International Journal of Data Science and Analytics
https://doi.org/10.1007/s41060-018-0147-5

REGULAR PAPER

Multi-domain and multi-view networks model for clustering hospital admissions from the emergency department

Nouf Albarakati[1,2] · **Zoran Obradovic**[1]

Received: 6 October 2017 / Accepted: 26 July 2018

Abstract
As the healthcare industry continues to generate a massive amount of medical data, healthcare organizations integrate data-driven insights into their clinical and operational processes to enhance the quality of healthcare services. Our preliminary hospital clustering analysis (Albarakati and Obradovic, in The IEEE 29th international symposium on computer-based medical systems (CBMS), IEEE, 2017) studied hospitals monthly admission behavior for different diseases. Results showed consistent behavior when disease symptoms similarity is considered. This study extends our preliminary work to include other aspects of disease data and the fusion of different views of disease data. It is an original approach that tackles clustering complex networks using a combination of multi-view and multi-domain clustering models while imposing data on the clustering goal from both medical and non-medical domains simultaneously. The objective of the study is to determine the effect of disease networks on characterizing the underlying clustering structure of 145 disease-specific hospital networks, each consisting of up to 152 hospitals. This is achieved by extracting two different views of disease networks. One disease network view based on similarity of symptom profiles was extracted from a 20 million medical bibliographic literature records. Another disease network view based on monthly hospitalization distribution was extracted from over 7 million discharge records data obtained from the California State Inpatient Database for years 2009–2011. Patient admission records included both medical and sociodemographic information. These multiple views were analyzed separately and were also integrated in a joint model that combined the two views. It is shown that the fusion of multi-view disease networks of monthly hospitalization distributions explained the hidden common structure shared among multiple hospital-specific disease networks. The group homogeneity measures for obtained hospital clusters ranged between 33 and 60% with average close to 50%. However, integrating multiple views of disease networks extracted from different domains, i.e., from literature and medical databases, better revealed the underlying clustering structure of disease-specific hospital networks. The group homogeneity measures for this multi-domain setting ranged between 38 and 76% with average close to 60%.

Keywords Hospital clustering · Disease-based · Multi-view · Multi-domain · Homogeneity analysis

1 Introduction

With the continuous improvement in Electronic Health Record, EHR, implementations, the healthcare industry continues to generate a massive amount of medical data [7]. On the other hand, big data analytics and their applications rapidly advanced to provide a means to analyze the growing volumes of data [15]. Most of the healthcare organizations integrate data-driven insights into their clinical and operational processes to enhance the quality of healthcare planning and decision making. In order to optimize the use of healthcare resources, many clustering algorithms have been developed, improved and applied over many decades to better detect hospital clusters [2;4;14].

The descriptive and cluster analysis was applied on healthcare data to identify clusters of hospitals that share similar properties while considering various factors and different configurations [18]. Clustering hospitals based on their monthly admission behavior gives healthcare facili-

✉ Zoran Obradovic
zoran.obradovic@temple.edu

Nouf Albarakati
nouf@temple.edu

1 Center for Data Analytics and Biomedical Informatics, Temple University, Philadelphia, USA

2 Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Published online: 07 August 2018

ties insight to adjust their plans and policies to society and patient needs [2;4;19]. Identifying various patterns of hospital admission helps to understand the effect of different factors on admission process [14;18]. For instance, diagnosis of patients is vital in understanding variation in admission rates among hospitals. However, other hospital-level and community-level factors were associated with admission rates, but these varied across conditions diagnosed [3;8]. Hospitals that show similar behavior on their admission distributions do not necessarily show similar behavior when the principle conditions diagnosed during the admission process are considered. Moreover, a single hospital would have different admission distributions for different diagnoses as there are some conditions accounted more than others for hospital admission [10].

As illustrated in Fig. 1, there are two different ways to build a hospital network: not considering diseases diagnosed in admission or considering these diagnoses. Using an example of clustering six hospitals into two clusters, the first setting is clustering hospitals based on general admission data for the hospitals where diseases diagnoses are not considered (Fig. 1a). The second setting clusters hospitals based on disease-specific admission data (Fig. 1b). That is, there is a different hospital network for each disease. The similarities among hospitals in these disease-specific hospital networks are different, and therefore, hospital clusters are different.

In the first setting, all hospitals are assumed to share a single underlying clustering structure of their admission distribution. Therefore, clustering analysis would be done on a single big network that represents all hospitals under the study. Hospitals in the first setting would belong to one of the only two hospital clusters assumed in the example as shown in Fig. 1a. On the other hand, in the second setting, there would be multiple hospital networks. These networks have different distributions, where hospitals have different admission distributions for different diseases. Some hospitals have no representation in some of these networks; i.e., specialized hospitals do not appear in all disease-specific networks. In fact, different disease-specific hospital networks may share multiple underlying clustering structures; i.e., some disease-specific hospital networks share similar admission distribution while other disease-specific hospital networks vary. A hospital may belong to one cluster in some of these networks while it belongs to the other cluster in others, as shown in Fig. 1b.

However, disease symptoms play a critical role in clinical diagnosis and, hence, admission decision. Therefore, symptoms of these diagnosed diseases would offer complementary information to improve clustering results. Our preliminary hospital clustering analysis [1] showed consistent behavior when disease symptoms similarity is considered. As each of the disease-specific hospital networks represents all hospitals that have a specific disease diagnosed during admission process, it was useful to model the similarity among different hospital networks using disease symptoms similarity network. Figure 2 exhibits an example of a disease network that corresponds to and models the similarity among different hospital networks illustrated in Fig. 1b.

Our preliminary study [1] investigated the underlying clustering structure among different disease-specific hospital networks by integrating methods developed at two other studies that have high impact in their fields. It utilized the human symptoms disease network [20] to generate symptom-based disease network. Also, it applied a multi-domain clustering method NoNClus that captures multiple underlying clustering structures across different networks [13]. This method integrated both disease network and disease-specific hospital networks into a Network of Networks model. It uses the disease network as a super-network where every node is a supernode that represents a corresponding disease-specific

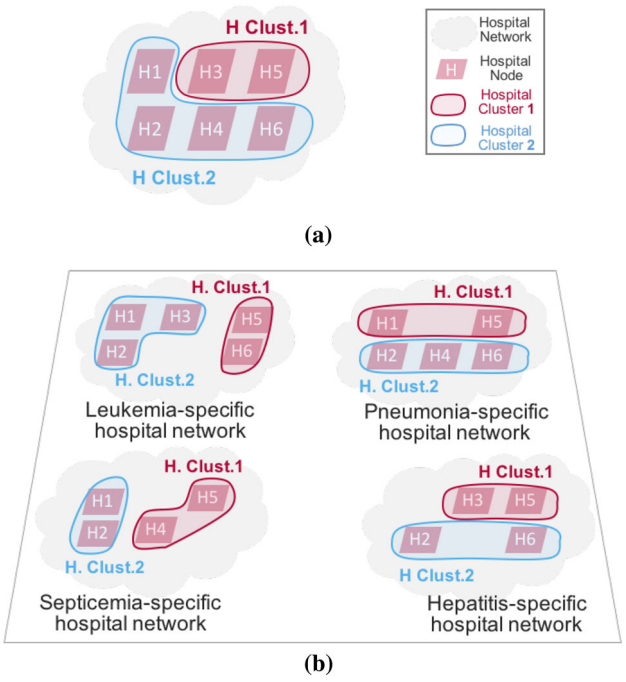

Fig. 1 Two settings of clustering hospitals. **a** Hospital network. **b** Disease-specific hospital networks

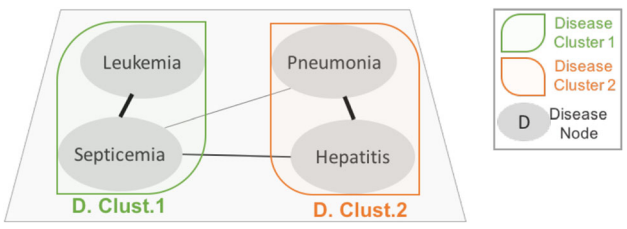

Fig. 2 Disease similarity network. Each node corresponds to the disease-specific hospital network in Fig. 1b

hospital network. Therefore, the NoNClus method utilizes the disease network to regularize the clustering of disease-specific hospital networks. This preliminary study showed that some of the hospital networks behave similarly if the diseases they represent share similar symptoms.

