Cross-Network Clustering and Cluster Ranking for Medical Diagnosis

Jingchao Ni[1], Hongliang Fei[2], Wei Fan[2] and Xiang Zhang[1]
[1]College of Information Sciences and Technology, Pennsylvania State University, [2]Baidu Research Big Data Lab
[1]{jzn47, xzhang}@ist.psu.edu, [2]{hongliangfei, fanwei03}@baidu.com

***Abstract*—Automating medical diagnosis is an important data mining problem, which is to infer likely disease(s) for some observed symptoms. Algorithms to the problem are very beneficial as a supplement to a real diagnosis. Existing diagnosis methods typically perform the inference on a sparse bipartite graph with two sets of nodes representing diseases and symptoms, respectively. By using this graph, existing methods basically assume no direct dependency exists between diseases (or symptoms), which may not be true in reality. To address this limitation, in this paper, we introduce two domain networks encoding similarities between diseases and those between symptoms to avoid information loss as well as to alleviate the sparsity problem of the bipartite graph. Based on the domain networks and the bipartite graph bridging them, we develop a novel algorithm, CCCR, to perform diagnosis by ranking symptom-disease clusters. Comparing with existing approaches, CCCR is more accurate, and more interpretable since its results deliver rich information about how the inferred diseases are categorized. Experimental results on real-life datasets demonstrate the effectiveness of the proposed method.**

I. Introduction

Existing computational methods on medical diagnosis are mostly developed on a Quick Medical Reference (QMR) graphical model [1], [2]. In these approaches, symptoms and diseases are regarded as two sets of nodes forming a bipartite graph. In this graph, each pair of correlated disease and symptom are connected by an *association edge*, which can be weighted by the corresponding correlation level [3]. The diagnosis problem is to infer a probability distribution for the disease nodes given a subset of the symptom nodes. Despite their effectiveness, the performance of these methods are limited by the following problems:

First, by using the bipartite graph, existing approaches basically assume diseases (symptoms) are independent (or conditionally independent) with each other. However, in practice, diseases (symptoms) are always correlated, e.g., "cold" relates to "influenza", "bronchitis" relates to "asthma", etc. Certain diseases can even cause other diseases to present [1]. Therefore, ignoring such dependencies inevitably results in severe information loss. Moreover, the associations between diseases and symptoms are usually far from complete due to the limited medical records [3]. Only using these sparse associations in the bipartite graph to perform diagnosis thus often results in unreliable outcomes of these methods.

Another drawback of existing methods is the poor interpretability of their outputs. For example, given symptoms "cough", "expectoration" and "sore throat", existing methods may retrieve a ranking list {"cold", "asthma", "influenza", "bronchitis", ...}. Although the top several diseases are relevant, some of them are very different in the underlying etiologies, such as "cold" and "asthma". Mixing them in a single list is quite confusing and can mislead following therapies since different diseases can be totally different in drug uses [4]. Thus considerable expert labors are still needed to identify the real causes from such lists.

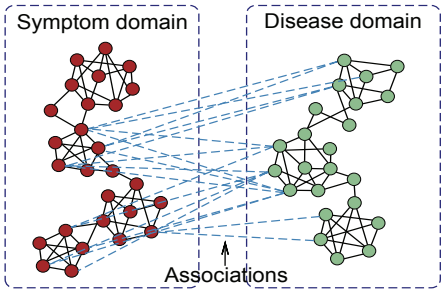


Fig. 1. **An example of the symptom-disease networks.**

In this paper, we address the first problem by introducing two *domain networks* to represent the similarities (or dependencies) between symptoms and those between diseases. Fig. 1 illustrates a symptom similarity network (left) and a disease similarity network (right), where nodes represent symptoms and diseases respectively, and each edge can be weighted by how similar two symptoms (diseases) are. Then, we refer to the bipartite graph between symptoms and diseases as the cross-domain *association network*, which bridges different domains by the association edges. It is worth to note that the rich dependency information in the domain networks can supplement the sparse association network, hence can substantially alleviate the problems caused by sparsity.

