

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**5,000**

Open access books available

**125,000**

International authors and editors

**140M**

Downloads

Our authors are among the

**154**

Countries delivered to

**TOP 1%**

most cited scientists

**12.2%**

Contributors from top 500 universities



**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



# Learning Health-Care Worker Networks from Electronic Health Record Utilization

*You Chen*

## Abstract

The health-care system is a highly collaborative environment where health-care workers collaborate to care for patients. Health-care organizations (HCOs) design and develop various types of staffing plans to promote collaboration among health-care workers. The existing staffing plans describe the cooperation at a coarse-grained level, such as team scheduling. They seldom consider connections among health-care workers and investigate how health-care workers receive and disseminate information, which is essential evidence to inform actionable staffing interventions to improve care quality and patient safety. In this chapter, we introduce how to apply network analysis methods to electronic health record (EHR) utilization data to learn connections among health-care workers and build networks to describe teamwork in a fine-grained level. The chapter includes: (i) a brief description of the EHR utilization data, (ii) approaches to learn connections among health-care workers, (iii) building health-care worker networks, (iv) developing survey instruments to validate health-care worker networks, (v) introducing sociometric measurements to quantify network structures and positions of health-care workers in the networks, (vi) using statistical models to test associations between teamwork structures and patient outcomes, and (vii) listing examples to learn health-care worker networks in an HCO and a specific setting, including neonatal intensive care unit and trauma.

**Keywords:** network analysis, methodology, collaboration, care team, patient outcome, electronic health record, data-driven, data mining, bottom-up, health-care worker network, health-care organization, sociometric measurement, audit logs, statistical model, survey instrument, network structure

## 1. Introduction

The United States health-care system has been moving to patient-centered care by incorporating different levels of collaborations, including those occurring within a health-care organization (HCO) or between HCOs [1, 2]. A classic model [3] proposed to understand patient-centered care divides the health-care system into four nested levels: (1) the individual patient; (2) the care team made up of health-care workers (e.g., clinicians, pharmacists, social workers, and utilization managers) to care for patients; (3) the HCO (e.g., hospital, clinic, and nursing home) that supports the development and work of care teams by providing infrastructure and complementary resources; and (4) the political and economic

environment (e.g., regulatory, financial, payment regimes, and markets) that support hospital collaborations with other HCOs and payers on population health management. To promote patient-centered care, HCOs create infrastructures and develop staffing strategies to encourage collaboration among health-care workers to care for patients [4, 5]. Collaboration among health-care workers can improve care quality (e.g., reducing readmission rates) [6], patient safety (e.g., preventing medical errors) [7], and patient outcome (shortening length of stay) [8–10].

Staffing plans describe collaboration at a macro-level. For instance, an intensive care unit (ICU) may use an intensivist-centered care team (closed model) or an ad hoc group consisting of nurses, nurse practitioners, and physicians (open model) to care for critically ill patients [11]. The macro-level staffing strategies seldom specify how health-care workers connect and how they receive and disseminate information to care for patients. Thus, it is difficult for HCOs to monitor those top-down staffing strategies implemented in clinical practice. Without the micro-level knowledge of teamwork (e.g., health-care worker connection), it is challenging for HCOs to assess their staffing strategies to identify inefficient and ineffective parts for further collaboration optimization.

Measuring connections among health-care workers is very challenging due to complex clinical workflows and dynamic structures of teamwork [12, 13]. That is also one of the reasons why HCOs do not specify connections among health-care workers in their staffing plans. Recent studies show connections among health-care workers can be learned from their activities in electronic health record (EHR) systems [14–19]. EHR systems are a platform used by health-care workers to diagnose patients and exchange diagnostic results [20, 21]. In modern health-care environments, an increasing number of health-care workers utilize EHR systems as the primary tool to diagnose patients and exchange health information [22]. Therefore, the volume and scale of the EHR system utilization data have been increasing exponentially in recent years, which provide abundant resources for researchers to learn collaborations through the EHR system utilization [14–19].

In this chapter, we provide a network analysis of the EHR system utilization data to learn teamwork structures and specify connections among health-care workers. We believe the chapter can provide researchers a new way to model teamwork/collaboration in health care. We anticipate the data, methods, and applications introduced in this chapter will be of interest to the teamwork in health-care readership, particularly those focused on network analysis, secondary data analysis, EHR utilization, and care teams.

## **2. EHR system utilization data**

EHR systems provide a platform for care coordination across a diverse collection of health-care workers [22–25]. Coordination activities occurring in EHR systems play an increasingly important role in the establishment of high-efficient health-care worker collaboration networks. Various studies, including our prior research, have leveraged health-care worker activities in EHR systems to infer patterns of collaboration [9, 10, 14–19]. The proportion of care activities performed via EHR systems has steadily increased with the adoption of EHR through meaningful use of incentives [22, 26].

Health-care worker activities occurring in EHR systems have been documented in the form of audit logs. When a provider accesses or moves between modules in the EHR interface, such as moving from Progress Notes to Order Entry, a record of these activities are documented, including the time the event occurred, the health-care worker and patient IDs, and the computer location. Audit logs include all

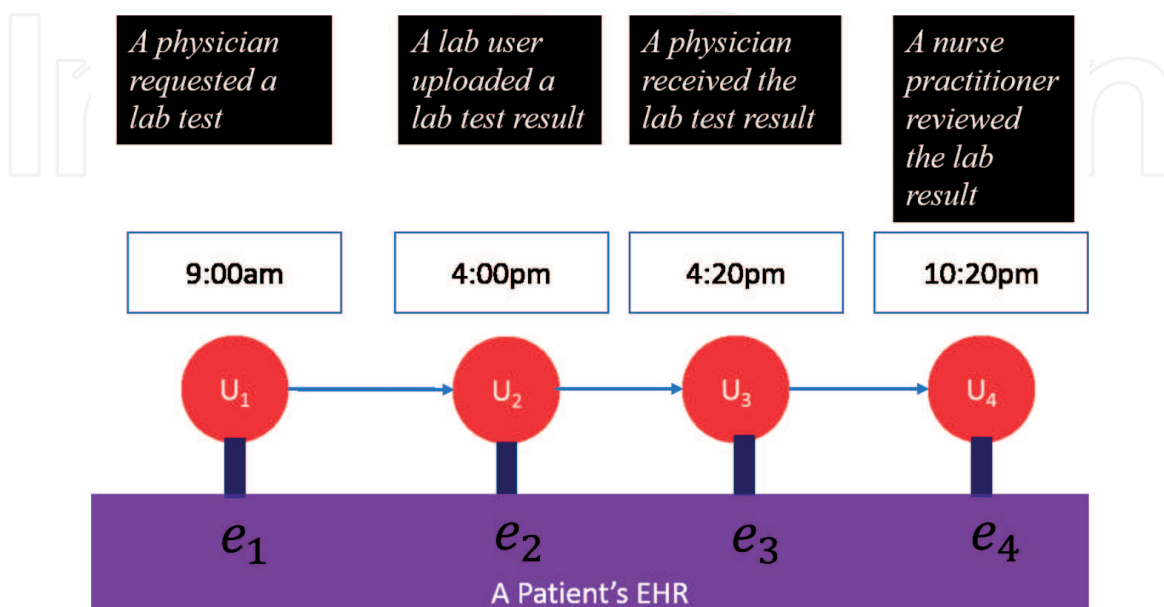
health-care worker interactions to EHRs of patients, which provides an opportunity to study connections among health-care workers. The continuous data collection of the EHR audit logs provides robust, readily available data. Since health-care worker activity is documented in the EHR in near real time, it is free from recall bias and variation introduced when health-care workers are retrospectively surveyed to describe their activities in EHRs.