Findings of our preliminary study spark the interest to explore more hidden factors that better reveal the underlying clustering structure of hospital networks. It is given that patient factors had been proven to be one of the important factors that explain variation in hospital admission rate [3;8]. Moreover, sociodemographic patient factors are accounted for about 45% of the variation in admission rates [16]. Equally important, this study will be limited on patients admitted from emergency rooms as they are now the primary source of hospitalization in the USA [17].

This study extends the preliminary hospital clustering analysis [1] to include other aspects of disease data extracted from the monthly admission distribution and both medical and social data of patients. Besides the expanding data aspects, Network Fusion for Composite Community Extraction (NF-CCE) [6] method is used to detect clusters of diseases shared by multi-view disease networks that represent different aspects of relations between disease.

The main contribution of our study reported in this paper is the original approach of clustering complex networks using a combination of multi-view and multi-domain clustering models while imposing data on the clustering goal from both medical and non-medical domains simultaneously. It is a novel characterization of the effect of multi-view of disease networks on the underlying clustering structure of 145 multi-domain disease-specific hospital networks. The different views of disease network were extracted from a big medical bibliographic literature database and from patient admission records that include both medical and sociodemographic information. It is shown that the fusion of multi-view disease networks extracted from EHRs explains the hidden common structure shared across all hospital networks. Conversely, integrating multi-view disease networks extracted from different distributions better revealed the underlying clustering structure of disease-specific hospital networks.

This paper is organized as follows: Sect. 2 illustrates some background details regarding a disease network and clustering approaches related to this work. Multi-domain and multi-view clustering methods as well as the heterogeneous data model are explained in Sect. 3. In Sect. 4, the experiment settings are described while the results are presented and discussed in Sect. 5. The work is concluded in Sect. 6.

2 Background and related work

Different settings of hospital clustering analysis were studied on numerous factors separately or jointly. This study integrated multiple disease similarity network aspects with a complementary clustering algorithm to introduce a hospital clustering model that is able to optimize clustering analysis of multi-domain disease-specific hospital networks.

2.1 Disease networks

Since the disease symptoms are critical in diagnosis at admission time, Zhou et al constructed the human symptoms disease network (HSDN) as a weighted disease network generated using Medical Subject Headings (MeSH) terminology along with a big medical bibliographic literature database, PubMed [20]. MeSH terminology was used to index all articles in PubMed for over four thousand disease terms and over three hundred symptoms terms. Then, the association between diseases and symptoms was identified where every disease was described by a vector of related symptoms. Similarity between vectors of diseases was calculated as cosine ranging from 0 with no shared symptoms to 1 which means both diseases shared identical symptoms. However, no patient records are utilized in generating this disease network.

On the other hand, it is known that sociodemographic patient factors are correlated with hospital admission rates [16]. Although the disease similarity network extracted from patient health records has not been used in guiding clustering hospital admission, it was utilized successfully in improving knowledge on disease mechanisms and predicting disease comorbidities [11]. Moreover, studies showed that factors of patients admitted from emergency rooms are the most important in explaining admission variation as ERs are the primary source of hospitalization in the USA [17].

2.2 Data model and network clustering

More advanced heterogeneous information networks that integrate different data forms and sources is shown to improve clustering analysis. Multi-view and multi-domain approaches emerged to discover hidden patterns within such data.

One of these approaches is merging different multi-domain networks into a multilayered Network of Networks architecture [9]. A large-scale network is composed of several sub-networks, and the interconnectivity between these sub-networks is modeled as the top-layer super-network. This super-network is crucial to the information distribution of these sub-networks [12].

NoNClus is a framework proposed for clustering in multi-domain networks that has a Network of Networks structure [13]. Also, it allows multiple underlying clustering structures across different networks. NoNClus models the clustering structure in the top-layer network, which can be used to guide the clustering structures in different sub-networks that every node in the super-network represents. That is, it partitions the

multiple sub-networks while respecting the clustering structure obtained initially from the super-network.

On the other hand, Network Fusion for Composite Community Extraction (NF-CCE) [6] is a multi-view clustering method used to reveal the communities shared by networks representing different aspects of relations between different views of the networks.

The combination of multi-view NF-CCE and multi-domain NoNClus clustering models was used in clustering disease-specific hospital networks. These multi-domain networks were constructed using Network of Networks data model to incorporate the interconnectivity among diseases that each network represents. The detailed explanation of this novel integration is explained in the following section.

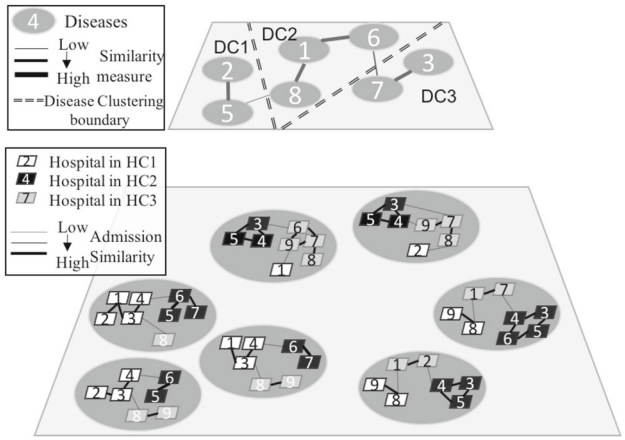

Fig. 3 Disease network of hospital networks data model

3 Methodology

This study aims to examine the effect of multi-view of disease networks on clustering multi-domain disease-specific hospital networks. Both multi-domain and multi-view clustering algorithms were jointly utilized to analyze monthly hospital clustering and to characterize the hidden common structure shared among all hospital networks.

The multi-views of disease network were extracted from a big medical bibliographic literature database and from patient admission records that include both medical and sociodemographic information.

3.1 Multi-view clustering

Network Fusion for Composite Community Extraction (NF-CCE) [6] is a multi-view clustering method that assumes a single underlying clustering among all views. It is used to discover the communities shared by networks representing different aspects of relations between different views of the networks. The NF-CCE is the Collective Nonnegative Matrix Factorization (CNMF). It collectively factorizes adjacency matrices representing different views of the network. All views must have an exact number of nodes.

This method works in two steps: First, nonnegative low-dimensional factors are obtained under column orthonormal constraints by using Symmetric Nonnegative Matrix Factorization (SNMF) for each network view. Second, these low-dimensional factors are fused into a common representation, using a collective matrix factorization model. All adjacency matrices of the multi-view networks are collectively decomposed into a common network.

3.2 Multi-domain clustering algorithm

NoNClus is a method for clustering multi-domain networks that have a Network of Networks structure [13]. Also, it allows multiple underlying clustering structures across different networks. To reveal non-overlapping clusters, NoNClus models the clustering structure in the top-layer disease network, which can be used to guide clustering structures in different disease-specific hospital sub-networks at the bottom layer. This is achievable because each network represents a disease node in the super-network through Network of Networks data model. Figure 3 shows the multi-domain disease network of hospital networks data model where the NoNClus is used as the multi-domain clustering method.

NoNClus works in two phases. In the first phase, NoNClus method starts by partitioning the top-layer disease network under column orthonormal constraints using a symmetric nonnegative matrix factorization. Nonnegative low-dimensional factors define the probability for each disease node to belong to one of the main clusters. In the second phase, the low-dimensional factor matrix of the disease network, obtained in phase 1, is incorporated as a guided regularization to get a factor matrix of network for each of the disease-specific hospital networks. However, NoNClus is developed to handle domain-specific networks that may have different number of nodes and clusters by minimizing the two-term objective function. The first term in the objective function deals with clustering the disease-specific hospital networks individually based on a similarity matrix of hospitals admission for each disease. The second term regularizes the factor matrix of each disease-specific hospital network, using main clustering structure defined in the factor matrix of the main disease network and the underlying clustering structure of domain-specific networks of the main cluster [13].