To address the second problem, we propose to conduct diagnosis at the cluster-level, in contrast to the traditional node-level diagnosis. For example, the following list, {{1st class: "cold", "influenza"}, {2nd class: "bronchitis", "asthma"}, ...}, is more interpretable than a mixed one since it shows how similar diseases are clustered while the clusters are ordered by their possibilities to the symptoms. In this way, people can easily understand what categories of diseases are relevant at a high level so they can zoom in certain category to locate the true causes more efficiently. Hence, allowing such a clustered structure in the ranking list can help reduce the risk of false identifications. In this paper, we refer to this novel problem as the *cross-network cluster ranking* problem, which cannot be solved by adapting the existing graph clustering algorithms [5]–[7] or co-clustering algorithms [8].

To handle the new problem, in this paper, we develop a new algorithm CCCR for joint **C**ross-network **C**lustering and

Cluster **R**anking, based on the dual network structure shown in Fig. 1. Generally, CCCR infers a probability distribution for the disease clusters given a symptom cluster. Detecting disease clusters can show users how similar diseases are categorized. Symptoms are also considered at the cluster-level since related symptoms often occur together. In particular, CCCR jointly clusters domain networks and ranking inferred clusters so that the dual procedures are mutually reinforced. Empirical results demonstrate CCCR can work effectively on real-life symptom-disease networks.

II. Problem Definition

In this paper, we introduce our method in a generalized manner so that it is theoretically applicable on any number of domain networks.

We represent the i^{th} *domain network* by its adjacency matrix $\mathbf{A}^{(i)} \in \mathbb{R}_+^{n_i \times n_i}$, where n_i is the number of nodes in domain i. Each entry $\mathbf{A}^{(i)}_{xy}$ measures the similarity between nodes x and y in domain i. Suppose we have g domains, for any pair of domains (i, j), nodes in the two domains may be linked by an *association network* $\mathbf{B}^{(ij)} \in \mathbb{R}_+^{n_i \times n_j}$. The entry $\mathbf{B}^{(ij)}_{xy}$ measures the association strength between node x in domain i and node y in domain j.

The Problem. Our goal is two-fold: (1) we aim to assign each node in each domain to a specific cluster; (2) for each cluster u in one domain, we want to assign each cluster in other domains a score to represent its relevance to cluster u.

III. CrossCR Algorithm

In this section, we introduce our CCCR algorithm. For domain network clustering, we employ a doubly stochastic matrix decomposition approach due to its superiority in clustering real-world sparse networks [5]. For cluster ranking, we develop a second-order random walk model to infer the cross-domain conditional probabilities of clusters. The two constituents are then integrated in a joint optimization problem to reinforce each other. Finally, a robust iterative solution is developed to solve the problem.

Domain Network Clustering. Suppose there are k_i clusters in domain network $\mathbf{A}^{(i)}$, let $\mathbf{H}^{(i)} \in \mathbb{R}_+^{n_i \times k_i}$ be a cluster membership matrix with $\mathbf{H}^{(i)}_{xu} = P(u|x)$ indicating the probability that node x belongs to cluster u. Then a *doubly stochastic approximation* to the domain network $\mathbf{A}^{(i)}$ is defined by [5]

$$\hat{\mathbf{A}}^{(i)}_{xy} = \sum_{u=1}^{k_i} \frac{\mathbf{H}^{(i)}_{xu}\mathbf{H}^{(i)}_{yu}}{\sum_{z=1}^{n_i}\mathbf{H}^{(i)}_{zu}} \tag{1}$$

where y and z are different node variables.

The clustering problem is to infer $\mathbf{H}^{(i)}$ by minimizing the approximation error $\mathcal{D}_{KL}(\mathbf{A}^{(i)}||\hat{\mathbf{A}}^{(i)})$ using KL-Divergence $\mathcal{D}_{KL}(\cdot||\cdot)$. This is equivalent to minimize

$$\begin{aligned} \mathcal{J}^{(i)}_A = &- \sum_{(x,y)\in E^{(i)}} \mathbf{A}^{(i)}_{xy} \log \hat{\mathbf{A}}^{(i)}_{xy} - (\alpha - 1)\sum_{xu} \log \mathbf{H}^{(i)}_{xu} \\ &\text{s.t. } \mathbf{H}^{(i)} \geq 0,\ \mathbf{H}^{(i)}\mathbf{1}_{k_i} = \mathbf{1}_{n_i} \end{aligned} \tag{2}$$

where $E^{(i)}$ represents the set of all edges in $\mathbf{A}^{(i)}$, $\mathbf{1}_{k_i}$ is a column vector of length k_i with all entries as 1. The second term is added by [5] to control the sparsity of $\mathbf{H}^{(i)}$. α ($\alpha \geq 1$) is a controlling parameter. In Eq. (2), the equality constraints are enforced to preserve the probabilistic interpretation of $\mathbf{H}_{xu}$.

