HHS Public Access
Author manuscript
IEEE Trans Vis Comput Graph. Author manuscript; available in PMC 2016 January 01.

Published in final edited form as:
IEEE Trans Vis Comput Graph. 2016 January ; 22(1): 71–80. doi:10.1109/TVCG.2015.2467555.

VisOHC: Designing Visual Analytics for Online Health Communities

Bum Chul Kwon,
University of Konstanz

Sung-Hee Kim,
University of British Columbia

Sukwon Lee,
Purdue University

Jaegul Choo,
Korea University

Jina Huh, and
Michigan State University

Ji Soo Yi
Purdue University

Bum Chul Kwon: bumchul.kwon@uni-konstanz.de; Sung-Hee Kim: kim731@cs.ubc.ca; Sukwon Lee: sukwon@purdue.edu; Jaegul Choo: jchoo@korea.ac.kr; Jina Huh: jinahuh@msu.edu; Ji Soo Yi: yij@purdue.edu

Abstract

Through online health communities (OHCs), patients and caregivers exchange their illness experiences and strategies for overcoming the illness, and provide emotional support. To facilitate healthy and lively conversations in these communities, their members should be continuously monitored and nurtured by OHC administrators. The main challenge of OHC administrators' tasks lies in understanding the diverse dimensions of conversation threads that lead to productive discussions in their communities. In this paper, we present a design study in which three domain expert groups participated, an OHC researcher and two OHC administrators of online health communities, which was conducted to find with a visual analytic solution. Through our design study, we characterized the domain goals of OHC administrators and derived tasks to achieve these goals. As a result of this study, we propose a system called VisOHC, which visualizes individual OHC conversation threads as collapsed boxes–a visual metaphor of conversation threads. In addition, we augmented the posters' reply authorship network with marks and/or beams to show conversation dynamics within threads. We also developed unique measures tailored to the characteristics of OHCs, which can be encoded for thread visualizations at the users' requests. Our observation of the two administrators while using VisOHC showed that it supports their tasks and reveals interesting insights into online health communities. Finally, we share our methodological lessons on probing visual designs together with domain experts by allowing them to freely encode measurements into visual variables.

Index Terms

Online health communities; visual analytics; conversation analysis; thread visualization; healthcare; design study

1 Introduction

Online health communities (OHCs) are becoming a central hub where patients and caregivers can learn and share health-related information and also gain emotional support during their health management process. An increasing number of people are joining OHCs [43] and health care organizations promote them as part of their patient-support services [30]. Studies showed OHCs play an important role in providing patients and caregivers with social support and improved health outcomes [9, 32].

OHC administrators, however, are charged with the challenging task of maintaining sustainable and successful OHCs for their members. Administrators support their members in actively participating in the community, maintaining lasting relationships, acquiring helpful information, and receiving as much support as possible from the rest of the community. At the same time, they need to prevent the propagation of any misinformation or trolling messages by monitoring new trends and heated topics. Thus, OHC administrators can benefit from understanding the quality of individual postings–*diversity, dynamism, and depth of conversation*–to better serve their communities. However, their current practices of monitoring conversations are limited, since they have to rely on simple quantitative measures, such as the number of postings and the number of new users, expressed in bar or pie charts. To understand the qualitative aspects of the threads, OHC administrators read through individual messages every day to grasp the main themes of the conversations.

Visual analytics can help OHC administrators automatically analyze diverse measurements by allowing them to explore the conversations visually. Prior studies investigated the use of visual analytic tools to explore OHC user postings in terms of temporal trends and sentiments [7] and relationships between members and postings through network analysis [16]. Studies on online communities also provided insights that facilitate an understanding of OHC conversations, through means such as topic analyses of discussion threads [25] and authorship networks on top of topics [17]. Although these approaches are inspiring, their design is not intended for OHC administrators, who face adverse events, such as suicidal and emotional messages, as well as patients requiring urgent help. Further work is needed to characterize the domain problems of OHC administrators, namely, an investigation of *how diversity, dynamism, and the depth of conversations in OHCs can be visually explored*, and to address their unmet needs, building on previous approaches for online communities.

In this study, we aimed to investigate visual analytic solutions for OHC administrators by conducting a design study. Our design study, which included interviews with two administrators of two different OHCs, elicited the goals and tasks of the OHC administrators. During the study process, we designed and developed a visual analytics application, called “VisOHC,” using a real data set of an OHC.

The contributions of this paper are as follows:

- We characterize the goals of OHC administrators, newly discovered through this design study in which two such administrators participated, and derive visualization tasks to achieve the goals;
- We develop unique measures and entities specific to OHCs and interactive visual designs to support the administrators' tasks;
- We share methodological lessons for a design study on how to allow domain experts to explore new measures and visual designs in a dashboard-style combinatorial prototype.

2 Related Work

In this section, we review literature that motivates our study and inspire our design choices.

2.1 Online Health Communities

Patients benefit from reliable health information and peer patients' social support [12]. Social sites on the Web, such as OHCs, provide opportunities to easily deliver such resources to patients. Health-related Web sites, such as webmd.com, not only provide health information, but also enable patients to connect with others experiencing similar challenges. Such peer patient interaction generates the unique resources often missing in conventional clinical settings, fulfilling the informational and emotional needs of the patients [2, 5]. Accordingly, hospitals and health care practices increasingly adopt social media Web technologies as a means to deliver high-quality patient care [2]. Examples include Johns Hopkins Hospital and the Cleveland Clinic, which incorporated patients' discussion forums as part of their patient-support services [30]. Evidence showed integrating such social media Web technologies with health care services empowers patients and facilitates the tailoring of health care practices to patient needs [12, 27].

As much as OHCs show promising advantages for patient care and support, challenges remain in terms of making them successful and sustainable. Only a small fraction of OHC users actively participate and maintain long-term use [29, 34]. Newcomers to OHCs find it difficult to actively participate [8]. Accordingly, researchers studied the factors that influence long-term participation. Receiving emotional support was found to have a high association with long-term participation in an online cancer community [39]. In addition, age, duration of disease, and interaction with peers and clinicians were significant factors in determining the degree of OHC participation [1]. As such, the responses that members receive, together with their innate characteristics and situational context, can influence their active participation in OHCs. Active participation in an OHC then provides more opportunities for patients to reach better health outcomes.

To allow active, sustainable participation and to maximize the benefit of OHCs to patients, moderation is critical. Studies on online communities show the importance of moderation in facilitating conversations, in that it prevents any negative messages from spreading and provides resources when members need them [10]. Administrators' roles in OHCs include those of administrators of regular online communities, such as facilitating conversations and preventing trolling. However, handling the sensitive domain of health still presents a unique challenge. OHC administrators' need to monitor adverse events, e.g., suicidal posts, understand the heavily discussed themes to provide any resources that might better support the current needs of the members, prevent misinformation from spreading, and ensure that members are not isolated. In an effort to facilitate this goal, Huh et al. [20] attempted to directly predict whether a particular thread in OHCs needs an administrator's help by formulating it as a binary classification problem. An OHC administrator's response to a patient's posts is highly associated with subsequent peer patient interaction [18]. Even when their administrators fulfill these important roles, there is no easy solution for effectively managing OHCs.

Visual analytics tools could be a promising solution for addressing different measures that would facilitate OHC administrators' understanding of user participation and thus help them maintain their communities. It has been demonstrated that visual analytic tools [41] can potentially help users easily monitor and understand large-scale data. Nevertheless, only a few studies have investigated the use of visual analytics to explore OHC data. For example, Chee et al. [7] suggested potential visualization themes, e.g., clustering, history, and sentiment. However, their work was limited to providing suggestions and did not evaluate potential target user groups. More recently, Heer and Perer [16] introduced the Orion system and presented a demonstration using large scale OHC data, e.g., members and postings of MedHelp.org. Orion is a useful tool for analyzing relationships between members and postings in a huge amount of OHC data. However, it was not designed for OHC administrators, whose tasks, such as searching for undiscovered dimensions related to creating successful OHCs, are unique, as described in Section 3.4.

2.2 Visual Analytics Approaches for Online Communities

Since there are few visual analytics studies on OHC administrators, we expanded our literature review to visual analytics for general online communities and social media in order to find inspiring examples.

In social media analysis, topic modeling and visualization can help users understand the main themes of textual data without reading all of them, e.g., [14, 25]. To present extracted topics, various text visualization techniques, such as Word Cloud and its variants, e.g., [37, 24], have been proposed. In addition to topic extraction, some researchers have attempted to encode temporal information. Cui et al. [11] employed a trend chart to encode the changes in document sentiment over time. Xu et al. [40] emphasized competition between different topics with inter-crossing multiple trend lines. These approaches can be effective for visualizing the overall trend of the whole OHC.

