HHS Public Access

Author manuscript

Proc 2023 7th Int Conf Med Health Inform ICMHI 2023 (2023). Author manuscript; available in PMC 2024 April 18.

Published in final edited form as:

Proc 2023 7th Int Conf Med Health Inform ICMHI 2023 (2023). 2023 May ; 2023: 104–111. doi:10.1145/3608298.3608319.

A Study of Healthcare Team Communication Networks using Visual Analytics

Hsiao-Ying Lu,
University of California at Davis, Davis, CA, USA

Yiran Li,
University of California at Davis, Davis, CA, USA

Brittany Garcia,
UC Davis Health, Sacramento, USA

Shin-Ping Tu,
UC Davis Health, Sacramento, USA

Kwan-Liu Ma
University of California at Davis, Davis, USA

Abstract

Cooperation among teams or individuals of healthcare professionals (HCPs) is one of the crucial factors towards patients' survival outcome. However, it is challenging to uncover and understand such factors in the complex Multiteam System (MTS) communication networks representing daily HCP cooperation. In this paper, we present a study on MTS communication networks constructed with real-world cancer patients' Electronic Health Record (EHR) access logs. We adopt a visual analytics workflow to extract associations between semantic characteristics of MTS communication networks and the patients' survival outcomes. The workflow consists of a *neural network learning* phase to classify the data based on the chosen input and output attributes, a *dimensionality reduction and optimization* phase to produce a simplified set of results for examination, and finally an *interpreting* phase conducted by the user through an interactive visualization interface. We provide the insights found using this workflow with two case studies and an expert interview.

Keywords

Healthcare; Electronic Health Records; Machine Learning; Interpretability; Visual Analytics; Networks

hyllu@ucdavis.edu .

1 INTRODUCTION

Electronic health records (EHRs) serve a central role in connecting virtual care teams, which are defined as groups of healthcare professionals (HCPs) who provide care to the same patients, work at different times and locations, and support their work with technology-mediated communication. Through EHRs, virtual care teams develop a distributed communication system for encoding, storing, and retrieving patient information, which can be examined as a communication network representing the information sharing among HCPs or the interactions between HCPs and the patients' EHRs. Understanding this communication network and factors influencing its effectiveness (e.g., patients' survival outcome) is crucial for the practice of virtual care teams.

Although the new communication system built upon EHRs is rapidly changing the way HCPs collaborate, research on the optimization of effective communication and information processing in such an ecosystem has been limited [11, 36]. In critical domains such as cancer care delivery, there's still little research on the relationship between the cancer care team composition, function, and effectiveness [31, 35]. Network analysis has been successfully applied to understand how interactions and communication occur in teams, and to design more effective communication networks [14, 15, 26]. However, as the communication networks become much more complex, it is ineffective to manually inspect the network structures and statistics. Machine learning (ML) models are able to achieve efficient and outstanding performance in complex network predictions [38], but the lack of interpretability of these black-box models limits their integration in critical domains such as cancer care delivery.

In this paper, we present a study on communication networks constructed from real-world cancer patients' EHR access logs. To understand and promote effective communication and teamwork in cancer care team systems, we first construct a communication network from the access log data, encoding the connections between HCPs and each patient's EHR. After extracting attributes from patients and the communication network, we employ a visual analytics workflow [23], which supports interpretable machine learning based network analysis to determine the association of MTS communication structures and patient survival. This paper presents our preliminary findings from applying this workflow to studying cancer care multiteam systems (MTS) of breast, colorectal and lung cancer patients using EHR data. Specifically, our contributions include:

- A new method for the construction of the MTS communication networks, which represents the interactions between HCPs and patients' EHR data.
- Application of a visual analytics workflow which enables interpretable machine learning on the communication network analysis.
- Interpretable insights found on the associations between the communication network structure, patient basic information, and patients' survival outcome.

Our contributions of this work include: the MTS communication network construction, and the interpretable associations found to be related to the patient survival outcome. patients of four cancer sites (breast, colon, lung, and rectum/rectosigmoid) using EHR data.

2 RELATED WORKS

Following the growing availability of healthcare data, we have seen the development of visual analytics methods and systems to support clinical research [8, 29, 33], decision making [2, 13], and studying disease progressions [9, 16]. Furthermore, machine learning has shown its promise for EHR data analysis. Tree-based methods [4, 6, 28, 32] are wildly used for predictive analysis. Recurrent neural network (RNN) based methods such as LSTM [21] are designed for temporal classification and can thus be used to make predictions on clinical data in terms of event sequences or time series. While deep learning methods like RNN are able to provide high prediction performance, these methods typically need large amounts of training records. More recently, there is a strong interest in developing interpretable machine learning to support clinical data analysis [5, 12, 19, 20, 25, 34]. Notably, RetainVis [17] introduced an interpretable and interactive visual analytics tool for EHR data modeling.

Our work focuses on the construction of communication networks from EHR access logs and the interpretable machine learning based network analysis. Therefore, in the rest of this section, we especially review previous works on constructing networks from EHR data and employing machine learning in network analysis.

Network Construction from EHR Data.