Cross-Network Cluster Ranking. Next, we propose a second-order random walk model to infer cross-domain cluster ranking scores. For the ease of presentation, we first consider two domains $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, then we generalize our method to multiple domain networks.

Given two domain networks, we first augment them by two sets of *latent nodes* $\mathcal{U} = \{u\}_{u=1}^{k_1}$ and $\mathcal{V} = \{v\}_{v=1}^{k_2}$ to represent the latent clusters in $\mathbf{A}^{(1)}$ and $\mathbf{A}^{(2)}$, respectively. The augmented network consists of six components: three original networks $\mathbf{A}^{(1)}$, $\mathbf{A}^{(2)}$, $\mathbf{B}^{(12)}$, and three added completed bipartite graphs $\{\mathbf{A}^{(1)}, \mathcal{U}\}$, $\{\mathbf{A}^{(2)}, \mathcal{V}\}$, $\{\mathcal{U}, \mathcal{V}\}$.

Using the augmented network, we can regard $P(u|x)$ (i.e., $\mathbf{H}^{(1)}_{xu}$ in Eq. (2)) as a one step random walk transition probability from a node x in $\mathbf{A}^{(1)}$ to a latent node u. Similarly, we have $P(v|y)$ from a node y to a latent cluster v in domain $\mathbf{A}^{(2)}$. Moreover, we have the cross-domain transition probability $P(y|x)$ from node x in $\mathbf{A}^{(1)}$ to node y in $\mathbf{A}^{(2)}$, which can be estimated by $P(y|x) = \mathbf{B}^{(12)}_{xy} / \sum_{z=1}^{n_2} \mathbf{B}^{(12)}_{xz}$.

Given these empirical probabilities, we want to estimate $P(v|u)$, which represents the importance of a cluster v in $\mathbf{A}^{(2)}$ given a cluster u in $\mathbf{A}^{(1)}$. We observe the above mentioned transition probabilities inherently form two kinds of *second-order* random walk paths, from a node x to a latent node v:

(1) Real path ($x \rightsquigarrow y \rightsquigarrow v$): $P_r(v|x) = \sum_{y=1}^{n_2} P(v|y)P(y|x)$

(2) Latent path ($x \rightsquigarrow u \rightsquigarrow v$): $P_l(v|x) = \sum_{u=1}^{k_1} P(v|u)P(u|x)$

We call the path as a "real path" (or a "latent path") because the bridge node y (or u) is a real (or latent) node. Since the estimated cluster-level probability $P(v|u)$ should well explain the generation of node-level associations $\mathbf{B}^{(12)}_{xy}$ (or $P(y|x)$), we propose to use the latent path transition probability P_l to approximate the real path transition probability P_r. That is, we want to minimize the approximation error $\mathcal{D}_{KL}(P_r||P_l)$, which gives the following loss function for all pairs of (x, v):

$$-\sum_{x=1}^{n_1}\sum_{v=1}^{k_2} P_r(v|x) \log P_l(v|x) \tag{3}$$

Formally, we define $\mathbf{S}^{(12)} \in \mathbb{R}_+^{k_1 \times k_2}$ with $\mathbf{S}^{(12)}_{uv} = P(v|u)$, and $\tilde{\mathbf{B}}^{(12)}$ to be the row normalized version of $\mathbf{B}^{(12)}$, i.e., $\tilde{\mathbf{B}}^{(12)}_{xy} = \mathbf{B}^{(12)}_{xy} / \sum_{z=1}^{n_2} \mathbf{B}^{(12)}_{xz}$. Then, by enforcing a stochastic constraint on $\mathbf{S}^{(12)}$, i.e., $\mathbf{S}^{(12)} \geq 0$ and $\mathbf{S}^{(12)}\mathbf{1}_{k_2} = \mathbf{1}_{k_1}$, Eq. (3) can be rewritten in a matrix form as

$$\mathcal{J}^{(12)}_R = -\sum_{xv} \underbrace{(\tilde{\mathbf{B}}^{(12)}\mathbf{H}^{(2)})_{xv}}_{\text{real paths}} \log \underbrace{(\mathbf{H}^{(1)}\mathbf{S}^{(12)})_{xv}}_{\text{latent paths}} \tag{4}$$

which is our loss function for cross-domain cluster ranking, where we want to estimate $\mathbf{S}^{(12)}$.

