Developing an Interactive Social Network Analysis Tool to Explore Activity Patterns on an Online Health Knowledge Sharing Platform

Danny T.Y. Wu, PhD, MS\textsuperscript{1,2}, Wu-Chen Su, MS\textsuperscript{1}, Paul Murdock\textsuperscript{1}, Divyay Gupta\textsuperscript{3}, Aniket Mahurkar\textsuperscript{3}, Nameeta Raj\textsuperscript{3}, Supraja Sivaramakrishnan\textsuperscript{3}, Daniel J. McLinden, EdD\textsuperscript{1,2}, Dong-Gil Ko, PhD\textsuperscript{3}

\textsuperscript{1}Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH; \textsuperscript{2}Department of Pediatrics, University of Cincinnati, Cincinnati, OH; \textsuperscript{3}Carl H. Lindner College of Business, Operations, Business Analytics & Information Systems Department, University of Cincinnati, Cincinnati, OH; \textsuperscript{4}Medical Operations Department, Cincinnati Children’s Hospital Medical Center, Cincinnati, OH;

Abstract

ICN Exchange is an online knowledge sharing platform that allows users such as health care staff members, clinicians, patients, and their families to identify like-minded people and share quality improvement tools and educational information. While this platform provides a unique opportunity to explore patterns of online knowledge sharing activities in the healthcare industry, conducting exploratory data analysis can be cumbersome and ineffective due to the complexity of the data. Therefore, we developed an interactive social network analysis tool and preliminarily evaluated the usability of this tool through the System Usability Scale. The results confirm the high usability of the tool, and indicate a few areas of improvement. We will continue to refine this interactive tool to ensure its ability to support exploratory data analysis and hypothesis generation, and to collaborate with researchers in different domains to expand the scope of the tool as a general exploratory data analyzer for document-based knowledge sharing activities.

Introduction

The ImproveCareNow (ICN) is a sustainable collaborative chronic care network aiming to improve the outcome of pediatric patients with inflammatory bowel disease (IBD), such as Crohn disease and ulcerative colitis\textsuperscript{1,2}. ICN has evolved since 2007 from a pilot improvement collaboration of eight centers to a national wide network with over 100 centers primarily in the United States but now includes care centers in three other countries\textsuperscript{3}. One particular challenge in this large geographically dispersed learning health network is to ensure that knowledge (e.g., best improvement methods, tools, ideas) is accessible and available to an appropriate individual when needed. To facilitate such knowledge sharing, ICN Exchange (www.icnexchange.org) was developed by a development team at Cincinnati Children’s Hospital Medical Center (CCHMC, an ICN network care center) in collaboration with leaders and members of the ICN network, and then launched in April 2013 to provide a comprehensive and scalable knowledge sharing solution. Based on a “Pinterest” type of framework, health care staff members, clinicians, patients, and their parents can benefit from identifying like-minded people and from sharing quality improvement tools, care management procedures, and other useful information.

With the steady growth over the past few years, ICN Exchange (aka the “Exchange”) has accumulated thousands of knowledge sharing activities that includes adding new content (i.e., pinning content), searching for existing content, downloading, curating existing content by putting them onto personal boards (repining), and commenting on content. These activities create digital traces in the online community and provide a unique opportunity to explore patterns of online knowledge sharing activities in a healthcare network. These behavior data also have great potential to further contribute to knowledge sharing and management literature. We are particularly interested in three types of knowledge sharing activities in the Exchange, including 1) Repin, 2) Comment, and 3) Download. These activities and the associated members of the Exchange are modeled in a social network to identify “central” persons and groups using standard network centrality measures\textsuperscript{4}. Due to the complexity and high dimensionality of the data, e.g., users can be grouped by their roles or the center they belong to, an advanced interactive social network analysis (SNA) tool is needed to effectively support exploratory data analysis and hypothesis generation. In addition, this tool should have the ability to perform text analysis and mining on the documents to provide more granular views of the knowledge sharing activities.

Unfortunately, to the best of our knowledge, an appropriate SNA tool does not exist. Some tools adopt a general design\textsuperscript{5,6}, others are designed for specific purposes\textsuperscript{7}, rendering the use of these tools suboptimal and ineffective. Also, most tools are limited in their ability to interact with datasets and to combine network analysis with text
mining techniques. The closest tool we found was Netlytic⁸, which unfortunately requires multiple datasets to be generated from our high dimensional data for network visualization. Therefore, in this study, we developed our own SNA tool with a focus on interactivity, interface usability, and the combination of network and text analysis. We then tested the usability of the tool⁹ to inform the next cycle of development. The ultimate goal of this project is to provide assistance for exploring knowledge sharing activities in healthcare and further contribute to knowledge sharing and management theories and models.