The activities performed by health-care workers stem from six primary sources [10], including conditions (e.g., assigning a diagnosis), procedures (e.g., intubation), medications (e.g., prescription), notes (e.g., progress note writing), orders (e.g., laboratory test ordering), and measurements (e.g., measuring respiratory rate).

**Figure 1** shows an example to illustrate health-care worker activities in EHR systems. Each event, such as *requesting a lab test*, includes a health-care worker, an EHR, and the time stamp. The four events depicted in the example demonstrate the hidden collaborations between health-care workers. For instance, the physician ordered a lab test and shared the order with the lab user; next, the lab user conducted the laboratory test and shared the test results of the patient with a health-care worker in the physician office; finally, the nurse practitioner reviewed and analyzed the results.

### 3. Transforming utilization data into matrices

Events document interactions of health-care workers to EHRs of patients, but they do not capture the direct connections among health-care workers. As shown in **Figure 1**, the four health-care workers performed events to EHRs of a patient, and they are not directly connected. We leverage events to measure the hidden connections among health-care workers. A hidden connection between two health-care workers is defined based on their interactions with the EHRs of patients. We call a hidden connection as an indirect relationship, because the two health-care workers do not communicate directly, but care for the same patients via performing actions to their EHRs. For instance, a physician ordered a lab test and sent the order to a

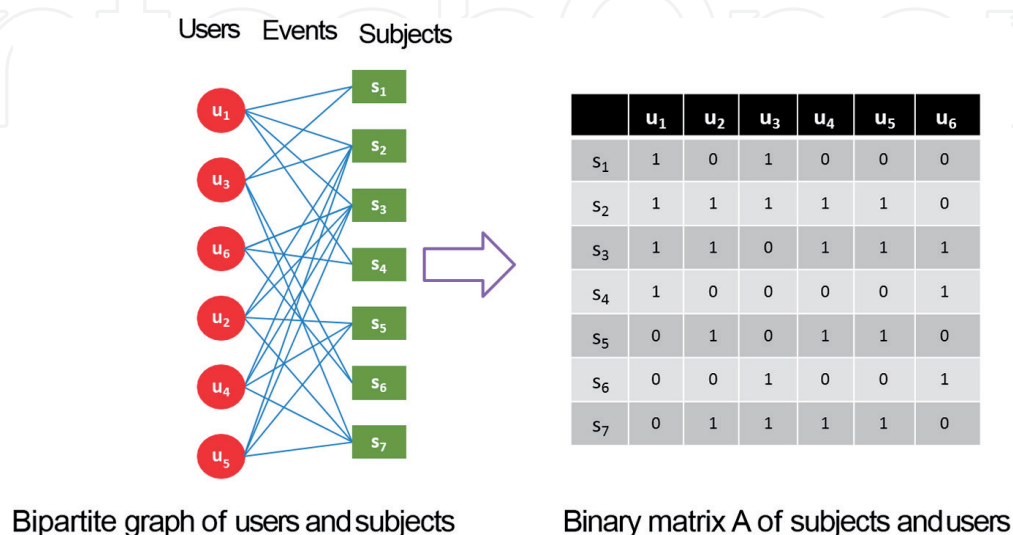


**Figure 1.** An example to illustrate data elements in EHR audit logs. Four health-care workers performed their actions to EHRs of a patient at different time stamps on the same day.

laboratory test user. The physician and the lab user have a hidden relationship that is built upon the lab test order. Hidden relations are essential knowledge to characterize processes of health information sharing and dissemination among health-care workers in EHR systems, which can potentially impact teamwork, and the following care quality and patient safety.

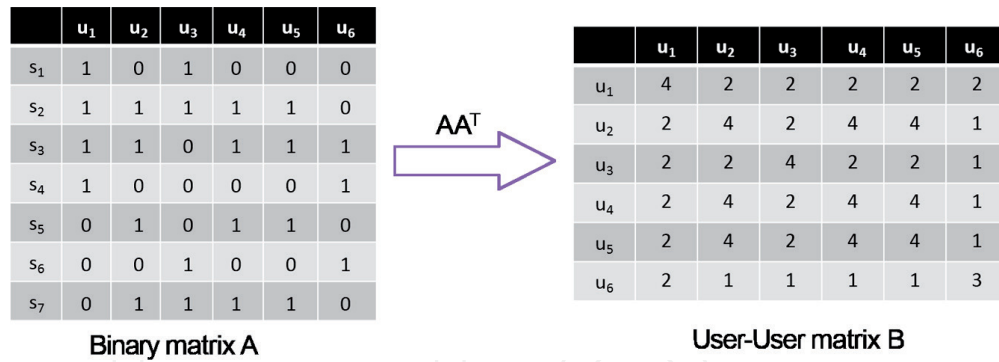
We use a bipartite graph of EHR users (health-care workers) and EHRs of subjects (patients) to represent events a user performed to EHRs of a subject. **Figure 2** shows an example of a bipartite graph, and a binary matrix to characterize interactions of six users to EHRs of seven subjects. In the example depicted in the figure, we use a binary matrix to represent if a health-care worker performed events to EHRs of a subject within a period (e.g., hour, day, week, or length of stay). Researchers can determine the period and whether using a binary value or the number of events to represent interactions of a health-care worker with EHRs of a patient according to their research purpose. To simplify our process, we use a binary matrix  $A$ , as shown in **Figure 2**. As mentioned above, if two health-care workers performed events to EHRs of the same patients within a period (e.g., a day), then there exists a hidden relationship between them. For instance,  $u_1$  and  $u_3$  both performed events to EHRs of  $s_1$  and  $s_2$ . Thus, in the binary matrix,  $A(1,1)$ ,  $A(1,3)$ ,  $A(2,1)$ ,  $A(2,3)$  are all ones. To transforming health-care works' interactions to EHRs to connections among health-care workers, we use binary matrix multiplication. For instance, the relationship between  $u_1$  and  $u_3$  can be learned by multiplying matrix  $A$  and its transpose matrix  $AT$ . The results of matrix multiplication are shown in matrix  $B$  in **Figure 3**.  $B(1,3)$  or  $B(3,1)$  represents the number of subjects whose EHRs were managed by both  $u_1$  and  $u_3$ . From **Figure 2**, we can see the number of subjects co-managed by both  $u_1$  and  $u_3$  is 2, which is equal to  $B(1,3)$  or  $B(3,1)$ . The larger the cell values in matrix  $B$ , the more strength of the relationship between health-care workers.

Matrix  $A$  represents the interactions of health-care workers to EHRs of subjects, and  $B$  describes the relationships between health-care workers. We show a simple way (matrix multiplication) to learn  $B$  from  $A$ . There are many alternative or advanced approaches that can be applied to matrix  $A$  (binary or nonbinary version) to measure hidden relationships between health-care workers. Examples of such methods include term-frequency, inverse documentary frequency (TF-IDF) [27], principal component analysis (PCA) [28], and similarity measurements (e.g.,



**Figure 2.** Events performed by health-care workers ( $u_i$ ) to EHRs of patients ( $s_j$ ) are represented by a bipartite graph (left) and corresponding binary matrix (right). In the right subfigure, if a health-care worker,  $u_i$ , performed events to EHRs of a patient,  $s_j$ , then the cell value  $A(i,j)$  in the matrix will be 1, otherwise 0.





**Figure 3.** Using the product of binary matrix A and the transpose matrix  $A^T$  to calculate the number of subjects whose EHRs are managed by a pair of users. Each cell value  $B(i,i)$  in the diagonal represents the number of subjects whose EHRs are managed by  $u_i$ . Each cell value  $B(i,j)$  ( $i \neq j$ ) represents the number of subjects whose EHRs are co-managed by both  $u_i$  and  $u_j$ .

cosine, Kullback-Leibler (KL) divergence, edit distance, and Jaccard distance) [29]. For instance, if the size of the matrix is big (a large number of subjects or health-care workers), we can apply PCA to it to reduce the dimensionality first, and then measure relationships for pairs of health-care workers based on the principal components.