A Unified Objective Function. As mentioned earlier, a principled way to infer the clustering of nodes and ranking of

clusters is to jointly train the objective functions in Eq. (2) and Eq. (4), which allows the mutual enhancement of the two procedures. By doing so, and generalizing the concept to any pair of domains i and j, we reach a joint optimization problem as following

$$\begin{aligned} \min\ & \mathcal{J}(\{\mathbf{H}^{(i)}\},\{\mathbf{S}^{(ij)}\}) = \sum_{i=1}^{g}\mathcal{J}_A^{(i)} + \beta\sum_{ij,\ i\neq j}\mathcal{J}_R^{(ij)} \\ \text{s.t. } & \mathbf{H}^{(i)} \geq 0,\ \mathbf{H}^{(i)}\mathbf{1}_{k_i} = \mathbf{1}_{n_i} \\ & \mathbf{S}^{(ij)} \geq 0,\ \mathbf{S}^{(ij)}\mathbf{1}_{k_j} = \mathbf{1}_{k_i},\ \forall 1 \leq i,j \leq g,\ i \neq j \end{aligned} \tag{5}$$

where β is a parameter to balance between the domain network clustering and the cross-domain cluster ranking. When $\beta = 0$, Eq. (5) degenerates to g independent network clustering. Intuitively, the more reliable the association networks, the larger the value of β.

Priority of Nodes in Each Cluster. Previously, we have considered how to order nodes at the cluster-level, but neglected the ordering of nodes within each inferred cluster. Next, we derive a strategy to prioritize nodes within each cluster by their importances to that cluster.

Once we have obtained the cluster membership probabilities $P(u|x)$ (or $\mathbf{H}_{xu}^{(i)}$) in domain i from Eq. (5), we can calculate the probability $P(x|u)$ by using the Bayes formula and expansion rule, which gives

$$P(x|u) = \frac{P(u|x)P(x)}{\sum_{z=1}^{n_i}P(u|z)P(z)} = \frac{P(u|x)}{\sum_{z=1}^{n_i}P(u|z)} \tag{6}$$

The above equation can be simply represented by $P(x|u) = (\mathbf{H}^{(i)}(\mathbf{D}_H^{(i)})^{-1})_{xu}$ where $\mathbf{D}_H^{(i)}$ is a k_i-by-k_i diagonal matrix with $(\mathbf{D}_H^{(i)})_{uu} = \sum_{z=1}^{n_i}\mathbf{H}_{zu}$.

Here, $P(x|u)$ indicates the "importance" of node x to cluster u in domain i. From Eq. (6), a "center" node of cluster u has a higher $P(x|u)$ than a "border" node. Thus, we can rank the entries in each column of $\mathbf{H}^{(i)}(\mathbf{D}_H^{(i)})^{-1}$ to obtain the most representative nodes in each cluster.

Learning Algorithm. The objective function in Eq. (5) is not jointly convex in all variables, hence we take an alternating minimization approach, i.e., the objective function is alternately optimized w.r.t. one variable while fixing others. The procedure repeats until a stationary point is achieved.

Solution to H. Let $\mathcal{J}(\mathbf{H}^{(i)})$ be the objective function in Eq. (5) w.r.t. $\mathbf{H}^{(i)}$ when fixing other variables as constants. Then the Lagrangian function of $\mathcal{J}$ in Eq. (5) w.r.t. $\mathbf{H}^{(i)}$ is

$$\mathcal{L}_H(\mathbf{H}^{(i)},\boldsymbol{\lambda}^{(i)}) = \mathcal{J}(\mathbf{H}^{(i)}) + \sum_{x=1}^{n_i}\lambda_x^{(i)}(\sum_{u=1}^{k_i}\mathbf{H}_{xu}^{(i)} - 1) \tag{7}$$

where $\boldsymbol{\lambda}^{(i)} = (\lambda_1^{(i)},...,\lambda_{n_i}^{(i)})^T$ are Lagrangian multipliers.

