Auditing Medical Records Accesses via Healthcare Interaction Networks

You Chen[1], Steve Nyemba[1], Bradley Malin[1,2]
[1] Dept. of Biomedical Informatics, Vanderbilt University, Nashville, TN; [2] Dept. of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN

Abstract

Healthcare organizations are deploying increasingly complex clinical information systems to support patient care. Traditional information security practices (e.g., role-based access control) are embedded in enterprise-level systems, but are insufficient to ensure patient privacy. This is due, in part, to the dynamic nature of healthcare, which makes it difficult to predict which care providers need access to what and when. In this paper, we show that modeling operations at a higher level of granularity (e.g., the departmental level) are stable in the context of a relational network, which may enable more effective auditing strategies. We study three months of access logs from a large academic medical center to illustrate that departmental interaction networks exhibit certain invariants, such as the number, strength, and reciprocity of relationships. We further show that the relations extracted from the network can be leveraged to assess the extent to which a patient's care satisfies expected organizational behavior.

Introduction

Electronic medical record (EMR) systems were initially designed to minimize the complexity associated with storage, retrieval, and reasoning over a patient's health information.[1] Modern computing technologies have radically transformed EMR system capabilities to enable timely, cost-effective care.[2] Despite the benefits that EMR systems afford healthcare providers, they also expose patient information to potential privacy vulnerabilities. Over the past decade there have been many notable examples of patients' records being exploited, or intruded upon, by various employees of healthcare organizations (HCOs). The majority of the cases reported in the media have focused on high-profile patients (e.g., movie stars), but there are also cases of intrusion for patients with much lower public interest. It has been argued that such intrusions could be addressed through the application of standard information security strategies, such as role-based access control.[3] These strategies are designed to reduce the accessibility of patient information to those who need to know (or have a well-qualified reason) for accessing the record. EMR developers have recognized the potential for such technologies and have embedded them into systems currently deployed at HCOs in the US and abroad. Yet, despite the inclusion of such protection mechanisms, they are only applied on a limited scale by the administrators of these systems. More often, they are invoked to limit access to certain functions (e.g., e-prescribing or uploading of clinical notes) than to manage the relationship between specific care providers and patients.

We suspect that a primary reason for limited information security implementation in EMRs is complexity. Healthcare systems are composed of interactions of various types of individuals, some of which provide primary care to the patient (e.g., physicians and nurses), whereas others provide support for business operations (e.g., schedulers and billers). The complexity of care is such that it is not uncommon for over one hundred HCO employees to interact with a patient's medical record during their visit.[4] This complexity is compounded by the dynamic nature of healthcare, where patients and HCO employees rotate in and out of the system at varying rates. While HCO employees provide provide care in a team-based process[5], it is often difficult to determine who the members of a team are and who will need access to what information at which time. Fortunately, to support the dynamic nature of modern teams, EMRs have adopted collaborative capabilities to facilitate interaction between teammates and coordinate care.[6,7]

We hypothesize that authenticated HCO employees will exhibit predictable behavior, akin to that observed in social computing technologies (e.g., Twitter[8]) The goal of this paper is to investigate the extent to which such behavior exists, as well as how it may be applied to determine when a certain patient's record has been utilized in a manner inconsistent with expected collaborative models. To accomplish this goal, we introduce a method to construct a *global* network that summarizes the relationships of an HCO's departments, as inferred by EMR users' interactions with patients' records. We show the global network can be applied to assess if a *local* network of care providers acting over a specific patient record constitute a low likelihood (or suspicious) scenario. To investigate the capabilities of our method, we perform an empirical analysis with three months of EMR access logs from the Vanderbilt University

Medical Center. We demonstrate that the global network is stable in terms of strength of relationships among, and reciprocity between, departments over time. Moreover, we verify that the local networks exhibit similar stability, such that changes in network features suggest strange behavior. To help illustrate our findings, we provide several examples of the types of patients detected by our system.