3.3 Heterogeneous data model

The data model, used in this study and shown in Fig. 3, is a Network of Networks model. It is represented in two layers

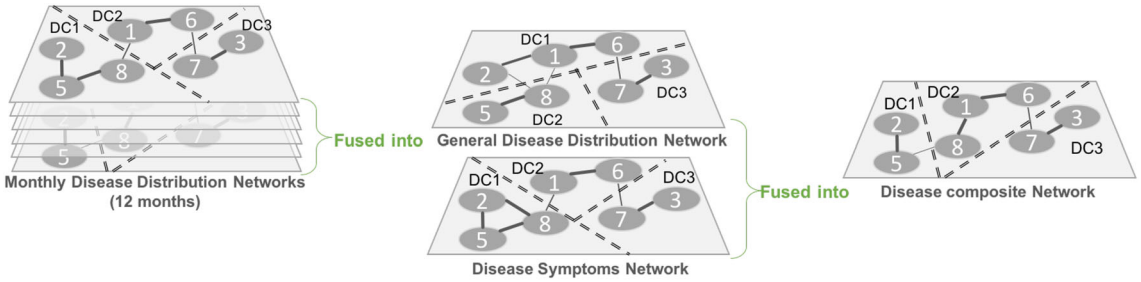

Fig. 4 Network fusion of multi-view disease networks for model 2 and model 3

of weighted networks: the top layer is a super-network of disease network that represents a disease similarity network. Every disease node in the top disease network represents a disease-specific hospital sub-network in the bottom layer. This layout helps to model the interconnectivity among these networks as it is obtained from the top-layer super-network. Every disease-specific hospital network in the bottom layer represents monthly hospital admissions for a specific disease among hospitals.

Three different settings for the top-layer disease network were introduced as multi-view networks. Each of these networks was used in a different model to investigate the multi-view effect on disease network characterization of the underlying clustering structure.

Model 1: Disease symptom network This is a single-view disease symptoms similarity network extracted from HSDN that is constructed using MeSH terminology [13]. However, the hospital admission networks for each disease were constructed using the California State Inpatient Database and CCS code was used to code diseases. Matching between the CCS codes and the MeSH terminology was done manually [5]. The average of similarities was used in some cases where matching was not one-to-one.

Model 2: Disease distribution network The top-layer multi-view networks represent the monthly disease distribution. This network was extracted from the California State Inpatient Database. All patient records were aggregated monthly based on disease where the distribution of patients for each disease was acquired. Medical data include age, admission on weekend, procedures, day of the procedure, number of chronic diseases and primary payer. Sociodemographic data included gender, race, median household income for patientś ZIP code and the four-category urban–rural designation for the patientś county of residence. The similarity among disease nodes was measured monthly using Kullback-Leibler divergence similarity. It measures how one probability distribution diverges from a second expected probability distribution. A set of 12 disease networks were produced for every year in the dataset. Fusion of these networks into the single network is done by the multi-view Network Fusion for Composite Community Extraction (NF-CCE) [6].

Model 3: Disease composite network The third disease network is a composite of both the disease symptoms similarity network that is extracted from a big medical bibliographic literature database and a fused monthly disease distribution network that is extracted form patient medical records at admission which include both medical and sociodemographic information. This is a dual clustering as constraints are imposed on the clustering goal from both medical and non-medical domains simultaneously. Fusion of these networks into the single network is done by the multi-view Network Fusion for Composite Community Extraction (NF-CCE) [6].

4 Experiment

The Network of Networks data model is a multi-domain network structure that is used in this study to investigate the effect of integrating the multi-view top-layer disease network on the underlying clustering structure of 145 bottom-layer disease-specific hospital networks.

This section explains the three proposed models. These models vary in the top-layer network. Model 1 has a single-view disease network while two settings of network fusion of multi-view disease networks are used in models 2 and 3. Figure 4 illustrates the network fusion of multi-view disease networks for model 2 and model 3. However, the bottom-layer disease-specific hospital networks are identical in the three models. Toward the end of this section, two measures of clustering homogeneity used to assess clustering results of all models are explained in detail.

4.1 Disease symptom network for model 1

The top-layer single-view disease symptoms network was extracted from HSDN [20] to represent diseases in CCS code instead of MeSH terminology. It is a super-network of 145 nodes in the Network of Networks data model. Each of these

nodes represents a disease included in this study. As some diseases have incomplete information in the study period and other diseases have no matching MeSH term, the total number of CCS disease codes included in this network is 189 [5]. However, we have reduced this number to 145 diseases/nodes after eliminating diseases that were represented in less than 50% of ERs in California hospitals.

4.2 Disease distribution network for model 2

The top-layer disease network was also extracted from the California State Inpatient Database (SID) as part of the Healthcare Cost and Utilization Project (HCUP) provided by the Agency for Healthcare Research and Quality (AHRQ). A network for 145-disease admission was aggregated monthly from the emergency department records between 2008 and 2011 from over 7 million single-patient discharge records included in the California SID. The similarity among disease nodes was measured using Kullback-Leibler divergence similarity for each month separately. Fusion of the multi-view monthly networks into a single disease distribution network is done for each year separately using the multi-view Network Fusion for Composite Community Extraction (NF-CCE) [6].

4.3 Disease composite network for model 3

The third disease composite network is constructed as a composite of the disease symptoms similarity network acquired for model 1 and a fused monthly disease distribution network acquired for model 2. Since the fused monthly disease distribution network generated in model 2 is different for every year, the resulted disease composite network for model 3 is also different for every year.

4.4 Disease-specific hospital networks

The proposed disease-specific hospital clustering model was evaluated on 145 networks, where each network has up to 152 nodes. Hospital data used in this study are also extracted from the California SID as part of (HCUP) provided by the (AHRQ).

For each disease included in this study, the data of hospitals monthly admission distribution from emergency department between 2008 and 2011 are aggregated for the principle diagnosis using over 7 million single-patient discharge records included in the California SID. Kullback-Leibler divergence was used to measure how the monthly admission distribution diverges among hospitals for every year separately. The total number of hospitals used in this study was 152 hospitals out of 500 California hospitals where hospitals with insufficient number of admission records for some diseases over the study period were excluded. In particular, a hospital was excluded from the study when it was represented in less than 50% of disease-specific hospital networks.

4.5 Homogeneity measures

Clustering homogeneity is used to evaluate our three models with respect to two measures, singular homogeneity and group homogeneity.

The singular homogeneity among different hospital networks represents the percentage of hospitals that belong together to the same clusters across different networks. For example, in case of two hospital clusters, assume that three hospitals (A, B and C) exist in both hospital networks 1 and 2. Assume further that in network 1 hospitals A and B belong to hospital cluster HC1 and hospital C belongs to HC2, while in network 2 hospitals A and C belong to cluster HC1 and hospital B belongs to HC2. In such a case, the singular homogeneity value between networks 1 and 2 is the total number of hospitals that belong to the same hospital clusters across both networks, which in our case is only hospital A, divided by the number of hospitals that exist in both networks, which in our example are hospitals A, B and C. Therefore, in this case singular homogeneity between two networks is computed as (the number of hospitals that belong to the HC1 across both networks + the number of hospitals that belong to the HC2 across both networks) divided by the number of hospitals in both networks. So, for the previous example, singular homogeneity between networks 1 and 2 is $(1 + 0)/3 = 1/3$ or 33%.

The group homogeneity among different hospital networks is defined as the percentage of the largest group of hospitals that belong to same cluster across different networks. For example, in case of 2 hospital clusters, if there are four hospitals (A, B, C and D) in both hospital networks 1 and 2, and A, B and D belong to cluster HC1 while C belongs to HC2 in network 1 and hospitals A and C belong to cluster HC1 while B and D belong to HC2 in network 2, the group homogeneity value between networks 1 and 2 is the maximum number of hospitals that belong to the same hospital clusters across both networks, which is either hospital A that belongs to HC1 at both networks or hospitals B and D that belong together to HC1 in network 1 and belong together to HC2 in network 2, divided by total number of hospitals in networks 1 and 2, which are hospitals A, B, C and D. Therefore, group homogeneity between these two networks is computed as maximum of different set of hospitals grouped together in both networks 1,2 divided by total number of hospitals in both networks. For the previous example, group homogeneity between networks 1 and 2 is $\max(1, 2)/4 = 2/4$ or 50%.

5 Results

The Network of Networks data model utilized in this study is a large-scale network composed of a top-layer super-network and bottom-layer sub-networks. In our model, each sub-network is called a disease-specific hospital network. It represents hospitals that have a certain monthly admission distribution for a specific disease. Some hospitals have no representation in some of these disease-specific networks as these hospitals have not treated these diseases in that period.

However, the interconnectivity between these sub-networks is modeled as the top-layer disease super-network. There are three different views for this disease network, and therefore, there are three interconnectivity models, and hence, three different clustering results for these diseases. Every disease node in the top-layer disease network is a supernode that represents a disease-specific hospital network at the bottom-layer set of networks. Hospitals are clustered into three hospital clusters at each one of the disease-specific networks.