## 4. Building networks

### 4.1 Relationship measurement

There are two types of relationships: directed and undirected between health-care workers. Directed relation emphasizes on the ordered relations, for instance, the connections from health-care worker A to B, and B to A that are different. To learn directed connection from the utilization data, we use time stamps of events to describe the ordered relationships. As shown in **Figure 1**, lab test ordering occurred ahead of lab test results uploading. Thus, the relationship between the physician who ordered the lab test and the lab user who uploaded the test results is directed. Upon the directed relations, we can create direct networks. We will use an example to illustrate the creation of directed networks of health-care workers concerning the management of each patient. Examples of undirected networks are used to describe structures of collaborations among health-care workers within a unit or across a HCO.

### 4.2 Directed health-care worker networks

As mentioned above, we define actions performed by a health-care worker in EHR systems as events. Events affiliated with EHRs of a patient constitute a sequence of information flow. We provide a simple scenario to understand the series of information flow as follows.

*“The night respiratory therapist documents an increased need for oxygen in a patient’s EHRs” → “the daytime nurse documents the patient’s vital signs and notes that the patient has tachypnea” → “on rounds the nurse practitioner and attending review the recorded vital signs focusing on the need for more oxygen and elevated respiratory rate” → “the physician prescribes a diuretic.”*

In this example, the nurse practitioner and attending’s comprehension of the patient’s condition grew with each update to the EHR. Health-care workers depend

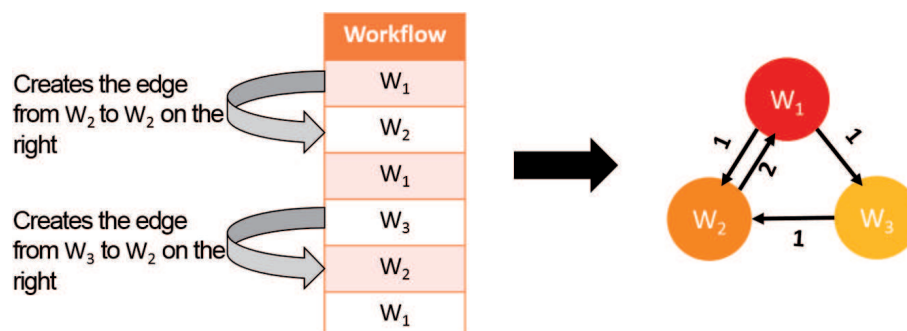
on their colleagues to provide information for clinical updates as they are essential to health-care workers' decision-making. As mentioned above, we call this virtual worker-worker interaction a hidden connection. A hidden connection does not mean a face-to-face interaction occurred, but rather, there existed the potential for the neighboring health-care workers to directly exchange information on the patient's condition via the EHR and arrive at the same conclusion, which in this scenario was to prescribe medication for pulmonary edema. We build networks that represent the hidden connections facilitating the dispersion of patient-related information. We call them patient-level health-care worker networks because they are composed of all health-care workers that treated a common patient.

To start, we create a simplified sequence dataset by condensing consecutive events by the same health-care worker into a single event. In this scenario, we can filter the self-loop relationships of health-care workers. In a network or a graph, a self-loop relationship is an edge that connects a vertex/node to itself. For example, health-care worker  $W_1$  made three EHR events consecutively to EHRs of a patient, and we condensed them into one event; one could interpret the simplified sequence as a workflow in EHR. Based on the sequences, we identified relationships between health-care workers whenever their events occurred consecutively (health-care worker  $W_2$  used the patient's EHR after health-care worker  $W_1$ ). We characterized each hidden connection with the frequency by which they occurred.

**Figure 4** shows an example of how we build a health-care worker network from a patient's sequence. As shown in **Figure 4**, the health-care worker  $W_1$  interacted with the EHR before health-care worker  $W_2$ , so the arrowhead on the right points to health-care worker  $W_2$ . The edge weight is the number of times the hidden interaction occurred. Note an edge exists if an interaction occurred at least once. While an observed interaction was not guaranteed to be an exchange of information, it did have the potential to be one.

### 4.3 Undirected health-care worker networks: Care for a group of patients

The structure of teamwork learned from a single patient is hard to represent the pattern of collaboration concerning the management of a group of patients. In this section, we introduce the creation of undirected health-care workers for the management of a group of patients. We assume health-care workers participating in the care of the same patients (performed events to EHRs of the same patients) on the same day have a relationship. Based on such an assumption, we can create a binary matrix (as shown in **Figure 3**) to describe whether a health-care worker performed events to EHRs of a patient. The cell value 1 is for Yes, and 0 for No. Based on the binary matrix of health-care workers and EHRs of patients, we can use the matrix multiplication, as shown in **Figure 3**, to get the daily relationships between pairs of



**Figure 4.**  
An example to learn a health-care worker network from a patient's EHR sequence.

health-care workers. Each non-diagonal cell value shows the number of patients; any two health-care workers both performed events to their EHRs on the same day. Two factors determine the strength of the relationship between two health-care workers. The first is the number of patients the two workers performed events to their EHRs on the same day, and the second is the number of days when the two workers performed events to EHRs of the same patients. We build a health-care worker network for a group of patients by using the relationships which are cumulatively added based on the number of days and patients.

We use a simple scenario to explain health-care worker networks built upon a group of patients. Assuming a medical intensive care unit (MICU) adopted a new scheduling strategy in a pandemic (e.g., COVID-19), and the health-care organization plans to investigate the changes in the structure of collaboration among health-care workers before and after the adoption of the new scheduling strategy. In this scenario, we use 8 months (4 months before and after adopting the new scheduling strategy) of EHR utilization data to learn the changes. To implement the study, we create two groups: critically ill patients admitted to the MICU before and after the new scheduling strategy adopted. To ensure the studied two groups share similar confounding factors (e.g., demographics and health conditions), we can use propensity score matching to create them. We use events performed by health-care workers to EHRs of the two groups of patients to measure relationships between health-care workers before and after the adoption of the scheduling strategy, respectively. Based on the relationships, we can build two health-care worker networks: before and after the adoption of the new scheduling strategy. The differences in the structures of the two networks can be measured using sociometric measurements, which are introduced in the following sections.