This manuscript is organized as follows. First, we explain the data extraction, system development, implementation, and evaluation methods. Second, we demonstrate the design of the prototype system and the evaluation results. We then discuss the benefits of our interactive SNA tool for supporting exploratory data analysis and hypothesis generation. Finally, we discuss our limitations as well as highlight future opportunities for expanding the tool to other domains as a general exploratory data analyzer for evaluating document-based knowledge sharing activities.

Methods

Data Extraction

Records of three knowledge sharing activities (Repin, Comment, and Download) were directly extracted from the MySQL database of the Exchange (Figure 1), facilitated by one of the senior authors (DJM) due to his leadership and connections in the Exchange. As of December 2016, the Exchange had more than 1,200 users classified into 12 user groups: Physician, Improvement/Research Coordinator, Parent, Patient, etc. Additional information depicting the volume of data by each knowledge sharing activity type among the active users (those who have had at least one such activity) is provided in Table 1.

![Figure 1. ICN Exchange homepage (www.icnexchange.org)](image)

<table>
<thead>
<tr>
<th>Activity Type</th>
<th>Date Range</th>
<th>Active Users</th>
<th># Activities</th>
<th>Avg. per user</th>
<th>Max. per user</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repin</td>
<td>04/2013~12/2016</td>
<td>240</td>
<td>2,564</td>
<td>10.68</td>
<td>35</td>
</tr>
<tr>
<td>Comment</td>
<td>04/2013~12/2016</td>
<td>335</td>
<td>1,991</td>
<td>5.94</td>
<td>118</td>
</tr>
<tr>
<td>Download</td>
<td>07/2015~12/2016</td>
<td>556</td>
<td>115,600</td>
<td>207.91</td>
<td>630</td>
</tr>
</tbody>
</table>

System Development

The prototype system was developed following a systematic approach. First, one of the senior authors (DGK) formed a student-based task force (DG, AM, NR, and SS) and made a significant effort to ensure data quality and understand business logic. The first author (DTW) participated in this task force and provided suggestions and feedback regarding data analyses and modeling strategies. The task force developed a set of computer scripts to analyze the data in an
exploratory manner. The first author observed the conduction of the exploratory data analysis and recorded key variables of interest. The task force generated a final report to summarize the findings derived from the exploratory data analysis. One challenge identified in the observations was the ability to efficiently and effectively explore the complex and multi-dimensional data. The first two authors (DTW and WCS) then independently designed the prototype SNA tool based on the observations and the task force’s final report. It is worth noting that the prototype tool utilizes network centrality measures, including Degree, Betweenness, and Closeness centrality⁴, to facilitate the visual analytics of network structures and interactions.

**Implementation**

The prototype system was implemented using R-studio and Shiny. The latter is an open source R package that provides an elegant and powerful web framework for building web applications using R language. The activity records were organized and stored in a MySQL database. The network visualization and its interactive features was made through existing R libraries (network and network3D) and custom jQuery scripts. Text in the documents were extracted and analyzed using Python NLTK library and clustered by K-means with Cosine Similarity using Python scikit-learn library to form document categories. The prototype system was hosted in a development server maintained by the Research IT group at the CCHMC and can only be accessed within the CCHMC and the University of Cincinnati (UC) computer network.

**Pilot Evaluation**

The prototype tool was preliminarily evaluated through the System Usability Scale (SUS)¹⁰. While the SUS questions remained unchanged, textual feedback was requested for each question along with overall feedback. Five participants were recruited, including four master students from the task force and an undergraduate research assistant who helped with the literature review of the project. A short tutorial was given to the participants to introduce the data and the tool functionality. The participants then individually scored the tool in an Excel spreadsheet without seeing the final adjusted SUS score being calculated in the background (hidden columns).

**Results**

**Interface Design**

Figure 3 shows the user interface design of our prototype SNA tool. The design can be organized into three components: 1) Data filter, 2) Social network relationship explorer, and 3) Data searcher. In the data filter, SNA tool users can focus on certain social interactions by applying multiple criteria such as date range and frequency. For example, evaluating social connections of its members based on their download behaviors could involve setting a filter on members who downloaded more than ten times in the year 2016. SNA tool users can also merge knowledge share activities based on user groups or their health institute/center (Subject dropdown). They can also highlight subjects (nodes) based on their “centrality” in the network. Since the knowledge sharing activities are directional, e.g. one user downloads a file “from” another, the centrality of a node is dependent on its in- or out-degree (or both). The data filters were chosen based on the variables frequently used in the previous observations.

The relationship explorer has three tabs. Each one represents a knowledge sharing activity and a directed network constructed based on the selected options in the data filter. The network graph has two layouts: dynamic and static. Figure 3 shows the dynamic layout where SNA tool users can zoom in a network graph and drag each node. The nodes are grouped and color-coded based on their level of in- or out-degree. In Figure 3, one can easily identify the largest node in the graph: the light blue node. This node has the most interactions with other nodes, with the legend indicating 145 edges. In contrast, the static graph does not offer the interactive features but instead provides a simplified view that may increase the interpretability of a network graph. An example of a static graph is shown in Figure 4.