5.1 Clustering of disease networks

The clustering structure in the top-layer disease network is used to guide clustering structures in different disease-specific hospital sub-networks at the bottom layer. Therefore, this subsection describes the clustering result of the three different disease networks.

Disease symptom network for model 1 The top-layer single-view disease symptoms network was extracted from HSDN [20]. It is a super-network of 145 nodes in the Network of Networks data model. The 145 diseases are grouped into three main clusters. The first cluster has 33 diseases while the second and the third clusters have 57 and 54 diseases, respectively. Symmetric nonnegative matrix factorization is used to partition the top-layer disease network under column orthonormal constraints. Nonnegative low-dimensional factors define the probability for each disease node to belong to one of the main clusters. The top five diseases with the highest probability to belong to each of the three clusters are listed in Table 1.

As evident from Table 1, each of the three clusters groups diseases that have similar symptoms. In particular, the first cluster dominantly contains diseases that are pregnancy and childbirth related. Diseases in the second cluster are related to respiratory and gastrointestinal disease. The last set of diseases has the wider range of diseases that might have general similar symptoms such as inflammation, infection, nonspecific chest pain, other hereditary and nervous system conditions.

Disease distribution network for model 2 The top-layer disease network obtained by the second model is also a network of 145 disease admissions. It was aggregated monthly from the emergency department records between 2008 and 2011 from over 7 million single-patient discharge records included in the California SID. Then, the multi-view monthly networks were fused into a single 145-disease distribution network which is done for each year separately. Table 2 lists the top five diseases with the highest probability to belong to each of the three clusters in 2008, and it also shows to which clusters these diseases were grouped over the following three years (2009, 2010 and 2011) with the corresponding probabilities.

In this table, the different clusters showed the severity to be admitted into ED as monthly distribution where disease symptoms are not similar. The first cluster found in the 2008 data includes a variety of diseases such as acute and unspecified renal failure, aspiration pneumonitis, congestive heart failure and other liver diseases. This set of diseases has similar monthly distribution for all hospitals. Conditions associated with these diseases are well defined, and severity of these conditions is similar, often requiring immediate intervention. The second cluster found in the 2008 data has not very specified conditions in its category, but quite serious conditions such as other complications of pregnancy, other upper respiratory diseases and other fractures. The third cluster found in the 2008 data contains diseases that have similar monthly admission but express symptoms very different than the previous two. This cluster includes epilepsy, asthma and abdominal pain.

Disease composite network for model 3 The top-layer disease network in the third model (145 diseases) is constructed as a composite of the disease symptoms similarity network acquired for model 1 and a fused monthly disease distribution network acquired for model 2. Table 3 shows the top five diseases with the highest probability to belong to each of the three clusters in the disease composite network in 2008. As earlier, the table also reports the most likely clusters for these diseases in 2009, 2010 and 2011 with the corresponding probabilities.

As shown in Table 3, diseases in the three clusters have different levels of severity in symptoms and in the need to be admitted into ED. The first cluster includes severe conditions that typically require fast intervention by being admitted to the ED such as open wounds of head, neck and trunk, complication with pregnancy, and gangrene. The second cluster seems to be less severe in symptoms and therefore, the less priority to be admitted such as ovarian cysts, other upper respiratory infections and diabetes mellitus without complication. The last cluster is related to some conditions that do not show severe symptoms that need immediate intervention or enough priority to be admitted compared to the other clusters such as secondary malignancies, pathological fracture and cancer of either colon, bronchus or lung.

Table 1 Top five diseases with their probability to belong to the three main clusters in the disease symptoms network for model 1

Probability	Main cluster	CCS code	Disease name
1.00	1	183	Hypertension complicating pregnancy, childbirth and the puerperium
1.00	1	186	Diabetes or abnormal glucose tolerance complicating pregnancy; childbirth; or the puerperium
0.951	1	190	Fetal distress and abnormal forces of labor
0.951	1	192	Umbilical cord complication
0.951	1	189	Previous C-section
1	2	63	Diseases of white blood cells
1	2	125	Acute bronchitis
1	2	126	Other upper respiratory infections
1	2	133	Other lower respiratory disease
1	2	134	Other upper respiratory disease
1.00	3	79	Parkinson's disease
1.00	3	81	Other hereditary and degenerative nervous system conditions
1.00	3	82	Paralysis
1.00	3	102	Nonspecific chest pain
1.00	3	225	Joint disorders and dislocations; trauma related

5.2 The comparison of disease clustering

Two different views of disease network were extracted from a big medical bibliographic literature database and from patient admission records that include both medical and sociodemographic information. The effect of applying the three models for characterization of different views of disease networks is compared in this section. It shows the different underlying clustering structures for different views. Also, the fusion of the two different views of the disease network in model 1 and 2 reveals another hidden pattern of the disease clustering in model 3.

The underlying clustering structures of three different groups of similar-symptom diseases using the three different models are shown in Table 4. Diseases that are shown up as one of the top five diseases in each of the three clusters of the three different models are chosen for this comparison. The list of these disease is ordered in a way to group the one that has similar behavior over different models.

We observed that diseases with similar symptoms have different underlying clustering structures in disease admission distribution network for all the three different clusters of similar symptoms disease in model 1. However, there are a high probability for diseases with similar symptoms to be grouped in the same cluster in the disease composite network (M3) regardless of the low probability for different monthly admission groups they belong to. For example, this is the case for diseases with CCS codes 185, 189 and 192 and also for diseases 125, 154 and 134 as well as for diseases 102 and 25. This means that the fusion of multi-view networks produces a latent variable that can explain the common structure shared across layers; i.e., it helps to reveal the hidden underlying clustering structure for better clustering results.

5.3 Clustering of hospital sub-networks

Hospital clustering for model 1 Although the results of clustering hospitals vary for different disease-specific hospital networks, the underlying clustering structure is more similar for the disease-specific hospital networks that represent

Table 2 Top five diseases with their probability to belong to three main clusters in the disease distribution network for model 2 in 2008 compared to the most probable clusters for these diseases in years 2009, 2010 and 2011

CCS code	Disease name	2008 Prob	2008 Clust	2009 Prob	2009 Clust	2010 Prob	2010 Clust	2011 Prob	2011 Clust
157	Acute and unspecified renal failure	0.515	1	0.536	3	0.530	3	0.463	2
135	Intestinal infection	0.475	1	0.497	2	0.486	3	0.519	3
129	Aspiration pneumonitis; food/vomitus	0.461	1	0.551	3	0.517	3	0.437	2
108	Congestive heart failure; non-hypertensive	0.455	1	0.514	3	0.470	3	0.464	2
151	Other liver diseases	0.475	1	#N/A	#N/A	0.465	3	0.405	2
181	Other complications of pregnancy	0.481	2	#N/A	#N/A	0.511	1	0.426	2
233	Intracranial injury	0.480	2	0.495	1	0.487	1	0.496	1
134	Other upper respiratory disease	0.476	2	0.456	1	0.476	1	#N/A	#N/A
231	Other fractures	0.464	2	0.506	1	0.486	1	0.502	1
160	Calculus of urinary tract	0.460	2	#N/A	#N/A	0.462	1	0.430	2
154	Noninfectious gastroenteritis	0.542	3	0.512	2	0.539	2	0.508	3
83	Epilepsy; convulsions	0.537	3	0.518	2	0.516	2	0.516	3
142	Appendicitis and other appendiceal conditions	0.527	3	0.533	1	0.444	2	0.538	1
128	Asthma	0.527	3	0.534	2	0.504	2	0.522	3
251	Abdominal pain	0.522	3	0.503	2	0.568	2	0.501	3

#N/A, not one of the top 5 diseases in that year

disease supernodes belonging to the same main cluster at the top-layer disease network as it is concluded in our preliminary work [1]. All data reported in the following subsections are for 2008. The results for 2009–2011 are similar and are therefore omitted.

Table 5 shows the summary of hospitals clustering analysis for each disease-specific network for model 1. It is organized into three sections. The first five rows exhibit the data for five disease-specific hospital networks that represent the top five diseases belonging to the first main cluster of disease symptoms network of model 1. Next section, which is the next five rows, shows the data for five hospital networks that represent the top five diseases belonging to the second main clusters of disease symptoms network of model 1. And last five rows in Table 5 are the data of disease-specific hospital networks that represents the top five diseases belonging to the third main clusters of disease symptoms network of model 1.