Previous studies have applied various network construction methods to examine interactions among HCPs. One commonly adopted method is to construct patient-sharing networks using claims or administrative data [1, 10, 18, 27] and EHR access logs [30], where HCPs are connected based on shared patient record access or referrals. The strength of patient-sharing ties is measured by the number of patients to whom two physicians provided care during a specific period, implying stronger collaboration or information-sharing connections between physicians if they shared more patients. Different from previous approaches, we emphasize on the interactions between HCPs and EHR data, and construct the communication network based on such interactions.

Machine Learning in Network Analysis.

Machine learning methods in network analysis mostly focus on network representation learning (NRL), which aims to generate a set of low-dimensional vectors (also called representation) that captures important characteristics of networks, nodes, or links [38] for easier performing the downstream network analysis tasks, such as node classification and link prediction. Various NRL methods are developed for heterogeneous EHR networks [3, 22, 37]. However, these work did not address the interpretability aspects of the learned representations. One commonly used method is composite variable crafting, which searches for a smaller set of input variables that have clear relationships to the representations

of interest. For instance, Gleicher [7] performed an exhaustive search for the selection of variables, and assigned a weight to each variable based on its relationship with the representations. In our study, we utilize a workflow that incorporates a machine learning based NRL to uncover relations of interest in MTS communication networks to obtain interpretable representations and analyze this representation using visualizations and composite variable construction.

3 DATA AND DERIVED NETWORKS

The raw data is collected from an academic medical center, and was approved by the institutional review board. 525 patients diagnosed with stage 2 or 3 cancer among four different sites (breast, colon, lung, and rectum/rectosigmoid) are included in the collection. For each patient, the data contains his/her *basic information* as well as the *access logs* of his/her EHR data. The basic information includes each patient's demographics (e.g., age and sex), comorbidities, treatments and survival outcome (alive/dead). Meanwhile, the EHR access logs are composed of time-stamped EHR access events sampled from three months before to one year after the diagnosis date. During each event, one HCP (e.g., physician) performs an action (e.g., add prescriptions) to the patient's EHR data.

Based on the EHR access logs, we construct communication networks for each patient encoding the interactions between HCPs and EHR data. We then extract the *network-related* attributes and the *patient-related* attributes, which will be used in the machine learning and visual analytics steps further described in section 4.

3.1 Network Construction

For each patient, we construct a bipartite heterogeneous communication network, *the original network*, based on the EHR access logs. The network contains two sets of nodes: the *provider nodes* representing HCPs and the *action nodes* representing different actions in the access logs; each edge in between represents that a provider had once performed an action to the patient's EHR data. Different provider nodes are identified through the combinations of their types (e.g., physician or nurse), titles (e.g., MD or RN), residency, and specialties (e.g., cardiology), whereas different action nodes are identified by their names (e.g., "Medication note saved" and "Medication refill added"). The weight of each edge between a provider node and an action node is the average frequency per day the provider performs the action.

Although the original network distinguishes different providers and actions in the lowest level, it is difficult to align the attributes extracted from different networks because of these distinctions. For example, healthcare communication experts may be interested in the number of edges (degree) coming from each provider node. However, the provider nodes' degrees in different networks cannot compare with each other, as the numbers and the identifications of provider nodes vary across different patients. Therefore, a higher-level grouping of providers and actions that is uniform across different patients is needed to extract attributes that are comparable between different networks.

In our study, we group the provider nodes into 2 categories, and the action nodes into 10 categories. The providers are categorized by their titles. We filter out titles that are not observed in every patient's access logs and the ones with too few connections, and the remaining two titles are MD and RN. The filtering criteria is highly dependent on the dataset. We chose to keep MD and RN because their quantity are comparable. For the grouping of actions, we directly use the categories provided in the raw data, and keep the ones that are observed for every patient. The selected action categories and the corresponding example nodes in the original network are listed in Table 1, and Figure 1 shows two examples of the original network, where the nodes are colored by the 12 categories of providers and actions.

3.2 Attribute Selection

Here, we describe the extraction of both patient-related and network-related attributes used in the visual analytics workflow in section 4.

The patient-related attributes are directly extracted from each patient's basic information. This includes the patient's status (survival outcome and cancer stage), demographics (age and sex), treatments (whether a certain treatment is given), and comorbidities (whether a certain comorbidity is present). These attributes and their possible values are described in detail in Table 2a. These values are considered different status of a patient, which means larger magnitude will not be treated more importantly and will not introduce bias in the model.

The network-related attributes (Table 2b) are extracted from the original communication networks with the higher-level grouping information of the provider nodes and the action nodes. We first extract the size, degrees, and clustering coefficients of nodes within each group. For degrees and clustering coefficients, we use the summarization statistics (average, minimum, maximum, and standard deviation) to represent the overall attributes for each node group. Furthermore, the number of connected components is computed to demonstrate the structure of the whole original network. Finally, to portray the connections between each specific provider group and each action group, we compute the densities upon the subnetworks extracted from the original network, connecting nodes within one provider group and nodes of one action group. Figure 2 presents two examples of subnetworks extracted from each of the original networks in Figure 1. These two subnetworks both connect nodes of the Physician MD group with nodes of the Patient Clinical Info group. As there are 2 provider groups and 10 action groups, $2\times10=20$ densities are generated for each of the provider-action combinations.

4 VISUAL ANALYTICS WORKFLOW

To study the MTS communication networks, we directly employ a visual analytics workflow designed for extracting associations between different structural and semantic characteristics of the multivariate networks [23] as shown in Figure 3. This work particularly addresses the challenge of balancing the trade-off between machine learning model accuracy and results

interpretability. It enables intuitive understanding of the important factors around patients that affect their survival outcomes.