In contrast, some other researchers focused on a subset of the whole content. For example, ConVis [17] visualized a single thread of conversation, instead of the whole online community. By doing this, a user can follow how conversation evolves among participating members, which could provide a much richer context as compared to showing only some key phrases. In another study by Kerr et al. [22], email threads were visualized using Thread Arcs. In this visualization, each email was represented as a dot, and a relationship between two pieces of email (“reply to”) was represented as an arc connecting two dots. This compact visualization was able to visualize an email thread in a single line. Angus et al. [4] narrowed the scope of visualization to a conversation between a doctor and a patient; however, their study did not consider conversation patterns among multiple participants.

In addition, some previous studies focused on persons and their social network, rather than contents. Kochtchi et al. [23] extracted names from a newspaper article and visualized a network among the names. Each edge of the network had a textual description of the relationship, e.g., “financial adviser of,” to help the user understand the nature of the relationships. In addition, in PEARL, developed by Zhao et al. [42], the emotional style was extracted from an individual's postings on social media, e.g., Twitter, and presented over time. These approaches are relevant because extracting profile information could help administrators understand OHC members' motivation and situation.

In addition, we were also inspired by other types of relevant visualization systems. Interestingly, series of geometric figures, e.g., box, triangle, and circle, have been used to visualize occasional events. For example, boxes and triangles were used to represent each event instance in LifeLines, a visual analysis tool for electronic patient records [33, 38]. In addition, icons were also used in Jinsight [13] and Poem Viewer [3] to visualize instances of a class in Java programming and phonetic units in a poem, respectively. These approaches can help OHC administrators see the dynamics of conversations.

In our literature review, we found copious studies that provide insightful examples of analyzing online communities and social media messages using diverse sets of techniques. However, a gap remains between the tasks and current approaches: i) no studies have addressed the domain goals and tasks of OHC administrators; ii) the techniques seem to focus on analyzing trends from a large quantity of text, which is in conflict with our domain task, focusing on individual threads and capturing diversity in them. In this respect, significant additional effort to understand OHC administrators is necessary in order to design customized visualization for supporting their tasks and requirements.

3 Gathering Goals and Tasks

In this section, we describe the procedure we used for extracting the tasks and requirements faced by OHC administrators and present the findings that were reflected in our design process.

3.1 Working with Administrators

This project was initiated as a collaboration of visualization researchers and a co-author of this paper, who is studying OHCs. The objective of our collaboration was to understand the use of OHCs and enhance their usage to fulfill their purpose, i.e., to help patients. A previous study [19] was conducted with the participation of 14 subjects to understand OHC users' different behaviors in information searching and participating in OHC activities. In this prior study, the motivation for this research emerged, which is to help OHC administrators manage members' diverging participation and the information gathering activities that occur in their communities. We recruited two OHC administrators, who managed their respective OHCs, as participants (hereafter, P1 and P2) for this study.

In this design study, we created a working prototype, an earlier version of our final prototype, based on the findings of the previous study. To show the diverse measurements of discussion threads, we created a "combinatorial prototype," a dashboard-style prototype, which enables users to flexibly map measurements to visual variables, such as width, colors, and vertical order, of boxes representing individual threads. The design details are explained in Section 4. In the interview with the participants, we clarified the domain goals and discovered new tasks using the prototype. Then, we advanced the prototype based on feedback from the participants. Since the final design subsumes the earlier one, we describe the final design with the design rationale, as well as the participants' validation with their quotes, in following sections.

3.2 OHC Discussion Thread Data

We first clarify our OHC data set structure. The following data structure was common across three OHCs with which the two teams of participants were familiar. In our OHC data set, OHC discussion threads were collections of questions and replies posted by registered members. A **thread** consists of a **question** entered by community members and its associated **replies**. A question and a reply share common quantitative attributes: content length, posting date and time, and participation of a questioner and replier. A thread provides another set of quantitative attributes, including the total time duration of conversations (time between first and last posting), authorship of questions and replies, and community member identification.

The data set we used for development was retrieved from the WebMD diabetes community. We chose the diabetes community because diabetes is a chronic disease that needs lifetime care, and patients suffering from diabetes engage in various types of conversation, including emotional and informative discussions. In addition, administrators indicated that when fulfilling their tasks they focused on each community separately, and diabetes communities were the most lively and biggest in terms of the number of daily conversations and users. For the development of our prototype, we collected 2124 posts (117 Questions and 2007 Replies) posted from December 23, 2008 to January 21, 2014. Each thread comprised 17.15 replies on average (min: 0; max: 91). Each question consisted of 929.92 characters (min: 43; max: 3925), and each reply consisted of 682.30 characters (min: 1; max: 4332) on average. We used this data set for our study.

3.3 Domain Goals and Problems

The main goal of the participants was to maintain a successful and lively community that fulfills the goals of its members. P2 further described the success of OHCs as "*maintaining patients' long-term loyalty [in the community]*.", which reflected the participants' shared understanding of the factors that lead to successful and lively online health communities. OHC members' long-term commitment to the community is translated as finding value in their communities, as P1 noted, "*We want to make sure problems are being solved [in OHC]*."

However, our participants did not have an established notion as to which discussion threads add value to their members' participation and provide a lively atmosphere in their communities. An initial step for the participant was to build a framework for evaluating the success of communities. P2 revealed, "*We have done our best to try to understand communities from the quantitative side. But honestly the qualitative side is really challenging.*" They explained that their analyses were generally based on the "eyeballing" approach–reading posts until they grasp the trends. The participants revealed the urge to find common factors that define productive discussions; P2 added "*[explaining examples of successful communities] those comments show a sort of community rapport being built throughout the discussion*." Therefore, the OHC administrators' goals were to understand conversations in OHCs qualitatively as well as quantitatively and to derive hypotheses for explaining what constitutes a successful OHC.

In our interviews, the participants revealed problems related to their ability to look at OHC members from diverse perspectives. The participants had a few working hypotheses about the existence of dimensions in their community discussions, for example, statistical dimensions, such as the number of replies, posting dates, and the number of members who replied, and health-related dimensions, such as age, medication, and food : "*There are diverse sets or types of discussion because people come from all these different backgrounds. So that [aspect] leads us to how we are thinking about quantifying these types of discussions*," P2 said. More importantly, they attempted to find hidden relationships between the measures, as P1 mentioned: "*Our experiences lead us to thinking about 'Does a discussion with a large number of responses [make] a better discussion or does a discussion with no response have something to do with its content?' This [OHC analysis] is not a single dimension analysis.*" The participants then tried to discover a set of unique values of measures that produce "high-value conversations" that impart additional useful knowledge to the community, which required that the participants verify the content. Finally, they wanted to confirm their hypotheses by finding similar conversations. It is difficult to assess high-level similarities between threads without automated support. As P2 mentioned, "*It is hard for us to assess similarities without aggressively tagging posts*". These challenges opened up opportunities for introducing our visual analytics applications, supporting the following domain purposes.

- Explore conversation and content dimensions of conversations.
- Find the correlation between the dimensions of conversations.
- Verify findings by reading the content.
- Find similar/dissimilar cases in other conversations.

3.4 Task Analysis

Our interviews with the participants revealed a set of tasks that need to be supported to enable them to reach their domain goals. In particular, during the interviews a preliminary version of our working prototype, which visualized all the questions and replies as boxes, was used.

In the interview process, the participants discussed new observations and revealed insights about OHC data, which helped us find a new dimension of conversations, e.g., conversation pattern analysis, for their analysis. We characterized three types of tasks that the participants wanted to perform to understand conversations in OHCs.

The four goals indicate that identifying the emerging patterns of individual threads is a central unit of analysis, with which participants start and finish their analysis. This provided an important requirement for/constraint on our design decisions. In contrast to the approaches that use thematic views to summarize the entire community, e.g., that suggested in [7], our design maintained the unity of each thread to contain the unique patterns of individual threads. Such patterns should be defined by diverse indicators that reveal the thread characteristics so that the administrators can estimate similarities between threads.

T1. Discover and compare characteristics of individual threads—We found that both the quantitative and qualitative characteristics of the discussion threads need to be visualized. The quantitative aspects of threads are derived from the descriptive statistics of the thread and the relationships between the questioner and the repliers in a thread. Other qualitative aspects are extracted from the content of each post.