Background

Insider Threat Detection

Auditing the actions of authenticated EMR users (i.e., insiders) is critical to detect intrusions. Over the past several years, various types of auditing strategies have been proposed. For instance, several supervised machine learning approaches (e.g., logistic regression and support vector machines) have been applied to detect suspicious accesses.[9,10] These approaches exhibit high discriminatory power, but it is challenging to deploy these techniques broadly because they require significant adjudication from privacy experts and pre-existing knowledge of what constitutes a suspicious access. As an alterative, several unsupervised learning strategies have been suggested. Recently, a rules-based approach was proposed to "explain" accesses to medical records.[11,12] According to this model, legitimate accesses to EMRs are associated with a clinical, or operational, reason specified by the user. Yet, there are many accesses to patients' records which are unaffiliated with clinical documentation. Moreover, even when there is documentation, it is not necessarily apparent how an access relates to a broader care process. Extending the collection of unsupervised strategies, another class of models have been proposed to detect anomalous accesses through the analysis of social structures gleaned from EMRs.[13,14,15] These methods assess if a user's behavior is dissimilar from user groups who interact through patients' records. The work in the current paper differs because we focus on the detection of anomalous patients, as opposed to anomalous users.

Relational Networks in Health Systems

The modeling of HCO relational structures and business processes is necessary for more than the protection of information security.[16] For instance, such efforts can assist HCO administrators facilitate organizational change management and redesign inefficient business processes.[17] Additionally, characterizing organizational dynamics via social network analysis can support process evaluation and organizational improvement.[18] For instance, Merrill and collaborators[19] applied a social network analysis of departmental communications from a public health perspective and found that it was a viable option for learning the constituents' dynamics and affecting change.

From the perspective of security, research with access logs has focused more on *what* users view, than *who* is viewing the health records of whom. It is critical to uncover the relationships between individuals in an HCO and the types of medical record access workflows. Recently, relational network analysis techniques were designed to extract interaction networks of users and HCO departments based on common patient record accesses[4]. While our work is similar in that it also uses relational networks, there are several significant differences. First, earlier work characterized the number of patient records a pair of departments accessed in common by counting over the departments making the access.By contrast, we characterize the number of common patients by counting the patients accessed, which is more directly and accurate. Second, prior research illustrated the general relations of users and departments, but did not investigate the extent to which such relations were stable. As such, prior work did not address how such relations could be applied to assess if the set of specific accesses associated with a particular patient record were anomalous.

Methods

This section introduces a strategy to model the interactions of HCO departments in a relational network. It begins with an overview of the dataset for this study and then delves into details of the relational modeling process.

Dataset

This study was conducted using three months of access logs from the StarPanel EMR system of the Vanderbilt University Medical Center (VUMC).[20] The system has been in operation for over 15 years and is well-ingrained in healthcare operations. It functions as the primary point of patient information management and integrates data from various clinical domains. It houses over 300,000,000 observations on over 1.7 million patients. For this study, the logs were supplemented with users' VUMC department affiliations, as documented by human resources. This study was

conducted in an offline manner under approval of the institutional review board.

Table 1 reports the average number of departments, users, patients, and unique accesses that transpire within different timeperiods. A unique access is defined as ⟨*user, patient, time*⟩, which indicates that a user accessed a patient at least one time during the timeperiod.[1]

Average Number of:	**Within 1 Week**	**Within 2 Weeks**	**Within 3 Weeks**	**Within 4 Weeks**
Departments	453	468	471	473
Users	9,187	9,953	10,195	10,316
Patients	99,103	151,400	189,729	214,073
Unique accesses	401,520	786,106	1,147,838	1,409,351

Table 1: Summary statistics for the VUMC access logs over various timeperiods.

A Global Network of Departments

To construct the global network of department interactions, we extend the relational learning methods of Malin and colleagues.[4] The network is directed and weighted to represent the conditional probability that an employee of an HCO department accesses a patient record, given that an employee of a particular HCO department accessed the record. We develop several measures to characterize departmental relationships based on the network structure, which we present informally before providing their mathematical definition. First, we use *certainty* to characterize the strength of departments' interactions over time. This measure is designed to assess the extent to which changes in the network influence departments' affinity towards one another. Second, we use *reciprocity* to measure the extent to which departments exhibit similar behavior with respect to one another. For instance, *Anesthesiology* and *Emergency Medicine* reciprocate if they tend to work with each other more often than other departments. These measures allow for the characterization of interactive policies; i.e., the business processes of the HCO. We hypothesize that such processes should be relatively stable in healthcare environments.