Step 1.

The workflow begins with the attribute selection. Based on the attributes described in subsection 3.2, the workflow uncovers how patients' survival outcomes are related to their surrounding conditions.

Step 2.

A network representation is generated by a neural network (NN) trained to classify the patients' survival outcomes based on the input attributes. We train a 4-layer Multi-Layer Perceptron (MLP), which is a machine learning architecture constructed by fully connected layers. The hidden layer neuron size are 128, 64, 64 and 2 neurons in the output layer for the binary classification. By learning how to classify, this NN model consequently encodes the relations between the survival outcome and the input attributes in the representation.

Step 3.

This step compresses the network representation's dimension from many to one (1D representation) while preserving the encoded relations. This makes the remaining steps simpler to complete and the related results easier to interpret. This 1D representation is one value (e.g., 0.5) to reflect each patient's selected attributes. This reflection is learned through Step 2 and compressed to 1D in this step. As shown in Figure 4-b1, a beeswarm-plot-like visualization presents the 1D network representation to help assess its quality by seeing how well the two classes are separated. The horizontal axis shows the distribution. The wider vertical span only means that there are more patients in that class.

Step 4.

This step evaluates and ranks each selected attribute's contribution to the construction of the 1D representation. Each row corresponding to one attribute shows a scatterplot where the *x*- and *y*-axes show the SHAP and attribute values, respectively and each dot represents one patient. The obtained ranks can be considered as the recommendation levels of the inclusion of the corresponding attributes for the composite variable construction in Step 5. To measure the attribute contributions, we use the SHAP (SHapley Additive exPlanations) method [24]. As shown in Figure 4-b1, the positive side of the 1D representation is the death class. Hence, for example, if instance A having age 60 and SHAP value 3.5, instance A is pushed towards the positive side of 1D representation by the degree of 3.5 (higher chance of death). In other words, the SHAP value indicates how much having a corresponding attribute value contributes to moving an instance toward a positive direction of the 1D representation.

Step 5.

Finally, based on the recommendation levels and interests, the analyst manually selects a small set of attributes and runs a composite variable construction algorithm. The optimized composite variable (e.g., the y-axis in Figure 4-b2) maximally resembles the 1D

representation (e.g., the x-axis in Figure 4-b2) and consequently is the linear combination of the small set of attributes that best correlates the distribution of the 1D representation. This composite variable provides an intuitive explanation of how it is related to the survival outcome. For example, in Figure 4-b2, the y-axis is a variable composed of *time_span* (the number of days from diagnosis to the last contact) and *age* with the weight −0.9 and +0.5 respectively. This suggests the two attributes have opposite (sign of weights) effects on the survival outcome and by different degrees (magnitude of weights).

Step 6.

The UI in Figure 4 visualizes the information related to each of the previous steps. By interactively reviewing these pieces of information, the analyst can gain analytical insights based on their analysis interests.

5 CASE STUDIES

We first use the case study shown in Figure 4 to explain how we can make interpretations based on the analysis results.

5.1 Study 1

From Figure 5a, we can observe the attribute *time_span* has a trend from upper left to lower right. This means the higher the *time_span* value is, the more chance the patient would survive (the negative side of the 1D representation as explained in workflow Step 4), whereas the attribute *age* has the opposite effect.

We then select the two attributes to run the composite variable construction and the result is shown in Figure 4-b2. We first note that the correlation value between the composite variable and the 1D representation is positive, which means that the higher the composite variable value is, the larger the 1D representation value is (higher chance of death). This correlation can also be observed in the scatterplot that the dense area in the survival class (blue) located in the bottom left corner and the death class (red) located in the upper right corner.

In addition, we can see the sign of the weights for the two attributes matches the observation in Figure 5a. For example, the negative sign for *time_span* suggests that this attribute needs to be flipped to have a positive correlation so the original value of *time_span* is negatively correlated to the 1D representation value. In other words, the higher the *time_span* value is, the lower the 1D representation value is (higher chance of survival). Moreover, from the magnitude of the weights, we find that the attribute *time_span* might be more important than *age* ($|-0.9| > |+0.5|$) in terms of affecting the survival outcome.

Last but not least, the correlation value between this composite variable and the 1D representation is larger than the correlation between either *time_span* or *age* and the 1D representation. This means the composite variable is more related to survival outcome than either of *time_span* or *age*. Considering the two variables together gives more accurate

prediction of the patient's chance of survival compared to only taking into account individual attributes.

These findings allow us to draw an assumption that when the time from diagnosis date to the last contact date (*time_span*) is shorter combined with the fact that the patient's age is sufficiently high, it is more likely the patient would die. This composite variable can be suggesting the phenomenon that older patients tend to have more severe disease and comorbidities so that the contact interval for them is shorter due death. On the other hand, longer *time_span* may reflect different situations, less severe disease, fewer comorbidities, or disease remission after treatment so the HCPs followed up their conditions for a longer period of time.