T1.1. Understand conversation patterns: Conversation patterns refer to the trend of dynamics in conversation that emerge from showing the order of conversations between different OHC members in terms of the duration of conversations or hibernation. Such patterns are generated from the sequence of replies within a single thread. Below, we list low-level tasks to analyze conversation patterns.

To understand the effect of the long-term participation of members, the first step was to observe the behavior of leaving a post in a thread. This task included understanding the relationship between the questioner and the repliers in the thread, e.g., the number of people involved in the thread, and understanding when the replies were answered. These measures reflected whether the questioner was actively engaged and whether the questioner also received appropriate support from the other users. Such support from other users is known to be an important factor in determining whether the users revisit OHCs [1]. The following are example questions that the administrator could ask: “Does the questioner participate in the conversation?” or “How many OHC members are involved in the conversation?”

The participants were also interested in finding whether questioners received solutions–at least a productive discussion on the matter in which they were interested, which is an important success factor. The temporal trend was often used as a predictive factor for defining the productivity of a conversation. For instance, if a question received few replies over a long period of time, then the questioner may not have been satisfied. Participants were also interested in finding whether there were different types of questions that were asked and answered based on the user's interest gained from a temporal perspective. The following are example questions that the administrator can ask. “Which threads have many replies in a short period of time? What are the topics in these threads?” or “Did the thread have a longer than usual hibernation period and was then revived?”

T1.2. Understand the contents of the posts: Among the large volume of posts generated every day, participants wanted to find interesting threads that required more attention. As mentioned above, helping users have a successful experience in the community is important. In particular, as the topics are health-related, stories of severe health conditions or unsatisfied experiences were often considered primary targets to which more attention should be given. Conversely, conversations having positive voices indicated that the users had successful experiences, and the participants wanted to study the details of their success. The participants were also interested in the unique topics that helped them understand emerging topics and new trends, which included new treatments. Finally, the contents in the post were distinguished by stories that were fact-based or contained more personal stories. The type of personal information, medical conditions, or treatment experience of the members of the community was an important means to understanding them. For instance, the administrator can ask: "Which threads discuss unique topics more than others?"

T1.3. Understand the correlations between diverse measures: Finding correlations was a critical task for the participants in that it helped them form hypotheses about the success factors for OHCs. For example, participants asked "Do threads with negative questions tend to have a large number of replies?" Participants wanted to find the correlations between multiple measures of individual threads.

T2. Scale up and drill down—After investigating individual discussion threads, the participants formed a hypothesis, which they wanted to test by scaling up to a big picture and showing similar threads in it. This similarity analysis was followed by drilling down to a subset of threads to examine outliers, as well as to confirm their hypotheses. In this series of tasks, the participants sought common patterns in a group of similar threads and the subset of common attributes in which they were similar to each other.

T2.1. Find similar threads: Similar threads were defined by thread characteristics, as described above. The need to find similar threads was formed in two different manners. First, the participants discovered interesting threads and wanted to view similar threads. Second, participants looked for specific conditions of threads and wanted to query threads with these conditions. For example, this task answered questions, such as "I want to see threads that include a negative question with more than 30 replies."

T2.2. Drill down to details: After observing a group of threads, the participants wanted to drill down to a subset of threads to answer why these trends occurred. This often led to new insights, which encouraged the participants to look for similar threads. This iterative process continued and formed knowledge about communities. In this task, participants answered questions, such as "I want to understand in which attributes these threads are similar or different from each other."

T3. Confirm findings from text—Finally, to confirm the hypothesis or gain a better understanding of the threads, the participants wanted to read the actual content. The participants were also interested in the characteristics of member demographics, which show the context of their discussions, such as diagnosis, critical life events, and their disease treatments. In other words, they wanted to profile member characteristics informally by checking relevant statements in the text. The kind of a question that can be asked during the confirmation process is “If a thread has a negative sentiment, is the user in a critical health condition?”

4 Designing VisOHC

In this section, we describe our design choices concerning VisOHC and their rationale based on the participants' tasks and goals.

4.1 Overall Design Process

To support our participants' tasks described in Section 3.4, we first needed to identify and extract the entities and measures in which the participants were interested. Our team worked together to brainstorm design ideas and give feedback to one another, and implemented the idea in a working prototype through a so-called “ideate” process [28]. That is, our prototype system visualized various initial measures mainly via flexible encoding options using box colors. Using the prototype, the participants could examine the usefulness of such entities and measures and provide new ideas for refining them. We describe the usefulness of providing a flexible prototype in the design process in Section 6. In the following sections, we explain the entities and measures that we derived in collaboration with the participants.

4.2 Extracting Information and Measures from OHC

In general, various information can be extracted from both the structured, e.g., author info, posting time, and unstructured portions, e.g., thread contents, of data available in OHC threads. However, particular types of information may be of interest to the participants, including health-related events, e.g., diagnosis (T1.2).

4.2.1 Thread Contents

Extracting keywords and phrases of interest: One of the most time-consuming tasks that our participants needed to perform was to understand thread contents by reading through them. Thus, we focused on developing measures that quickly summarize the thread contents in various fashions. To this end, we started by extracting entities and phrases that constitute themes related to the participants' interest.

However, standard entity extraction libraries, e.g., Stanford NLP, could not easily distinguish the keywords specific to OHCs. Hence, we used a pre-trained machine learning algorithm that extracts only medical-related keywords from patient-authored text [26]. After removing general stop words, our initial keyword set consisted of 4,633 words, which were then used to further compute meaningful measures and to highlight such keywords in different views.

Furthermore, OHC threads can contain unique phrases related to disease management and patient characteristics. The participants said they often identified unique phrases and/or words that indicated the context of OHC members. We derived an initial set of patient characteristics of interest to the experts by reading OHC posts. We progressively found six categories based on the initial categories provided by our participants: personal voice, diagnosis, life event, weight, age, and nutrition and medication. To detect the categories in the text, we created a set of phrases and/or keywords that explain each of these categories by creating bi-grams and tri-grams based on users' needs. Starting with initial phrases as seeds, we gradually populated the list.

The following list shows exemplary phrases used to detect the six categories of patient characteristics. **Personal Voice** described personal opinions, interpretations, and experiences. To find subjective expressions, we used the regular expression patterns that match the following phrases, e.g., I {*think*, *experience*, *etc*.}, *my* {*personal*} {*opinion*, *experience*, *etc*.}, {*to*, *for*} *me*. **Diagnosis** and **life event** contained following phrases, e.g., {*preposition*} {*numeric expressions in digits or words*} {*time units*}. **Age, weight**, and **nutrition and medication** contained numerical expression and units, e.g., for age: {*numeric expressions in digits or words*} *years* {*old*, *young*}.

Characterizing thread contents via measures: We described entities and phrases used for several different purposes above. A main usage of characterizing thread contents was to provide the participants with a summary of them. The participants wanted to identify the characteristics of thread contents in the following three perspectives.

First, OHC administrators wanted to find which threads discuss unique topics (T1.2). To serve this need, we developed the **Anomaly Score**, which is computed for each thread. In detail, we start with the bag-of-words vector representations of threads only using 4,633 medical-related keywords. Next, we compute a keyword-wise anomaly score as the inverse of the total frequency of a particular keyword. Then, within the bag-of-words vector of each thread, we multiply the frequency of each keyword by its corresponding keyword-wise anomaly score. The anomaly score of each thread is then computed as the total sum of these weighted frequency values of all the terms shown in each thread. Finally, we exponentially scale this score to strengthen its ability to discover rare threads. Overall, our participants liked Anomaly Score because they could answer questions, such as "*Is this a different topic that the community has been addressing?*"

Second, another perspective about a post was whether its contents were mainly about personal stories or fact-driven contents. Personal stories were of interest to our participants as they contained the member's own subjective interpretation and personal experiences, as compared to the posts that contained facts with strong evidence, e.g., scientific papers. The participants were curious about the impact of such subjectivity on discussion threads, because they often attempt to encourage OHC users to speak in their personal voices: "*We tell people to share their own personal and work experiences, talk about them, and show examples that they have direct experience with, as a way of engaging people*," P2 said. The **Personal Voice Score** reflects the degree of such subjectivity, and this score is computed as the number of Personal Voice entities and phrases described above for each post. We found that this personal voice score was helpful in guiding our participants to those threads with high/low score values so that they could inspect whether misinterpretation was being propagated (T1.2).

