Bicentric Visualization of Pediatric Asthma Care Process Activities

Rahul C. Basole, Hyunwoo Park, Vikas Kumar, Mark L. Braunstein, James Bost, Duen Horng Chau, and Minsuk Kahng

Fig. 1. The prototype next generation of the Bicentric Care Activity Visualization tool features: (a) Selection of two focal activities (highlighted in yellow) and their desired properties; (b) Dynamic filtering by patient characteristics; (c) Selection of providers; (d) Visualization pane with zoom+pan and controls. This example shows clinical activities related to an Inhalation Spacing Device (on the left) and Thyroglobulin Antibody (on the right).

Abstract—Health care delivery processes consist of a complex set of clinical activities. We have designed and developed a graph-based set visualization layout – called the bicentric layout – that allows simultaneous depiction of care activity sets, tiers, and connections. We demonstrate the applicability and value of our layout for clinical decision support using EMR data from 5,784 pediatric asthma exacerbation emergency department visits.

Index Terms—Visual analytics, graph-based set visualization, EHR, process mining, asthma, emergency care, pediatric hospital

1 INTRODUCTION

Effective techniques are needed to analyze and visualize the vast amounts of clinical data in electronic health records (EHRs) for specific purposes. Specifically, interactive visualization could help clinicians digest, summarize and more effectively use raw EHR data, which is almost boundless in its complexity and volume [6]. It could allow clinicians to explore patient data in novel ways and synthesize and test hypotheses on-the-fly. This could provide insights about the current state of care and how it can be made more effective and efficient. It could also inform ideas for more targeted, lower-risk future studies, such as resource-intensive clinical trials. However, clinical data poses challenges that make it harder to visualize than data from other domains. These include temporal complexity; noisiness; missing values; and a variety of data types [4].

Numerous previous studies have provided meaningful insights for visualizing clinical data (for a review, see [5]). These differ largely in characteristics such as the number of patient records visualized at once, the data types considered (e.g. event, categorical, or numerical data), and the supported interaction features.

We chose to centrally focus on characteristic events that identify a patient as belonging to a specific clinical or etiological group. For example, a pediatric asthma patient who has received blood cultures and antibiotics is likely suspected of having an infectious etiology, while a patient who has received an anti-histamine is likely suspected of an allergic cause. We believe this approach can be useful for more simply
inferring etiologies and diagnoses from large clinical datasets.

More specifically, we hypothesize that by comparing characteristic events we can usefully group patients based on clinical factors for purposes such as planning clinical research studies or in evaluating the differences in care processes based on suspected disease etiology.

Many of the studies reviewed in [5] and others have a “Select” or “Filter” functionality to select patients based on demographic characteristics or clinical events. Here, we introduce a novel, graph-based, bicentric activity browser adapted from a previous, non-clinical study [3] that allows clinicians to see variations in their approach to clinically significant groups of patients. We apply our visualization method to EMR data obtained from Children’s Healthcare of Atlanta (Childrens) for asthma-related pediatric emergency department (ED) visits.

2 Data

Our dataset includes all ED visits over a 13-month period for which asthma was the primary diagnosis. For each of these 5,784 visits we obtained information regarding administrative events, clinical respiratory test events, laboratory test events and medication administration events with their date/timestamps. We also received detailed demographic, charge, and provider-related information for each visit. For this study we focused only on the visualization of laboratory and medication-related events for patients grouped based on laboratory tests or medications. We ignored administrative events since they are performed for almost all patients. The data was received as comma separated value (csv) files split into several tables as a relational database. The visit.csv file contained 5,785 visit observations and had 143 attributes, including demographic information and administrative timestamps. The medications.csv and labresult.csv files contained information regarding medication and lab-related date/timestamps, respectively.

After cleaning and integrating the EHR data we constructed an aggregate activity network graph.1

3 Visualization Layout and Design

Guided by our objective of understanding the activity network of pairs of clinical activities, we adapted the bicentric layout presented by [2]. This layout uses ideas from set visualization [1] and provides an effective representation of two focal nodes as well as their shared direct and indirect activities ordered by tier. A conceptual representation of this layout is shown in Figure 2.

The two focal nodes of interest (A and B: the activities that define the two patient groups) are positioned at a distance d apart and highlighted using a transparent yellow halo effect. Two concentric circles with radii $d/2$ and $d$ are drawn around each focal node and are where other (non-focal) nodes are placed. The inner circle contains the focal node's 1st-tier nodes (1-step away neighbors). Similarly, the outer circle contains 2nd-tier nodes (2-steps away). Thus, an activity that co-occurs with a 1st-tier node but not with a focal node would be 2-steps away from the focal node in the outer circle. An edge between two nodes represents the co-occurrence of those two activities in a patient visit. For instance, if a certain patient was treated by medications A and B, an edge is drawn between A and B.

The thickness of an edge corresponds to the number of patients that received the two activities during the same visit.