#### **4.4 Undirected health-care worker networks: care for patients within an HCO**

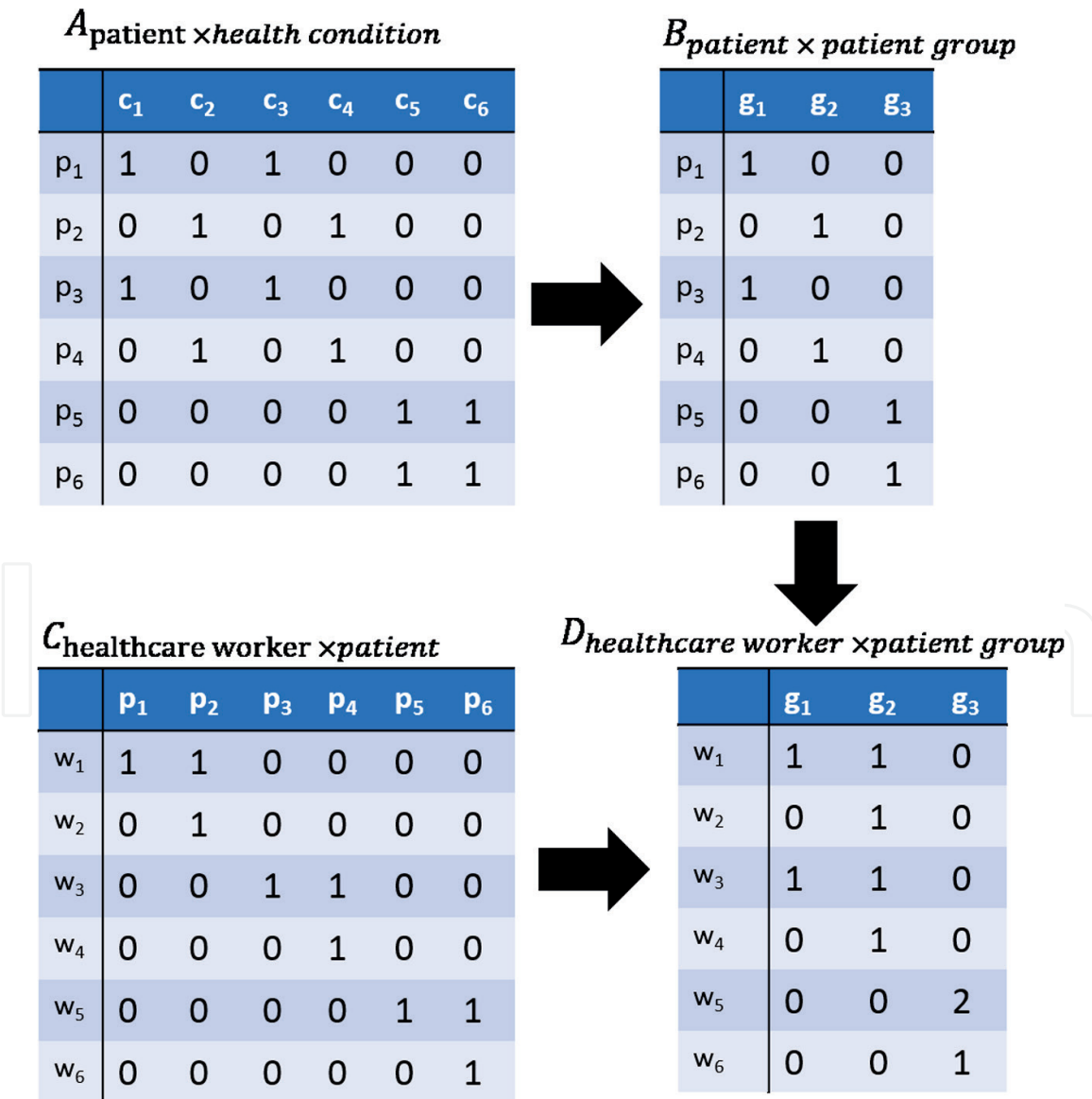
When learning a collaboration network at the level of a health-care organization, the number of patients and health-care workers investigated will be much bigger, and the relationships between health-care workers will become more complex. If we have a large number of patients, then it may complicate the measuring of the relationships between health-care workers. For instance, if we investigate 10,000 health-care workers and 1,000,000 patients, then the size of the health-care worker-patient matrix is 10 K by 1 M. There is a necessity to reduce the dimensionalities of the matrix to ensure it is appropriate for the following approaches to measure relationships between health-care workers. As mentioned above, PCA can be applied to the matrix to reduce dimensionalities. After the dimensionality reduction, we can use similarity measurements (e.g., cosine similarity or KL divergence) to calculate the relationships between health-care workers, which are used to build networks of health-care workers. If PCA is unable to represent the variance of the data in the matrix, an alternative way is to transform the matrix of health-care workers and patients into a higher level. Instead of building networks of health-care workers, we can create networks of operational areas (e.g., medical intensive care unit, and burn center). Also, we can cluster patients into groups according to their phenotypes and transform the matrix of health-care workers by patients into operational areas by patient groups. Based on the new transformed matrix, we measure relationships between operational areas and build a collaboration network of operational areas.

**Figure 5** shows an example to illustrate the process of transforming interactions between health-care workers and EHRs of patients into interactions of health-care workers to EHRs of groups of patients. Patient groups can be learned by conducting phenotyping algorithms on patient health conditions and

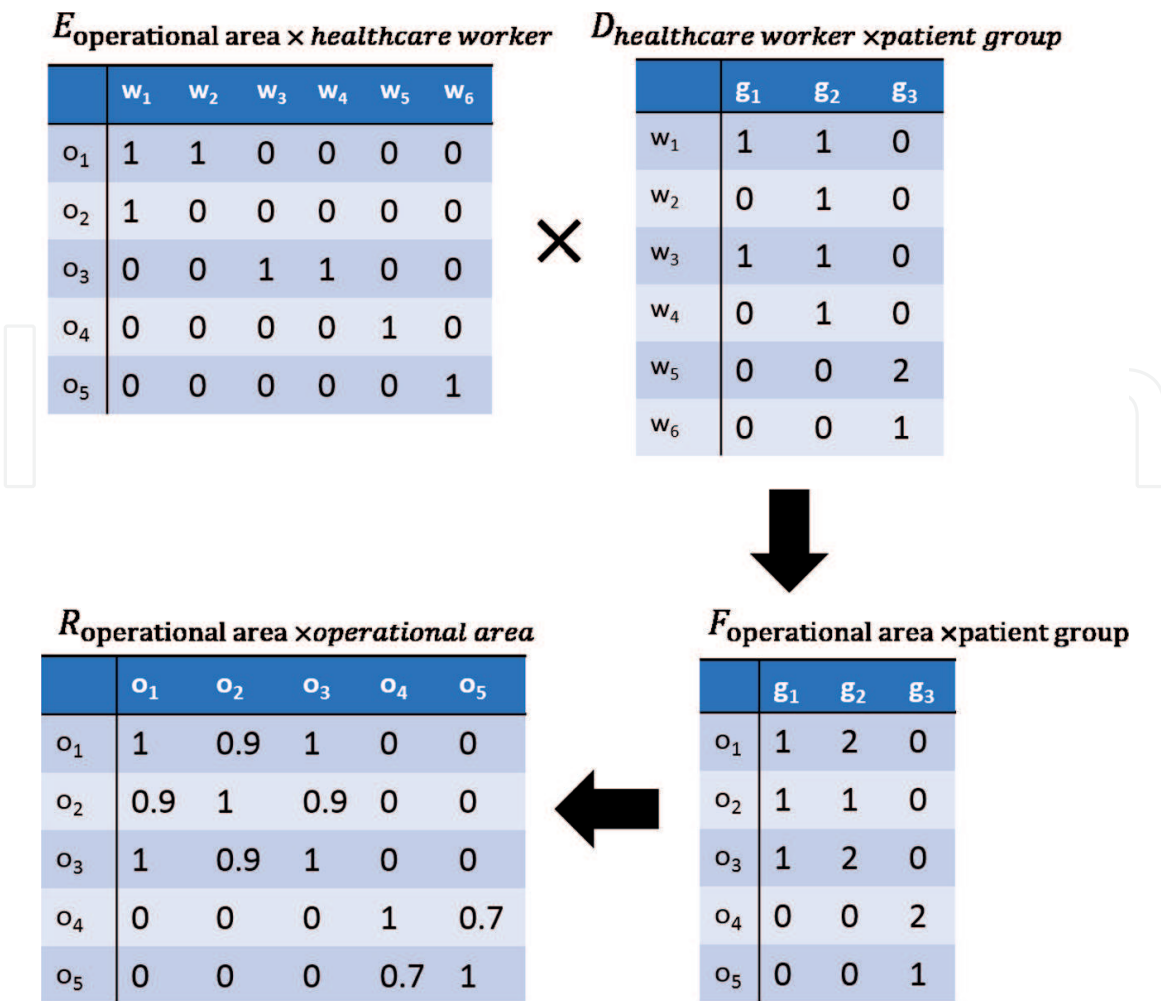


demographics. For instance, a typical topic modeling algorithm – Latent Dirichlet Allocation (LDA) can be used to learn topics to represent phenotypes of each patient. Based on the phenotypic topics, patients can be clustered into groups. As shown in **Figure 5**, the transformation from  $A_{patient \times health\ condition}$  to  $B_{patient \times patient\ group}$  can be implemented by using LDA.