Third, it was found that the participants considered the sentiment of a thread very important. The sentiment can be used to explore the tone of a questioner or a replier in a thread (T1.2). More interestingly, the sentiment can reveal the intensity of the conversation in a thread, as well as the urgency of the questioner looking for an answer. That is, a negative sentiment sometimes indicated that the patient was desperately looking for an answer to her problem. On the other hand, a positive sentiment was often found in the threads describing successful stories of the patients. Motivated by these findings, we developed the **Sentiment Score** indicating the degree of positive/negative sentiments for each post. This score is computed by summing the sentiment scores of keywords shown in each post [31]. When computing this score, one design choice was whether to include only the medical-related keywords or all the keywords (after removing stop-words). After testing both strategies and discussions with the participants, we found that the former sometimes gave somewhat biased sentiment scores, and therefore, we decided to use the latter strategy.

4.2.2 Conversation Pattern of Threads—The participants wanted to explore the conversation in each thread. Our interview revealed several aspects of conversation dynamics (T1.1), and such conversation patterns were characterized in various manners. We derived three different measures based on the interview for each thread to describe its conversation pattern. The first measure is the **reply count per person**, which shows the number of participants in the conversation and the number of times they entered replies. P1 shared, "*I think there's a nature. I wonder if there's a way to categorize responses. Does this discussion bring in new people or does it just go back and forth between a small group?*" The second measure is the **questioner participation count**, which shows the number of times a questioner provided feedback to repliers. P2 shared, "*Whoever posted originally, if they come back, that does seem to feed the discussion.*" The third measure is the **total duration of a conversation**, which shows the duration between the first and the last entry.

In addition to the questions answered by these measures, participants wanted to address more complicated questions related to the conversation patterns by inspecting the details of OHC members' participation patterns. For instance, these measures could not answer questions such as when users join and leave a conversation and whether the conversation one-to-one, one-to-many, or many-to-many at a particular time segment of a conversation. To answer these questions, we developed visualizations for further exploring the conversation patterns, as discussed in Section 4.3.1.

4.3 Interactive Visualization Design

Now, we describe our visual analytics system design, which effectively presents the above-identified information via interactive visualization modules. As shown in Fig. 1, our system consisted of three views, Thread View (Fig. 1(D-E)), Similarity & Histogram View (Fig. 1(B-C)), and Text View (Fig. 1(G)), together with control user interfaces (Fig. 1(A)), which provide versatile mapping capabilities of extracted information and measures to diverse visual encoding channels. Text View, Thread View, and Similarity & Histogram View are connected via brushing and linking. Our participants found this capability useful for quickly moving from one view to another to smoothly learn the overview and details of the threads of interest.

4.3.1 Thread View: Boxes Trailing a Box—Thread View is a key visualization module of our system. For its visual design, we adopted a visual metaphor of online community threads. Specifically, Thread View visualizes an original question, which initiates a thread, as a Question Box, trailed by its replies as smaller boxes (Reply Boxes) aligned horizontally in their posting order (see Figure 2). This design choice provided us with a unique set of advantages, design spaces, and constraints. First, it shows the conversation pattern of threads, e.g., the number of replies per questions, and the inherent dependencies of replies on their questions in an intuitive manner. In addition, this visual metaphor created various visual encoding channels, as Figure 2 shows. For example, we can use the box elements, such as box width, vertical order, and color, to map measures (see Section 4.3.1). In addition, we augmented circles to represent questioners and repliers; the circles are used as an interaction handle to show user-wise information, e.g., a Word Cloud of all the questions and replies written by a user. By mapping multiple measures to boxes, Thread View allows users to compare threads and test correlations between the measures within and between threads (T1.3).

However, this design created constraints on an overview analysis. Since the design illustrates individual threads in a horizontal slice of a viewport, it is most effective for exploration and detailed analyses, but is not effective for overview tasks. If our domain tasks follow the traditional visual information seeking mantra, we might as well visualize an overview of the entire thread, such as, in terms of topics, e.g., in [17]. In fact, at an initial stage, we also considered a ThemeRiver-style temporal trend visualization [15] using topic modeling [6], which should provide a topical summary of the entire thread. Nevertheless, our analysis of user tasks revealed that, rather than starting with an overview of the topical trends, the participants wanted to review each thread one by one, checking its first post. For this reason, we excluded such a topical overview visualization in our design.

From this exploration process of individual threads, the participants found clues in an individual or a subset of threads of interest and then formed and confirmed their hypotheses by identifying similar threads– *find trees of interest and understand trees in the context of forest*. Thus, revealing conversation patterns, OHC measures, and other descriptive measures of threads took a higher priority in our case. Nonetheless, for a top-down style of analysis, we also provided Similarity & Histogram View, which provides the grouping of threads in a user-defined context, as discussed in the next section.

Mapping Measures to Question Boxes: In Thread View, the entire horizontal space is available for mapping OHC measures to each thread. We allowed the participants to choose one measure among descriptive statistics, e.g., a character length or a reply count, which was then mapped to the box width. In addition, the participants could also choose to map the measures of the conversation patterns to the box width, such as the reply count per person, questioner participation count, and total duration of a conversation (See Section 4.2.2 for description). When participants selected a measure, the box width was expanded accordingly to the left so that all the question boxes were horizontally aligned at the edge on the right side, followed by their reply boxes to the right. In addition, the participant could also sort the vertical order of boxes by a measure of choice. The ability to map multiple measures to the visual variables of boxes allows users to explore the correlations among these measures (T1.3).

Next, we used box colors to show the characteristic of threads: Sentiment Score, Anomaly Score, and Personal Voice Score (see Section 4.2.1 for description). We differentiated these scores from other scores for the following reasons. (i) The absolute values of scores do not add a significant amount of interpretation; (ii) the scores are used for gaining an overall impression while scanning multiple threads; (iii) the scores are occasionally used to make a correlation with other measures mapped to the box width, e.g., whether longer questions tend to be negative. Hence, we represented these measures in colors, which constitute a different perceptual channel. We confirmed the usefulness of providing two different types of measures separately in colors and box width/vertical order to analyze the correlations between characteristics of threads (T1.3): "*Being able to look at things like what the sentiment of a discussion [is] and sort it by [number of] replies made that view very interesting. Some of the more negative discussions were actually generating the most conversations in the community.*", P1 said. We selected three color scales, red-to-green for Sentiment Score, white-to-yellow for Anomaly Score, and white-to-blue for Personal Voice Score, to highlight the differences among the scores and strengthen the visual salience of high values of the scores in which users were interested (see Figure 3).

Mapping Measures to Reply Boxes: Reply boxes are uniformly shaped and aligned in the order of posting date to the right. The box width of reply boxes is fixed so that the reply counts are shown by the position of the last reply of the thread.

Our participants confirmed the conversation pattern of threads provided a cue for building hypotheses. First, we created three views to understand the users' conversation structure in a thread (T1.1). These three views explained below, Replier Circle & Beam, Replier Marks, and Replier Timeline, show different aspects of the conversation structure by highlighting reply counts, order between repliers, and the sequence and duration of replies. Using these views, users can visually inspect the patterns of conversation within each thread.

Replier Circle & Beam draw circles, representing authors, from which semi-transparent beams shoot toward their replies, as Figure 4 (a) shows. Each circle is drawn per replier once at the horizontal position above its first reply. In this view, participants can view the first entry point of repliers by the number of circles as well as the reply counts per person. This view creates a visual clutter as beams start crossing each other. Nevertheless, the participants liked the view because the visual clutter generated by crossings indicates the amount of replies written by the person.

Replier Marks draw horizontal marks on top of the reply boxes of those who reply more than once in a given thread. In other words, an absence of replier marks on a thread indicates that no replier returned to the thread after writing a reply. The vertical position of the mark represents the replier's identification: the order of the first entry to the thread. Hence, the marks show various patterns of conversation saliently, answering, (e.g., "Was the conversation one-to-one or many-to-many at a given period of time?", or "How did conversation participants change?"). For example, Figure 4 (b) shows that four repliers engaged in conversation throughout the thread while three others each left only one reply. Likewise, the distribution, quantity, and vertical position of marks together form unique conversation patterns of discussion threads. Our participants found many interesting “signatures of conversation structure” using the view, as explained in Section 5.

Replier Timeline shows the posting date of replies using the horizontal position as a timeline (see Figure 4 (c)). This view shows visual collapses of replies based on their similarity in terms of posting dates. We animated the transition to show which replies were mapped into which time segments. In particular, this view creates the live period as condensed boxes and the hibernation period as an empty space between replies. A user can interpret the duration of a conversation as a reflection of the level of the thread participants' engagement or the length of time the discussions were heated (see Section 4.2.2). It also indicates the interesting development of discussions over a period of time. For instance, some threads included late replies after a few years of hibernation; some threads included multiple replies within a short period of time. Users can easily detect these differences visually.

