CPViz: Visualizing Clinical Pathways Represented in Higher-Order Networks

***Junghoon Chae*[*], Byung H. Park[*], Minsu Kim[*], Everett Rush[*], Ozgur Ozmen[*], Makoto Jones[†‡], Merry Ward[†], Jonathan R. Nebeker[†‡]**
[*] ***Oak Ridge National Laboratory***
[†] ***Veterans Administration***
[‡] ***University of Utah School of Medicine***

Abstract

To improve clinical care practice, it is important to understand the variability of clinical pathways executed in different contexts (e.g., pathways in different geographical locations, demographics, and phenotypic groups). A common way of representing clinical pathways is through network-based representations that capture trajectories of treatment steps. However, first-order networks, which are based on the Markovian property and the de facto standard model to represent transitions between steps, often fail to capture real trajectories. This paper introduces a visual analytic tool to explore and compare pathways represented in higher-order networks. Because each higher node in the network is a subtrajectory (i.e., partial or full history of treatment steps), the tool can display true sequences of treatment steps and compute the similarity of the two networks in a space of higher-order nodes. The tool also highlights areas in which the two networks are similar and dissimilar and how a certain subtrajectory is realized differently in different pathways. The paper demonstrates the tool's usefulness by applying it to multiple antidepressant pharmacotherapy pathways for veterans diagnosed with major depressive disorder and by illustrating heterogeneity in prescription patterns across pathways.

Introduction

Clinical pathways (CPs) are typically structured healthcare plans designed to implement evidence-based clinical guidelines, medical algorithms, and protocols [12]. Intended to improve the quality of personalized care, establish cost-effective and evidence-based care management, and standardize care procedures, CPs have become increasingly important for clinical process optimization and communication between different stakeholders in clinical process management. To improve CPs or enforce a new policy, it is important to capture CP variability in different contexts, such as hospitals in different geographical locations or demographic subgroups, and perform comparative analysis for various outcome measures (e.g., cost, survival rate).

A comparison of pathway pairs (i.e., pathway variants) is often conducted by measuring topological similarities in their graph representations (Figure 1 (Top)). The Markov property-based (first-order dependency) network is mostly used to model a pathway and portrays a compact and intuitive representation. However, it does not show complete trajectories (i.e., treatment histories), and thus often presents the wrong impression about the treatment process. This paper adopts the higher-order network [14] to represent a CP, in which a trajectory of a treatment sequence appears as a node (Figure 1 (Bottom)).

A higher-order network invariably has a more complex structure than a first-order network. To handle the increased complexity, we created the Clinical Pathway Visualization system (CPViz). Given a number of CPs to compare, CPViz compares distances between all pairs and offers interactive functionalities to explore similar and dissimilar treatment patterns between a given pair of pathways. As an example, we describe how CPViz is used to conduct a comparative analysis of ten US Department of Veterans Affairs (VA) antidepressant pharmacotherapy treatment pathways for an Operation Enduring Freedom (OEF) and Operation Iraqi Freedom (OIF) cohort diagnosed with major depressive disorder (MDD) [1].

Background and Related Work

In this section, we briefly describe how clinical pathways can be represented as higher-order networks. Also, we discuss previous studies focusing on visualizing clinical pathways and clinical events. We conclude the section with a review of existing work regarding higher-order network visualization.

Background: Higher-Order Network for Clinical Pathways

A network, $G = (V,E)$, is a graph with vertices, V, as objects and edges, E, as links between objects. In a first-order network representation of a CP, V is a set of single treatment steps (e.g., taking 50 mg of an antidepressant daily for 2 months). In a higher-order network representation, $v \in V$ may represent a sequence of treatment steps (e.g., starting with 25 mg of an antidepressant daily for 2 months followed by increasing the dose to 50 mg and continuing for 3 months). Formally, an n^{th}-order node, v, denotes a path through $(s_1,...,s_{n-1},s_n)$, where s_i is the i-th step in the path. Although a node can represent a sequence of steps, all other properties can be considered the same as the first-order network. For example, a path from one node (h,i) to the next node j (i.e., steps h–i) is denoted as $(h,i) \rightarrow j$, and its transition