**Figure 6** shows an example to illustrate the process of transforming interactions between health-care workers and EHRs of patient groups into the interactions between operational areas and EHRs of patient groups, which are further leveraged to measure relationships between operational areas. The transformation from  $D_{health-care\ worker \times patient}$  and  $E_{operational\ area \times health-care\ worker}$  into  $F_{operational\ area \times patient\ group}$  is implemented using matrix multiplication.  $E_{operational\ area \times health-care\ worker}$  represents the affiliations of health-care workers to operational areas. Similarity measurements can be applied to  $F_{operational\ area \times patient\ group}$  to learn relationships between pairs of operational areas  $R_{operational\ area \times operational\ area}$ . Collaboration networks of operational areas can be built upon the  $R_{operational\ area \times operational\ area}$ . To learn stable relationships between operational areas, we may need to create the matrix  $C_{health-care}$



**Figure 5.** An example to illustrate the process of transforming the interactions between health-care workers and EHRs of patients ( $C_{health-care\ worker \times patient}$ ) into interactions between health-care workers and EHRs of patient groups ( $D_{health-care\ worker \times patient\ group}$ ).



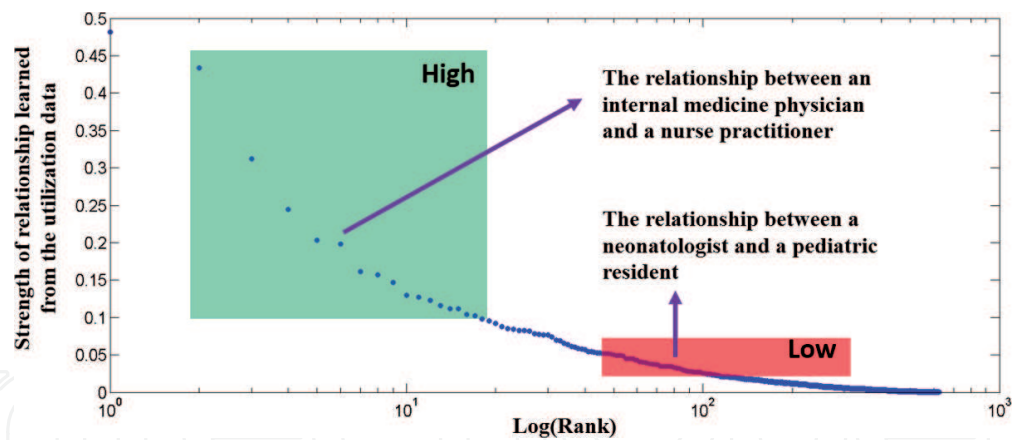
**Figure 6.** An example to illustrate the process of transforming interactions between health-care workers and EHRs of patient group ( $D_{\text{health-care worker} \times \text{patient group}}$ ) into relationships between operational areas ( $R_{\text{operational area} \times \text{operational area}}$ ).

$worker \times patient$  by setting a longer window size, such as 1 week/month rather 1 day we used in the previous examples. A study shows it requires at least 4 weeks to get stable relationships between operational areas by using interactions between health-care workers and EHRs of patients [30].

## 5. Validating relationships among health-care workers learned from the EHR system utilization

Concerns over the trustworthiness of the results of automated learning methods are not limited to the health-care worker network learned in this chapter. Instead, this is a problem that manifests when any knowledge is learned from the secondary analysis of EHR data. Researchers always need to review the knowledge learned from the data for their plausibility. As we mentioned above, the relationships between health-care workers learned from the utilization data are indirect. In other words, they are not explicitly documented by health-care organizations. To use networks built upon such relationships to describe or interpret structures of collaborations among health-care workers, we need to validate the relationships.

To do so, we design and deploy an online survey to assess the plausibility of relationships among health-care workers. **Figure 7** depicts an example of 626



**Figure 7.**

*Relationships between pairs of health-care workers ranked by their strength. Each of the two shaded areas represents relationships with high and low strengths, respectively. Each node in the graph represents a relationship.*

relationships ranked on a log scale and the strength of the relationships. This is clearly more relationships than a human can evaluate without fatigue, and so we need to sample a small number of them for respondents to assess. For instance, we can randomly select 20 relationships: 10 of high, and 10 of low strength. A survey can be designed to evaluate a specific hypothesis of the form: hospital employees can correctly distinguish between relationships of high and low strengths.

A survey contains a series of questions. The hospital employees who respond to the survey are presented with questions of the form: “*An internal medicine physician performed actions to the record of patient John Doe. How likely is it that an internal medicine nurse practitioner performed actions to the same patient’s record?*”. Respondents are not presented with the strength of the relationship between internal medicine physicians and nurse practitioners. The respondents are asked to choose one of five candidate answers: “Not at all likely,” “Slightly likely,” “Moderately likely,” “Very likely,” and “Completely likely.” In order to conduct a survey analysis through statistical models, we can convert these answers into integer values in the range 1–5 (e.g., “Not at all likely” is mapped to 1).

A set of respondents will answer each question in the survey. We can conduct a pretest to obtain feedback from the experts to refine the surveys and estimate the required number of experts via a power analysis. REDCap, which is a secure web application for building and managing online surveys and databases [31], can be used to implement the online survey. Details of the plausibility validation of the relationships between health-care workers can be found in our previous works [30, 32]. If we can verify with statistical significance that the learned relationships are often in line with the expectations of hospital employees, then we can suggest that collaboration networks of health-care workers, as well as strategies built on such networks, may be reliable and scalable.

## 6. Sociometric measurements

Sociometric measurements include network- and node-level metrics. The network-level metrics such as size, graph density, reciprocity, triads, average path length, clustering, cohesion and density, core-periphery, centralization, diameter, and K-core are used to characterize the structure of a network; while the node-level metrics such as degree, closeness, betweenness, eigencentrality, and eccentricity are

used to describe the characteristics of each node in the network. In this section, we explain those measurements in the health-care worker networks.

## 6.1 Network-level metrics

**Diameter.** The diameter is defined as the number of steps in the longest path in the network. There are two types of paths for any two nodes in the network. The first one is the shortest path, which is defined as the smallest number of steps between the two nodes, and the other one is the longest path, which has the largest number of steps between the two nodes. The network diameter is the number of steps between the two nodes, who have the largest number of steps in their longest path. Given two networks of health-care workers, if the diameter of the first one is larger than the second, then the information sharing and dissemination among health-care workers in the first network requires more steps.

**Density and cohesion.** Graph density is defined as the total number of edges within the network, divided by the number of edges that could exist. The cohesion of a network is described by the diameter and the average path length. The average path length is the average of the steps between all the nodes in the network. The low diameter or low average path length indicates a cohesive network with little clustering. Usually, when density increases, the average path length decreases because high-density network provides many paths along which to connect nodes. Studies show the relationship between density and average path length is nonlinear [33]. Density values above 0.5 indicate networks have many redundant paths between nodes, and it is hard to identify structures of networks [33]. If density values are very low, then there will be no network structures. To learn structures of health-care worker collaboration networks, we may need to prune the networks by using density values (e.g.,  $<0.5$ ). For instance, we can filter edges whose weight strength is low to decrease the density values of networks.