5.2 Study 2

In Figure 5b, we select the two MTS communication network structural attributes to run the composite variable analysis. We find that the *provider_size_0* is negatively correlated with the 1D representation and otherwise for the *provider_degree_max_1*. The *provider_size_0* means the number of type 0 (physician MD) providers in the patient's treatment network. The *provider_degree_max_1* is the largest count of access logs made among the provider type 1 (nurse). Therefore, the degree measure reflects more about the time they have spent on the EHR system and the actual work they have done in a day, whereas the size measure shows us the count of participating providers.

We can interpret this result as the insufficient number of physician MDs in a treatment team (*provider_size_0*) might potentially be harmful to the patient's survival rate when the nurses in this treatment network spend more time on the system (*provider_degree_max_1*). A potential interpretation may be a situation that the nurses' workload is too much or they are obligated to be spending too much time documenting or reviewing information in the EHR.

6 EXPERT INTERVIEW

To further validate our interpretations, we conducted an informal interview with an expert in the medical field. The expert (E1) is a Professor of the Department of Internal Medicine and the Chief of Division of General Internal Medicine, Geriatrics and Bioethics. E1 conceptualized this study and the interview was conducted through a video conference setup, where the case studies were presented.

E1 commented "The age attribute having a positive correlation with survival is consistent with the research and literature using other statistical methods." E1 was surprised that cancer stage is not one of the top contributing factors, as extensive research has shown cancer stage is a risk factor of patient survival. E1 added that we need to incorporate cancer type (i.e., breast, colorectal, lung) and that additional analysis of access log data with HCPs individually identified by a study ID and actual EHR access dates (i.e. no random date shifts) is needed.

Finally, E1 noted that the analysis terms such as 1D representation and SHAP value are not very intuitive for HCPs and health researchers to understand at first glance. Hence, we tried to use more paragraphs to explain the terms and how to interpret them in this paper.

7 DISCUSSION AND FUTURE WORKS

Our work constructs the MTS communication networks with unified provider and action types across patients. Although this grouping allows us to analyze different patients with the same set of attribute semantics, it lose the detailed information within each group. These details might hold non-negligible contribution to the patients' survival outcomes. Owing to the same concern, we filter out the provider titles that are not pervasive in all patients' access logs or does not have the balanced amount of records compared to other provider titles. This prevents the analysis process from being biased by the data imbalance. Yet, those providers categories are still part of the communication. This work overlooks the effects from these filtered providers on patients. For extracting network-related attributes, we select size, degrees, and clustering coefficients of nodes within each group. These network measures statistically brief the MTS networks. We manually pick these based on the observation of the network characteristics. However, if users find other network measures important to be included, this analysis workflow is flexible to adjust to different sets of measures of interests.

The preliminary results show composite variables combing the network-related and patient-related attributes. The results explain the importance between different attributes and the composite contribution to the patients' survival outcomes. For future works, we plan to collect a more comprehensive access log data to avoid the information loss from filtering and to consult more experts in order to evaluate the hypotheses in actual clinical practice.

8 CONCLUSION

We present the application of a visual analytics workflow to the MTS communication network to explore the associations between the patients' surrounding conditions and cancer patients' survival outcomes. The workflow generates expressive as well as interpretable analytical results, which were evaluated by a medical expert. Considering also the characteristics of patients' treatment teams communication networks, the results match the observation in clinical practices on the age attribute, but not suggest the same importance of the cancer stage attribute as the previous studies. This different importance suggested by the workflow introduces new opportunities to examine the existing research literature with the considerations of MTS communication networks constructed with different focuses. With our promising and interpretable results, this work potentially contributes to a wide range of applications that involve analyses of large, complex health record data.

ACKNOWLEDGMENTS

We would like to thank the UC Davis Health Innovation Technology Data Center of Excellence team. This work was supported by grant R01CA270454 from the National Cancer Institute. Contents of this manuscript are solely the responsibility of the authors and do not represent the official view of the National Cancer Institute.

REFERENCES

[1]. Barnett ML, Christakis NA, O'Malley J, Onnela JP, Keating NL, and Landon BE. 2012. Physician patient-sharing networks and the cost and intensity of care in US hospitals. Med Care 50, 2 (2012), 152–60. 10.1097/MLR.0b013e31822dcef7 [PubMed: 22249922]

[2]. Berner Eta S. and La Lande Tonya J.. 2016. Overview of Clinical Decision Support Systems. Springer Int. Publishing, Chapter 1, 1–17.

[3]. Cai Derun, Sun Chenxi, Song Moxian, Zhang Baofeng, Hong Shenda, and Li Hongyan. 2022. Hypergraph contrastive learning for electronic health records. In Proceedings of the 2022 SIAM International Conference on Data Mining (SDM). SIAM, 127–135.

[4]. Chen Tianqi and Guestrin Carlos. 2016. XGBoost: A scalable tree boosting system. In Proc. SIGKDD. 785–794.

[5]. Choi Edward, Mohammad Taha Bahadori Jimeng Sun, Kulas Joshua, Schuetz Andy, and Stewart Walter. 2016. RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. In Proc. NIPS. 3504–3512.

[6]. Friedman Jerome H.. 2001. Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics 29, 5 (2001), 1189–1232.

[7]. Gleicher Michael. 2013. Explainers: Expert explorations with crafted projections. IEEE Trans Vis Comput Graph 19, 12 (2013), 2042–2051. [PubMed: 24051770]

[8]. Guo Rongchen, Fujiwara Takanori, Li Yiran, Lima Kelly M, Soman Sen, Tran Nam K, and Ma Kwan-Liu. 2020. Comparative visual analytics for assessing medical records with sequence embedding. Visual Informatics 4, 2 (2020), 72–85.