4.3.2 Word Cloud View—We found that it was useful to provide the participants with a summary within each thread in the form of a Word Cloud (see Figure 5). To achieve this, we tested several candidate sets and weighting strategies for these keywords. The results showed that the best approach according to the participants' feedback was the set of medical-related keywords (described in the previous section), together with a term frequency-inverse document frequency (tf-idf) weighting scheme.

4.3.3 Similarity View & Histogram View—Similarity View & Histogram View support the tasks described in T2.1 and T2.2, which required the administrators to find a group of similar threads based on a user-defined specification of measures and to inspect similar threads in detail. For this purpose, we constructed a vector containing eight different measure values for each thread and then computed the pair-wise cosine similarities between these vectors. Then, using these similarities as input, we ran t-SNE [36], a state-of-the-art dimension reduction method for visualization, to project threads onto a two-dimensional space, which we called Similarity View (see Figure 6 (a)). Additionally, we provide a histogram of each measure of all the threads, which we call the Histogram View (see Figure 6 (b)). The two views are connected via linking and brushing so that users can filter threads by specifying regions of interest or ranges of values in measures in either of the views. Users can also include or exclude any of eight measures to re-compute cosine similarities and update Similarity View accordingly. This interaction allows users to focus on particular measures when exploring similar threads.

Similarity View & Histogram View helped participants “*identify the [new] kinds of beneficial threads.*” The participants found many new combinations of measures that created interesting groupings. Even more interestingly, they often found outliers that contradicted their initial assumptions about online health communities. P1 said, “*Those outliers were interesting. [To understand] what was different about that, I would need to go into reading the threads and find interesting things.*” In combination, these views generated interesting insights.

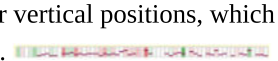
4.3.4 Text View—Text View shows the content written by questioners and repliers. The replies are indented to distinguish between questions and replies, and replies can be shown/hidden per the user's request. As Figure 7 shows, Text View highlights the sentences containing the extracted entities and phrases on the subject of diagnosis, life event, weight, age, and nutrition and medication (see Section 4.2 for details) per participants' request. In addition, the view shows the first sentence in order to provide a quick glance to find the summary. P2 reported that scanning various OHC entities marked on Text View (T3) was "showing the 'why' of this [hypothesis]."

5 Observation of Administrators

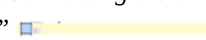
We observed two administrators to understand how they analyze OHC conversations and find insights using VisOHC. We first provided a 15-minute video tutorial on the use of VisOHC three days before the observation. They were then directed to our Web site to explore all the functions. In the following paragraphs, we describe their workflow using the tool and the elements that they found helpful.

Our tool is designed to support individual tasks, as described in previous sections. While it is difficult to draw conclusions on the basis of working with two administrators, we observed a common workflow between the two. The workflow follows a sequence of tasks starting from formulating questions about threads or exploring patterns of threads (T1), to forming hypotheses and exploring the data by scaling up and drilling down (T2), and, finally, confirming or rejecting their hypothesis by reading the text (T3). This workflow, where all interactions centered around the thread view, was consistent with the support that we attempted to provide through our tool. In the following, examples of the administrators' analysis processes are introduced. Related tasks are given in parenthesis.

First, P1 started with a specific question about the aspects of questions that attracted more replies. His hypothesis was that this might have a correlation with three scores. He ran a similarity computation using three scores (Sentiment, Anomaly, Personal) (T1.3). Then, he defined a range in reply counts (more than 50 replies). This showed seven threads in the Similarity View, where six of these were close together (T2.1). By checking the colors of the boxes, the participant confirmed that six of the threads had a medium-to-high anomaly score and a personal score with negative sentiments (T2.2). By reading the six threads, he confirmed that the six threads had commonalities: personal failure stories attract many replies (T3).

Next, P1 was curious why the seventh thread, which had a low anomaly score and personal score with a positive sentiment, also attracted many replies. The question was on the subject of recommendations to eat fruit and couched in an enthusiastic tone. He quickly changed the box width to show the total duration of conversation, which revealed that the duration of the conversation was short (T2.2). The Replier Mark confirmed that the view contained many marks located at the bottom having similar vertical positions, which indicated that the thread was between a small group of participants.  (T2.2). By reading the replies, the administrator confirmed that the conversation started with food stories, but somehow diverged into Denzel Washington movies, sharing photos, and a spider story between three people, which constituted more than a half of the replies (T3). He confirmed that question length alone cannot predict productive conversations, because long threads often include such social conversations within a small group. Through the analysis, the administrator could find two hypotheses about the kinds of stories that attracted more members: they contain people's personal failures and trivial stories irrelevant to health. This knowledge can be useful for engaging more members in conversations.

P2 started to explore different measures with the VisOHC without a specific question. She first performed some correlation tests among different dimensions and sorted the threads by different measures (T1.3). She discovered some interesting insights from threads sorted by reply counts (T2.2): *"When ordered by reply count,* 7 *of the top 10 discussions had a negative sentiment but the number [of threads with negative sentiment] decreased to 4 when sorted by reply counts per person."* Then, the administrator interpreted the results to gain new insights into such trends: *"It seems negative posters come back often and increase the negative sentiment of the threads."*

During the interview, the administrators also shared other insights they learned. P2 formed a hypothesis about a correlation between question length and replies and the analysis revealed the hypothesis was wrong: *"We usually advise people to write 200 characters or less. I was surprised that question length did not seem to correlate with replies. We have a hypothesis that shorter questions/initial posts are easier for members to engage with, but that does not seem to be the case."* P1 described his theory about the occurrence of a long hibernation with a late spike in threads, *"What we typically find about stories is that we have some late comers stumbling upon [old threads]. There are particular reasons for this; probably, someone Googled and stumbled on this thread and left a note without knowing the thread was from a while ago. This [event] sometimes reopens the thread:"* , which is confirmed (see Figure 8).

Having used VisOHC, the administrators expressed that they felt the tool empowered them to test which subjects are more likely to engage more people. They praised the work flow provided by VisOHC; as P2 mentioned "I *saw this interesting pattern [in Thread View] and it let me look at a cohort of these patterns [in Similarity View]."* They revealed that their analysis successfully resulted in insights, which flawlessly fed into next analysis, thereby enabling them to discover new testable hypotheses in OHC threads. P1 described the strength of VisOHC as *"[It] leaves me hungry for what's next and continuously on the verge of big insights."* Through our observation, we confirmed that VisOHC supported the administrators' task sequences and helped them answer questions about maintaining an OHC.

6 Lessons Learned

In this section, we first describe our next steps for improving VisOHC, followed by design considerations and recommendations for designing visual analytics for OHCs. We conclude with lessons we learned by conducting a design study with the participation of domain experts.

6.1 Design Considerations for Visual Designs of VisOHC

Having reviewed the tasks of the OHC administrators and our participants' evaluation of the visual designs of VisOHC, we highlight the lessons we learned about the improvements that should be made to VisOHC. Table 1 shows the summary of the mapping between tasks and visual designs. For T1.1 task, three views–Replier Mark, Replier Circle & Beam, and Replier Timeline–drew a high level of interest from our participants, because of their novel visual designs. Our participants further provided suggestions for increasing the views' visual clarity. For instance, we should provide zoom-in features to make patterns more visible. Our participants found Word Cloud and Text View to be effective for understanding the summary and context of threads (T1.2). Since the opinion of OHC administrators as to what issues are important can vary depending on the situation, we should allow administrators to map the colors and sizes of Word Cloud differently to reflect important content within certain threads or all the threads. To further support administrators' heavy interpretation work involving digesting OHC content, adding the ability to annotate Text View would enable administrators to mark new types of entities or take notes on their analysis progress. To improve correlation tasks (T1.3) in which multiple comparisons among threads are made, we can enable administrators to slice and dice a subset of Boxes in threads to help them perform more sophisticated analyses. We can further run clustering analysis to color the results in Similarity View. For T2.1 and T2.2, we saw areas for improving the administrators' tasks by providing automatic suggestion of similarities and pattern-based searching abilities.

6.2 Visual Analytics for OHC

Designing Visual Analytics for OHC Administrators—Knowledge of domain goals and tasks facilitate the design of visual analytics applications for OHCs. OHC administrators aim to identify success factors in discussion threads by exploring diverse sets of measures. The goals of their tasks require that interesting patterns be discovered within individual threads in detail. In this study, we defined and created domain specific measures to characterize the patterns, such as personal voice score and anomaly score, that characterize OHC threads. As our design process shows, many other unique measures can be developed and used depending on the specialty of the OHC. An interesting next step would be to collect hypotheses and patterns from administrators and to derive a predictive model of OHCs (“How many replies will discussion threads with certain attributes have”).