**Core-periphery.** Core-periphery structures are networks in which there is a group of nodes that are densely connected to one another (the core) and a separate group of nodes loosely connected to the core and loosely connected to each other (the periphery). It is not uncommon to find core-periphery networks in the health-care domain. In the NICU, nurses, neonatologists, and anesthesiologists work in a core network [9]. In contrast, otolaryngology residents, endocrinology physicians, and hematology physicians collaborate in a periphery network [9].

**Centralization.** The typical calculation of centralization is as:  $\sum_{i=1}^{i=n} (\max(v_i) - v_i) / (n^2 - 3n + 2)$ , where  $v_i$  is the centrality score (e.g., degree, betweenness, and closeness) of a node in the network, and  $n$  is the total number of nodes in the network [34]. In the centralized network, one or a few people hold a position of power and control in the network. An alternative way to calculate centralization is to measure the standard deviation of the node centrality scores. A large standard deviation indicates a lot of variation in the individual centrality scores, and hence a centralized structure. In contrast, a small standard deviation suggests little variation and hence a decentralized structure. In a network of health-care workers, if we can identify workers with high centrality scores in the centralized network, then we can further investigate how those workers share and disseminate information in the EHR systems. Do they act like broadcasters to reach many health-care workers quickly, or do they act as gatekeepers to slow down the information sharing and dissemination?

**Clustering coefficient.** A clustering coefficient is a measure of the degree to which nodes in a graph tend to cluster together. The metric can be defined at the network- and health-care worker levels. The network-level clustering coefficient



gives an overall indication of the clustering in the network. The network-level clustering coefficient is measured as:  $\# \text{ of closed triplets} / \# \text{ of all triplets}$ , where a triplet is three nodes that are connected by either two (open triplet) or three (closed triplet) undirected edges. If a health-care worker network has a high clustering coefficient, then health-care workers are connected in dense pockets of interconnectivity. There are two types of network structures that connect clustered subgroups. The first is the bridge structure, in which the clustered subgroups are connected by bridges (intermediates), and the second one is the centralized structure, in which central health-care workers connect the subgroups.

**Reciprocity.** Reciprocity is used to characterize the symmetry in relationships between health-care workers. In network science, the reciprocity is measured in directed networks. A typical approach [35] to calculate the network reciprocity is:  $\sum_{i \neq j} (a_{i,j} - \bar{a})(a_{j,i} - \bar{a}) / \sum_{i \neq j} (a_{i,j} - \bar{a})^2$ , where  $a_{i,j}$  is one if a link from  $i$  to  $j$  exists, and 0, otherwise.  $\bar{a} = \sum_{i \neq j} a_{i,j} / n \times (n-1)$ , where  $n$  is the number of health-care workers in the network. If a network has a higher value of reciprocity, then the greater likelihood of health-care workers to be mutually linked in information receiving and dissemination in the network.

**K-core.** The K-core is a subset of the network, in which each health-care worker within the K-core is connected to at least K other workers. A health-care worker in the K-core sub-network is considered as one of the cores in the whole network.

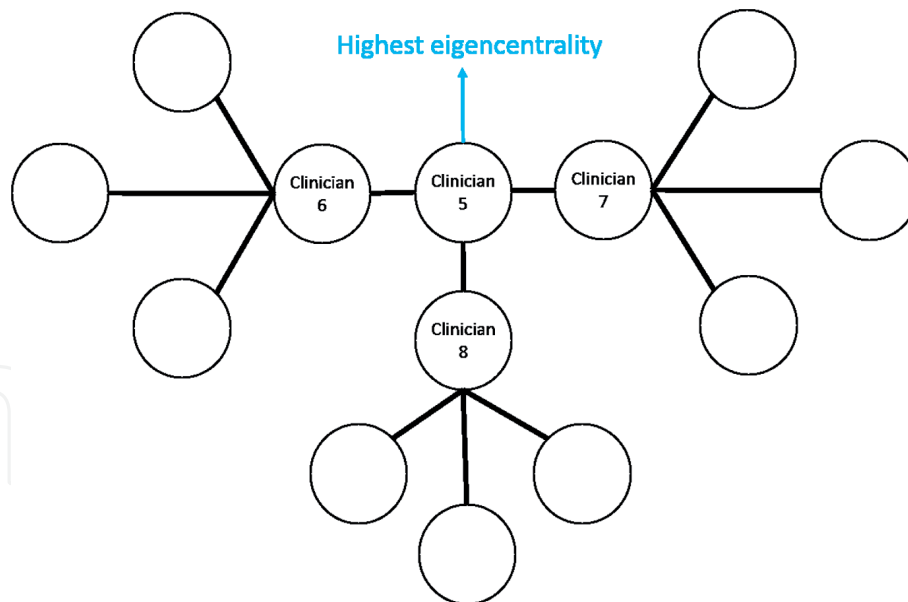
## 6.2 Health-care worker-level metrics

**Degree.** The degree of a health-care worker is the total number of edges connected. The weighted degree is the sum of the weights of connected edges. In the health-care worker network, the weight of an edge can be the strength of the relationship.

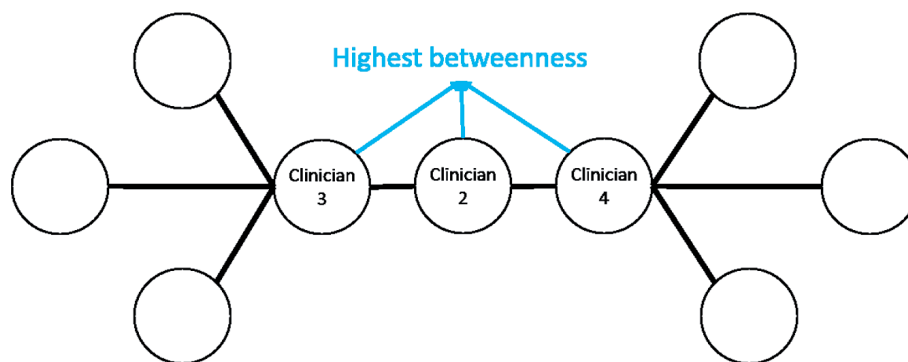
**Clustering coefficient.** The clustering coefficient of a health-care worker is the proportion of connections among their adjacent health-care workers divided by the number of connections that could possibly exist between them. One can think of the clustering coefficient as quantification of how close a health-care worker's neighbors are to be a clique of clinicians (e.g., a small group of clinicians, with shared interests in common patients). A health-care worker with a large clustering coefficient is the one who shares patients with health-care workers who also share patients with each other [9].

**Betweenness.** The betweenness is defined as the number of shortest paths between two health-care workers that pass through the specific health-care worker. Betweenness refers to whether a health-care worker lies on the path of others who are not directly connected. A health-care worker with a broad skillset could frequently be in a high-betweenness position. For instance, in **Figure 8**, clinicians 2, 3, and 4 have the largest number of shortest paths going through them. Betweenness reflects a health-care worker's access to diverse communication channels about evidence-based practice. A high betweenness worker cares for a wide spectrum of patients.

**Eigencentality.** Eigencentality is used to quantify the influence or leadership of a health-care worker on the collaboration and coordination among health-care workers in the network. A health-care worker with a high eigencentality is connected to workers who have high eigencentality. An example of health-care workers with high eigencentality is shown in **Figure 9**. A high eigencentality health-care worker acts as a leader in the sharing of patients in the network.



**Figure 8.**  
*Examples of health-care workers with the highest betweenness.*



**Figure 9.**  
*An example of a health-care worker with the highest eigencentrality.*

## 7. Statistical models to test hypotheses related to network structures

Most of the research studies in health care are hypothesis-driven. One of the goals of the network analysis in health care is to provide evidence on network structure to assist in the designing and development of teamwork-based hypotheses. Various hypotheses can be developed between sociometric measurements and clinical outcomes, including delayed ICU admission, ICU readmission, medication error, adverse event, length of hospital stay (LOS), mortality risk, and health-care cost.