To formalize these measures, we convert the access logs into a matrix form. Let P, U, and D be the set of patients, users, and departments, respectively and let $|\cdot|$ represent the cardinality of a given set of elements. Given a certain timeframe, the corresponding patient record access transactions are translated into two matrices A and B. The first matrix A has size $|P| \times |U|$ and $A(i,j) = 1$ if patient $p_i \in P$ was accessed by user $u_j \in U$; 0 otherwise. The second matrix B has size $|U| \times |D|$ and $B(j,k) = 1$ if user $u_j \in U$ was affiliated with $d_k \in D$; 0 otherwise. To address the fact that a user may be affiliated with multiple departments, B is normalized by its row-sums (e.g., if a user is affiliated with 3 departments, then the corresponding scores in this row are 0 and 0.33), the result of which we refer to as B_N. The relations between the patients and the HCO departments is thus represented as $G = AB_N$.

Relational Certainty

The global network of interdepartment interactions is represented by a graph $Net = (D, E)$, where E corresponds to the set of ordered department pairs (i.e., $d_i \rightarrow d_j$ and $d_j \rightarrow d_i$). The weight of the edge $e_{ij} \in E$ corresponds to a *certainty score*, which is defined as

$$Cert(d_i \rightarrow d_j) = \begin{cases} \frac{Q(i,j)}{Q(i,i)} & \text{if } i \neq j \\ \frac{\sum_{k=1}^{|P|} \theta(G(k,i))}{\sum_{k=1}^{|P|} \psi(G(k,i))} & \text{if } i = j \end{cases} \quad (1)$$

where $\theta(x)$ and $\psi(x)$ are indicator functions that return 1 if $x > 1$ and $x > 0$, respectively; 0 otherwise. $Q = {G_b}^{\mathsf{T}} G_b$ is a matrix of size $|D| \times |D|$, and the cell value $Q(i,j)$ represents number of common patients accessed by departments d_i and d_j. G_b is a binary matrix, which is retrieved by using $G_b(i,j) = \psi(G(i,j))$.

Notice equation 1 handles two cases. The first corresponds to when the departments are different (i.e., $i \neq j$). In this case, $Q(i,j)$ represents the number of patients that departments d_i and d_j accessed in common, while $Q(i,i)$

[1]Different users utilize the EMR system at different rates. As a consequence, the number of accesses to a particular patient could be artificially inflated due to system design. To mitigate bias, we use the number of distinct, as opposed to the total number, of accesses.

represents the number of patients accessed by department d_i. As an example, imagine there are three users. The first two users are affiliated with *Anesthesiology* and accessed patients p_1 and p_2, whereas the third user was affiliated with *Emergency Medicine* and accessed patient p_1. As a result, $Cert(d_{Anesthesiology} \rightarrow d_{EmergencyMedicine}) = 1/2$, and $Cert(d_{EmergencyMedicine} \rightarrow d_{Anesthesiology}) = 1/1$.

The second case corresponds to the situation when the departments are the same (i.e., $i = j$). If we apply the previous computation, the certainty would always be equal to 1, which is clearly incorrect. Thus, to account for this case, the equation is constructed such that the numerator reports the number of patients accessed by at least two users from the same department, while the denominator reports the number of patients accessed by at least one user in the department. The ratio is an approximation of the rate with which patients are accessed by multiple users in the same department. We adopt this approximation because, to the best of our knowledge, there is no exact computation for this relation.

Relational Reciprocity
The certainty measure approximates the likelihood that department d_j will access a patient's record given department d_i accessed the same record. To characterize the symmetry in the relationship between the departments, we measure the reciprocity of the network. We build upon the definition of *link reciprocity*, which measures the tendency of vertices to form mutual connections.[21] We modified the computation to measure symmetry in the certainty between each department pair, such that reciprocity of the global network is defined as:

$$Reciprocity = \frac{\sum_{\forall d_i, d_j \in D, i \neq j} |(Cert(d_i \rightarrow d_j) - a) \times (Cert(d_j \rightarrow d_i) - a)|}{\sum_{\forall d_i, d_j \in D, i \neq j} (Cert(d_i \rightarrow d_j) - a)^2} \tag{2}$$

where a defines the average certainty for all edges in the network:

$$a = \frac{\sum_{\forall d_i, d_j \in D, i \neq j} Cert(d_i \rightarrow d_j)}{2 \times |E|}. \tag{3}$$