Use a visual metaphor for text containers (it's effective sometimes)—Our Thread View design was inspired by the representation of conversation threads on the Web, where a question post has a box shape and the replies that follow are indented. Using such a metaphor created an immediate connection between the visual representations and the real world. We hypothesize that our visual metaphor may help users internalize patterns shown on threads and externalize their knowledge, reflecting findings from studies on visual metaphor [44]. We believe that matching measures with appropriate visual variables requires careful selection, because some measures can be perceived better through a shape's length (as used for the number of replies), while others can be perceived better through color (as used for the anomaly score). This is our hypothesis, and therefore, future work is required to verify it. It will also be interesting to study the compatible combination of measures and visual variables to make visual metaphors more "compatible," as Ziemkiewicz and Kosara stated [44]. The metaphor could also lower the barrier for understanding several measures encoded on top. Of course, such a metaphor creates design constraints. Conversely, the metaphor itself created new design space and creative ways to show patterns, in our case sequential patterns of conversations structure represented by replier Circles and Beams, Replier Marks, and Replier Timeline. Discovering a metaphor to contain the primary data variable could improve the usability of visual analytics applications.

Detail to Overview and Detail in Overview—Our experts used workflows that enabled them to start finding patterns in detailed view (Thread View), and then find similar ones in an overview (Similarity View). Meanwhile, they also checked the distribution of dimensions in the overview (Histogram View) to understand the overview as well. Even if they frequently used overviews, they did so to find the next interesting threads that they wished to inspect. For administrators, treating separate threads as individuals is critical, because aggregating threads may hide obscure patterns that are of interest to them. The administrators tended to focus on individual cases first and then attempt to find similar cases whether such interestingness can be predicted with some other measures. Thus, showing details of conversation and similarities in overview in a juxtaposed location could support seamless analysis processes.

Revealing the characteristics from text—While working on text data in other projects, we frequently encountered situations where existing preprocessing techniques, e.g., entity extraction, were not effective, because the domain experts wanted to see very specific subsets of such entities. In our experience, topic modeling and sentiment analysis were frequently useful to some extent, but were not compatible with the users needs. We learned that administrators wanted to learn more about the people in their communities, e.g., why they write certain questions and answers. From an in-depth review of their tasks, we derived the important attributes of threads, e.g., personal voice, that they wanted to extract. The anomaly score is a reverse measure of the popularity of topics, but it shows exactly what administrators wanted to see. Thus, we learned that characterizing users' tasks and fitting our measures to their needs are critical.

6.3 Methodological Lessons

Having conducted this design study with domain experts, we retrospectively review some positive actions that led to our design solution.

Pave the cow path—We focused on identifying a unit of domain analysis such as a thread. The unit of domain analysis is frequently observed as a central language that describes knowledge and insights in data. In our case, a thread is an appropriate interpretation layer to which administrators refer while describing the findings. This observation led us to set design requirements and the overall workflow around the Thread View, which had a positive influence on the administrator's analysis. Users can vividly express their findings and relate what they know to what they see. Abstracting the central unit requires that we carefully analyze domain tasks and listen to experts' communications.

Working with front line analysts, as well as translators, who have common understanding and knowledge, can create an opportunity to be a 'fly-on-the-wall.'

The combinatorial prototyping—When exploring visual designs, researchers frequently encounter difficulty in effectively prototyping the appropriate level of consideration space to inspire domain experts [35]. Paper prototypes often fail to give a tangible idea of the appearance of the prototype when implemented, in particular, when the prototype includes a considerable amount of interaction and data abstraction. We encountered this problem when we had conceived a basic idea of Thread View and the various OHC measures to extract patterns from text. Our solution was to provide measures and encoding options that could be easily manipulated. With the simplest default selections (no encoding for most design variables), users were asked to build their own visualizations by encoding measures for various variables. This construction process was limited but sufficiently flexible to inspire fruitful feedback and creative ideas for new measures similar to that which Huron et al. showed [21]. This self-trial process helped the users progressively learn the visual encoding process of various measures. In addition, it is effective for evaluating effective views, measures, and patterns for specific tasks.

7 Conclusions

In this design study, we characterized and derived the domain goals and tasks that OHC administrators attempt to achieve by analyzing multiple threads. Throughout the iterative design processes, our visual analytics application captured hidden dimensions of threads and enabled the administrators who participated in the study to explore, compare, and find similar threads using multiple views and interactions. Our study indicates the importance of capturing the unit of analysis, which determined the direction of our main view and work flow design. We also shared our lessons on design study methodology learned by providing a combinatorial prototype in which users could freely brainstorm and validate combinations of measures and visual encodings. In future work, based on our administrators' findings and hypotheses, we aim to investigate predictive analysis for establishing a success model of OHC.

Acknowledgments

We thank John Crowley, Marie Connelly, Aaron Beals, and Jessica Ludvigsen for participation and invaluable feedback. Research reported in this paper was in part supported by National Library of Medicine of the National Institutes of Health under award number K01LM011980.

References

1. Aarts JW, Faber MJ, den Boogert AG, Cohlen BJ, van der Linden PJ, Kremer JA, Nelen WL. Barriers and Facilitators for the Implementation of an Online Clinical Health Community in Addition to Usual Fertility Care: A Cross-Sectional Study. Journal of Medical Internet Research. 2013; 15(8):e163. [PubMed: 23996964]
2. Aarts JWM, Haak Pvd, Nelen WLDM, Tuil WS, Faber MJ, Kremer JaM. Patient-focused Internet interventions in reproductive medicine: a scoping review. Human Reproduction Update. 2012; 18(2):211–227. [PubMed: 22108381]

3. Abdul-Rahman A, Lein J, Coles K, Maguire E, Meyer M, Wynne M, Johnson CR, Trefethen A, Chen M. Rule-based visual mappings with a case study on poetry visualization. Computer Graphics Forum. 2013; 32(3pt4):381–390.

4. Angus D, Watson B, Smith A, Gallois C, Wiles J. Visualising conversation structure across time: Insights into effective doctor-patient consultations. PloS one. 2012; 7(6):e38014. [PubMed: 22693629]

5. Baker TB, Gustafson DH, Shaw B, Hawkins R, Pingree S, Roberts L, Strecher V. Relevance of CONSORT reporting criteria for research on eHealth interventions. Patient Education and Counseling, pages. 2010:77–86.

6. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. Journal of Machine Learning Research. 2003; 3:993–1022.

7. Chee, B.; Karahalios, KG.; Schatz, B. Social visualization of health messages; IEEE 42nd Hawaii International Conference on System Sciences; 2009. p. 1-10.

8. Chiu TM, Eysenbach G. Stages of use: consideration, initiation, utilization, and outcomes of an internet-mediated intervention. BMC Medical Informatics and Decision Making. 2010; 10(1):73. [PubMed: 21092275]

9. Cobb NK, Graham AL, Byron MJ, Niaura RS, Abrams DB. Online Social Networks and Smoking Cessation: A Scientific Research Agenda. Journal of Medical Internet Research. 2011; 13(4)

10. Collison, G.; Elbaum, B.; Haavind, S.; Tinker, R. Facilitating online learning: Effective strategies for moderators. ERIC; 2000.

11. Cui W, Liu S, Wu Z, Wei H. How hierarchical topics evolve in large text corpora. IEEE Transactions on Visualization and Computer Graphics. 2014; 20(12):2281–2290. [PubMed: 26356942]

12. Dancet, EaF; Empel, IWHV.; Rober, P.; Nelen, WLDM.; Kremer, JaM; D'Hooghe, TM. Patient-centred infertility care: a qualitative study to listen to the patient's voice. Human Reproduction. 2011; 26(4):827–833. [PubMed: 21317152]

13. De Pauw, W.; Jensen, E.; Mitchell, N.; Sevitsky, G.; Vlissides, J.; Yang, J. Software Visualization. Springer; 2002. Visualizing the execution of java programs; p. 151-162.

14. Diakopoulos, N.; Zhang, A.; Elgesem, D.; Salway, A. Identifying and analyzing moral evaluation frames in climate change blog discourse; Proc International Conference on Weblogs and Social Media; 2014.

15. Havre S, Hetzler E, Whitney P, Nowell L. Themeriver: visualizing thematic changes in large document collections. IEEE Transactions on Visualization and Computer Graphics. 2002; 8(1):9–20.