### 7.1 Relationships of sociometric measurements with clinical outcomes

Structures of teamwork among health-care workers can be quantified by using both network- and node-level sociometric measurements. It has been recognized that structures of teamwork are associated with clinical outcomes. To inform actionable staffing interventions, we can develop hypotheses for each of the sociometric measurements and validate their relationships with clinical outcomes. For each inpatient stay (ranging from their admission to discharge), we can create a network to describe the structure of teamwork among health-care workers during the patient stay. Hypotheses can be designed based on the network. An example

of the hypotheses can be: the clustering coefficient of a network is associated with LOS. Statistical models can be leveraged to test the hypotheses. The distributions of most network measurements are not Gaussian distributed, so we can use rank-based approaches to measure associations between the measurements and clinical outcomes. For instance, we can use the Spearman rank-order correlation to measure the association between the clustering coefficient and LOS. If we want to investigate multiple sociometric measurements or add confounding factors (patient demographics, the severity of sickness), we can use advanced statistical models, such as a proportional-odds (PO) logistic regression model.

The PO model can be thought of as a set of logistic regression models, where each model describes the log-odds of LOS (continuous variable) being higher than some threshold  $j$  (rather than lower than or equal to), and where  $j = 1, 2, \dots, J$  represents all possible thresholds by which LOS can be dichotomized, and  $J$  is equal to the number of unique outcome values minus one. The set of models is collapsed into a single model, via the proportional odds assumption that coefficients for predictor variables are the same across the threshold values. Even when this assumption is not met, a coefficient from the proportional odds model can be thought of as a weighted average of coefficients across all the threshold-specific logistic regression models.

Some outcomes, such as ICU readmission, delayed ICU admission, or mortality risk are categorical variables. In that case, we can use the Mann Whitney U test or analysis of variance (ANOVA) to test the differences in the sociometric measurements between networks. The hypotheses can also be developed between node-level measurements (e.g., betweenness, eigencentrality, and degree) and clinical outcomes. For instance, critically ill patients who were cared for by more high-betweenness nurses were significantly less likely to die in their ICU stays.

## **7.2 Changes in structures of health-care worker networks**

Analyzing changes in the collaboration network structures and measuring relationships of the changes with outcomes are very important research questions in the teamwork in health care. When a health-care organization adopts a new staffing intervention (e.g., creating a new team scheduling), they will need to assess and monitor the changes in the behavior of collaboration among health-care workers before and after the interventions and how such changes impact clinical outcomes. Getting feedback from the adoption of a new staffing intervention can provide evidence to identify weak and ineffective parts to do further optimization. For instance, ICUs adopt staffing interventions in the COVID-19 pandemic, and the responses may change the structure of collaboration. We can use network analysis approaches to analyze the changes in the network structures from pre-COVID-19 to intra-COVID-19 and measure the relationships of such changes with clinical outcomes such as ICU readmission or delayed ICU admission. Examples of hypotheses can be: neonatology physicians have higher betweenness after the staffing intervention or the health-care worker network after the staffing intervention has a larger diameter (the difficulty of sharing patients increases). Since sociometric measurements are not Gaussian distributed in many situations, we can apply a Mann-Whitney U test to measure the significance of the difference.

## **8. Applications**

We introduce three applications to show how we use network analysis to identify care teams in a health-care organization, measure associations between

collaborations and length of stay in the trauma setting, and assess health-care worker networks for the management of surgical neonates, respectively.

### **8.1 Care team identification**

We applied network analysis to the EHR utilization records of over 10,000 hospital employees and 17,000 inpatients at a large academic medical center during a 4-month window [19]. The study aimed to learn collaboration structure across the entire health-care system, and thus it built networks of departments (higher level) rather than the networks of health-care workers. Each node in the network is a department. LDA models were used to cluster patients into groups. As shown in **Figures 5 and 6**, matrix multiplications were conducted to transform the matrix of health-care workers and patients into the matrix of departments and patient groups. Connections among 317 departments were inferred from the department-patient group matrix. We identified 34 collaborative groups of departments [19]. Each of the groups is a subnetwork and could be considered as a care team across various types of departments. The results suggested that, although the over 17,000 patients exhibited over 1400 different types of phenotypes, the health-care workers treating them tended to work in only 34 collaborative groups. When the 34 groups were presented to health-care experts via online surveys, 27 (79.4%) of 34 were confirmed as administratively plausible. Of those, 26 teams depicted strong collaborations, with a clustering coefficient  $>0.5$ .

### **8.2 Length of stay and trauma team structures**

We started the network analysis of trauma team structures by creating a matrix of ~5000 health-care workers and EHRs of ~5500 patients based on the EHR system utilization data [10]. The difference is we applied a spectral co-clustering methodology to the matrix to infer groups of patients and clusters of health-care workers simultaneously. By using the co-clustering algorithm, we created three trauma patient groups, each of which has a corresponding network of health-care workers. For each network of health-care workers, we calculated sociometric measurements to quantify their structures. Length of stay was used as the outcome. The association between a sociometric measurement (e.g., clustering coefficient) and length of stay was measured by using statistical models incorporating various confounding factors (e.g., demographics and admission dates). We found a remarkably clear distinction in LOS: those patients experiencing the largest quantity of collaborations between health-care workers had the shortest LOS, while those subject to fewer collaborations (i.e., supported by less well-integrated care teams), spent much longer in hospital, indicating greater financial cost as well, of course, as pain, distress, and inconvenience to the patient [10].

### **8.3 Length of stay and NICU team structures**

We extracted EHR data of 15 NICU gastrostomy patients from the day prior to the patient's surgery day until postoperative day 30. The study aims to validate the associations between health-care worker networks and post-surgical length of stay (PLOS) [36]. For each patient ICU stay, we built a directed network to show how information was received and disseminated among health-care workers in the NICU. For each patient's stay, we created a simplified sequence dataset by ordering health-care worker actions based on their time stamps starting from the day prior to the patient's surgery until postoperative day 30 or the patient's discharge date. Based on the sequences, we identified connections between health-care workers whenever



their actions occurred consecutively. We learned 15 patient-level health-care worker networks. We used the sociometric measurements, including in-degree, out-degree, and betweenness, to quantify the structures of each patient-level network.

We modeled patient PSLOS with each structure measurement controlling for patient age and weight using a proportional-odds logistic regression model. Study results show health-care workers, whose patients had lower PSLOS, tended to disperse patient-related information to more colleagues within their network than those, who treated higher PSLOS patients ( $P = 0.0294$ ). Our results demonstrate in the NICU that improved dissemination of information may be linked to reduced PSLOS.

## 9. Conclusions

This chapter provides an introduction of a network analysis of secondary EHR system utilization data to learn health-care worker networks. We introduce five main components when applying network analysis to team structures and clinical outcomes: (i) matrix multiplication to build connection among health-care workers, (ii) survey instruments to validate the plausibility of the learned connections among health-care workers, (iii) sociometric measurements to characterize network structures, (iv) hypothesis development to connect network structures with clinical outcomes, and (v) statistical models to test the hypotheses. Finally, we use three examples to show the application of network analysis in health care. In short, EHR data provide an efficient, accessible, and resource-friendly way to study teamwork using network analysis tools.

## Acknowledgements

The research studies introduced in this chapter were supported, in part, by the National Library of Medicine of the National Institutes of Health under Award Numbers K99LM011933, R00LM011933, and R01LM012854.

## Conflict of interest

The authors declare no conflict of interest.