[9]. Guo Shunan, Jin Zhuochen, Gotz David, Du Fan, Zha Hongyuan, and Cao Nan. 2018. Visual progression analysis of event sequence data. IEEE Trans. on Visualization and Computer Graphics 25, 1 (2018), 417–426.

[10]. Hackl F, Hummer M, and Pruckner GJ. 2015. Old boys' network in general practitioners' referral behavior? J Health Econ 43 (2015), 56–73. 10.1016/j.jhealeco.2015.06.005 [PubMed: 26184383]

[11]. Horwitz Leora I. and Detsky Allan S.. 2011. Physician Communication in the 21st Century: To Talk or to Text? JAMA 305, 11 (2011), 1128–1129. 10.1001/jama.2011.324 [PubMed: 21406650]

[12]. Ji X, Shen H, Ritter A, Machiraju R, and Yen P. 2019. Visual Exploration of Neural Document Embedding in Information Retrieval: Semantics and Feature Selection. IEEE Trans. on Visualization and Computer Graphics 25, 6 (2019), 2181–2192. [PubMed: 30892213]

[13]. Jin Zhuochen, Cui Shuyuan, Guo Shunan, Gotz David, Sun Jimeng, and Cao Nan. 2019. CarePre: An Intelligent Clinical Decision Assistance System. ACM Trans. on Computing for Healthcare (2019).

[14]. Katz Nancy, Lazer David, Arrow Holly, and Contractor Noshir. 2004. Network Theory and Small Groups. Small Group Research 35, 3 (2004), 307–332. 10.1177/1046496404264941

[15]. Kilduff M and Brass DJ. 2010. Organizational Social Network Research: Core Ideas and Key Debates. Academy of Management Annals 4, 1 (2010), 317–357. 10.1080/19416520.2010.494827

[16]. Kwon Bum Chul, Anand Vibha, Severson Kristen A, Ghosh Soumya, Sun Zhaonan, Frohnert Brigitte I, Lundgren Markus, and Ng Kenney. 2019. DPVis: Visual Exploration of Disease Progression Pathways. arXiv:1904.11652 (2019).

[17]. Kwon Bum Chul, Choi Min-Je, Kim Joanne Taery, Choi Edward, Kim Young Bin, Kwon Soonwook, Sun Jimeng, and Choo Jaegul. 2018. RetainVis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. IEEE Trans. on Visualization and Computer Graphics 25, 1 (2018), 299–309.

[18]. Landon BE, Keating NL, Barnett ML, and et al. 2012. Variation in patient-sharing networks of physicians across the united states. JAMA 308, 3 (2012), 265–273. 10.1001/jama.2012.7615 [PubMed: 22797644]

[19]. Li Oscar, Liu Hao, Chen Chaofan, and Rudin Cynthia. 2017. Deep Learning for Case-Based Reasoning through Prototypes: A Neural Network that Explains its Predictions. arXiv:1710.04806 (2017).

[20]. Li Yiran, Fujiwara Takanori, Choi Yong K, Kim Katherine K, and Ma Kwan-Liu. 2020. A visual analytics system for multi-model comparison on clinical data predictions. Visual Informatics 4, 2 (2020), 122–131.

[21]. Lipton Zachary Chase, Kale David C., Elkan Charles, and Wetzel Randall C.. 2016. Learning to Diagnose with LSTM Recurrent Neural Networks. In Proc. Int. Conf. on Learning Representations.

[22]. Liu Zheng, Li Xiaohan, Peng Hao, He Lifang, and Philip S Yu. 2020. Heterogeneous similarity graph neural network on electronic health records. In 2020 IEEE International Conference on Big Data (Big Data). IEEE, 1196–1205.

[23]. Lu Hsiao-Ying, Fujiwara Takanori, Chang Ming-Yi, Fu Yang-chih, Ynnerman Anders, and Ma Kwan-Liu. 2023. Visual Analytics of Multivariate Networks with Representation Learning and Composite Variable Construction. Visual Informatics forthcoming (2023).

[24]. Lundberg Scott M and Lee Su-In. 2017. A Unified Approach to Interpreting Model Predictions. In Proc. NIPS, Vol. 30.

[25]. Ming Yao, Xu Panpan, Qu Huamin, and Ren Liu. 2019. Interpretable and Steerable Sequence Learning via Prototypes. Proc. ACM SIGKDD Int. Conf. on Knowledge Discovery & Data Mining (2019).

[26]. Monge Peter R and Contractor Noshir S. 2003. Theories of communication networks. Oxford University Press, New York, NY.

[27]. Pollack CE, Weissman G, Bekelman J, Liao K, and Armstrong K. 2012. Physician social networks and variation in prostate cancer treatment in three cities. Health Serv Res 47, 1 Pt 2 (2012), 380–403. 10.1111/j.1475-6773.2011.01331.x

[28]. Ross Quinlan J. 1983. Learning Efficient Classification Procedures and Their Application to Chess End Games. In Machine Learning: An Artificial Intelligence Approach. Springer, 463–482.

[29]. Rogers Jen, Spina Nicholas, Neese Ashley, Hess Rachel, Brodke Darrel, and Lex Alexander. 2018. Composer: Visual Cohort Analysis of Patient Outcomes. In Proceedings of 9th workshop on Visual Analytics in Healthcare.