16. Heer J, Perer A. Orion: A system for modeling, transformation and visualization of multidimensional heterogeneous networks. Information Visualization. 2014; 13(2):111–133.

17. Hoque E, Carenini G. Convis: A visual text analytic system for exploring blog conversations. Computer Graphics Forum. 2014; 33(3):221–230.

18. Huh, J. Clinical Questions in Online Health Communities: The Case of “See Your Doctor” Threads; The ACM Conference on Computer Supported Cooperative Work & Social Computing; 2015. p. 1488-1499.

19. Huh, J.; Kwon, BC.; Choo, J.; Kim, SH.; Yi, JS. Coddlers, scientists, adventurers, and opportunists: Personas to inform online health community development. ArXiv e-prints. 2015. http://arxiv.org/abs/1506.09191

20. Huh J, Yetisgen-Yildiz M, Pratt W. Text classification for assisting moderators in online health communities. Journal of biomedical informatics. 2013; 46(6):998–1005. [PubMed: 24025513]

21. Huron S, Jansen Y, Carpendale S. Constructing Visual Representations: Investigating the Use of Tangible Tokens. IEEE Transactions on Visualization and Computer Graphics. 2014; 20(12): 2102–2111. [PubMed: 26356924]

22. Kerr B. Thread Arcs: an email thread visualization. IEEE Symposium on Information Visualization, 2003. 2003:211–218.

23. Kochtchi A, von Landesberger T, Biemann C. Networks of names: Visual exploration and semi-automatic tagging of social networks from newspaper articles. Computer Graphics Forum. 2014; 33(3):211–220.

24. Koh K, Lee B, Kim B, Seo J. ManiWordle: Providing Flexible Control over Wordle. IEEE Transactions on Visualization and Computer Graphics. 2010; 16(6):1190–1197. [PubMed: 20975158]
25. Liu S, Wang X, Chen J, Zhu J, Guo B. TopicPanorama: A full picture of relevant topics. IEEE Visual Analytics Science and Technology. 2014:183–192.
26. MacLean DL, Heer J. Identifying medical terms in patient-authored text: a crowd sourcing-based approach. Journal of the American Medical Informatics Association. 2013; 20(6):1120–1127. [PubMed: 23645553]
27. Malik SH, Coulson NS. Computer-mediated infertility support groups: An exploratory study of online experiences. Patient Education and Counseling. 2008; 73(1):105–113. [PubMed: 18639409]
28. McKenna S, Mazur D, Agutter J, Meyer M. Design Activity Framework for Visualization Design. IEEE Transactions on Visualization and Computer Graphics. 2014; 20(12):2191–2200. [PubMed: 26356933]
29. Murray E, Burns J, May C, Finch T, O'Donnell C, Wallace P, Mair F. Why is it difficult to implement e-health initiatives? A qualitative study. Implementation Science. 2011; 6(1):1–11. [PubMed: 21208425]
30. Nambisan P. Information seeking and social support in online health communities: impact on patients' perceived empathy. Journal of the American Medical Informatics Association. 2011; 18(3):298–304. [PubMed: 21486888]
31. Oelke D, Hao M, Rohrdantz C, Keim DA, Dayal U, Haug L, Janetzko H. Visual opinion analysis of customer feedback data. IEEE Visual Analytics Science and Technology. 2009:187–194.
32. Oprescu F, Campo S, Lowe J, Andsager J, Morcuende JA. Online Information Exchanges for Parents of Children With a Rare Health Condition: Key Findings From an Online Support Community. Journal of Medical Internet Research. 2013; 15(1):e16. [PubMed: 23470259]
33. Plaisant C, Mushlin R, Snyder A, Li J, Heller D, Shneiderman B. LifeLines: using visualization to enhance navigation and analysis of patient records. The AMIA Symposium. 1998:76–80.
34. Richardson CR, Buis LR, Janney AW, Goodrich DE, Sen A, Hess ML, Mehari KS, Fortlage LA, Resnick PJ, Zikmund-Fisher BJ, Strecher VJ, Piette JD. An Online Community Improves Adherence in an Internet-Mediated Walking Program. Part 1: Results of a Randomized Controlled Trial. Journal of Medical Internet Research. 2010; 12(4)
35. Sedlmair M, Meyer M, Munzner T. Design Study Methodology: Reflections from the Trenches and the Stacks. IEEE Transactions on Visualization and Computer Graphics. 2012; 18(12):2431–2440. [PubMed: 26357151]
36. van der Maaten L, Hinton G. Visualizing data using t-SNE. Journal of Machine Learning Research. 2008; 9:2579–2605.
37. Viegas F, Wattenberg M, Feinberg J. Participatory Visualization with Wordle. IEEE Transactions on Visualization and Computer Graphics. 2009; 15(6):1137–1144. [PubMed: 19834182]
38. Wang T, Plaisant C, Shneiderman B, Spring N, Roseman D, Marchand G, Mukherjee V, Smith M. Temporal Summaries: Supporting Temporal Categorical Searching, Aggregation and Comparison. IEEE Transactions on Visualization and Computer Graphics. 2009; 15(6):1049–1056. [PubMed: 19834171]
39. Wang, YC.; Kraut, R.; Levine, JM. The ACM 2012 Conference on Computer Supported Cooperative Work. ACM; 2012. To Stay or Leave?: The Relationship of Emotional and Informational Support to Commitment in Online Health Support Groups; p. 833-842.
40. Xu P, Wu Y, Wei E, Peng TQ, Liu S, Zhu J, Qu H. Visual Analysis of Topic Competition on Social Media. IEEE Transactions on Visualization and Computer Graphics. 2013; 19(12):2012–2021. [PubMed: 24051767]
41. Zhang L, Stoffel A, Behrisch M, Mittelstadt S, Schreck T, Pompl R, Weber S, Last H, Keim D. Visual analytics for the big data era – A comparative review of state-of-the-art commercial systems. IEEE Visual Analytics Science and Technology. 2012:173–182.
42. Zhao J, Gou L, Wang F, Zhou M. PEARL: An interactive visual analytic tool for understanding personal emotion style derived from social media. IEEE Visual Analytics Science and Technology. 2014:203–212.

43. Ziebland S, Wyke S. Health and Illness in a Connected World: How Might Sharing Experiences on the Internet Affect People's Health? Milbank Quarterly. 2012; 90(2):219–249. [PubMed: 22709387]

44. Ziemkiewicz C, Kosara R. The Shaping of Information by Visual Metaphors. IEEE Transactions on Visualization and Computer Graphics. 2008; 14(6):1269–1276. [PubMed: 18988973]

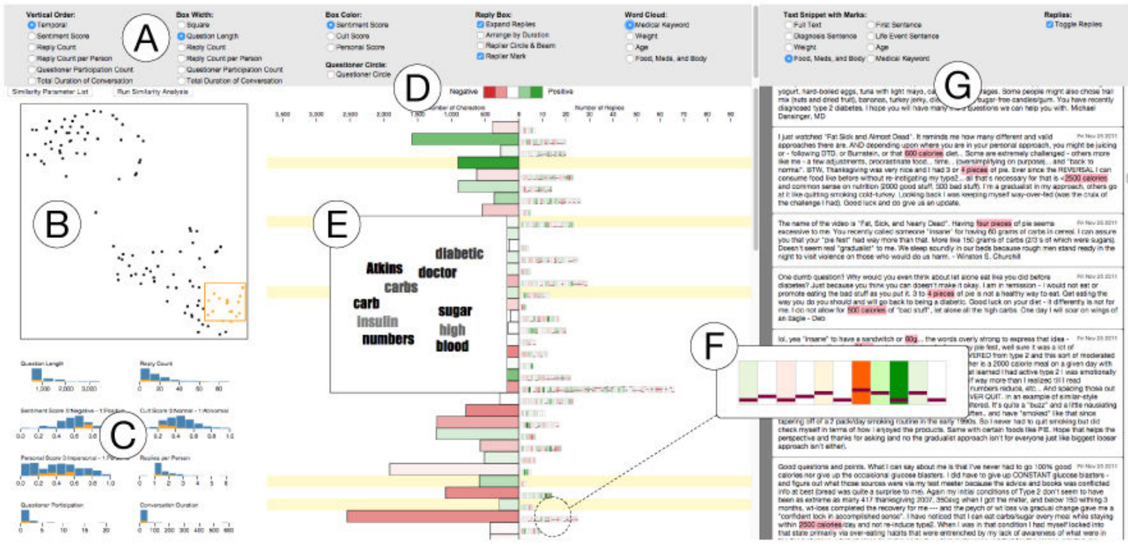


Fig. 1.

