

# Visualization of Longitudinal Clinical Trajectories using a Graph-based Approach

Filip Dabek  
NICoE  
Walter Reed, Bethesda  
filip.j.dabek.ctr@mail.mil

Jian Chen  
Dept. of Computer Science  
and Electrical Engineering  
University of Maryland, UMBC  
jichen@umbc.edu

Alexander Garbarino  
Dept. of Computer Science  
and Electrical Engineering  
University of Maryland, UMBC  
garba1@umbc.edu

Jesus J Caban  
NICoE  
Walter Reed, Bethesda  
jesus.j.caban.civ@mail.mil

## ABSTRACT

The adoption of electronic health records (EHRs) and the increased participation of hospitals and clinics in health information exchange systems have resulted in unique longitudinal data that describes a patient's clinical trajectory. In-depth analysis of that information is important to better understand the general course to recovery or the evolution of a particular disease. Unfortunately, modelling and understanding the progression of a disease is still a challenging task given that most often patients take vastly different paths after being diagnosed with a specific disease or condition. The different trajectories patients follow, the individual temporal events that each patient goes through, the uncertainties associated with clinical diagnoses, and the irregular time intervals between clinical diagnoses present challenges to researchers trying to analyze the common trajectories of a set of  $N$  patients. This paper presents a graph-based visualization method to interactively analyze the longitudinal clinical trajectory of a group of patients. The system allows users to select a specific set of events or conditions, filter the data based on different thresholds, and compare different cohorts while using an interactive virtual space that expands as the user continues to analyze and explore the data. The system has been tested with a dataset of over 89,000 patients and 8.7 million clinical events.

## Keywords

Data visualization, visual analytics, longitudinal events, healthcare data

## 1. INTRODUCTION

The rapid advancements in medical sciences and the increased adoption of electronic health records (EHRs) have

resulted in rich longitudinal data available for most patients with valuable information about the progression of different conditions. The course of most diseases are quite variable, but understanding their trajectory can provide key information about the effectiveness of treatments, insight about potential prognosis, and information about the state of the patient. Unfortunately, due to the complexity and challenges associated with the analysis of sparse longitudinal clinical data, the EHR data is not often used within clinical settings to study trajectories of similar patients. Some of the contributing factors that have limited the amount of information that is extracted from the longitudinal data of EHR systems have been the vastly different paths that similar patients can follow, the irregular intervals at which particular events occur within comparable patients, the comorbidity of diseases many patients experience, the uncertainty associated with clinical diagnosis and coding, and the lack of tools to quickly analyze the information.

At the individual patient level, the lack of effective tools to review longitudinal events and trajectories often forces clinicians into spending a significant amount of time and resources reviewing low-level, unrelated data points while trying to construct a mental model of the patient. In addition, when trying to simultaneously review a cohort of  $N$  patients with the intention of finding common patterns, the scale of the information is so large and the incidental or random conditions for particular patients are so strong that often the relevant patterns and trajectories are overlooked.

Most of the existing approaches to analyze clinical trajectories of patients have focused on finding the top trails followed by a particular cohort and using that information to develop clinical decision support systems to predict whether a new patient will develop a condition [10]. The use of visual analytics applied to healthcare data can enable new insights in the interpretation of complex clinical trajectories and scenarios.

This paper presents a graph-based method to interactively analyze the longitudinal clinical trajectory of a cohort of patients. The system reads from a database of clinical events associated to a particular cohort of patients, estimates the different events, and generates a single graph illustrating all of the possible transitions. The system then allows users to select a specific set of events or conditions, filter the data based on different thresholds, and compare different cohorts

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

WOODSTOCK '97 El Paso, Texas USA

© 2015 ACM. ISBN 123-4567-24-567/08/06...\$15.00

DOI: 10.475/123.4

all while using an interactive virtual space that expands as the user continues the data analysis and exploration. We have implemented a system using data from over 89,000 patients and 8.7 million clinical events and have shown a usage scenario for comparing groups of patients before and after a clinical event.

## 2. BACKGROUND

In the clinical realm, existing approaches to understanding the clinical trajectory of patients have focused on producing a single trajectory graph for understanding a population of patients. In many approaches researchers have attempted to use a popular algorithm, k-means clustering, to identify the unique groups of patients and label them according to the group that they belong to (i.e. chronic, recovery, worsening, etc.) [10, 15, 1].