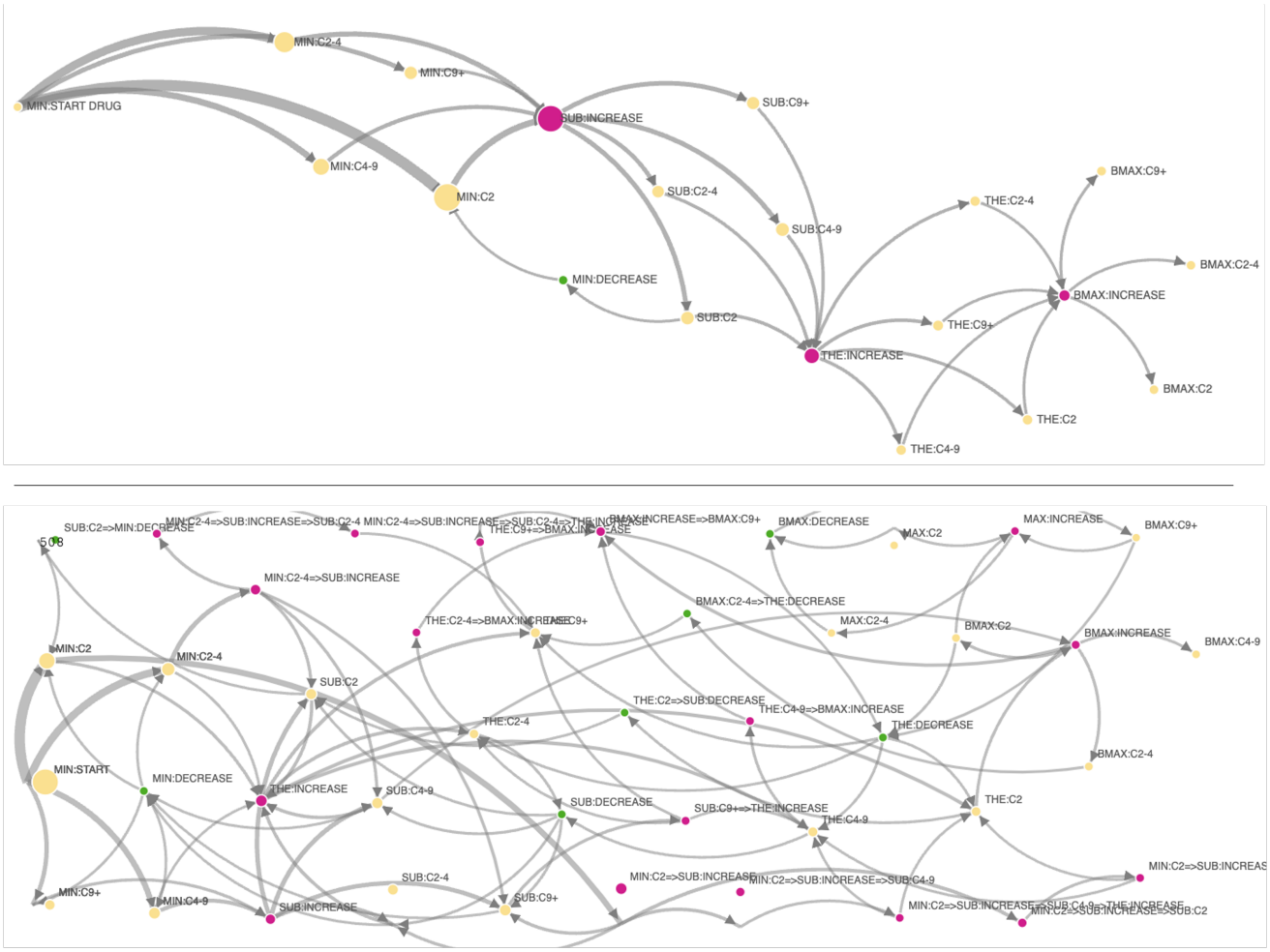

Figure 1. *A CP represented in two different network types: first-order network (top) and higher-order network (bottom). The first-order version denotes a single treatment step as a node and the treatment transitions as edges, and the higher-order version denotes a sequence of treatment steps as a node and the dependencies of the treatment steps as edges.*

probability is

$$P((h,i) \to j) = \frac{W((h,i) \to j)}{\sum_k W((h,i) \to k)}, \tag{1}$$

where $W(i \to j)$ is the sum of connections $i \to j$ found in the data.

Visual Analytics for Clinical Data

Most of previous studies on visualizing clinical pathways and clinical events focused on summarizing large-scale electronic health records (EHRs) as flow-based visualizations to highlight frequent patterns of clinical events [7, 9] in order to aid decisions for future health care plan [5, 8].

Perer *et al.* [9] preprocessed EHR data to extract and map information to hierarchical standard clinical diagnosis codes and detect frequent patterns from the pre-processed data using Sequential PAttern Mining using a bitmap representation (SPAM) [2]. The visualization interfaces based on Sankey Diagrams represent the frequent pathway events with which users explore paths of interest in detail. Guo *et al.* [7] introduced a visual analytics system that aligns clinical paths based on dynamic time wrapping and segments into more detailed stages to help illustrate the progression of disease in the context of a care plan. DecisionFlow [5] analyzes disease progresses and their outcomes in EHR by aggregating patients at each stage of the disease. It was also designed to handle varying sequences of events. DecisionFlow visualizes the aggregated symptoms and their average development time for the patients in color-coded paths using Sankey Diagram visualization.