Let the gradient of $\mathcal{J}(\mathbf{H}^{(i)})$ w.r.t. $\mathbf{H}^{(i)}$ be $\nabla_H^{(i)} = (\nabla_H^{(i)})^+ - (\nabla_H^{(i)})^-$, where $(\nabla_H^{(i)})^+$ and $(\nabla_H^{(i)})^-$ represent the positive and non-positive parts of $\nabla_H^{(i)}$, respectively. Then the following theorem gives the iterative solution to $\mathbf{H}^{(i)}$.

Theorem 1. *Let* $\lambda_x^{(i)} = (b_x^{(i)} - 1)/a_x^{(i)}$, *where*

$$a_x^{(i)} = \sum_{u=1}^{k_i}\frac{\mathbf{H}_{xu}^{(i)}}{(\nabla_H^{(i)})_{xu}^+},\quad b_x^{(i)} = \sum_{u=1}^{k_i}\mathbf{H}_{xu}^{(i)}\frac{(\nabla_H^{(i)})_{xu}^-}{(\nabla_H^{(i)})_{xu}^+} \tag{8}$$

It holds that $\mathcal{L}_H((\mathbf{H}^{(i)})^{new},\boldsymbol{\lambda}^{(i)}) \leq \mathcal{L}_H(\mathbf{H}^{(i)},\boldsymbol{\lambda}^{(i)})$, *by updating* $\mathbf{H}^{(i)}$ *according to Eq.* (9).

$$\mathbf{H}_{xu}^{(i)} \leftarrow \mathbf{H}_{xu}^{(i)}\frac{a_x^{(i)}(\nabla_H^{(i)})_{xu}^- + 1}{a_x^{(i)}(\nabla_H^{(i)})_{xu}^+ + b_x^{(i)}} \tag{9}$$

Proof: Omitted for brevity. ■

Solution to S. Similarly, let $\mathcal{L}_S(\mathbf{S}^{(ij)},\boldsymbol{\eta}^{(ij)})$ be the Lagrangian function of $\mathcal{J}$ in Eq. (5) w.r.t. $\mathbf{S}^{(ij)}$, where $\boldsymbol{\eta}^{(ij)} = (\eta_1^{(ij)},...,\eta_{k_i}^{(ij)})^T$ are Lagrangian multipliers. Let the gradient of $\mathcal{J}$ w.r.t. $\mathbf{S}^{(ij)}$ be $\nabla_S^{(ij)} = (\nabla_S^{(ij)})^+ - (\nabla_S^{(ij)})^-$. Then the following theorem gives the iterative solution to $\mathbf{S}^{(ij)}$.

Theorem 2. *Let* $\eta_u^{(ij)} = (d_u^{(ij)} - 1)/c_u^{(ij)}$, *where*

$$c_u^{(ij)} = \sum_{v=1}^{k_j}\frac{\mathbf{S}_{uv}^{(ij)}}{(\nabla_S^{(ij)})_{uv}^+},\quad d_u^{(ij)} = \sum_{v=1}^{k_j}\mathbf{S}_{uv}^{(ij)}\frac{(\nabla_S^{(ij)})_{uv}^-}{(\nabla_S^{(ij)})_{uv}^+} \tag{10}$$

It holds that $\mathcal{L}_S((\mathbf{S}^{(ij)})^{new},\boldsymbol{\eta}^{(ij)}) \leq \mathcal{L}_S(\mathbf{S}^{(ij)},\boldsymbol{\eta}^{(ij)})$, *by updating* $\mathbf{S}^{(ij)}$ *according to Eq.* (11).

$$\mathbf{S}_{uv}^{(ij)} \leftarrow \mathbf{S}_{uv}^{(ij)}\frac{c_u^{(ij)}(\nabla_S^{(ij)})_{uv}^- + 1}{c_u^{(ij)}(\nabla_S^{(ij)})_{uv}^+ + d_u^{(ij)}} \tag{11}$$