Many of the prior approaches to understanding the trajectory of a patient have focused on a specific disease with the intention of being able to predict whether a new patient will develop a similar diagnosis or not [8]. Moskovitch and Shahar analyzed a dataset of diabetic patients to the goal of identifying the most common patterns to be able to cluster patients' trajectories in a hierarchical manner with the ultimate goal of understanding the overall population [9]. Quaglini et al. developed a framework for making clinical decisions by involving the patient in the process of decisions by allowing them to fine tune the model's parameters based on their preferences, allowing for a system that is able to understand the patient more effectively [13]. Related to psychological conditions such as post-traumatic stress disorder (PTSD), Bryant et al. studied the different trajectories that PTSD patients follow including recovering and worsening [4, 5].

Other approaches to evaluating the trajectory of patients have looked for frequent patterns in EHR data by applying a "frequent sequence mining" algorithm to extract a set of frequent patterns and furthermore computing the correlation between each pattern and various outcome measures. The events, subsequences, and probability measures are then presented in a modified Sankey diagram allowing the user to interactively analyze the mined data [11, 12].

In the visual analytic space, most existing tools are restricted to presenting visualization results to a single view at a time, limiting data comparison and forcing humans to store information in their working memory during an analysis process [14]. Additionally, displaying data is limited by the physical size of the screen [2].

To overcome some of the limitations and challenges discussed in Section 1 and 2, most of the previous approaches remove the "uncommon" patients that do not follow the majority of the population. In addition, previous approaches present a single static graph for analysis with zero ability for an expert clinician to interact with and identify key patterns and trends.

## 3. DATASET

The dataset used for this project was a subset of our larger clinical collection that consists of over 15 million patients and billions of clinical encounters. A dataset with 8.7 million clinical records from 98,342 mTBI patients was used for this study. The data was organized using various Teradata/Aster services and retrieved using MapReduce. The

original dataset of 98,342 patients was further filtered to only include patients with more than thirty days of data and no history of moderate or severe TBI. The resulting subsets of 89,840 patients had 5.3 million TBI-related clinical encounters and 8.7 million clinical diagnoses. In this study, a TBI-related encounter was defined as a visit to the doctor regardless of inpatient/outpatient where the patient was treated with one or more of the conditions that are commonly known to be related to concussions such as behavioral disorder, sleep problems, cognitive deficiencies, and audiology complaints. Note that only TBI-related encounters were taken into consideration. The patients under consideration had an average of 59.06 encounters.

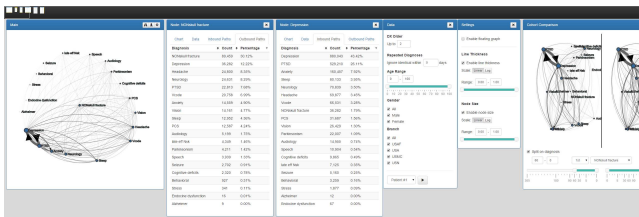
To build our system the dataset was defined to be  $P = \{P_1, P_2, \dots, P_n\}$  where  $P$  is a set containing each patient  $P$ . Each patient  $P_i$  had an associated sequence of encounters  $E_i \in \{E_1, E_2, \dots, E_m\}$  representing unique clinical appointments or hospital visits. Each encounter  $E_i$  had an associated set of diagnoses represented by  $D \in \{D_1, D_2, \dots, D_k\}$ . The diagnoses were then used to create the force-directed graph that is covered in Section 4.

## 4. APPROACH

With a vastly large clinical dataset, we sought to develop a system that would support interactive analysis such that users would not become lost in the data and would be encouraged to explore the data further. Therefore, we developed a web-based framework for interactive exploration of clinical encounters (Figure 1), packed with features that include:

1. **Bubble views:** applying the ideas of VisBubbles [7], our system extends the traditional static multiple view into a panel-based interface. Panels do not overlap and can be rearranged to help users perform dynamic analysis and switch views for comparative studies.
2. **Virtual working space:** our system applies the idea of "virtual screen space extension" [6] by providing a panning navigation bar at the top of the window and extending the current view space to a much larger continuous working space.
3. **Force-directed graph visualization:** in the center of our system we display the longitudinal events such as clinical diagnoses undergone by patients in a single force-directed graph. The nodes encode information about specific events (in our case a clinical diagnosis) and the edges encode information about the frequency of observing a transition between two events within the clinical trajectories of the patients under consideration.  
The forced-directed graph can be arranged based on a particular attribute such as the frequency of a specific event happening and can be used to open other panels for a deeper exploration into the data.
4. **Built-in Filters:** directly from the graph-based visualization representing a summary of all the patient trajectories, the user can open a panel to filter the data being analyzed. The user can filter based on attributes that can be added to the longitudinal sequence of events such as gender, race, age, military branch, and many others.