Measuring Global Network Evolution

To investigate the stability of the network over time, we measure how the network changes in terms of certainty and reciprocity. To illustrate how this analysis is performed, imagine that the access logs are partitioned into a set of contiguous, equally-sized time periods $[(t_0, t_1], (t_1, t_2], \ldots, (t_{n-1}, t_n]$. A set of networks N_1, N_2, …, N_n is then constructed, such that N_i corresponds to the time period (t_0, t_i]. By constructing the network in this manner, the networks cover increasingly larger time periods, each of which is anchored at the first date of the study. Given this construction, we measure the changes between the networks of adjoining time periods; i.e., N_i and N_{i+1}. We recognize this is an unconventional approach to dynamic network analysis; however, we believe it is appropriate for assessing network evolution in this context for several reasons. First, the business processes of an HCO are such that the extent to which departments interact is dependent on the types of patients passing through the healthcare system. Thus, we investigate how the departmental relationships established by earlier patients are influenced as new patients enter the system. Second, this model allows us to retain relationships between departments that may not interact in every time period.

Figure 1 provides an example of how the network can evolve in this model. Notice, N_1 consists of four departments, not all of which interact with each other. When the network transitions into N_2, we come across two changes. First, the certainty between department d_1 and d_2 becomes stronger, while the certainty between d_1 and d_3 becomes weaker. In this case, the strengthening of the relation occurs because the departments accessed more patients in common than not. Similarly, the weakening of the relation occurs because each department access patients not accessed by the other department. Second, a new relationship, between d_2 and d_4, has entered the system. In the final week, we come across a new relationship between d_1 and d_4.

Local Network Evolution
Local networks are constructed from the departments that interact through a given patient record. We characterize the interactions of a local network by certainty and reciprocity as derived from the global network.

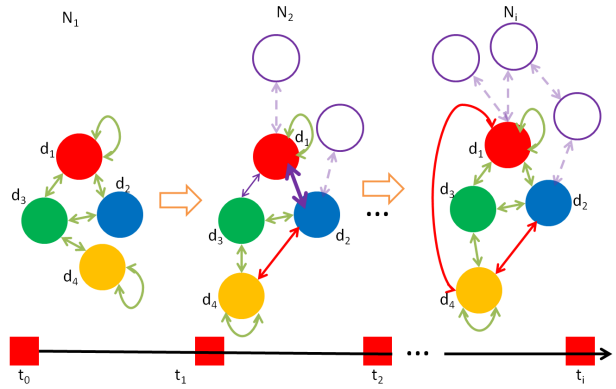

Figure 1: An example of the evolutionary processes for a global network.

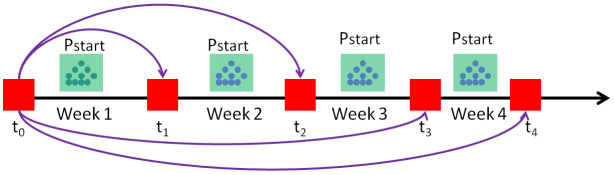

Figure 2: A depiction of network evolution for a set of initial patients.

Structurally, a local network is a complete graph $G_k = (D_k, E_k)$, where D_k is the set of departments that accessed a given patient $p_k \in P$ and E_k is the set of edges between the pairs of departments.

We define the local network score for each patient as a measure of the average certainty scores:

$$Score(p_k) = \frac{\sum_{\forall d_i, d_j \in D_k} Cert(d_i \rightarrow d_j)}{2 \times |E_k|} \tag{4}$$

Reciprocity in a local network is defined using equation 2, substituting D for D_k. The local reciprocity indicates the mutual interaction among departments associated with a specific patient.

Like the global network, we investigate the evolution of local network and reciprocity scores for all patients. Specifically, let P_{start} be the set of patients in the first week. We follow those patients across each week of interest and measure the changes in local network and reciprocity scores for P_{start} between each timeperiod (i.e., first week vs. second week vs. third week, etc.). This evolutionary process is depicted in Figure 2.

Experimental Design

For the purposes of this study, we divide the dataset into 3 sections, one per month. We construct a global network for each partition and categorize its evolution on a weekly basis. At the local level, we set P_{start} equal to the set of patients accessed in the first week and categorize the evolution of their networks on a weekly basis as well.

To compare the evolution of departmental relations on a common scale, we report the change in certainty relative to the original certainty:

$$Change(d_i \rightarrow d_j) = \frac{Cert_{w_{k+1}}(d_i \rightarrow d_j) - Cert_{w_k}(d_i \rightarrow d_j)}{Cert_{w_k}(d_i \rightarrow d_j)} \tag{5}$$

where w_k represents week $k \in \{1, 2, 3\}$. We model change in local network and reciprocity scores are relative to their original scores as well.

