, and M. Sarrafzadeh, "HeteroMed: Heterogeneous information network for medical diagnosis," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 763–772.

[6] W. L. Hamilton, R. Ying, and J. Leskovec, "Representation learning on graphs: Methods and applications," *IEEE Data Eng. Bull.*, vol. 40, no. 3, pp. 52–74, 2017.

[7] Y. Fang, W. Lin, V. W. Zheng, M. Wu, K. C.-C. Chang, and X.-L. Li, "Semantic proximity search on graphs with metagraph-based learning," in *Proc. IEEE 32nd Int. Conf. Data Eng. (ICDE)*, May 2016, pp. 277–288.

[8] J. F. Roddick, P. Fule, and W. J. Graco, "Exploratory medical knowledge discovery: Experiences and issues," *ACM SIGKDD Explor. Newslett.*, vol. 5, no. 1, pp. 94–99, Jul. 2003.

[9] C. Zhao, G.-Z. Li, C. Wang, and J. Niu, "Advances in patient classification for traditional Chinese medicine: A machine learning perspective," *Evidence-Based Complementary Alternative Med.*, vol. 2015, pp. 1–18, Apr. 2015.

[10] Z. Jiang, H. Liu, B. Fu, Z. Wu, and T. Zhang, "Recommendation in heterogeneous information networks based on generalized random walk model and Bayesian personalized ranking," in *Proc. 11th ACM Int. Conf. Web Search Data Mining*, Feb. 2018, pp. 288–296.

[11] X. Li, Y. Wu, M. Ester, B. Kao, X. Wang, and Y. Zheng, "Semi-supervised clustering in attributed heterogeneous information networks," in *Proc. 26th Int. Conf. World Wide Web*, Apr. 2017, pp. 1621–1629.
[12] S. Hou, Y. Ye, Y. Song, and M. Abdulhayoglu, "HinDroid: An intelligent Android malware detection system based on structured heterogeneous information network," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 1507–1515.
[13] C. Shi and S. Y. Philip, *Heterogeneous Information Network Analysis and Applications*. Cham, Switzerland: Springer, 2017.
[14] Z. Huang, Y. Zheng, R. Cheng, Y. Sun, N. Mamoulis, and X. Li, "Meta structure: Computing relevance in large heterogeneous information networks," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, pp. 1595–1604.
[15] Y. Shi, H. Gui, Q. Zhu, L. Kaplan, and J. Han, "AspEm: Embedding learning by aspects in heterogeneous information networks," in *Proc. SIAM Int. Conf. Data Mining*, 2018, pp. 144–152.
[16] J. Yu, M. Gao, J. Li, H. Yin, and H. Liu, "Adaptive implicit friends identification over heterogeneous network for social recommendation," in *Proc. 27th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2018, pp. 357–366.
[17] C. Shi, X. Kong, P. S. Yu, S. Xie, and B. Wu, "Relevance search in heterogeneous networks," in *Proc. 15th Int. Conf. Extending Database Technol. (EDBT)*, 2012, pp. 180–191.
[18] Y. Fan, S. Hou, Y. Zhang, Y. Ye, and M. Abdulhayoglu, "Gotcha–sly malware!: Scorpion a metagraph2vec based malware detection system," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 253–262.
[19] R. Ying, R. He, K. Chen, P. Eksombatchai, W. L. Hamilton, and J. Leskovec, "Graph convolutional neural networks for Web-scale recommender systems," in *Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2018, pp. 974–983.
[20] T. H. Nguyen and R. Grishman, "Graph convolutional networks with argument-aware pooling for event detection," in *Proc. AAAI*, vol. 18, 2018, pp. 5900–5907.
[21] J. Bastings, I. Titov, W. Aziz, D. Marcheggiani, and K. Sima'an, "Graph convolutional encoders for syntax-aware neural machine translation," 2017, *arXiv:1704.04675*. [Online]. Available: https://arxiv.org/abs/1704.04675
[22] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, Jul. 2018.
[23] A. Fout, J. Byrd, B. Shariat, and A. Ben-Hur, "Protein interface prediction using graph convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6530–6539.
[24] Z. Ying, J. You, C. Morris, X. Ren, W. Hamilton, and J. Leskovec, "Hierarchical graph representation learning with differentiable pooling," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 4800–4810.
[25] A. Sankar, X. Zhang, and K. C.-C. Chang, "Motif-based convolutional neural network on graphs," 2017, *arXiv:1711.05697*. [Online]. Available: http://arxiv.org/abs/1711.05697
[26] P. Kazienko, "Social network analysis: Selected methods and applications," in *Proc. DATESO*, 2012, p. 151.
[27] H. Zhao, Q. Yao, J. Li, Y. Song, and D. L. Lee, "Meta-graph based recommendation fusion over heterogeneous information networks," in *Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2017, pp. 635–644.
[28] J. Chang, J. Gu, L. Wang, G. Meng, S. Xiang, and C. Pan, "Structure-aware convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 11–20.
[29] M.-T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," 2015, *arXiv:1508.04025*. [Online]. Available: https://arxiv.org/abs/1508.04025
[30] H. Wan, M.-F. Moens, W. Luyten, X. Zhou, Q. Mei, L. Liu, and J. Tang, "Extracting relations from traditional Chinese medicine literature via heterogeneous entity networks," *J. Amer. Med. Inform. Assoc.*, vol. 23, no. 2, pp. 356–365, Mar. 2016.
[31] B. Xu, H. Shen, Q. Cao, Y. Qiu, and X. Cheng, "Graph wavelet neural network," in *Proc. ICLR*, 2019, pp. 1–13.
[32] C. Ruan, Y. Wang, Y. Zhang, J. Ma, H. Chen, U. Aickelin, S. Zhu, and T. Zhang, "THCluster: Herb supplements categorization for precision traditional Chinese medicine," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Nov. 2017, pp. 417–424.

CHUNYANG RUAN received the Ph.D. degree from the School of Computer Science, Fudan University, Shanghai, China. He is currently a Lecturer with the Department of Data Science and Big Data Technology, School of Economics and Finance, Shanghai International Studies University, Shanghai. His main research interests include graph embedding, graph neural networks, heterogeneous information networks, data analysis, and knowledge discovery.

YINGPEI WU is currently pursuing the Ph.D. degree with the School of Computer Science, Fudan University, Shanghai, China. His current research interests include heterogeneous information networks, graph neural networks, representation learning, and traditional Chinese medicine informatization.

YUN YANG received the B.S. degree in traditional Chinese medicine and the M.S. degree in internal medicine of traditional Chinese medicine from the Shanghai University of Traditional Chinese Medicine, Shanghai, China, in 2008 and 2012, respectively. She is currently pursuing the Ph.D. degree with Shanghai University of traditional Chinese Medicine of Internal medicine of traditional Chinese Medicine. Her research interests include the basic and clinical research of traditional Chinese medicine treatment of tumor and the combination of traditional Chinese medicine and artificial intelligence.

GUANGSHENG LUO (Member, IEEE) received the B.S. degree in computer science from the Huazhong University of Technology, Wuhan, China, in 2005, and the M.S. degree in computer science from Fudan University, Shanghai, China, in 2015. He is currently a Lecturer with the Department of Data Science and Big Data Technology, School of Economics and Finance, Shanghai International Studies University. His research interests include deep learning, smart financial, and automatic speech recognition.