Each row in this table has the following structure: the name of the disease-specific hospital network, the CCS code for the disease, the main cluster to which that disease belongs at the top layer, its probability to belong to that main cluster, and the remaining are about the percentage of hospitals in that disease-specific hospital network in different categories. These categories are: the percentage of hospitals that have no admission for the specific disease, the percentage of hospitals that have admitted at least one patient for the given disease out of the 152 hospitals included in this study, the percentage of hospitals that have admission for this disease in the first

Table 3 Top five diseases with their probability to belong to three main clusters in the disease composite network for model 3 in 2008 compared to the most probable clusters for these diseases in years 2009, 2010 and 2011

CCS code	Disease name	2008 Prob	2008 Clust	2009 Prob	2009 Clust	2010 Prob	2010 Clust	2011 Prob	2011 Clust
235	Open wounds of head; neck; and trunk	0.583	1	0.592	2	0.583	3	0.476	2
191	Polyhydramnios and other problems of amniotic cavity	0.565	1	0.532	2	0.543	3	#N/A	#N/A
225	Joint disorders and dislocations; trauma related	0.560	1	0.582	2	0.579	3	0.491	2
186	Diabetes or abnormal glucose tolerance complicating pregnancy; childbirth; or the puerperium	0.545	1	0.570	2	0.572	3	0.511	3
248	Gangrene	0.539	1	0.596	2	#N/A	#N/A	0.505	2
181	Other complications of pregnancy	0.506	2	0.570	3	0.479	2	#N/A	#N/A
172	Ovarian cyst	0.503	2	0.606	3	#N/A	#N/A	0.560	3
126	Other upper respiratory infections	0.500	2	0.545	3	0.471	2	#N/A	#N/A
49	Diabetes mellitus without complication	0.497	2	0.613	3	0.491	2	0.572	3
142	Appendicitis and other appendicular conditions	0.497	2	0.565	3	0.496	2	0.489	3
42	Secondary malignancies	0.520	3	0.460	1	0.502	1	#N/A	#N/A
114	Peripheral and visceral atherosclerosis	0.505	3	#N/A	#N/A	0.493	1	0.472	2
207	Pathological fracture	0.484	3	0.457	1	0.495	1	#N/A	#N/A
19	Cancer of bronchus; lung	0.469	3	0.445	1	0.496	1	#N/A	#N/A
14	Cancer of colon	0.467	3	0.425	1	0.499	1	#N/A	#N/A

#N/A, not one of the top five diseases in that year

hospital cluster, in the second hospital cluster and in the third hospital cluster, respectively.

For example, the first row of Table 5 reports the data for hospitals that had admission for hypertension complicating pregnancy condition (CCS code 183). This network represents the supernode with the same name at the top-layer disease symptoms network. This supernode belongs to main cluster 1 with high probability (100%). This disease-specific hospital network has 64% hospitals (out of 152 hospitals) where 39% of the available hospitals belong to hospital cluster 1, 32% belong to hospital cluster 2 and 29% belong to hospital cluster 3.

As observed from Table 5, the first set of disease-specific hospital networks that represent the top five diseases belong-

Table 4 Comparison of clustering results of disease networks in the three models (M1, M2 and M3). The numbers show the probability of the disease belonging to clusters (C1, C2, C3) in the different models (M1, M2 and M3)

CCS code	Disease name	M1 C1	M1 C2	M1 C3	M2 C1	M2 C2	M2 C3	M3 C1	M3 C2	M3 C3
192	Umbilical cord complication	0.951	0	0	0.387	0	0	0.566	0	0
189	Previous C-section	0.951	0	0	0.374	0	0	0.556	0	0
181	Other complications of pregnancy	0.951	0	0	0	0.481	0	0	0.506	0
114	Peripheral and visceral atherosclerosis	0.717	0	0	0	0.442	0	0	0	0.505
185	Prolonged pregnancy	0.951	0	0	0	0	0.366	0.573	0	0
127	Chronic obstructive pulmonary disease and bronchiectasis	0	0.865	0	0.459	0	0	0	0	0.378
122	Pneumonia (except caused by tuberculosis or STD)	0	0.800	0	0.459	0	0	0	0.377	0
134	Other upper respiratory disease	0	1.00	0	0	0.476	0	0	0.377	0
154	Noninfectious gastroenteritis	0	1.00	0	0	0	0.542	0	0.441	0
125	Acute bronchitis	0	1.00	0	0	0	0.442	0	0.420	0
225	Joint disorders and dislocations;	0	0	1.00	0.407	0	0	0.560	0	0
159	Urinary tract infections	0	0	0.500	0.451	0	0	0	0.378	0
102	Nonspecific chest pain	0	0	1.00	0	0	0.399	0	0.461	0
251	Abdominal pain	0	0	1.00	0	0	0.522	0	0.383	0
109	Acute cerebrovascular disease	0	0	0.678	0	0	0.369	0	0	0.434

ing to the first main cluster of disease symptoms network has between 31 and 64% of hospitals in each network out of 152 hospitals included in the study. The second set of networks has more hospitals in each network; they are between 91 and 99% of 152 hospitals. The third set of similar symptom disease-specific hospital networks has more varying hospital number across these networks. These networks have between 55 and 89% of the 152 hospitals. However, the percentages of hospitals in each hospital cluster in all different disease-specific hospital networks listed in Table 5 showed balanced clustering results.

Hospital clustering for model 2 Table 6 shows the summary of hospitals clustering analysis for each disease-specific network for model 2. Hospital clustering in model 2 is regularized by clusters of the disease distribution network. It is also organized into three sections. Each section shows the data for five disease-specific hospital networks representing the top five diseases that belong to one of the three main clusters of the disease distribution network for model 2. The structure of this table is similar to the structure of Table 5.

As observed from Table 6, the three sets of disease-specific hospital networks that represent the top five diseases belonging to the three main clusters of Disease Distribution Network have almost all hospitals in the study. There are between 91

Table 5 Hospital clustering analysis for the first model (M1)

Disease-specific hospital network	CCS code	Main Clust	Prob. Main Clust	No adm.	Hosp. adm.	Hosp. in Clus1	Hosp. in Clus2	Hosp. in Clus3
Hypertension complicating pregnancy; childbirth; and the puerperium	183	1	1.00	0.36	0.64	0.39	0.32	0.29
Diabetes or abnormal glucose tolerance complicating pregnancy; childbirth; or the puerperium	186	1	1.00	0.57	0.43	0.34	0.26	0.40
Umbilical cord complication	192	1	0.95	0.69	0.31	0.40	0.26	0.40
Previous C-section	189	1	0.95	0.64	0.36	0.40	0.33	0.34
Fetal distress and abnormal forces of labor	190	1	0.95	0.66	0.34	0.33	0.27	0.27
Acute bronchitis	125	2	1.00	0.02	0.98	0.32	0.41	0.27
Other upper respiratory infections	126	2	1.00	0.05	0.95	0.31	0.30	0.40
Other lower respiratory disease	133	2	1.00	0.01	0.99	0.33	0.36	0.31
Other upper respiratory disease	134	2	1.00	0.09	0.91	0.32	0.35	0.33
Diseases of white blood cells	63	2	1.00	0.07	0.93	0.33	0.29	0.38
Paralysis	82	3	1.00	0.45	0.55	0.29	0.40	0.31
Parkinson's disease	79	3	1.00	0.29	0.71	0.32	0.37	0.31
Nonspecific chest pain	102	3	1.00	0.00	1.00	0.32	0.28	0.40
Other hereditary and degenerative nervous system conditions	81	3	1.00	0.11	0.89	0.33	0.33	0.34
Joint disorders and dislocations; trauma related	225	3	1.00	0.19	0.81	0.35	0.34	0.31

Each row represents a disease-specific hospital network

Table 6 Hospital clustering analysis for the second model (M2)