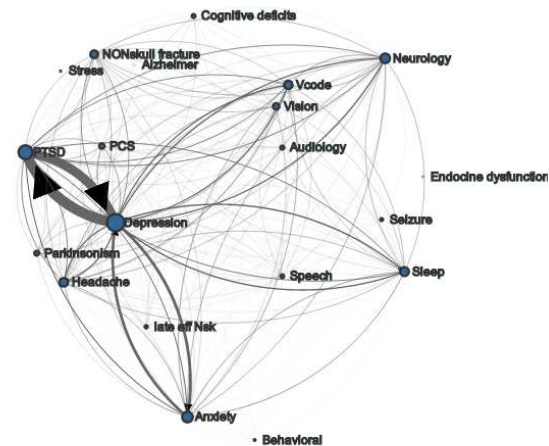
- Overview panels:** the exploration panel with the graph-based visualization supports looking at a detailed overview of the data in a tabular form to enable users to better understand the data being presented. For example, from the graph-based visualization the user can load to overview panel of a particular event (e.g. concussion) and obtain information about what's the event that leads towards concussion the most and the events that follow concussion the most.
- Cohort comparison:** the visualization system allows users to open a new force-directed graph and split the screen so side-by-side comparison of different cohorts can be accomplished.
- Interactive Interface:** the visualization system has been developed so the user can quickly interact with the data, the properties of the longitudinal events, and even select a subset of nodes (i.e. events) and open a new (simpler) graph-based visualization of the data. The system has been optimized to support caching of files on the server side, thus allowing the application to quickly load datasets that have thousands of patients and millions of events without any kind of delays to the end-user.
- Patient specific analysis:** the framework has been designed to support the visualization, analysis, and comparison of the trajectory of a particular patient against the group. The system support a "play" functionality that illustrate within the graph the different events and transitions a particular patient underwent.



**Figure 1:** A view of our full interface with a user interacting using multiple panels. From left to right: the main graph interface of the entire dataset, two panels with detailed path information for nodes, a data panel to filter the data, a settings panel to manipulate the visualization, and a cohort comparison panel to compare the dataset before and after a first concussion.

Our interface was developed using Javascript and HTML5, with the aid of D3 for drawing and rendering visualizations [3]. Ajax is used to communicate with a PHP server which serves pre-processed data from a database in a JSON format. The system as a whole was built in a modular fashion to allow for future features in interactive data analysis.

We chose to represent our dataset in a force-directed graph because as discussed in Section 2, different approaches have been proposed to identify and analyze the common trajectory of a group of patient by showing the common paths or by clustering the patients into separate groups, but being able to understand and visualize an entire population in a single visualization has not received too much attention.



**Figure 2:** The drawing of the force-directed graph that is rendered by our system where the node sizes correlate to the number of times that diagnosis occurred over all patient, and the path thicknesses correlate to the number of times that the patients took the path from the source node to the target node. For example, in this graph the majority of patients moved from Depression to PTSD as well as in reverse from PTSD to Depression.

To build our graph that is shown in the main panel of our system we first determine all the unique clinical diagnosis on our dataset, treated them as temporal events, and created a node for each possible diagnosis,  $D$ , which resulted in a total of 20 nodes. Then we drew a path/force with an arrow from each node to every other node, representing a patient moving from one diagnosis to the next within their clinical trajectory. At each node we stored the number of times that the diagnosis occurred in our dataset and at each path we stored the number of times that unique transition (i.e. Anxiety  $\rightarrow$  Depression) occurred in our dataset. In the next section we will describe the manipulations and encodings that we provided to the user to alter the graph and make patterns and trends be identifiable, such as is shown in Figure 2.