Higher-Order Network Visualization

To the best of our knowledge, a very few works exist on the visualization of higher-order networks. Processing a rich set of information and complex dependencies in higher-order networks are major obstacle to pattern discovery and interpretation. HoNVis [13] delivered a significant contribution in this area. With a global shipping network as an example, it demonstrated how an interactive exploration of higher-order networks could help a decision process. Multiple coordinated visualizations allowed users to quickly identify patterns of interest, and formation and evolution of higher-order dependencies. HOTVis [10] proposed a dynamic graph visualization algorithm that utilizes higher-order graphical

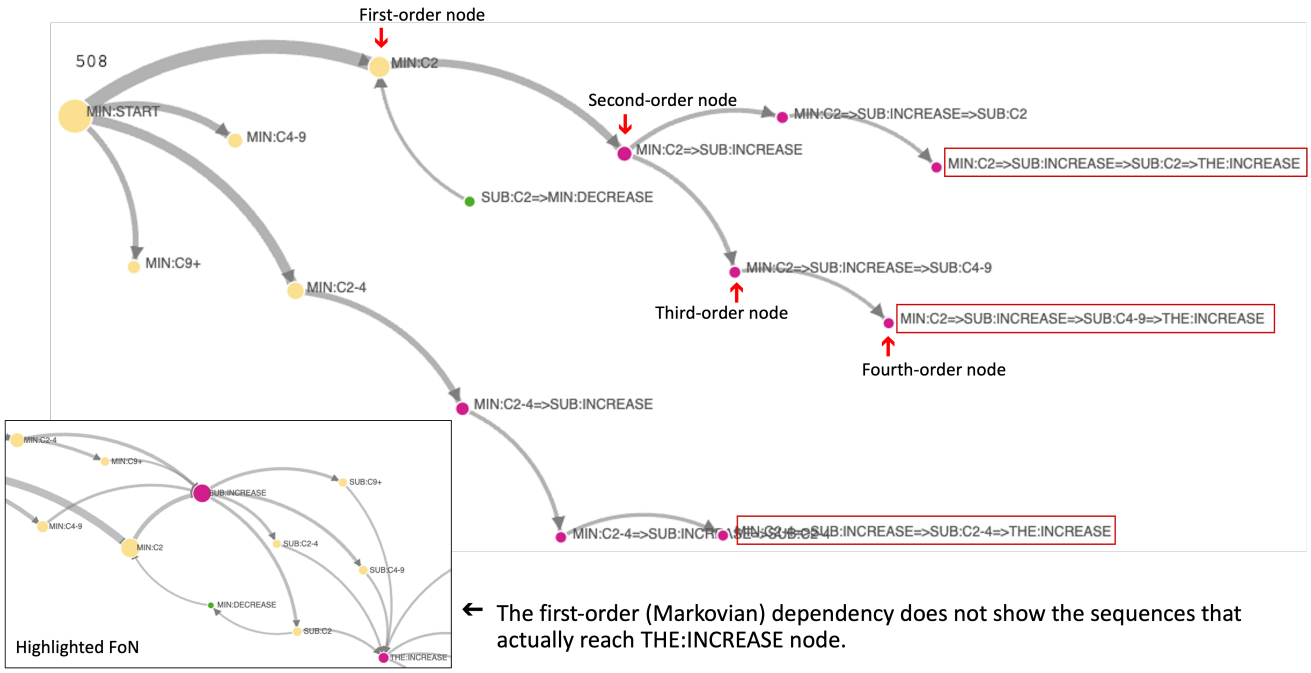

Figure 2. *Exploration of higher-order dependencies (treatment sequences) in a CP using CPViz. The dominant sequences from MIN:C2 and MIN:C2-4 to THE:INCREASE are shown. The graph on the bottom-left, which is part of a first-order network, shows difficulty in tracing the sequences.*

models of causal paths in temporal data based on time slices. The generated time-aware static visualizations of temporal graphs can highlight patterns in the underlying temporal data. CCVis [6] also utilized a higher-order network construction algorithm to extract the critical sequences that lead to different transition probabilities. Their algorithm extracted the critical activity sequences to describe the online learning behavior patterns of students. In addition, coordinated multiple views provide an effective overview of vast amounts of behavioral data as well as detailed comparisons of individual student behaviors.

Design Requirements

The overall goal of exploring and comparing CPs represented as higher-order networks is to clearly visualize complex dependencies and identify similarities between multiple CPs. To achieve the goal, multiple requirements should be achieved. We had many discussions with domain experts, such as clinicians and health services researchers. We identified four design requirements below.

- **R1. Exploring Higher-order Networks:** The visualization system should support effective exploration of higher-order networks. It should reduce the visual clutter issue as complex in structure of a higher-order network. Also, it should support discovering important higher-order dependencies.
- **R2. Identifying Salient Patterns:** The visualization system should support finding salient nodes (treatment or a sequence of treatments) and specific treatment patterns (e.g., aggressive or conservative).
- **R3. Finding (Dis)Similar CPs:** The visualization system should enable users to find (dis)similar CPs. Given a set of CPs, the system should provide the similarities of all CP pairs.
- **R4. Comparing CPs in depth:** Given the similarity of all pairs, the visualization system should be able to perform a deep comparative analysis of the selected pair. Users need to know how they are (dis)similar each other, such as different dependency patterns.