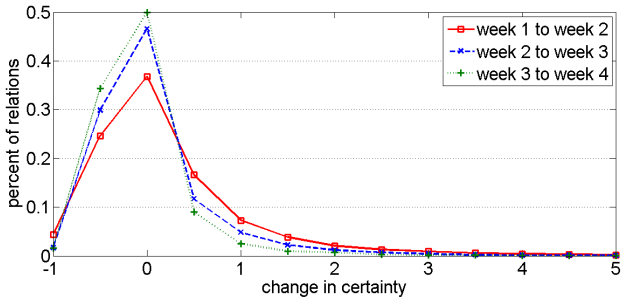

Figure 3: The distribution of certainty changes across four weeks.

Department (d_i)	**Department** (d_j)	**Min Certainty**	**Max Certainty**
Intradepartmental Relations			
4East OB/GYN	4East OB/GYN	0.74319	0.7669
Adult Emergency Medicine	Adult Emergency Medicine	0.74024	0.78453
Cancer Infusion Center	Cancer Infusion Center	0.73171	0.844
8N Inpatient Medicine	8N Inpatient Medicine	0.7197	0.80909
Newborn Nursery	Newborn Nursery	0.70406	0.72727
Interdepartmental Relations			
DOT Radiology	Orthopaedics	0.99621	1
Nursing Education and Development	Medical Information Services	0.95833	1
Main OR - Trauma/Renal	Medical Information Services	0.94444	1
Life Flight Event Medicine	Emergency Medicine	0.90805	1
Emergency Medicine Admin	Adult Emergency Medicine	0.91489	0.94186

Table 2: Strong relations between VUMC departments across four weeks.

Results

Evolution of the Global Network

Figure 3 shows the distribution of the certainty changes for the global network on a weekly basis, averaged over the three months of the study.

There are several important observations to highlight. First, there is variance in the changes across all weeks. This implies that the global network is not completely stable in terms of the strength of each relation. However, second, the total amount of change in certainty is decreasing from week to week. In fact, the distribution of change from the second to the third week (blue line) is approximately the same from the third to the fourth week (green line). This implies that, the strength of relations in the global network change, but at a relatively constant rate by the third week. Third, and most importantly, the total amount of change by this point is relatively small. Over 82.5% of the change resides in the [-0.25, 0.25] range.

While the network is evolving, a significant portion of the relations discovered in the first week appear to be stable. Almost 40% of the certainty scores, for instance, exhibit 0 change from week to week. Moreover, we observed that many of the relations learned in the first week are stable across the month. To illustrate this observation, Table 2 provides examples of the top five certainty scores for relationships at the intra- and interdepartmental level. At the intradepartmental level, many of these relationships correspond to clinical departments or wards in the hospital (e.g., *Adult Emergency Medicine*). At the interdepartmental level, more business related components of the HCO are represented, such as *Medical Information Services* and *Nursing Education and Development*.

Next, we turn our attention to the reciprocity in the system. Here, we find that the global network has an average reciprocity of 0.267 in the first week, which increases to 0.2814 in the second week, 0.2858 in the third week, and

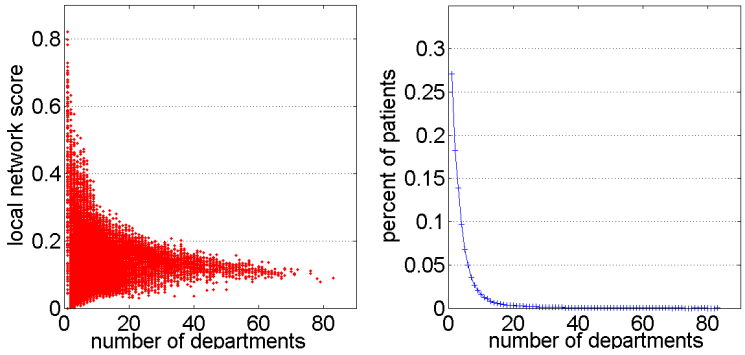

Figure 4: Relations between local network size and local network score *left)* and distribution of local network size *right)*.