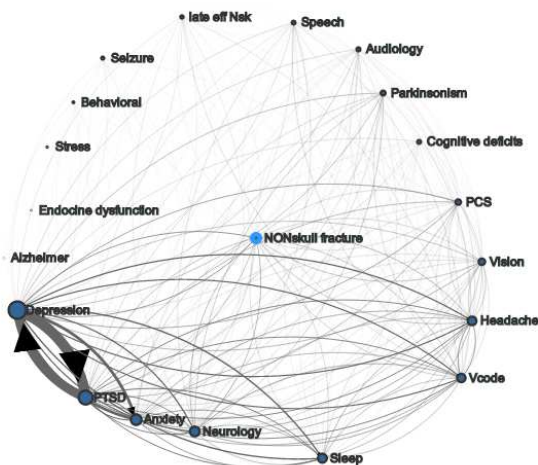
## 5. RESULTS

Our force-directed graph is able to depict millions of clinical diagnoses through various node and path encodings while providing users with various mechanisms and features to manipulate the graph and understand its complex underlying data. In this section we will show how users are able to interact with the graph and spawn new panels, use and manipulate the graph's various encodings, view detailed information at each node, filter the displayed data, and compare patient groups and varying periods of patient timelines using our cohort comparison tool.

Users are given the ability to move nodes around, allowing

for them to interact and identify potential groupings in the data. This method of dragging nodes around also allows them to transform a random force graph into something more meaningful, such as a circular graph. For example, shown in Figure 3 is a force-directed graph in which the user dragged the nodes to be sorted by in a clockwise manner by the size of each node and then moved one node into the center that they were focusing on to understand the relationships between the singular node and all other nodes in the graph. In this example, the user was concentrating on nonskull fractures (concussions) and analyzed the flow of diagnoses, to see what diagnoses lead to concussions and what diagnoses to expect after a concussion. This is critical information that a user is able to interactively extract from the graph without needing to analyze millions of patient records.

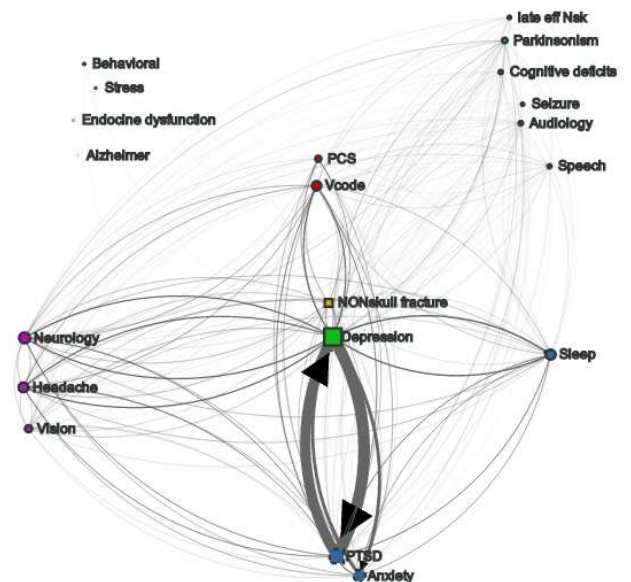
Building off of the ability to drag and drop nodes interactively, we have built a selection feature into our system where users are able to highlight a set of nodes and choose to open them in a new panel to further explore only a certain subset of nodes. Combining the ability to select a set of nodes to analyze along with the ability to spawn separate panels in a virtual workspace environment will enable users to keep track of their analysis and be able to continuously compare various subsets of data.



**Figure 3:** In this force-directed graph, the nodes were dragged in such a way that they are ordered clockwise by their size in descending order and the node for concussions (nonskull fracture) was moved to the middle. With this technique the relationships between all diagnoses and concussions can be thoroughly analyzed.

Next, to provide the user with information about the complex clinical data in a simple graph, we encode information into the nodes and paths to give an understanding of the amount of patients flowing through each area of the graph. First, we encode each the number of times that each diagnosis occurred over all patients into the size of the node,

such that larger nodes occur more often while smaller nodes rarely occur. Second, we encode each path with the number of patients that took corresponding path between nodes, such that thicker paths were traversed more often. Third, we give the user the ability to manipulate the graph further by altering the shape, color, and outline of nodes to their own encodings, as can be seen in Figure 4. Using these encodings in just a simple graph provides a user with a deeper understanding of the flow of patients between diagnoses.