Antidepressant Pharmacotherapy CP Model

Antidepressants are medications prescribed to treat MDD [4]. Our pathway model was designed to represent three aspects of pharmacotherapy sequences: dosage, ramping up/down of dosage, and duration of dosage. Based on this model, we designed a nomenclature for each step in a pathway using intuitive labels. For example, *SUB:INCREASE* means ramp up to subtherapeutic dosage, and *THE:C4-9* means continue the current therapeutic dosage for 4–9 months. To include different antidepressant medications in the same pathway, we converted all medications into Fluoxetine-equivalent doses. Table 1 shows all the labels we use for the study.

Similarity Measure

We define the distance metric between two CPs using their topological similarity and the fraction of patient cases that fall on nodes and edges. Let M be a square matrix, and let m_{ij} be the value of the i^{th} row and the j^{th} column that represents the transition probability of edge (v_i, v_j).

We compute the distance between two pathways using random walks. More specifically, from each pathway, we randomly select first-order nodes using proportions of their occurrences in data as initial probabilities, and then we produce random walks up to ten in length by traversing the network using the transition

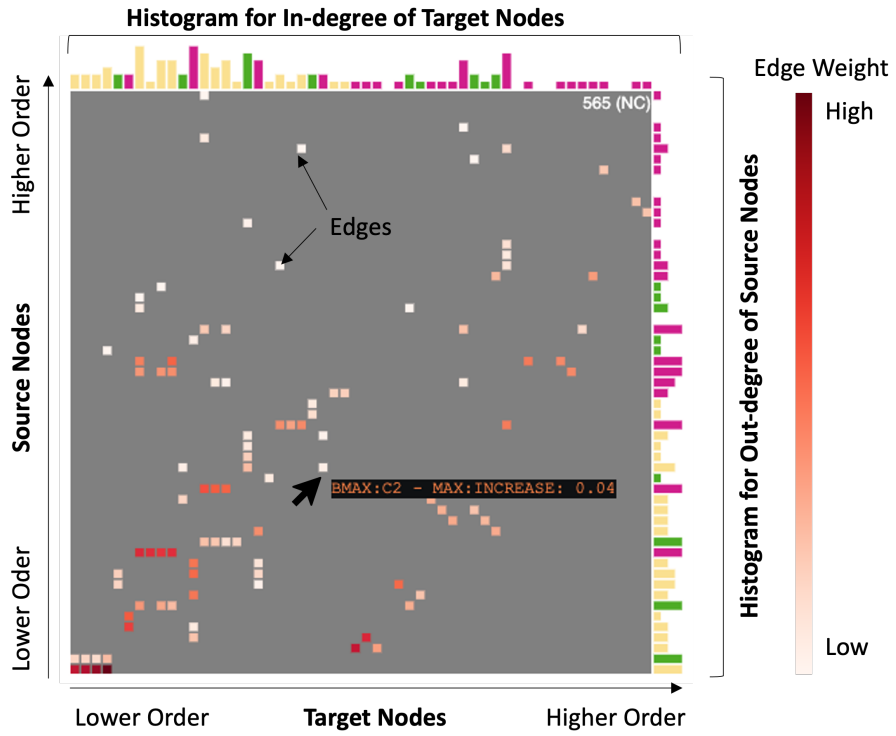

Figure 3. *The matrix view visualizes edges—intersections between rows and columns are displayed as squares in a 2D layout. Also, it shows edge weights and the in/out-degrees of target and source nodes.*

probability matrix, M. Formally, the distance between pathways A and B is defined as

$$D(A,B) = 1 - S(A,B) \tag{2a}$$

$$S(A,B) = \frac{\sum_i^{N_{AB}} A_i B_i}{\sum_i^{N_{AB}} {A_i}^2 \sum_i^{N_{AB}} {B_i}^2} \tag{2b}$$

Here, $D(A,B)$ is the distance between the two pathways, A and B, defined as one minus the similarity of two pathways, $S(A,B)$, where $S(A,B)$ is the cosine value of the angle between the two vectors derived from random walks on pathways A and B. The dimension of each vector is the union of random walks generated from two pathways, and values are the number of times they are generated. So, N_{AB} represents the number of unique random walks sampled at least once in pathways A or B. A_i and B_i represent the number of times that random walk i is sampled from A and B, respectively.

To place all distances among stations in a global context, an embedding space has been created by applying kernel PCA onto a matrix where each row corresponds to pairwise distances to all other stations from the given station. We tested various kernels including linear, poly, RBF, sigmoid, and cosine. We then computed Spearman's correlation between $D(A,B)$ and the euclidean distance within each embedding space, where larger correlation value means better similarity. As the result, the sigmoid kernel was selected to create the embedding space.