Disease-specific hospital network	CCS code	Main Clust	Prob. Main Clust	No adm.	Hosp. adm.	Hosp. in Clus1	Hosp. in Clus2	Hosp. in Clus3
Acute and unspecified renal failure	157	1	0.51	0.01	0.99	0.64	0.09	0.27
Intestinal infection	135	1	0.47	0.01	0.99	0.35	0.50	0.15
Aspiration pneumonitis; food/vomitus	129	1	0.46	0.02	0.98	0.45	0.15	0.40
Congestive heart failure; non-hypertensive	108	1	0.45	0.01	0.99	0.46	0.52	0.02
Other liver diseases	151	1	0.47	0.03	0.97	0.09	0.57	0.34
Other complications of pregnancy	181	2	0.48	0.06	0.94	0.55	0.15	0.30
Intracranial injury	233	2	0.48	0.07	0.93	0.13	0.33	0.55
Other upper respiratory disease	134	2	0.47	0.09	0.91	0.38	0.22	0.40
Other fractures	231	2	0.46	0.00	1.00	0.36	0.25	0.39
Calculus of urinary tract	160	2	0.46	0.01	0.99	0.33	0.34	0.33
Noninfectious gastroenteritis	154	3	0.54	0.01	0.99	0.28	0.40	0.32
Epilepsy; convulsions	83	3	0.54	0.00	1.00	0.16	0.51	0.33
Appendicitis and other appendiceal conditions	142	3	0.53	0.00	1.00	0.82	0.04	0.14
Asthma	128	3	0.53	0.00	1.00	0.24	0.61	0.15
Abdominal pain	251	3	0.52	0.00	1.00	0.16	0.54	0.30

Each row represents a disease-specific hospital network

Table 7 Hospital clustering analysis for the third model (M3)

Disease-specific hospital network	CCS code	Main Clust	Prob. Main Clust	No adm.	Hosp. adm.	Hosp. in Clus1	Hosp. in Clus2	Hosp. in Clus3
Open wounds of head; neck; and trunk	235	1	0.58	0.13	0.86	0.38	0.40	0.24
Polyhydramnios and other problems of amniotic cavity	191	1	0.57	0.63	0.36	0.43	0.34	0.27
Joint disorders and dislocations; trauma related	225	1	0.56	0.19	0.80	0.23	0.36	0.43
Diabetes or abnormal glucose tolerance complicating pregnancy; childbirth; or the puerperium	186	1	0.55	0.57	0.41	0.40	0.37	0.27
Gangrene	248	1	0.54	0.22	0.77	0.32	0.36	0.34
Other complications of pregnancy	181	2	0.506	0.06	0.93	0.35	0.35	0.31
Ovarian cyst	172	2	0.503	0.13	0.86	0.27	0.30	0.44
Other upper respiratory infections	126	2	0.5	0.05	0.93	0.24	0.44	0.33
Diabetes mellitus without complication	49	2	0.497	0.31	0.68	0.28	0.32	0.40
Appendicitis and other appendicular conditions	142	2	0.497	0.00	0.99	0.37	0.29	0.35
Secondary malignancies	42	3	0.52	0.03	0.95	0.28	0.28	0.46
Peripheral and visceral atherosclerosis	114	3	0.51	0.05	0.94	0.48	0.30	0.23
Pathological fracture	207	3	0.49	0.08	0.91	0.30	0.39	0.42
Cancer of bronchus; lung	19	3	0.47	0.05	0.94	0.38	0.38	0.24
Cancer of colon	14	3	0.48	0.05	0.94	0.43	0.34	0.27

Each row represents a disease-specific hospital network

and 100% of hospitals in each network out of 152 hospitals included in the study. However, the percentages of hospitals in each hospital cluster in all different disease-specific hospital networks listed in Table 6 showed unbalanced clustering results in most of these networks.

Hospital clustering for model 3 Results of clustering hospitals for each disease-specific network using the disease composite network for Model 3 are shown in Table 7. The first five rows report the data for the first main cluster in model 3 while data of the second main cluster and for the third main cluster are shown at rows 6–10 and 11–15, respectively.

Table 7 shows that the three sets of disease-specific hospital networks that represent the top five diseases belonging to the three main clusters of disease composite network of Model 3 have more than about 70% of the hospitals in the study except two networks in the first set. There are between 68 and 99% of hospitals in each network out of 152 hospitals included in the study. Only hospital networks of diseases with CCS codes 186 and 191 have 41% and 36% of hospitals, respectively. Moreover, the percentages of hospitals in each hospital cluster in all different disease-specific hospital networks listed in Table 7 showed balanced clustering results in most of these networks.

5.4 Clustering homogeneity

The clustering homogeneity is evaluated in the sense of belongingness. Two measures of belonging defined in Sect. 4.5 are calculated.

The first measure is the singular homogeneity that represents the percentage of hospitals that belong to the same clusters across different networks. A comparison of the three models is demonstrated in Fig. 5, where the X-axis and Y-axis represent the 145 disease-specific hospital networks that are listed in combination of main disease clusters. For instance, the disease-specific hospital networks that represent diseases at top-layer disease network and belong to the top-layer disease cluster 1 are listed first, and then the networks that represent diseases that belong to top-layer disease cluster 2 are listed second and so on. In this way, the effect of different disease networks can be presented in these figures. Every data point in these figures represents the singular homogeneity in both x and y disease-specific hospital networks. It is evident that according to singular homogeneity, Model 3 resulted in more compact and better separated groups of hospitals as compared to the results obtained by Models 1 and 2. This provides the initial evidence that it is beneficial to consider the disease network composed of both a disease symptom network and a fused network of the 12-monthly disease distribution networks.

The other measure considered in evaluating three clustering methods is group homogeneity (also defined in Sect. 4.5) that measures for a group of hospitals a fraction that belongs together to the same group across networks. In other words, group homogeneity is aimed to characterize hospitals that belong to the same group across different networks

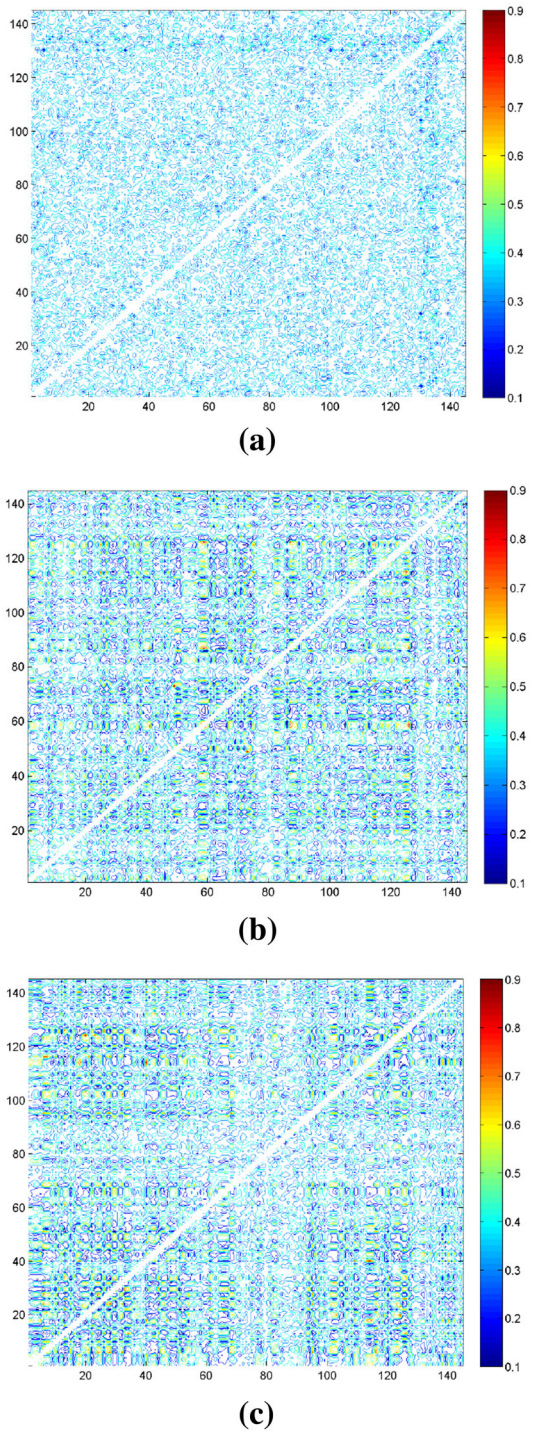

Fig. 5 Singular homogeneity for all models for 2009 where 145 hospital networks are shown at the X- and Y-axis. **a** Singular homogeneity for models 1. **b** Singular homogeneity for models 2. **c** Singular homogeneity for models 3