The data searcher provides detailed information about the selected nodes and their neighbors. For example, if the light blue node in Figure 3 is clicked on, the data searcher will automatically list all its neighbor nodes (child-uid), or all nodes that have a relationship with the light blue node based on the specified data filter criteria.
Figure 3. Interactive SNA tool with a dynamic layout

Figure 4. An example of the static layout

Evaluation

Five participants assessed the prototype SNA tool and returned the SUS form. The average SUS score is 79.5 (70, 75, 80, 82.5, and 90), indicating that the usability of the tool is considered as “above average”. By summarizing the scores of each question in the SUS form, the participants agreed that the tool overall is not cumbersome or inconsistent. Its functions are well integrated and laypeople could quickly learn to use this tool. Participants also agreed that the SNA tool is easy to use and would use the system frequently.
While the participants gave positive responses to the SNA tool and felt confident using it, they indicated this is largely based upon their understanding of the underlying data structure and business logic. Users without such an understanding may get lost in the beginning and not be able to interpret the network visualization correctly. For example, one participant stated that “the end user must be aware of the underlying data so that s/he does not misinterpret the visualization”. Another participant also expressed that he gained much understanding of the tool in the short tutorial that the first author instructed. One area of improvement lies within the min and max values of the frequency sliding bar. This sliding bar allows users to set thresholds to the frequency of the activities. However, the min and max values were not set based on the actual values in the dataset but instead on a set of arbitrary numbers that may include likely chosen thresholds from the designers’ perspective. This gap likely introduces inconsistency between the tool and the underlying data, and may hinder users’ understanding of the tool.

In addition, participants suggested several changes to improve the usability of the tool. First, the tool can provide definitions for the technical terms, e.g. centrality measures and meaning of the arrows. Second, the legend in the dynamic layout did not make sense. Neither the labels nor the color coding were easily understandable to the participants. Third, the participants seemed to have different browsing preferences. Some preferred the dynamic layout while others preferred the static layout. The tool should continue providing different ways to visualize a network and allow users to switch. Finally, some participants expressed concerns about scalability, although they thought the prototype tool responded well to user queries in a reasonable time frame during the evaluation. A loading message might be needed to provide immediate feedback to the SNA tool users in a larger dataset.

Discussion

ICN Exchange provided a unique opportunity to understand online knowledge sharing behaviors in the health domain, especially among healthcare staff members, clinicians, patients, and their parents. We focused on three key knowledge sharing activities: Repin, Comment, and Download. We modeled each set of activities as a directed social network. In order to effectively and efficiently explore meaningful patterns in these social networks, we developed an interactive analysis tool and preliminarily evaluated the usability of this tool through the System Usability Scale with textual feedback. This pilot evaluation shows that our tool is intuitive and easy to learn. Meanwhile, the participants highlighted the need to facilitate the understanding of the tool and its underlying data to fully unleash the power of the tool and to provide users more control with the visualization settings, such as the legend and the layout of the activity network visualization.

It is important to acknowledge that the plot study outcome described in this study was obtained in a preliminary design phase and it confirms the potential of our tool to support future analysis activities. We are still refining this tool in an iterative and incremental manner based on the evaluation feedback and will invite more domain experts to the second round of user study. There, we will focus on the ability of the tool to support exploratory data analysis and hypothesis generation.

We would like to emphasize that this SNA tool has potential to be used in a variety of domains where organizations conduct document-based knowledge sharing activities. It is particularly useful in a context where 80-20 rule applies in that the knowledge selected for the system may be the most common 20% likely to be of value to 80% of the users13. Firms who strategically package a small proportion of their knowledge to benefit the majority of users could avoid incurring high costs associated with collecting and packaging knowledge14.

This study has at least two limitations. First, the pilot evaluation has only five participants. Although this small number may have higher benefit-cost ratio as suggested by the Neilson Norman Group12, it constrains the reliability and the generalizability of the evaluation findings. Second, the participants are only proxy users of the prototype tool. While their feedback is still valuable due to their understanding of the project and the data, their feedback may not touch the pain points of the actual users (health care staff and researchers). Despite these limitations, we believe this pilot evaluation is a critical first step to the development of an interactive SNA tool focusing on document-based knowledge sharing activities in the healthcare industry. We also plan to collaborate with researchers in other domains to seek use cases to expand the scope of the tool.

Conclusion

We successfully developed a highly usable, interactive social network analysis tool that can be used to support future studies of online health knowledge sharing activities. We will continue to refine this tool to be an exploratory data analyzer combining network analysis and text mining techniques with an emphasis on usability.
References