Table 1. Labels used to denote treatment steps in a CP

Label	Description
MIN	Minimum dosage
SUB	Subtherapeutic dosage (below 20 mg)
THE	Therapeutic dosage (20–40 mg)
BMAX	Below max dosage (40–60 mg)
MAX	Max dosage (60–80 mg)
INCREASE	Increase dosage
DECREASE	Decrease dosage
C2	Continue for 2 months
C2-4	Continue for 2–4 months
C4-9	Continue for 4–9 months
C9+	Continue for 9+ months

Functional Aspects of CPViz

At a high level, CPViz provides two types of visual analytic functionalities: exploration of a single CP and comparison of multiple CPs. The first is to sift treatment sequences that involve steps of interest (e.g., all treatment sequences starting from step *MIN:C2* to step *MAX:INCREASE*). The second is to understand how multiple CPs can be identified and grouped as similar CPs and study how a given pair of CPs are similar or dissimilar.

Exploration of a Single Pathway

For browsing a given pathway, CPViz places a first-order network version of the pathway over the higher-order version (Fig-

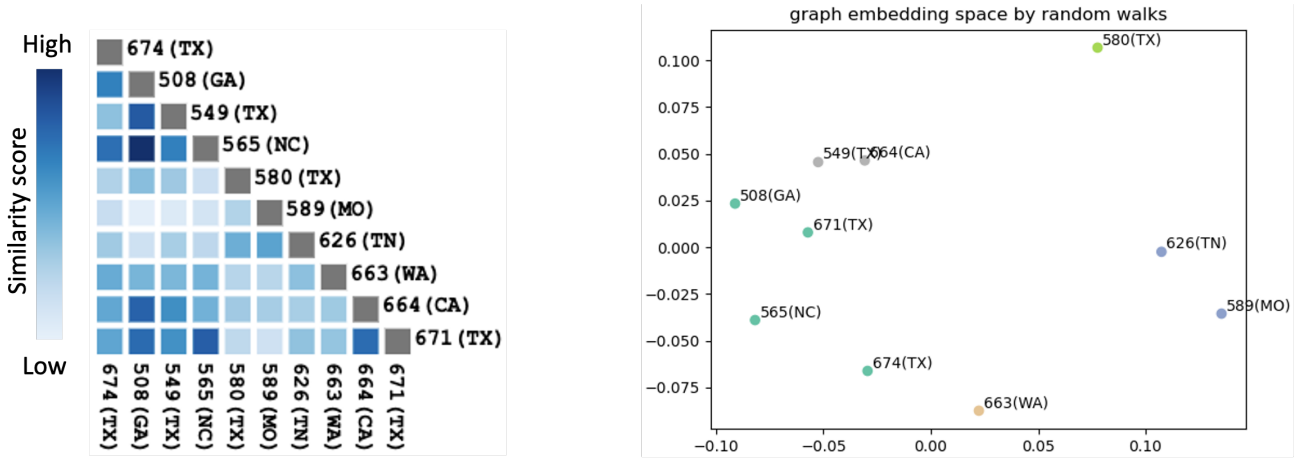

Figure 4. *A heat map that represents similarities between all pairs of stations calculated by random walks (Left). 2D embedding space where all stations are placed preserving their pairwise distances with the others (Right). The map is created using the random walk distances and kernel PCA (Right).*

ure 1). Users can select steps of interest on the first-order network, where steps are nodes, and the complete pharmacotherapy sequences (i.e., paths or sequences of steps) that include the selected steps are shown in the higher-order network. Figure 2 shows an example in which steps *MIN:C2, MIN:C2-4, MIN:C4-9, and MIN:C9+* are selected. We can see the two nodes, *MIN:C2* and *MIN:C2-4*, have fourth-order dependencies ending *THE:INCREASE* and the connected paths and intermediate higher-order nodes are visualized. This decreases the visual clutter issue significantly and enables efficiently extracting the traces from the selected nodes to the highest-order nodes (R1). To aid clinicians interpreting these paths, CPViz uses three different colors to denote the type of the last step in the node. The dark pink represents nodes with an *increase in dosage*, the green represents a *decrease in dosage*, and the yellow represents a *continuation of the same dosage*. The radius of a node and the thickness of an edge are proportional to the number of cases found in the data. We also adopted a force-directed layout [3], which brings together nodes with mutual connections to better track of paths of interest (R1 and R2).

Comparison of Multiple Pathways

Given a number of pathways, CPViz offers another visual analytic capability to compare pathways at two levels. First, it shows similarities of all pairs of pathways by a grid-based heat map, and places all pathways in a 2D embedding space so that clusters of similar pathways can be detected (see Figure 4). Second, when users select a square on the heat map, CPViz provides an in-depth comparison of two selected pathways by visualizing the edges that are both common and unique between them (see Figure 5).