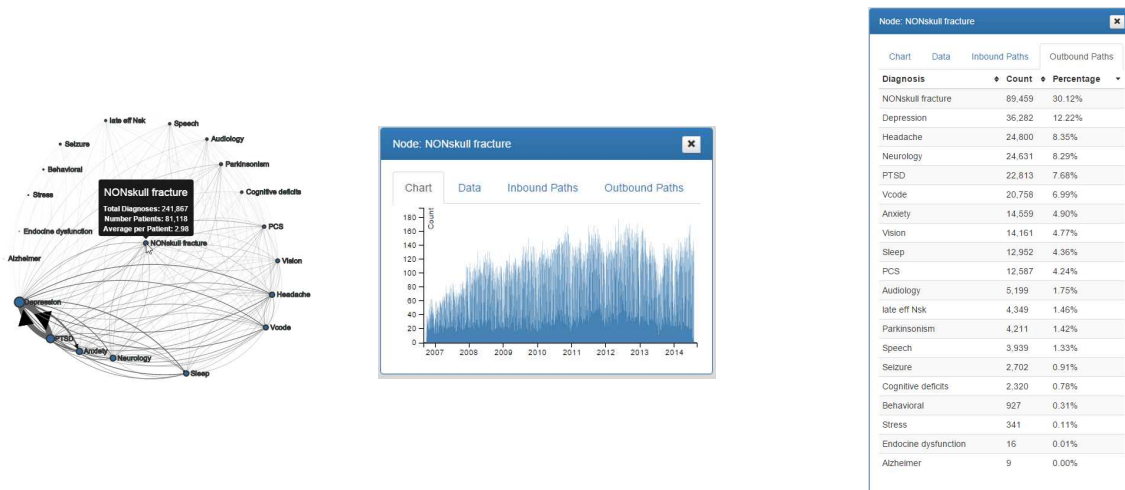


**Figure 4:** A force-directed graph where the user grouped the nodes and assigned various colors, shapes, and outlines to each.

Along with the encodings, users are able to hover their mouse over a node and/or a path to be presented with the raw data that is encoding into that aspect of the graph (Figure 5 Left). Furthermore, by clicking on a node users are presented with a new panel that contains an area chart, data, and inbound and outbound paths. The area chart displays the number of diagnoses over time, the data reiterates the information from the tooltip, and the inbound/outbound paths display a table with each path into and out of the node with their corresponding count and frequency, along with giving the user the ability to sort the table in ascending and decreasing order. This panel of information can be seen in Figures 5 Middle and 5 Right.

In addition to the various interactive methods of manipulating the graph, we allow users to filter the dataset and visualization that the graph is shown. As seen in Figures 5 Left and 5 Right, users are able to filter the data to only include certain ages, branches, genders, or encounters; while they are also able to modify the graph to only draw lines within a certain range of thicknesses and nodes within a certain range of sizes. These filtering methods allow for an exploration of only viewing paths that are relevant and minimize the amount of noise in the graph.

We present a usage scenario in the clinical realm where clinicians that are interested in comparing groups of patients and understanding the history of a patient before and after



**Figure 5: (Left)** When a user mouses over a node or a path, they are presented with detailed information in a tooltip about the number of times the specific node/path had occurred, the number of patients that took that route, and the average occurrences per patient.

**(Middle)** In this panel the information for a node has been displayed in tabs: a chart, raw data, and inbound/outbound paths. The selected tab, chart, displays an area chart of how many times the diagnosis occurred over time.

**(Right)** In this panel the user had selected the “Outbound Paths” tab which displays all of the paths from the selected node to every other node and the count and percentage of each path occurring. In addition, each column in the table of outbound paths is sortable in ascending and descending order.

an event. This comparison aids clinicians in identifying the susceptibility of a condition as well as the path to expect a patient to follow after the event, giving the clinician the ability to enact preventive care. With this our system is able to display this information using a Cohort Comparison panel which shows two force-directed graphs side by side for easy comparison, as shown in Figure 7. In this cohort comparison panel, the two graphs are aligned such that the nodes are in identical positions to make the visual analytic process of identifying the differences in graphs easier as the human brain has a hard time comparing two completely different images. Using this technique, users will be able to identify that a node in one graph is larger than another or that a certain group of patients follows through a different path flow than the other based on the line thicknesses. With the data panel that was previously explained, shown in Figure 5 Left, users are able to manipulate the data that is used for each individual graph allowing for comparisons such as: male vs female, army vs navy, 20-30 vs 40-50, as well as many other combinations which will aid in the process of identifying the differences in patients and the varying effects.