Relation Score	Departments Accessing Patient	# of EMR Users
0.0003	Children's Hospital Outpatient Clinic, VMG [Anonymized Town A]	3
0.0003	Gastroenterology Clinic, General Pediatrics	2
0.0004	Hearing And Speech, Rheumatology	4
0.0006	Pediatric Gastroenterology, Urology	5
0.002	Cancer Registry, Service Free Stipends	3
0.001	Adult Urology, [Anonymized Town B], Orthopaedics	2

Table 3: Patients with small local network scores.

0.2871 in the fourth week. The small change in reciprocity across weeks, suggests that the relationship between most departments is relatively unbalanced. This is a clear illustration of why a directed network is more appropriate for auditing purposes.

Evolution of the Local Network

The relation between local network score and number of departments is depicted in Figure 4. A large local network score suggests that the departments accessing the patient are highly related to each other, whereas a small score indicates the departments are, for the most part, unrelated. It can be seen that there is an inverse relationship between the number of departments accessing a patient record and the range of local network scores. This implies it is more difficult to audit patient records via a local network score when the record is accessed by a larger number of departments. Fortunately, as shown in the bottom of Figure 4, approximately 85% of the patients' records were accessed by less than 6 departments, which suggests that the majority of patients can be effectively audited. For illustration, we depict several patients with small local network scores in Table 3.

If a patient exhibits a small local network score, it does not necessarily indicate their record has been intruded upon. This is because local network scores provide a static view of the system and many patients are associated with multiple departments that have a weak, but steady, relationship. Rather, we suspect the stability of these scores are more useful for auditing purposes. The last patient in Table 3 provides a useful illustration of this scenario. This patient had a small local network score has a small local network score in the first week (0.001), but only a small change in the score over the course of a month. Specifically, the score sustained a change of 0.0016, 0.0013, and 0.0014 over the next three weeks. Notably, this low score may be due, in part, to [Anonymized Town B], a moderately sized clinic that has not referred many patients to the main hospital. By contrast, the last second patient has a small local network score (0.002), but also exhibits a large change over time. The change in local network score for this patient from first week to second week is 3.5. This is a massive change when considering the changes for all patients, as shown in Figure 5.

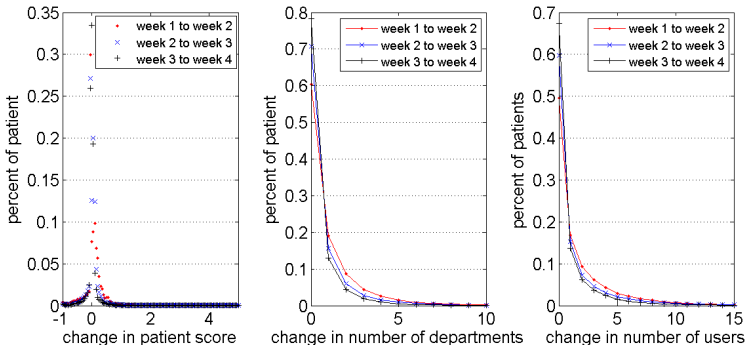

Figure 5: The distribution of changes in local network *left*) scores, *middle*) number of departments, and *right*) number of users for all weeks.

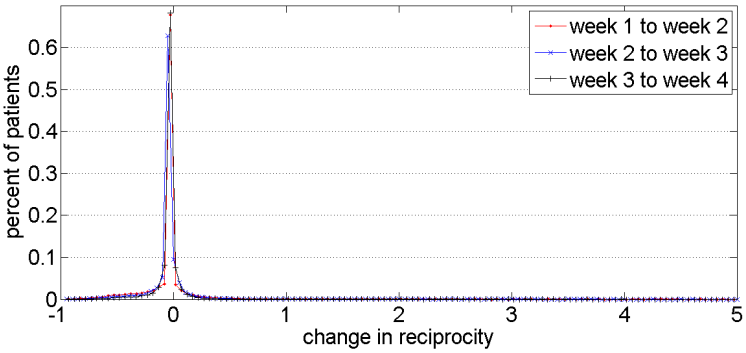

Figure 6: The distribution of changes in reciprocity for all local networks.

To investigate stability, we assessed the changes of local network scores, local network size (number of departments), and number of users associated with P_{start}. The results are depicted in Figure 5. It can be seen that over 98% of patients in P_{start} exhibit a score change of less than 0.05. Additionally, more than 95% of patients accumulated three or less departments and users. One reason this occurs is because outpatients are rarely seen by more than several departments over the course of a week. We also note that the changes in reciprocity for this group of patients is also quite small. Approximately 99% of patients have a change in reciprocity of less than 0.1, as depicted in Figure 6.