When comparing two pathways, CPViz displays a pathway's edges in a matrix, which we call the *matrix view*. As shown in Figure 3, a square located in the i^{th} row and the j^{th} column is an edge coming out of the i^{th} node from the vertically arranged nodes and going into the j^{th} node from the horizontally arranged nodes. The color of a square represents the edge weight, with the minimum value colored in white and the maximum value in dark red. Additionally, CPViz adds an in-degree histogram (number of incoming edges) on the top and an out-degree histogram (number of outgoing edges) on the right side. This helps identify hub nodes (nodes with a large neighbor) and outliers (R1 and R2). The coloring scheme for histogram bars is the same as in Figure 2. To aid visual analytics, CPViz shows labels as a tooltip when the mouse hovers over a node even though the tooltip is not shown in this paper. For edges, it shows tool tips that contain the source nodes, target nodes, and edge weights.

The matrix view is also used to compare two pathways. Given a set of pathways, CPViz illustrates all pairwise similarities in a heat map, which is shown as a lower-triangular form in Figure 4 (Left) showing a displacement of each pathway that preserves their similarities with all other pathways (R3). The colors of squares indicate similar scores. The dark blue means that the pair is relatively similar to each other, while the white color does a dissimilar pair. The grey squares are pairs of the same HoNs. Also, it shows the similarity metric and projection to a 2D space in Figure 4 (Right). The similarity measurement is explained in Section of Similarity Measure below.

Once a user selects a pair of pathways, three matrix views are displayed: two for each selected pathway and the third for their combined view. The third view displays the union edges from both pathways for comparative analysis in depth (R4), as shown in the rightmost view in Figure 5, in which the red squares represent the edges of the first pathway, and the blue squares represent the edges of the second pathway. The outlined squares indicate the edges that appear in both pathways. The histograms show the in-degree and out-degree values of the combined node set. A histogram bar for a node that appears in the both pathways is split by a black line to indicate the separate portions from each pathway.

Result

For the case study using CPViz, we collected data from OEF/OIF veterans with MDD from ten different VA facilities. We used data between January 1, 2006 and January 1, 2020 from the VA's Corporate Data Warehouse [11]. After processing data based on the labels in Table 1, we constructed pathways as higher-order networks. We conducted two types tasks using CPViz: (1) identification of true pharmacotherapy sequences and (2) visual

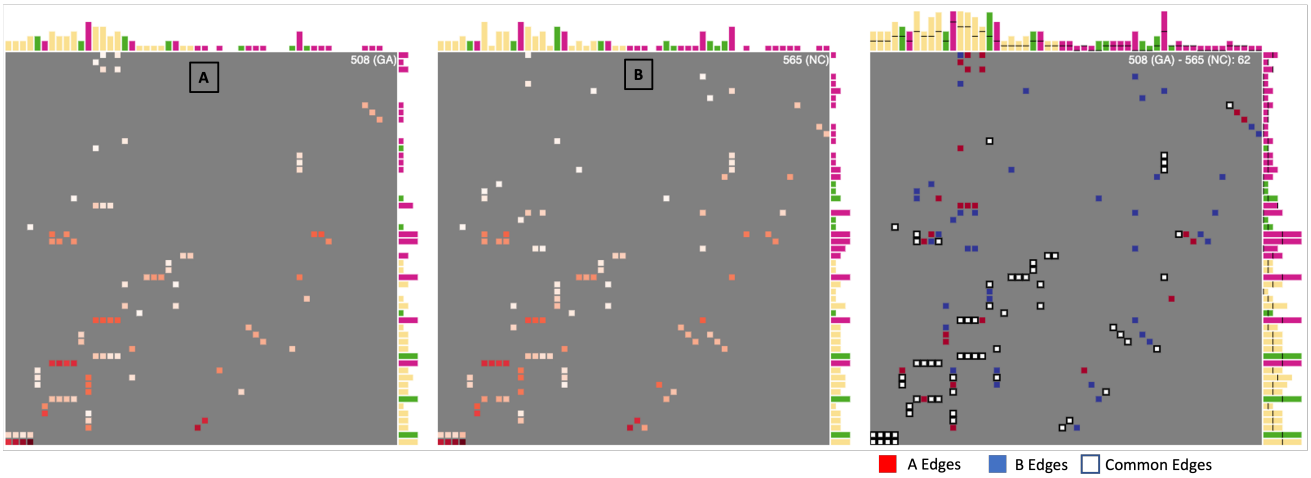

Figure 5. *CPViz for comparative analysis of two CPs. Three matrix views for the selected pair A (left) and B (center), and the combined view of A and B (right).*

comparative analysis of multiple pathways to highlight clinical differences.