[30]. Soulakis ND, Carson MB, Lee YJ, Schneider DH, Skeehan CT, and Scholtens DM. 2015. Visualizing collaborative electronic health record usage for hospitalized patients with heart failure. J Am Med Inform Assoc 22, 2 (2015), 299–311. 10.1093/jamia/ocu017 [PubMed: 25710558]

[31]. Stephen H Taplin Sallie Weaver, Salas Eduardo, Chollette Veronica, Edwards Heather M, Bruinooge Suanna S, and Kosty Michael P. 2015. Reviewing cancer care team effectiveness. Journal of oncology practice 11, 3 (2015), 239–246. [PubMed: 25873056]

[32]. Ho Tin Kam. 1995. Random decision forests. In Proc. Int. Conf. on Document Analysis and Recognition, Vol. 1. 278–282 vol.1.

[33]. Wang Chun-Fu, Li Jianping, Ma Kwan-Liu, Huang Chih-Wei, and Li Yu-Chuan. 2014. A Visual Analysis Approach to Cohort Study of Electronic Patient Records. In Proceedings of IEEE BIBM (to appear).

[34]. Wang Qianwen, Huang Kexin, Chandak Payal, Zitnik Marinka, and Gehlenborg Nils. 2022. Extending the Nested Model for User-Centric XAI: A Design Study on GNN-based Drug Repurposing. IEEE Transactions on Visualization and Computer Graphics (2022).

[35]. Weaver AC, Callaghan M, Cooper AL, Brandman J, and O'Leary KJ. 2015. Assessing Interprofessional Teamwork in Inpatient Medical Oncology Units. J Oncol Pract 11, 1 (2015), 19–22. 10.1200/jop.2014.001536 [PubMed: 25352390]

[36]. Wu Robert, Rossos Peter, Quan Sherman, Reeves Scott, Lo Vivian, Wong Brian, Cheung Mark, and Morra Dante. 2011. An evaluation of the use of smartphones to communicate between clinicians: a mixed-methods study. Journal of Medical Internet Research 13, 3 (2011), e59. [PubMed: 21875849]

[37]. Wu Tong, Wang Yunlong, Wang Yue, Zhao Emily, Yuan Yilian, and Yang Zhi. 2019. Representation learning of ehr data via graph-based medical entity embedding. arXiv preprint arXiv:1910.02574 (2019).

[38]. Zhang Daokun, Yin Jie, Zhu Xingquan, and Zhang Chengqi. 2018. Network representation learning: A survey. IEEE Trans Big Data 6, 1 (2018), 3–28.

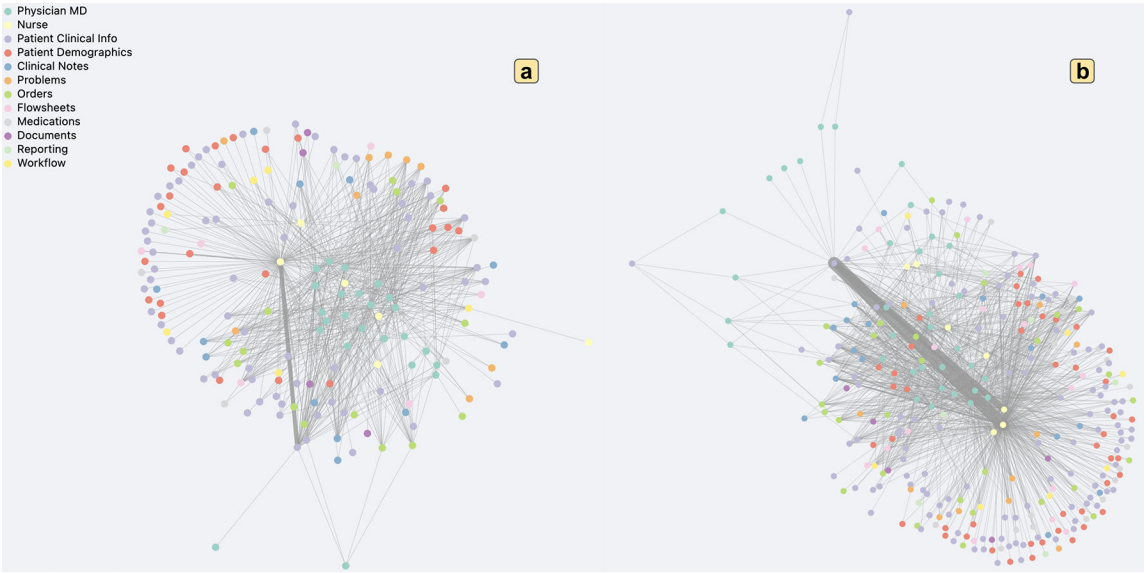


Figure 1:
Examples of the original network, with nodes colored by the higher-level categories. Here we sample one network of an alive patient (a) and the other of a dead patient (b).