A Case Study in Large Relation Score Changes

To illustrate how local network score changes relate to HCO semantics, Figure 7 depicts two specific patients.

First, let us follow patient p_1. In the first week, this patient has only one department *Vanderbilt Internal Medicine (VIM)* associated with their record. At this time, the intradepartmental certainty for this department (i.e., $d_{VIM} \rightarrow d_{VIM}$) was equal to zero. From a global perspective, during this week, *VIM* users accessed several different patients, but each patient was accessed by only one *VIM* user. In the second week, this p_1 again was accessed by only one department, *Internal Medicine*, however, the local network score grew significantly to 0.69. The reason for this change is that in this second week, some of the patients accessed by *VIM* in the previous week were accessed by other additional *VIM* users. As a result, the intradepartmental certainty rose, leading to a high local network score for p_1. In the third week, the patient was accessed by a new department *Med Peds*. At a global level, the certainty between *VIM* and *Med Peds* was small, so the patient's score decreased to 0.16, which held steady in the fourth week.

Next, we turn our attention to patient p_2. This patient accumulated a large number of departments over time. They sustained significant decreases in their patient score as a result. Like p_1, in the first week, p_2 had only one department *Cancer Infusion Center* affiliated their record. In the second week, the number of departments increased dramatically to 11 while the local network score decreased from 0.73 to 0.05 (i.e., a -0.93 change of local network score). In this week, the patient was accessed by various departments, including *Breast Center*, *[Anonymized Street Location]*,

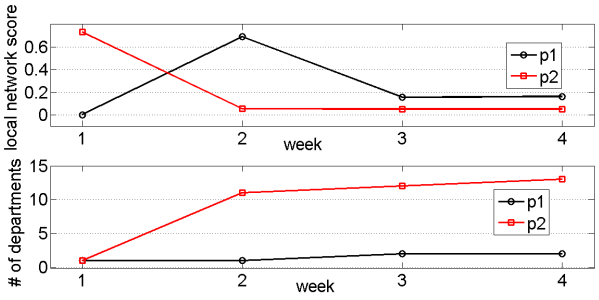

Figure 7: Depiction of changes for two specific patients regarding *top*) scores and *bottom*) total number of departments accessing their records.

Care/Eskind Diab Acces, *Disease Management Service*, *Eskind Diabetes - Adult*, *Free Stipends*, *Internal Medicine*, *VIM*, *VMG Physician Billing Services*, *Vanderbilt Home Care Primary*.

We can also measure a patient using a combination of the local network score and reciprocity. We suspect that patients who exhibit large swings in both of these measures during the same timeperiod are representative of strange organizational behavior. For p_2, it has -0.93 change of local network score and -0.79 change of local network reciprocity from the first to the second week.

Discussion

The global and local networks constructed in this paper appear to represent the business processes of HCO departments. However, such claims must be confirmed with employees knowledgable about the working of the medical center and its affiliated clinics. We hypothesized that an HCO would exhibit strong stability, which would be reflected in networks of smaller size. The results of our experiments confirmed this hypothesis. We observed that the distribution of variation of certainty and reciprocity have strong concentrations around a score of 0, which implies infrequent change.

If the global networks depict operational behavior of an HCO, then its characteristics can be applied to benchmark patients' local networks. Specifically, we can characterize how strange a patient's local network appears with respect to the global network. The results of the experiments suggested there are two gross groups of patients; those with small changes in relation and reciprocity scores and those who significant changes. The changes in the latter group do not justify grounds to claim the patient is an intrusion, but may provide a reason for an investigation that incorporates more nuanced domain knowledge.

As presented, our method has several limitations. In general, the limitations stem from the fact that the access logs studied did not incorporate semantics about the patients (e.g., physical location in a hospital). Ideally we should construct more diverse networks based on the status of the patient (e.g., diagnoses) as has been done in other work on network-based auditing.[14] Our case studies illustrate the potential for such information. The local network around p_2 in Figure 7, for instance, shows high changes in local network scores, but upon further investigation it appears there is no intrusion. Rather, it is likely a complex cancer patient, which could be confirmed by inspection of clinical documentation in the medical record. We plan to investigate how such information can embellish and improve our methods in future work.