Identification of Pharmacotherapy Sequences

Figure 2 illustrates CPViz identifying all pharmacotherapy sequences that start with the minimum dosage. It only shows paths of higher-order nodes that span through four first-order nodes following *MIN:START*: *MIN:C2*, *MIN:C2-4*, *MIN:C4-9*, and *MIN:C9+* from the CP of a facility in Georgia (station ID 508). As illustrated, the three paths that lead to an increase of the antidepressant to a therapeutic dosage (*THE:INCREASE*) are highlighted. A closer investigation shows that no such path exists through *MIN:C9+* nor *MIN:C4-9*. In other words, if a patient stays at the minimum dosage for more than 4 months, it is less likely that the patient will ramp up to the therapeutic dosage. In contrast, a network of Markovian dependency (i.e., first-order) fails to expose this pattern. To highlight this issue, the bottom-left corner of Figure 2, shows a part of the first-order version of the CP in which we can find paths to *THE:INCREASE* from both *MIN:C4-9* and *MIN:C9+*.

Visual Comparative Analysis (Most Similar Pair)

The heat map shown in Figure 4 (Left) shows the all-pairwise similarity scores of the ten pathways. We selected the pair with the highest similarity: station ID 508 of Georgia and station ID 565 of North Carolina. The rightmost matrix view shows few common edges at the upper-right area where the higher-order nodes are placed, but it shows many common edges around the lower-left area where the lower-order nodes are placed. We observed the same pattern in the majority of pairs, which suggests that the difference in patterns is small at the beginning of treatment, but the difference increases as treatment sequences move toward the end of the treatment step.

Visual Comparative Analysis (Least Similar Pair)

Next, we select another pair of pathways that shows the smallest similarity score: station ID 508 of Georgia and station ID 589 of Missouri. Figure 6 illustrates the matrix view that compares the two pathways. As suggested by the low similarity score, CPViz exposes very few common edges between the two pathways. Most of all, CP 589 has noticeably more higher-order nodes and edges than CP 508. A closer examination reveals that the two pathways actually differ (R4). We call out three regions in the matrix view to highlight the differences.

In region (1), the edges of CP 589 (blue) run diagonally upwards, whereas no such edges are found in CP 508 (red). This means that many lower-order nodes progress into higher-order nodes in CP 589. In other words, there are more unique paths that include higher-order ones in CP 589 than in CP 508. Region (2) shows that CP 589 has two target nodes with high in-degree values: *BMAX:INCREASE* and *MAX:INCREASE*, whereas the same nodes in CP 508 have small in-degree values. In region (3), there are some edges (paths) from minimum and subtherapeutic dosage treatments to below max and max dosage treatments. This means that CP 589 has more aggressive treatment sequences (i.e., ramping up to maximum dosage and bypassing therapeutic or subtherapeutic dosages) than what we see in CP 508.

Conclusion

Unlike pathways represented in the Markovian property (first-order network), pathways of higher-order networks portray both partial and complete histories because the nodes provide the actual trajectories of treatment sequences, although they are often too complex to comprehend. CPViz offers interactive visual analytic functionalities for higher-order networks to facilitate the exploration of a single pathway and the comparison of multiple pathways. We demonstrated that CPViz captured some treatment sequences from a higher-order CP, which was infeasible with first-order networks. We also showed that CPViz exposes heterogeneity in the prescription of antidepressants across different pathways by mapping and visualizing dependencies of connections between treatment sequences.

Acknowledgment

The work described here is sponsored by the US Department of Veterans Affairs. This research used resources from the Knowledge Discovery Infrastructure at the Oak Ridge National Laboratory, which is supported by the US Department of Energy's Office