The intersecting points of the concentric circles represent areas for positioning nodes corresponding to the various sets. Nodes directly connected to both focal nodes (i.e. are 1-step away) are placed in a cluster at the center of the focal dyad. On the other hand, nodes that co-occur with only one of the focal nodes are positioned on the semi-circle on the side of that focal node. We differentiate between nodes placed at the top and the bottom intersection points. Nodes at the top belong to the main component (denoted as $\in mc$, where mc is main component); all others within the corresponding set are placed at the bottom (and denoted as $\notin mc$). The main component refers to the largest weakly connected component of a graph, where all nodes

1For a diagram of the relational schema as well as detailed pseudocode of our activity graph construction, please see the Supplementary Information provided in [2].

within that component can reach each other through some paths. This differentiation enables us to further understand the cohesion pattern of subtier clinical activities. To reduce visual clutter and improve readability and aesthetics, we apply "no overlap" and node jittering rules.

The node diameter is sized proportionally to the square root of the number of patients receiving this activity. The larger the node, the more common the activity is in the context of asthma care for patients in this dataset. The visualization currently contains only two node types: medications (depicted in red) and lab results (depicted in blue). A control sliderbar allows us to set a minimum level of co-occurrence to eliminate rarely co-occurring activities thereby reducing visual "noise". In the following use case examples, we use a consistent threshold of 5 percent or more.

4 Use Cases

4.1 Visualizing an Activity and its Component

An asthma exacerbation may have an infectious or allergic etiology which affects its treatment. Moreover, infections divide into those with a viral etiology which can only be symptomatically treated and those with a bacterial etiology which can be treated with antibiotics. At the Childrens ED the C-Reactive Protein (CRP) laboratory test is often used to help differentiate viral from bacterial infections. In practice, only children with a high fever or who look toxic will likely have this test done. In our dataset 390 of 5,785 ED visits had a CRP determination. In Figure 3, we group patients on the left who received a CRP and we group patients who did not have a CRP on the right. In the middle of the two focal nodes one would expect to find general treatments for all asthma cases, and that is what we see. One would also expect that patients whose CRP was high would have additional lab tests ordered to more specifically define their problem and this is also seen in the cluster of nodes on the upper left, which shows a predominance of red laboratory test nodes. This cluster presumably represents physicians further "working up" these patients to verify a bacterial infection and determine the proper antibiotic by finding its source and identifying the organism. Patients on the right are apparently being treated for the asthma exacerbation and any associated allergic problems. This example suggests that, with the proper selection of focal nodes, our tool can find clinically significant patient clusters and their associated care.

In a future version of the tool we would expect to be able to look at this in more detail based on the actual results of one or more tests to look at the underlying clinical logic and/or care by physician to discover individual approaches to the same clinical challenge.

4.2 Identifying Shared Activities

In Figure 4 the focal nodes are patients with a CRP determination on the left and those actually treated for bacterial infection (with the most
Fig. 3. Identification of activities by clinical condition. Activities for patients suspected of an infectious problem are shown on the left, those for a likely allergic problem are shown on the right. In this case both sides represent patients with a suspected (left) or actual (right) infectious process, so one would expect that most other clinical activities would co-occur with both focal nodes and, indeed, virtually all other clinical activities are in the middle between the focal nodes.

4.3 Identifying Non-shared Activities
Conversely, in Figure 5, patients on the left had a CRP determination but those on the right were treated for an allergic problem (with the medications diphenhydramine (Benadryl) and/or epinephrine). Here one would expect few shared activities and, once again, this is the case with more specific laboratory tests to further investigate infection or the administration of antibiotics to treat it on the left and treatments for allergic problems or asthma symptoms on the right and with no activities co-occurring with both focal nodes in the middle.

5 Concluding Remarks
This is a preliminary exploration of a novel technique for usefully visualizing large amounts of clinical data to support more effective, cost efficient care as well as clinical research. There are obvious limitations to the dataset used. It represents only two aspects of care (laboratory tests and medications) for only one condition (pediatric asthma) in only one care venue (a single emergency department) and, as a result, care delivered over a limited period of time. However, despite these limitations, the results are easy to interpret, confirm and illustrate the anticipated logical clinical expectations and have the clear potential to provide new insights when used with a richer, more clinically complex dataset and a more advanced version of the tool.

We are in the process of finalizing a prototype activity visualization browser (as shown in Figure 1), which allows clinicians to interactively explore the bicentric layout of clinical activities. The browser consists of a visualization panel and multiple controls to dynamically filter the visualization by patient, provider, and care process characteristics. Changes in focal activity nodes result in an animated transition of all other activities enabling users to see the overall interconnectedness within the data. This next generation tool will provide more advanced features such as comparisons of different physicians (using the same focal activities) to determine how their care processes differ in order to determine those physicians that provide the most streamlined, efficient care and ultimately help to determine optimal clinical care guidelines. We are also planning to incorporate charge and patient financial class information to enable assessment of the cost effectiveness of different care processes.

While our results are preliminary and are based on a relatively limited dataset, they provide clear evidence that this technique could potentially allow for further exploration of care in such important scenarios as complex and very expensive high technology inpatient cases or the care of adult multi-chronic disease patients at multiple venues over long periods of time, arguably the single largest driver of U.S. healthcare costs.

Acknowledgments
This work was supported in part by Children’s Healthcare of Atlanta, the Raytheon Faculty Fellowship, the Tennenbaum Institute and the Institute for People and Technology at Georgia Tech.

References
Fig. 4. Clinically similar patients share most activities. Patients with suspected infections are shown on the left and those treated for an infection are on the right.

Fig. 5. Clinically different patients do not share activities. Activities for patients with a suspected infectious etiology are shown on the left and those with suspected allergic etiology or actual allergic problems are shown on the right.