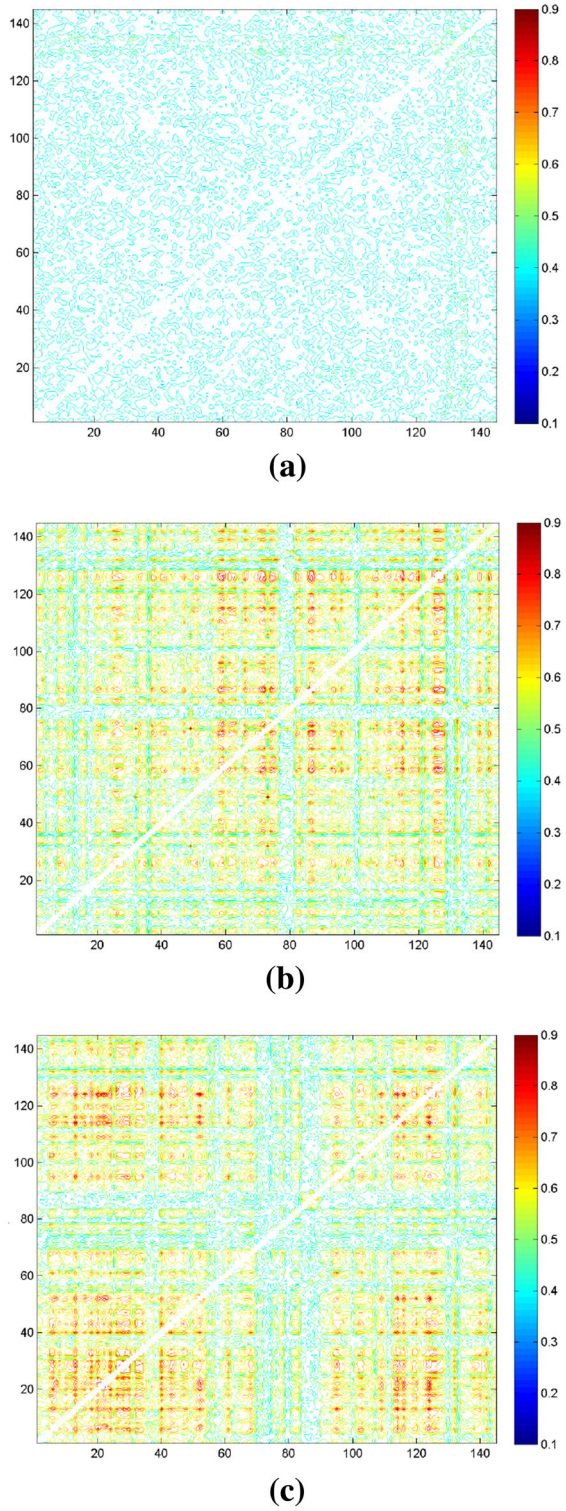

Fig. 6 Group homogeneity for all models for 2009 where 145 hospital networks are shown at the X- and Y-axis. **a** Group homogeneity for models 1. **b** Group homogeneity for models 2. **c** Group homogeneity for models 3

despite the cluster category. The result of comparison of the three models according to group homogeneity is shown in Fig. 6. Model 3 again outperforms the other two models in explaining and predicting the underlying clustering structure for ER hospitals admission distribution. The results obtained by Model 3 are more homogeneous among networks belonging to the same disease cluster as evident from Fig. 6. This means that the fusion of multi-view networks produces a latent variable that can explain the hidden common structure shared across layers while integrating results with multi-domain networks helps to reveal the underlying clustering structure.

Singular homogeneity, which is the percentage of hospitals that belong to same clusters across different disease-based hospital networks, was consistent for Model 3 over the 4 years (see Fig. 7). The other two models are also consistent but figures were omitted due to space limitation. Also, Fig. 8 shows consistency on group homogeneity for Model 3 over the 4 years. The other two models were also consistent but figures were also omitted due to space limitation since the results obtained by Model 3 were clearly more homogeneous for every four years considered.

5.5 The comparison of hospital clustering

In order to analyze the performance of the different three models, group homogeneity of the top five hospital networks in each main cluster at each model is listed in Table 8. Different table shows fraction of available hospitals in each network that belongs together to the same group across these networks. Each set of row tables represents the three sets of disease-specific hospital networks for each model. For example, at the disease-specific hospital networks of main cluster 1 at model 1, which is the top-left table, 40% of shared hospitals at the disease-183 hospital network and hospitals at disease-186 hospital network belong to the same cluster over these two networks taking into consideration that disease-183 hospital network has 64% of the 152 hospitals and disease-186 hospital network has 43% of the 152 hospitals.

Comparing the three models, the first model has group homogeneity measures ranging between 36 and 47%. The diagonal numbers show the percentage of hospitals in a single network. The second row shows the group homogeneity measures for the networks of the second models. Considering the large number of hospitals included into each network at model 2 as shown in the diagonal cells, the group homogeneity measures ranging between 33 and 60% with average measures close to 50% are considered low in comparison with the third model. In the third model, homogeneity measures range between 38 and 76% with average measures close to 60%.

These numbers provide evidence that the third model outperformed the first and the second models and revealed

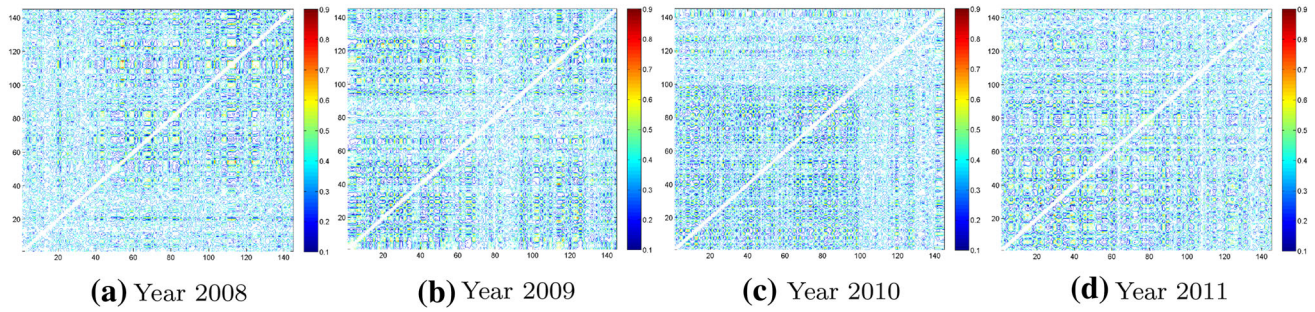

Fig. 7 Singular homogeneity for model 3 over 4 years where 145 hospital networks are shown at the X- and Y-axis

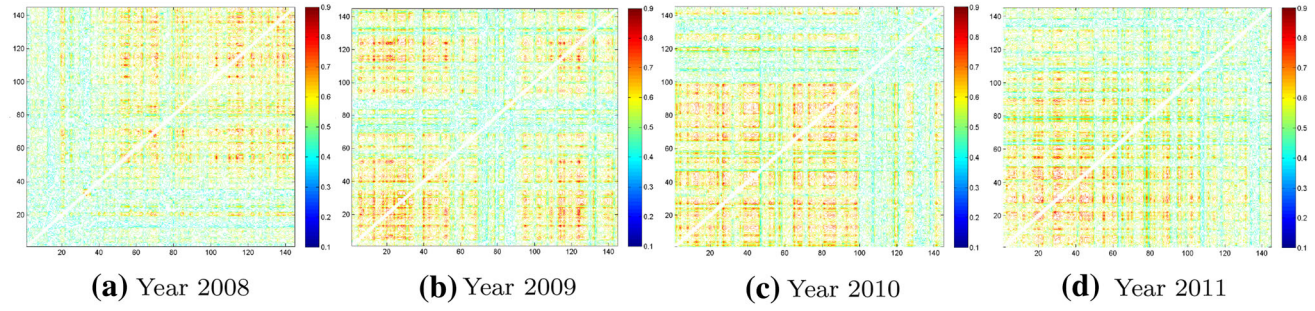

Fig. 8 Group homogeneity for model 3 over 4 years where 145-different hospital networks at the X- and Y-axis