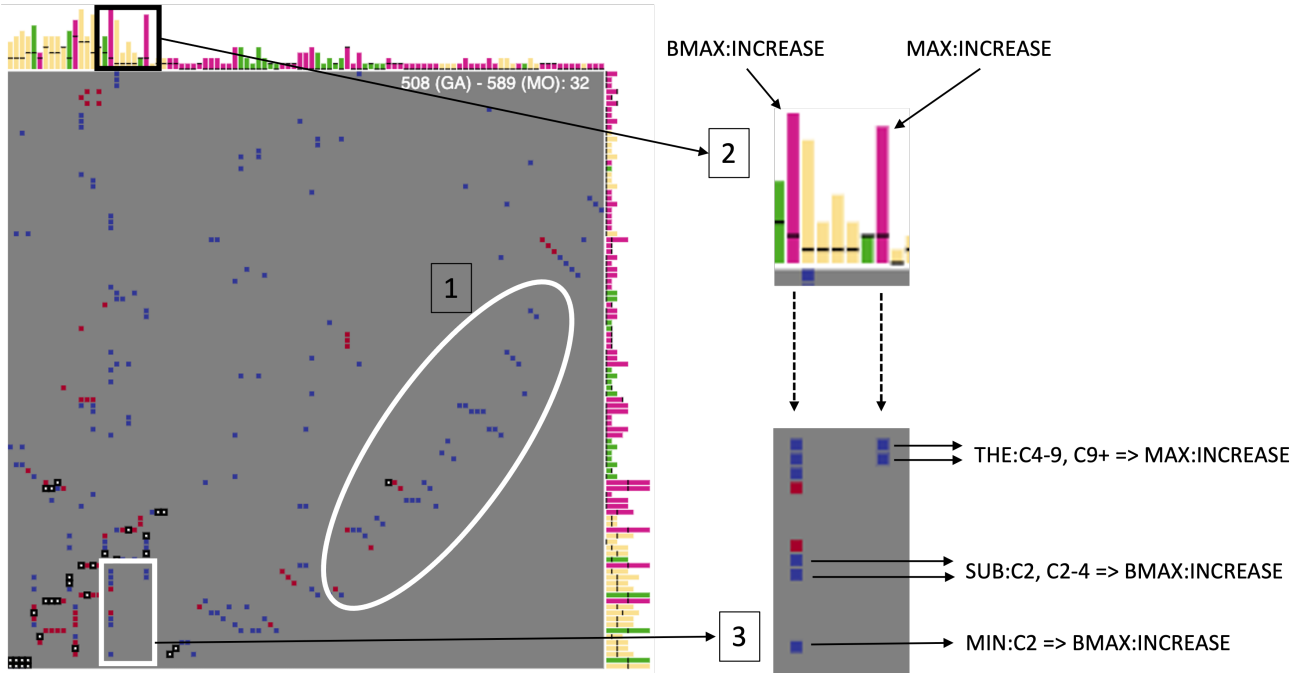

Figure 6. *Matrix view comparing pathways of two VA facilities: station ID 508 from Georgia (red squares) and station ID 589 from Missouri (blue squares). The view shows that CP 589 has more longer treatment sequences and more aggressive paths that increase dosage to the maximum dosage and bypass the therapeutic dosage.*

of Science under contract DE-AC05-00OR22725.

References

[1] Management of Major Depressive Disorder (MDD) (2016) - VA/DoD Clinical Practice Guidelines.

[2] Jay Ayres, Jason Flannick, Johannes Gehrke, and Tomi Yiu. Sequential pattern mining using a bitmap representation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 429–435, 2002.

[3] Thomas M. J. Fruchterman and Edward M. Reingold. Graph drawing by force-directed placement. *Softw. Pract. Exper.*, 21(11):1129–1164, November 1991.

[4] T. Furukawa, A. Cipriani, P. Cowen, S. Leucht, M. Egger, and G. Salanti. Optimal dose of selective serotonin reuptake inhibitors, venlafaxine, and mirtazapine in major depression: a systematic review and dose-response meta-analysis. *The Lancet Psychiatry*, 6(7):601 – 609, 2019.

[5] David Gotz and Harry Stavropoulos. Decisionflow: Visual analytics for high-dimensional temporal event sequence data. *IEEE transactions on visualization and computer graphics*, 20(12):1783–1792, 2014.

[6] Maggie Celeste Goulden, Eric Gronda, Yurou Yang, Zihang Zhang, Jun Tao, Chaoli Wang, Xiaojing Duan, G Alex Ambrose, Kevin Abbott, and Patrick Miller. CCVis: Visual analytics of student online learning behaviors using course clickstream data. *Electronic Imaging*, 2019(1):681–1, 2019.

[7] Shunan Guo, Zhuochen Jin, David Gotz, Fan Du, Hongyuan Zha, and Nan Cao. Visual progression analysis of event sequence data. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):417–426, 2019.

[8] Zhuochen Jin, Shuyuan Cui, Shunan Guo, David Gotz, Jimeng Sun, and Nan Cao. Carepre: An intelligent clinical decision assistance system. *ACM Transactions on Computing for Healthcare*, 1(1):1–20, 2020.

[9] Adam Perer, Fei Wang, and Jianying Hu. Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics*, 56:369–378, 2015.

[10] Vincenzo Perri and Ingo Scholtes. HOTVis: Higher-order time-aware visualisation of dynamic graphs. In David Auber and Pavel Valtr, editors, *Graph Drawing and Network Visualization*, pages 99–114, Cham, 2020. Springer International Publishing.

[11] Lauren Emilie Price, Kimberly Denise Shea, and Sheila M. Gephart. The veterans affairs's corporate data warehouse: Uses and implications for nursing research and practice. *Nursing administration quarterly*, 39 4:311–8, 2015.

[12] Deborah Seys, Luk Bruyneel, Svin Deneckere, Seval Kul, Liz Veken, Ruben van Zelm, Walter Sermeus, Massimiliano Panella, and Kris Vanhaecht. Better organized care via care pathways: A multicenter study. *PLOS ONE*, 12:e0180398, 07 2017.

[13] Jun Tao, Jian Xu, Chaoli Wang, and N. V. Chawla. HoNVis: Visualizing and exploring higher-order networks. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 1–10, Los Alamitos, CA, USA, apr 2017. IEEE Computer Society.

[14] Jian Xu, Thanuka L. Wickramarathne, and Nitesh V. Chawla. Representing higher-order dependencies in networks. *Science Advances*, 2(5):e1600028, 2016.