As clinicians not only interested in the differences in patient groups, but also the changes that occur in patients before and after an event, such as a concussion, we provide users with the ability to display graphs based on a period of a timeline for a patient. In the previous Figure 7 the user had chosen to split the patients’ timeline at the point of the first concussion (nonskull fracture) such that the left graph included the data prior to the concussion and the right graph included the data after the concussion. With that, the user also chose to only include the 60 days before the concussion while including the full 365 days after the concussion. The user is now able to see the difference in the type of diagnoses that patients commonly underwent pre and post concussion

through the various node sizes as well as the difference in diagnosis flows through analyzing the thicknesses of the paths.

## 6. CONCLUSION

In this paper we have presented our graph-based method for interactively analyzing the longitudinal clinical trajectory of a cohort of patients. With this system a user is able to effectively filter their dataset based on age, gender, branch, etc., manipulate the visualization through various encodings and dragging maneuvers, compare different cohorts and the before and after effects of an event; all the while using an interactive, virtual space compared to the standard static views that have been commonly developed. We have shown the initial effectiveness of our method on a clinical dataset of over 89,000 patients and 8.7 million clinical events through the generated graphs. In the future we look forward to developing our system further followed by a user study evaluating the insights gained by users. Some of our planned features include providing users with the ability to “play” a patient through the graph, where the nodes would highlight in the order of the diagnoses that the patient underwent, providing a means to subtract graphs to analyze the difference, and other methods of clinical exploration in a virtual space.

## 7. REFERENCES

- [1] A. Abu-Hanna and P. J. Lucas. Prognostic models in medicine. *Methods of Information in Medicine-Methodik der Information in der Medizin*, 40(1):1–5, 2001.
- [2] C. Andrews, A. Endert, B. Yost, and C. North. Information visualization on large, high-resolution displays: Issues, challenges, and opportunities.



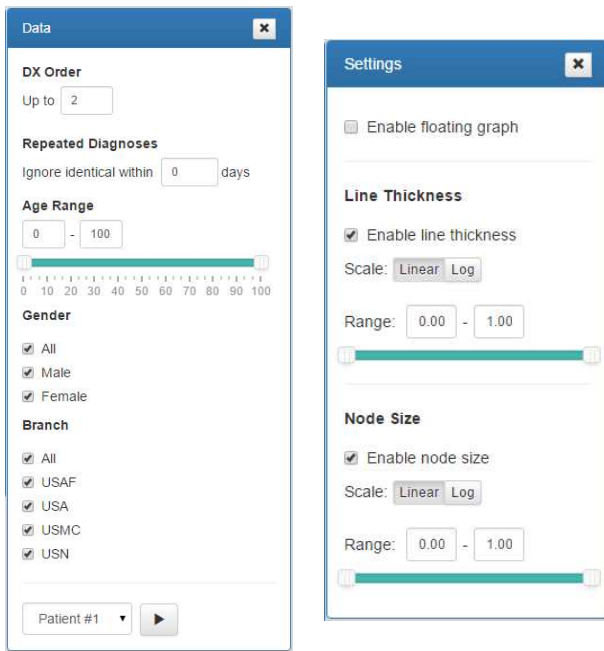


Figure 6: (Left) Users are able to filter the data that is used to render the force-directed graph by limiting the data to the most important diagnoses (dx order), ignoring repeated diagnoses (such as a diagnosis of depression over and over for many consecutive days), limiting to a certain age range, gender, or branch.

(Right) Users can manipulate the rendered graph to either make it static or “floating” where gravity exists and the nodes respond according to their weights. In addition, users can enable/disable, change the scale, and set a visibility range for the path thicknesses and node sizes. The visibility range removes any nodes or paths that are outside of the range selected, allowing the user to only focus on certain aspects of the graph.