Proof: Omitted for brevity. ■

In our CCCR algorithm, we first randomly initialize $\{\mathbf{H}^{(i)}\}$ and $\{\mathbf{S}^{(ij)}\}$ with row normalizations. Then we alternately update $\{\mathbf{H}^{(i)}\}$ by Eq. (9) and $\{\mathbf{S}^{(ij)}\}$ by Eq. (11) until convergence. According to Theorem 1 and 2, alternately updating $\mathbf{H}^{(i)}$ and $\mathbf{S}^{(ij)}$ will decrease the objective value in Eq. (5), as well as enforces stochastic constraints on them.

IV. Experimental Results

In this section, we present the experimental results on a real-life symptom-disease network dataset, which is collected from the largest medical website in China[1]. Specifically, it contains a disease similarity network of $9,721$ disease nodes and $29,332$ edges, a symptom similarity network of $5,093$ symptom nodes and $22,548$ edges, as well as an association network with $5,337$ symptom-disease associations.

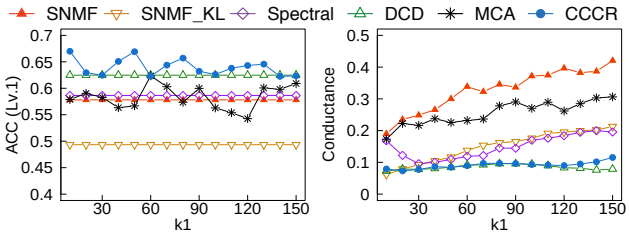
Throughout the experiments, we compare the performance of CCCR with single network clustering methods including (1) SNMF, i.e., symmetric non-negative matrix factorization using Euclidean distance [9]; (2) SNMF_KL, i.e., SNMF using KL-Divergence [9] (3) Spectral clustering (Spectral) [10]; (4) DCD, i.e., stochastic matrix decomposition approach [5], and a graph regularized co-clustering method, MCA [8], which can generate cluster-level relationships between domains.

Clustering Results. In this dataset, we have 17 disease classes covering 1447 diseases (14.89%). Thus we use purity accuracy (ACC) to evaluate disease clustering performance. For symptom clustering, since there is no ground truth, we use the widely used conductance as a quality measure, which is defined as [10] $\text{Cond}(\mathcal{C}) = |\partial(\mathcal{C})|/\min(\text{Vol}(\mathcal{C}),\text{Vol}(\bar{\mathcal{C}}))$, where $\mathcal{C}$ is a set of nodes, $|\partial(\mathcal{C})|$ is the number of edges with one endpoint inside of $\mathcal{C}$ and one endpoint outside of $\mathcal{C}$, $\text{Vol}(\mathcal{C})$ is the sum of node degrees in $\mathcal{C}$, and $\bar{\mathcal{C}}$ is the set of

[1] http://www.xywy.com/

TABLE I. TOP RANKED DISEASE CLUSTERS GIVEN BY CCCR.

symptom cluster	1st disease cluster (probability)	2nd disease cluster (probability)	3rd disease cluster (probability)
(1) bloating burp stomachache	reflux esophagitis gastroesophageal reflux (0.7877) gastritis	duodenal inflammation antral erosion (0.1351) superficial gastritis	– (<0.1000)
(2) eye fissure photophobia pupillary block	macular degeneration retinal detachment (0.5002) vitreous opacities	amblyopia hyperopia (0.1569) esotropia	conjunctivitis keratitis (0.1330) pink eye
(3) cerebral hemorrhage intracranial hemorrhage increased intracranial pressure	cerebral infarction brainstem infarction (0.5400) stroke	skull fracture epidural hematoma (0.1449) brain contusion	diabetes hypertension (0.1161) dyslipidemia



(a) ACC of disease clusters (b) Cond. of symptom clusters

Fig. 2. **The clustering performance comparison.**

nodes outside $\mathcal{C}$. Usually, a lower conductance implies a better cluster-like structure of $\mathcal{C}$.