## Author details

You Chen

Vanderbilt University Medical Center, Nashville, Tennessee, United States

\*Address all correspondence to: [you.chen@vanderbilt.edu](mailto:you.chen@vanderbilt.edu)

## IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Fix GM, VanDeusen Lukas C, Bolton RE, Hill JN, Mueller N, LaVela SL, et al. Patient-centred care is a way of doing things: How healthcare employees conceptualize patient-centred care. *Health Expectations*. 2018;**21**(1):300-307
- [2] Gusmano MK, Maschke KJ, Solomon MZ. Patient-centered care, yes; patients as consumers, no. *Health Affairs*. 2019;**38**(3):368-373
- [3] Ferlie EB, Shortell SM. Improving the quality of health care in the United Kingdom and the United States: A framework for change. *Milbank Quarterly*. 2001;**79**(2):281-315
- [4] D’Lima DM, Murray EJ, Brett SJ. Perceptions of risk and safety in the ICU: A qualitative study of cognitive processes relating to staffing. *Critical Care Medicine*. 2018;**46**(1):60
- [5] Upadhyay S, Weech-Maldonado R, Lemak CH, Stephenson AL, Smith DG. Hospital staffing patterns and safety culture perceptions: The mediating role of perceived teamwork and perceived handoffs. *Health Care Management Review*. 25 October 2019. DOI: 10.1097/HMR.0000000000000264. [Epub ahead of print] PMID: 31702706
- [6] Newman MW. Integrated and collaborative care: Quality improvement in action. *Psychiatric Annals*. 2017;**47**(7):374-377
- [7] Ma C, Park SH, Shang J. Inter- and intra-disciplinary collaboration and patient safety outcomes in US acute care hospital units: A cross-sectional study. *International Journal of Nursing Studies*. 2018;**85**:1-6
- [8] Reeves S, Pelone F, Harrison R, Goldman J, Zwarenstein M. Interprofessional collaboration to improve professional practice and healthcare outcomes. *Cochrane Database of Systematic Reviews*. 2017;**6**(6):CD000072
- [9] Chen Y, Lehmann CU, Hatch LD, Schremp E, Malin BA, France DJ. Modeling care team structures in the neonatal intensive care unit through network analysis of EHR audit logs. *Methods of Information in Medicine*. 2019;**58**(4-05):109
- [10] Chen Y, Patel MB, McNaughton CD, Malin BA. Interaction patterns of trauma providers are associated with length of stay. *Journal of the American Medical Informatics Association*. 2018;**25**(7):790-799
- [11] Chowdhury D, Duggal AK. Intensive care unit models: Do you want them to be open or closed? A critical review. *Neurology India*. 2017;**65**(1):39
- [12] Rosen MA, DiazGranados D, Dietz AS, Benishek LE, Thompson D, Pronovost PJ, et al. Teamwork in healthcare: Key discoveries enabling safer, high-quality care. *The American Psychologist*. 2018;**73**(4):433
- [13] Jagannath S, Sarcevic A, Marsic I. An analysis of speech as a modality for activity recognition during complex medical teamwork. In: *Proceedings of the 12th EAI International Conference on Pervasive Computing Technologies for Healthcare*; 21 May 2018. pp. 88-97
- [14] Durojaiye AB, Levin S, Toerper M, Kharrazi H, Lehmann HP, Gurses AP. Evaluation of multidisciplinary collaboration in pediatric trauma care using EHR data. *Journal of the American Medical Informatics Association*. 2019;**26**(6):506-515
- [15] Wu DT, Smart N, Ciemins EL, Lanham HJ, Lindberg C, Zheng K.

Using EHR audit trail logs to analyze clinical workflow: A case study from community-based ambulatory clinics. In: AMIA Annual Symposium Proceedings. Vol. 2017. San Francisco, CA: American Medical Informatics Association; 2018. pp. 1820-1827

[16] Amroze A, Field TS, Fouayzi H, Sundaresan D, Burns L, Garber L, et al. Use of electronic health record access and audit logs to identify physician actions following noninterruptive alert opening: Descriptive study. *JMIR Medical Informatics*. 2019;7(1):e12650

[17] Wang JK, Ouyang D, Hom J, Chi J, Chen JH. Characterizing electronic health record usage patterns of inpatient medicine residents using event log data. *PLoS One*. 2019;14(2):e0205379

[18] Chi J, Bentley J, Kugler J, Chen JH. How are medical students using the electronic health record (EHR)?: An analysis of EHR use on an inpatient medicine rotation. *PLoS One*. 2019;14(8):e0221300

[19] Chen Y, Lorenzi NM, Sandberg WS, Wolgast K, Malin BA. Identifying collaborative care teams through electronic medical record utilization patterns. *Journal of the American Medical Informatics Association*. 2017;24(e1):e111-e120

[20] Durojaiye AB. A novel approach for the investigation of multidisciplinary collaboration using social network analysis on electronic health record data [doctoral dissertation]. Baltimore MD: Johns Hopkins University; 2018

[21] Van Liew JR. Balancing confidentiality and collaboration within multidisciplinary health care teams. *Journal of Clinical Psychology in Medical Settings*. 2012;19(4):411-417

[22] Henry J, Pylypchuk Y, Searcy T, Patel V. Adoption of electronic health

record systems among US non-federal acute care hospitals: 2008-2015. *ONC Data Brief*. 2016;35:1-9

[23] Platform IT. Open or closed: A project proposal for investigating two different EHR platform approaches. *Context Sensitive Health Informatics: Sustainability in Dynamic Ecosystems*. 2019;265:207

[24] Ballaro JM, Washington ER. The impact of organizational culture and perceived organizational support on successful use of electronic healthcare record (EHR). *Organization Development Journal*. 2016;34(2):11-29

[25] Serbanati LD, Ricci FL. EHR-centric integration of health information systems. In: 2013 E-Health and Bioengineering Conference (EHB); 21 November 2013. IEEE. pp. 1-4

[26] Adler-Milstein J, Adelman JS, Tai-Seale M, Patel VL, Dymek C. EHR audit logs: A new goldmine for health services research? *Journal of Biomedical Informatics*. 2020;101:103343

[27] Chen K, Zhang Z, Long J, Zhang H. Turning from TF-IDF to TF-IGM for term weighting in text classification. *Expert Systems with Applications*. 2016;66:245-260

[28] Ku W, Storer RH, Georgakis C. Disturbance detection and isolation by dynamic principal component analysis. *Chemometrics and Intelligent Laboratory Systems*. 1995;30(1):179-196

[29] Xuecheng L. Entropy, distance measure and similarity measure of fuzzy sets and their relations. *Fuzzy Sets and Systems*. 1992;52(3):305-318

[30] Chen Y, Nyemba S, Malin B. Auditing medical records access via healthcare interaction networks. In: AMIA Annual Symposium Proceedings. Vol. 2012. American Medical Informatics Association; 2012. p. 93

[31] Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics*. 2019;**95**:103208

[32] Chen Y, Lorenzi N, Nyemba S, Schildcrout JS, Malin B. We work with them? Healthcare workers interpretation of organizational relations mined from electronic health records. *International Journal of Medical Informatics*. 2014;**83**(7):495-506

[33] Friedkin NE. The development of structure in random networks: An analysis of the effects of increasing network density on five measures of structure. *Social Networks*. 1981;**3**(1):41-52

[34] Freeman LC. Centrality in social networks conceptual clarification. *Social Networks*. 1978;**1**(3):215-239

[35] Garlaschelli D, Loffredo MI. Patterns of link reciprocity in directed networks. *Physical Review Letters*. 2004;**93**(26):268701

[36] Kim C, Lehmann C, Schildcrout J, Hatch D, France D, Chen Y. Provider networks in the neonatal intensive care unit associate with length of stay. In: *IEEE 5th International Conference on Collaboration and Internet Computing*. Los Angeles, CA: IEEE; 2019:127-134