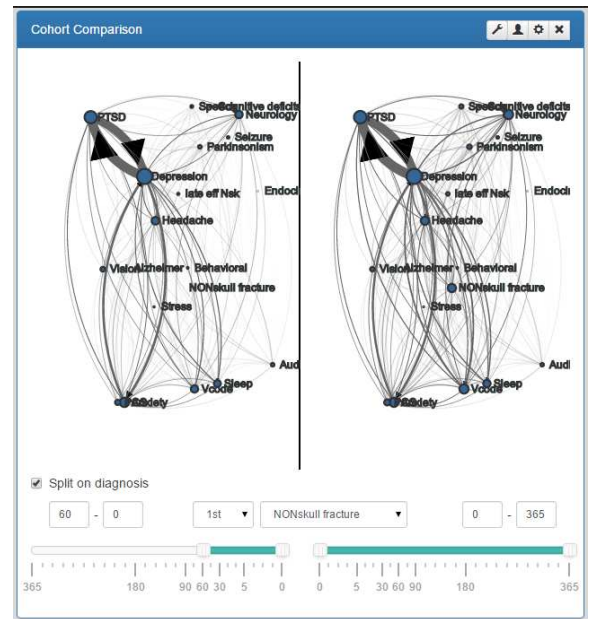


Figure 7: In this panel there exist two force-directed graphs where the nodes are rendered in the same coordinates to allow for easy comparison. The user is able to compare two different cohorts of users by filtering the data or they are also able to split all patient trajectories at the occurrence of a diagnosis to compare the difference in patients before and after the split point. This is critical in understanding the change in diagnoses and the trajectories that patients undergo after a diagnosis: for example the difference that a first concussion makes.

*Information Visualization*, page 1473871611415997, 2011.

- [3] M. Bostock, V. Ogievetsky, and J. Heer. D<sup>3</sup> data-driven documents. *Visualization and Computer Graphics, IEEE Transactions on*, 17(12):2301–2309, 2011.
- [4] R. A. Bryant, J. E. Marosszeky, J. Crooks, and J. A. Gurka. Posttraumatic Stress Disorder After Severe Traumatic Brain Injury. *American Journal of Psychiatry*, Nov. 2014.
- [5] R. A. Bryant, A. Nickerson, M. Creamer, M. O’Donnell, D. Forbes, I. Galatzer-Levy, A. C. McFarlane, and D. Silove. Trajectory of post-traumatic stress following traumatic injury: 6-year follow-up. *The British Journal of Psychiatry*, page bjp.bp.114.145516, Jan. 2015.
- [6] D. A. Henderson Jr and S. Card. Rooms: the use of multiple virtual workspaces to reduce space contention in a window-based graphical user interface. *ACM Transactions on Graphics (TOG)*, 5(3):211–243, 1986.
- [7] G. Li, A. C. Bragdon, Z. Pan, M. Zhang, S. M. Swartz, D. H. Laidlaw, C. Zhang, H. Liu, and J. Chen. Visbubbles: a workflow-driven framework for scientific data analysis of time-varying biological datasets (sap-0281). *SIGGRAPH Asia Posters. Hong Kong, China: ACM*, 2011.
- [8] I. Marinic, F. Supek, Z. Kovacic, L. Rukavina, T. Jendricko, and D. Kozaric-Kovacic. Posttraumatic Stress Disorder: Diagnostic Data Analysis by Data Mining Methodology. *Croat Med J*, 48(2):185–197, Apr. 2007.
- [9] R. Moskovitch and Y. Shahar. Medical temporal-knowledge discovery via temporal abstraction. *AMIA Annu Symp Proc*, 2009:452–456, 2009.
- [10] D. S. Nagin and C. L. Odgers. Group-based trajectory modeling in clinical research. *Annual Review of Clinical Psychology*, 6:109–138, 2010.
- [11] A. Perer and F. Wang. Frequence: Interactive Mining and Visualization of Temporal Frequent Event Sequences. In *Proceedings of the 19th International Conference on Intelligent User Interfaces, IUI ’14*, pages 153–162, New York, NY, USA, 2014. ACM.
- [12] A. Perer, F. Wang, and J. Hu. Mining and exploring care pathways from electronic medical records with visual analytics. *J Biomed Inform*, 56:369–378, Aug. 2015.
- [13] S. Quaglini, Y. Shahar, M. Peleg, S. Miksch,

C. Napolitano, M. Rigla, A. Pallás, E. Parimbelli, and L. Sacchi. Supporting Shared Decision Making within the MobiGuide Project. *AMIA Annu Symp Proc*, 2013:1175–1184, Nov. 2013.

- [14] M. Q. Wang Baldonado, A. Woodruff, and A. Kuchinsky. Guidelines for using multiple views in information visualization. In *Proceedings of the working conference on Advanced visual interfaces*, pages 110–119. ACM, 2000.
- [15] X.-H. Zhou, D. K. McClish, and N. A. Obuchowski. *Statistical methods in diagnostic medicine*, volume 569. John Wiley & Sons, 2009.