Fig. 2 shows the clustering performance of different methods w.r.t. varying number of symptom clusters k_1. Here the number of disease clusters k_2 is fixed for each method to achieve the best purity accuracy. In Fig. 2(a), we observe CCCR outperforms other approaches in most values of k_1, which indicates its superiority in detecting meaningful clusters in terms of domain knowledge. From the figures, we also observe several peak accuracies of CCCR, which may imply a hierarchical structure in the symptom network (or disease network) that results in multiple good choices of k_1. Fig. 2(b) shows the averaged conductances of the detected symptom clusters. Although DCD is comparable with CCCR in terms of conductance, which is very small with little room to improve, its purity accuracy is lower than CCCR. This means the clusters detected by CCCR make more sense in both medical context and network topology than other competing approaches.

Cluster Ranking Results. Next, we demonstrate that CCCR has better result interpretation by comparing its outcomes with a well-known QMR-DT algorithm, quickscore [1]. Table I presents the top 3 disease clusters given a symptom cluster detected by CCCR. We select 3 interesting symptom clusters as examples. The disease clusters are sorted in descending order by their probabilities. Each symptom (or disease) cluster is represented by its top 3 representative symptoms (or diseases), as discussed in Sec. III. Here, disease clusters with probabilities less than 10% are filtered out. Table II shows the top several diseases of quickscore for the symptoms in Table I. This algorithm only runs on the bipartite network and returns a single ranking list of diseases for the given symptoms.

Clearly, the lists in Table II mix diseases from different categories, such as (2) cataract and ocular trauma, which results in a fairly poor interpretation. Moreover, we have highlighted relevant diseases in each line by bold *italics*. As can be seen, only a few diseases in each list are relevant to the corresponding symptoms. Such false inferences are caused by

TABLE II. RESULTS OF A QMR-DT ALGORITHM.

Symptom #	Top ranked diseases
(1)	***gastritis***, cold, heart disease, fracture, epilepsy
(2)	***cataract***, ***uveitis***, ***ocular trauma***, ***keratitis***, ***pink eye***
(3)	***subarachnoid hemorrhage***, ***aneurysm***, ***hypertension***, cold

the limitation of using the bipartite network only. These results demonstrate the effectiveness of CCCR and the importance to take the rich dependencies between symptoms (diseases) into account when performing diagnosis.

V. CONCLUSION

In this paper, we introduce disease and symptom domain networks to address the information loss and association sparsity problems of traditional medical diagnosis algorithms. Based on the domain-association network structure, we develop a new algorithm CCCR to handle a new diagnosis problem, with the goal to better interpret the outcomes by allowing a clustered structure in the retrieved ranking list than a mixed ranking list. Experimental results on real-life datasets demonstrate the effectiveness of the proposed method.

VI. ACKNOWLEDGEMENT

This work was partially supported by the National Science Foundation grants IIS-1162374 and CAREER, and the National Institute of Health grant R01GM115833.

REFERENCES

[1] D. Heckerman, "A tractable inference algorithm for diagnosing multiple diseases," in *UAI*, 1989.

[2] D.-I. Curiac, G. Vasile, O. Banias, C. Volosencu, and A. Albu, "Bayesian network model for diagnosis of psychiatric diseases," in *ITI*, 2009.

[3] X. Zhou, J. Menche, A.-L. Barabási, and A. Sharma, "Human symptoms–disease network," *Nat. Commun.*, vol. 5, 2014.

[4] W. Wang, S. Yang, X. Zhang, and J. Li, "Drug repositioning by integrating target information through a heterogeneous network model," *Bioinformatics*, vol. 30, no. 20, pp. 2923–2930, 2014.

[5] Z. Yang and E. Oja, "Clustering by low-rank doubly stochastic matrix decomposition," in *ICML*, 2012.

[6] J. Ni, H. Tong, W. Fan, and X. Zhang, "Flexible and robust multi-network clustering," in *KDD*, 2015.

[7] J. Ni, W. Cheng, W. Fan, and X. Zhang, "Self-grouping multi-network clustering," in *ICDM*, 2016.

[8] R. Liu, W. Cheng, H. Tong, W. Wang, and X. Zhang, "Robust multi-network clustering via joint cross-domain cluster alignment," in *ICDM*, 2015.

[9] C. H. Ding, X. He, and H. D. Simon, "On the equivalence of nonnegative matrix factorization and spectral clustering." in *SDM*, 2005.

[10] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.