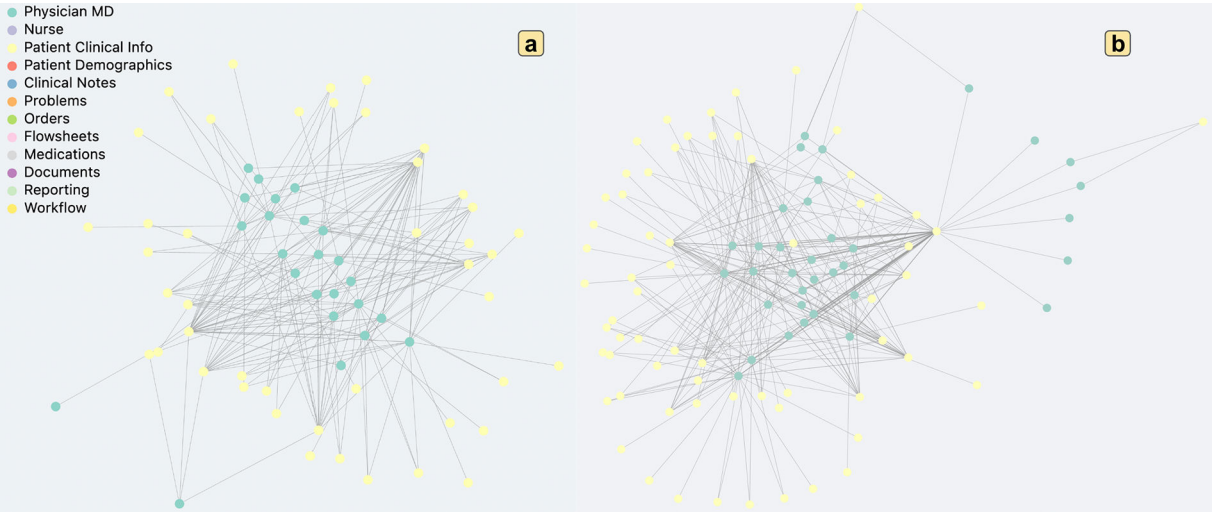


Figure 2:

Examples of the subnetworks extracted from the original networks in Figure 1. Each subnetwork contains nodes of only one provider group (e.g., Physician MD) and one action group (e.g., Patient Clinical Info).

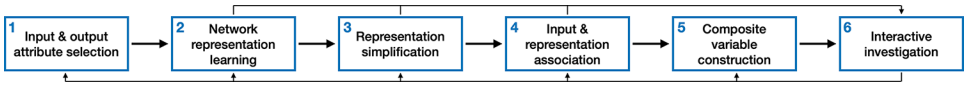


Figure 3:
The visual analytics workflow for understanding the relations in the MTS communication networks.