Table 8 Comparison of hospital clustering homogeneity in the three models

a. Hospital clustering homogeneity in the first model

	183	186	192	189	190		125	126	133	134	63		82	79	102	81	225
183	0.64	0.40	0.50	0.40	0.39	125	0.98	0.39	0.36	0.39	0.37	82	0.55	0.43	0.47	0.43	0.47
186	0.40	0.43	0.44	0.41	0.38	126	0.39	0.95	0.39	0.37	0.42	79	0.43	0.71	0.41	0.38	0.39
192	0.50	0.44	0.31	0.48	0.47	133	0.36	0.39	0.99	0.40	0.41	102	0.47	0.41	1.00	0.37	0.36
189	0.40	0.41	0.48	0.36	0.58	134	0.39	0.37	0.40	0.91	0.36	81	0.43	0.38	0.37	0.89	0.39
190	0.39	0.38	0.47	0.58	0.34	63	0.37	0.42	0.41	0.36	0.93	225	0.47	0.39	0.36	0.39	0.81

b. Hospital clustering homogeneity in the second model

	157	135	129	108	151		181	233	134	231	160		154	83	142	128	251
157	0.99	0.47	0.55	0.56	0.52	181	0.94	0.49	0.58	0.48	0.53	154	0.99	0.58	0.42	0.52	0.52
135	0.47	0.99	0.52	0.41	0.48	233	0.49	0.93	0.58	0.49	0.54	83	0.58	1.00	0.46	0.55	0.48
129	0.55	0.52	0.98	0.48	0.48	134	0.58	0.58	0.91	0.47	0.33	142	0.42	0.46	1.00	0.49	0.48
108	0.56	0.41	0.48	0.99	0.60	231	0.48	0.49	0.47	1.00	0.56	128	0.52	0.55	0.49	1.00	0.53
151	0.52	0.48	0.48	0.60	0.97	160	0.53	0.54	0.33	0.56	0.99	251	0.52	0.48	0.48	0.53	1.00

c. Hospital clustering homogeneity in the third model

	235	191	225	186	248		181	172	126	49	142		42	114	207	19	14
235	0.86	0.53	0.62	0.50	0.45	181	0.93	0.67	0.57	0.60	0.57	42	0.95	0.76	0.71	0.56	0.66
191	0.53	0.36	0.40	0.53	0.38	172	0.67	0.86	0.59	0.56	0.59	114	0.76	0.94	0.62	0.56	0.65
225	0.62	0.40	0.80	0.56	0.47	126	0.57	0.59	0.93	0.57	0.58	207	0.71	0.62	0.91	0.64	0.60
186	0.50	0.53	0.56	0.41	0.63	49	0.60	0.56	0.57	0.68	0.67	19	0.56	0.56	0.64	0.94	0.50
248	0.45	0.38	0.47	0.63	0.77	142	0.57	0.59	0.58	0.67	0.99	14	0.66	0.65	0.60	0.50	0.94

Tables for the three sets of disease-specific hospital networks obtained by the first model for the three main clusters are shown at the top (panel a) followed by the corresponding tables when using the second and the third models (panels b and c, respectively). Each table shows percentage of hospitals that belong to the same cluster at both networks

hidden patterns for the fused multi-view disease networks which enhanced hospital clustering. Therefore, hospitals grouped to the same cluster by the third model in a certain disease-specific hospital network share similar monthly admission distribution taking into account the similar disease symptoms that were considered in the fused disease composite network.

One of the implication of this study relates to enhancing the quality of healthcare services to facilitate assessment and timely detection for hospital needs in ED admission for certain diseases. For example, healthcare officials may optimize the use of resources among hospitals by taking into account discovered similarities among hospitals that belong to the same hospital cluster in a disease-specific hospital network.

On the same basis, the rapid development of ER at hospitals could be planned efficiently by providing appropriate infrastructure for similar hospitals that treated similar diseases with similar monthly admission distribution. Additionally, providers seeking to improve the quality and efficiency of healthcare could examine performance across similar hospitals in different clusters to provide a more precise characterization for better bed utilization and personnel distribution. Further extensions of the findings of this study could also provide a foundation and possibly a better platform to predict future resource and/or personnel needs. Analyzing data for a very similar hospital in a disease-specific hospital group would result in higher certainty regarding the prediction results. Also, this type of analysis could improve the efficiency of planning and facilitate risk assessment and timely outbreak detection for certain diseases or conditions.

6 Conclusion

The healthcare industry continues to generate a massive amount of medical data. Hospital-based clustering advances as one of the data-driven insights to be integrated into clinical and operational processes to enhance the quality of healthcare planning and decision making. Toward this objective, our study analyzed more than a million EHR records of patients admitted to emergency departments with no prior schedule at 152 hospitals over four years. Clustering hospital admission based on different diseases helps discover hospital patterns for different diseases. Major findings show different aspects of these patterns based on the different views of disease interconnectivity that were applied. The third model that has applied the fusion of multi-view networks from different domains, literature and medical databases and has integrated results with multi-domain hospital networks produced a latent variable that better revealed the hidden underlying hospital clustering structure. It outperformed other models in clustering homogeneity analysis measures with average measures close to 60% compared to 50% for model 2 and 41% for model 1.

Acknowledgements This research was supported in part by NSF BIGDATA Grant 14476570. Healthcare Cost and Utilization Project (HCUP), Agency for Healthcare Research and Quality, provided part of the data used in this study.

References

Albarakati, N., Obradovic, Z.: Disease-based clustering of hospital admission: disease network of hospital networks approach. In: The IEEE 29th International Symposium on Computer Based Medical Systems (CBMS). IEEE (2017)

Belciug, S., Gorunescu, F.: Improving hospital bed occupancy and resource utilization through queuing modeling and evolutionary computation. J. Biomed. Inform. **53**, 261–269 (2015)

Bourgeois, F.T., Monuteaux, M.C., Stack, A.M., Neuman, M.I.: Variation in emergency department admission rates in us children's hospitals. Pediatrics **134**(3), 539–545 (2014)

Delamater, P.L., Shortridge, A.M., Messina, J.P.: Regional health care planning: a methodology to cluster facilities using community utilization patterns. BMC Health Serv. Res. **13**(1), 333 (2013)

Glass, J., Ghalwash, M.F., Vukicevic, M., Obradovic, Z.: Extending the modelling capacity of Gaussian conditional random fields while learning faster. In: AAAI, pp. 1596–1602 (2016)

Gligorijevic, V., Panagakis, Y., Zafeiriou, S.: Non-negative matrix factorizations for multiplex network analysis. arXiv:1612.00750 (2016)

Groves, P., Kayyali, B., Knott, D., Van Kuiken, S.: The big data revolution in healthcare: accelerating value and innovation. McKinsey Q. **2**, 3 (2013)

Khojah, I., Li, S., Luo, Q., Davis, G., Galarraga, J.E., Granovsky, M., Litvak, O., Davis, S., Shesser, R., Pines, J.M.: The relative contribution of provider and ED-level factors to variation among the top 15 reasons for ED admission. Am. J. Emerg. Med. **35**(9), 1291–1297 (2017). https://doi.org/10.1016/j.ajem.2017.03.074

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J.P., Moreno, Y., Porter, M.A.: Multilayer networks. J. Complex Netw. **2**(3), 203–271 (2014)

McMahon Jr., L.F., Wolfe, R.A., Tedeschi, P.J.: Variation in hospital admissions among small areas: a comparison of Maine and Michigan. Med. Care **27**, 623–631 (1989)

Moni, M.A., Liò, P.: Network-based analysis of comorbidities risk during an infection: SARS and HIV case studies. BMC Bioinform. **15**(1), 333 (2014)

Ni, J., Tong, H., Fan, W., Zhang, X.: Inside the atoms: ranking on a network of networks. In: Proceedings of the 20th SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1356–1365. ACM (2014)

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

Phillip, P.J., Mullner, R., Andes, S.: Toward a better understanding of hospital occupancy rates. Health Care Financ. Rev. **5**(4), 53 (1984)

Raghupathi, W., Raghupathi, V.: Big data analytics in healthcare: promise and potential. Health Inf. Sci. Syst. **2**(1), 3 (2014)

Reid, F.D., Cook, D.G., Majeed, A.: Explaining variation in hospital admission rates between general practices: cross sectional study. BMJ **319**(7202), 98–103 (1999)

Sabbatini, A.K., Nallamothu, B.K., Kocher, K.E.: Reducing variation in hospital admissions from the emergency department for

low-mortality conditions may produce savings. Health Aff. **33**(9), 1655–1663 (2014)

Shay, P.D.: More ore than just hospitals: an examination of cluster components and configurations. Ph.D. dissertation, Health Services Organization and Research Department, Virginia Commonwealth University (2014)

Thomas, J.W., Griffith, J.R., Durance, P.: Defining hospital clusters and associated service communities in metropolitan areas. Socioecon. Plan. Sci. **15**(2), 45–51 (1981)

Zhou, X., Menche, J., Barabási, A.L., Sharma, A.: Human symptoms-disease network. Nat. Commun. **5**, 4212 (2014)