Conclusion

As health information technologies are deployed on a larger scale to support complex care pathways, it is increasingly important to protection patient information from misuse in the context of primary care environments. In this paper, we showed that healthcare organizations are dynamic systems, but that they exhibit certain invariants in the relationships between their subdivisions; i.e., departments. We demonstrated, using three months of access logs from a large academic medical center, that *global* networks (i.e., those constructed from all patients whose records were accessed)

can be leveraged to inform the behavior around specific patient records, or *local* networks. Although our network analysis methods permit the partitioning of patients into a large group of patients with low risk and a small group of high risk, we can not claim the patients have been intruded upon before incorporating additional semantics or human administrative review.

Acknowledgements

This research was supported by grants CCF-0424422 and CNS-0964063 from the NSF and R01-LM010207 from the NIH. We thank C. Gunter (University of Illinois at Urbana-Champaign) and D. Liebovitz (Northwestern University) for insightful discussions. We also thank D. Giuse from Vanderbilt University for compiling the dataset.

References

1. Hyrinen K, Saranto K and Nyknen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform*, 77:291–304, 2008.
2. National Research Council (US) Committee on Engaging the Computer Science Research Community in Health Care Informatics; Stead WW eds, Lin HS. *Computational Technology for Effective Health Care: Immediate Steps and Strategic Directions*. National Academies Press, Washington, DC, 2009.
3. Blobel B. Authorisation and access control for electronic health record systems. *Int J Med Inform*, 73:251–257, 2004.
4. Malin B, Nyemba S and Paulett J. Learning relational policies from electronic health record access logs. *J Biomed Inform*, 44:333–342, 2011.
5. Patel V, Cytryn K, Shortliffe E and Safran C. The collaborative health care team: the role of individual and group expertise. *Teach Learn Med*, 12:117–132, 2000.
6. Safran C. Electronic medical records: a decade of experience. *JAMA*, 285:1766, 2001.
7. Eysenbach G. Medicine 2.0: Social networking, collaboration, participation, apomediation, and openness. *J Med Internet Res*, 10:e22, 2008.
8. Honey C and Herring S. Beyond microblogging: conversation and collaboration via Twitter. In *Proc* 42^{nd} *Hawaii International Conference on System Sciences*, pages 1–10, 2009.
9. Boxwala A, Kim J, Grillo J and Ohno-Machado L. Using statistical and machine learning to help institutions detect suspicious access to electronic health records. *J Am Med Inform Assoc*, 18:498–505, 2011.
10. Kim J, Grillo J, Boxwala A, Jiang X, Mandelbaum R, Patel B et al. Anomaly and signature filtering improve classifier performance for detection of suspicious access to ehrs. In *AMIA Annu Symp Proc*, pages 723–731, 2011.
11. Fabbri D and LeFevre K. Explanation-based auditing. *Proc VLDB Endowment*, 5(1):1–12, 2011.
12. Fabbri D, LeFevre K and Hanauer D. Explaining accesses to health records. In *Proc ACM Workshop on Data Mining for Medicine and Healthcare*, pages 10–17, 2011.
13. Chen Y and Malin B. Detection of anomalous insiders in collaborative environments via relational analysis of access logs. In *Proc ACM Conference on Data and Application Security Security and Privacy*, pages 63–74, 2011.
14. Chen Y, Nyemba S and Malin B. Detecting anomalous insiders in collaborative information systems. *IEEE Transaction on Dependable and Secure Computing*, 9:332–344, 2012.
15. Chen Y, Nyemba S, Zhang W and Malin B. Specializing network analysis to detect anomalous insider actions. *Security Informatics*, 1(1):5:1–24, 2012.
16. Carley K. Computational organization science: a new frontier. *Proc Natl Acad Sci*, 99:7257–7262, 2002.
17. Ash J, Anderson N and Tarczy-Hornoch P. People and organizational issues in research systems implementation. *J Am Med Inform*, 15:283–289, 2008.
18. Carley K and Lee J. Dynamic organizations: organizational adaptation in a changing environment. *Advances in Strategic Management*, 15:269–97, 1998.
19. Merrill J, Bakken S, Rockoff M, Gebbie K and Carley K. Description of a method to support public health information management: organizational network analysis. *J Biomed Inform*, 40:422–428, 2007.
20. Giuse D. Supporting communication in an integrated patient record system. In *AMIA Annu Symp Proc*, page 1065, 2003.
21. Garlaschelli D and Loffredo M. Patterns of link reciprocity in directed networks. *Phys Rev Lett*, 93:268701, 2004.