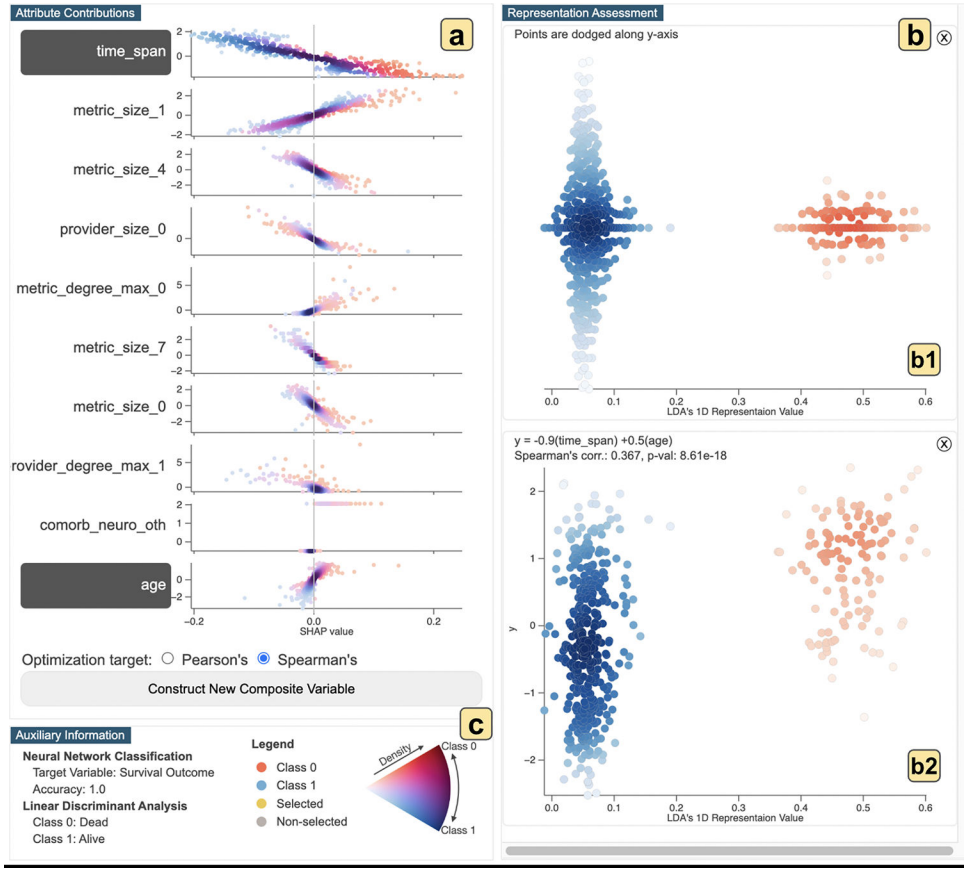
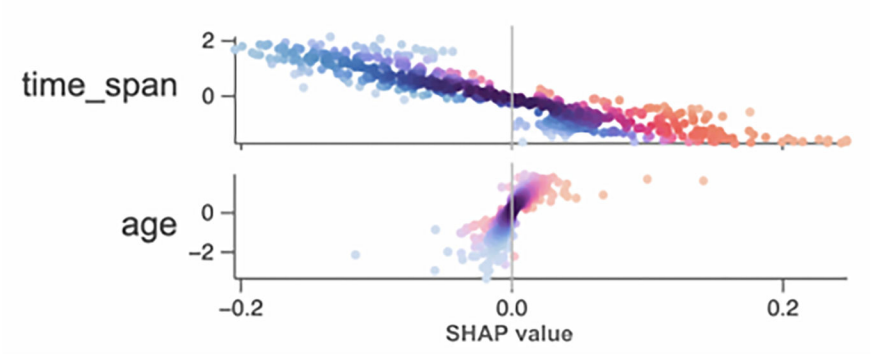
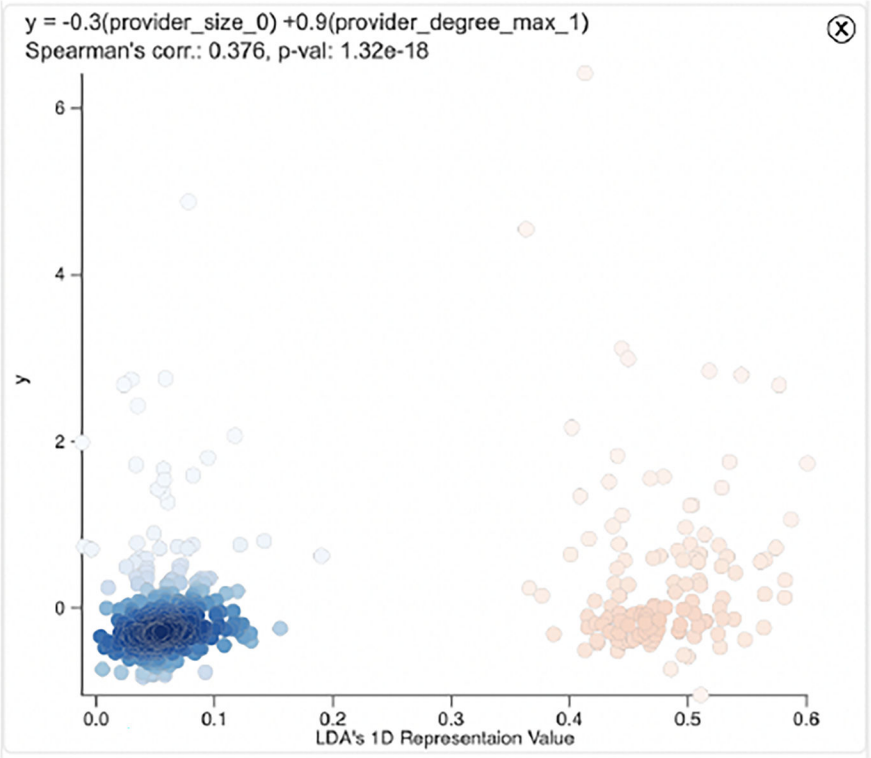


Figure 4:

The visual interface for facilitating interactive analysis using the workflow in Figure 3. (a) Relating to Step 4, this view visualizes the input attributes' contributions to the 1D network representation. (b) The composite variables generated in Step 5 are shown as a list of scatterplots. (c) Other auxiliary information is displayed, including the prediction accuracy of NNs trained for Step 2 and legends used across the views.



(a) Study 1: Opposite correlation observed in the *time_span* and *age* contribution scatterplots.



(b) Study 2: Composite variable analysis for attributes *provider_size_0* and *provider_degree_max_1*.

Figure 5:
Exploration results in two case studies.

Table 1:

The selected categories used for grouping providers and actions in the original network.

Name	Example Node
Patient Clinical Info	Report with patient data viewed
Patient Demographics	Edit Chart activity accessed
Clinical Notes	Clinical note viewed
Problems	Allergies activity accessed
Orders	Order placed
Flowsheets	Health Trends report viewed
Medications	Medication taking status modified
Documents	Annotated image printed
Reporting	Patient History Report accessed
Workflow	New user record created

Table 2:

Selected input and output attributes from the EHR networks.

Attribute	**Values**
Survival Outcome	0-Dead, 1-Alive
Stage	0-Stage 2, 1-Stage 3
Time Span	number of days from diagnosis to last contact
Access Log Span	number of days from first to last access log
Age	age at diagnosis
Sex	0-Male, 1-Female
Treatment Given	0-not given, 1-given
Systemic and Surgery	0-no systemic or surgery, 1-systemic or surgery
Treatment Summary	Chemotherapy, Immunotherapy, Hormone, and Radiation: 0-not given, 1-given
Comorbidities	38 comorbidities: 0-not present, 1-present
(a) Patient-related attributes and values	
Attribute	**Values**
Size	Size of each node group
Degree	Mean, minimum, maximum, and standard deviation of degrees in each node group
Clustering Coefficient	Mean, minimum, maximum, and standard deviation of clustering coefficients in each node group
Connected Components	Number of connected components
Density	Density of each subnetwork between a provider group and an action group
(b) Network-related attributes and values.	