The VisOHC system includes (A) Dashboard for manipulating visual encoding schemes and measures; (B) Similarity View displaying the cosine distances of the discussion threads with respect to user-selected measures, where users can find similar threads and filter threads by drawing an orange rectangle; (C) Histogram View displaying the distribution of eight measures of the discussion threads, where users can view the distribution of measures of selected threads and filter threads by specifying the range per measure; (D) Thread View displaying the threads as boxes where a color, a width, and a vertical order represent the sentiment, the question length, and the recency, respectively; (E) Word Cloud displaying medical keywords of a thread over which the user hovers the mouse; (F) Replier Marks displaying the unique number of participants as well as the sequence of replies; and (G) Text View displaying text with highlights related to food, medications, and the body.

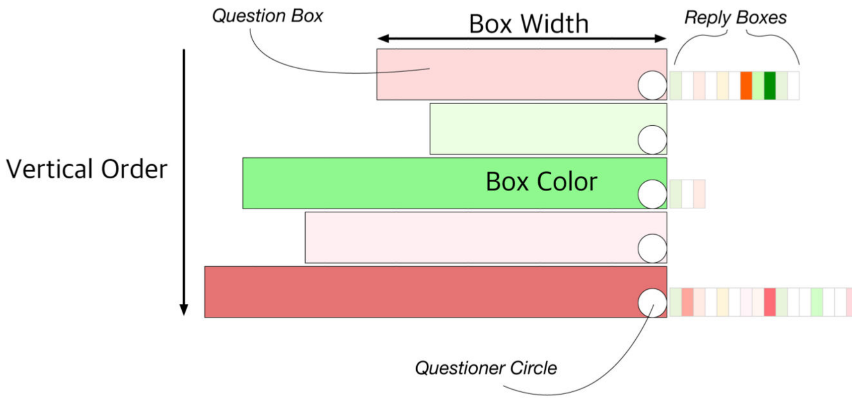


Fig. 2.
Question boxes can encode Anomaly Score, Personal Voice Score, and Sentiment Score in colors and other measures in box width and vertical order. Reply boxes are arranged immediately adjacent to their questions. Users can click on boxes to highlight the clicked questions/replies, mouse over circles to open Word Cloud, and click on circles to highlight the OHC member's postings.

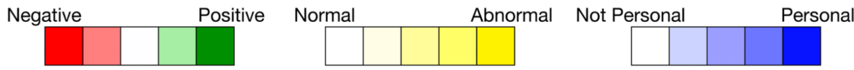


Fig. 3.
Color scales for Sentiment, Anomaly, and Personal Scores.

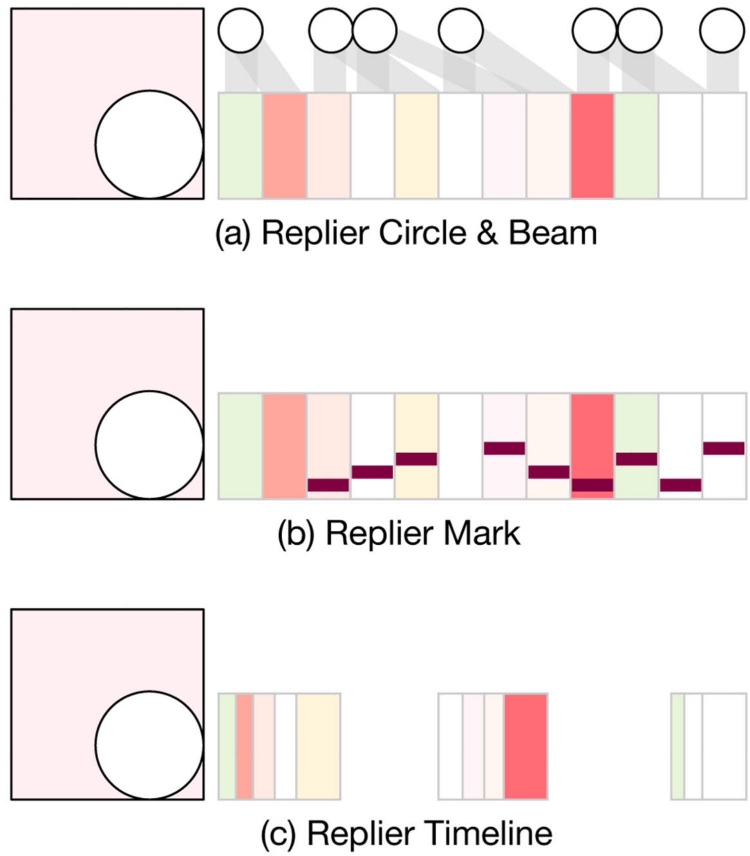


Fig. 4.

Three views of replier boxes: (a) Replier Circle & Beam shows the first reply of individual repliers as circles and their authorship of the subsequent replies as beams (gray lines); (b) Replier Mark shows marks for repliers who left replies more than once. The vertical position of marks shows the first entry order of the repliers; (c) Replier Timeline shows the posted times of replies in a horizontal position.

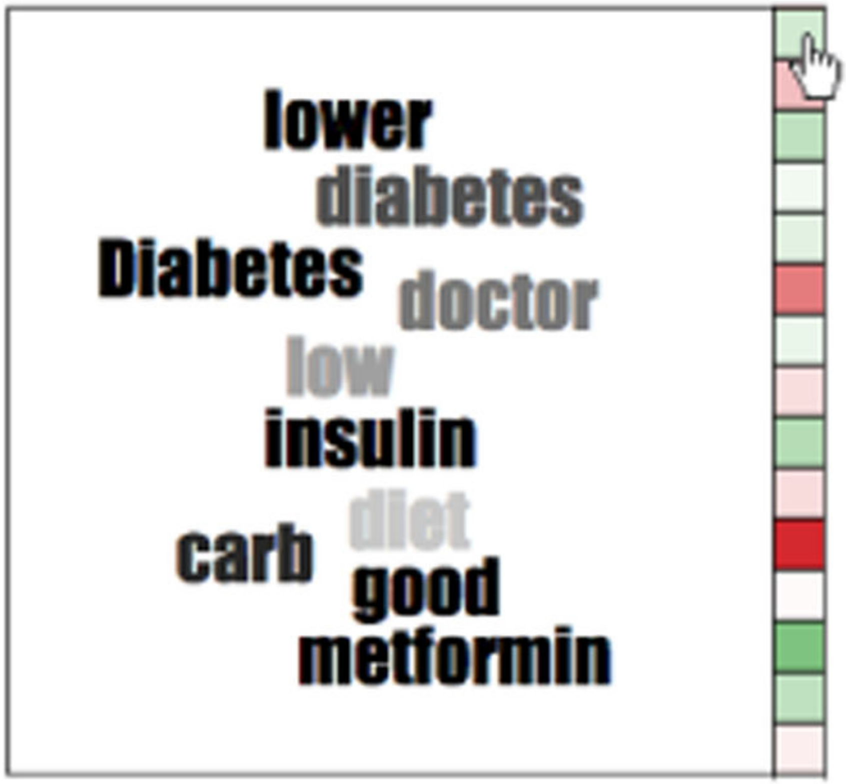
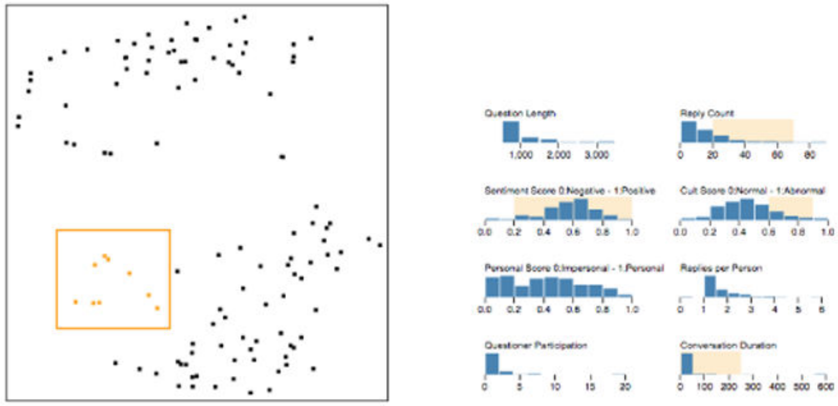


Fig. 5.

Users can open Word Cloud for questioners and repliers, as well as their individual questions and answers.



(a) Similarity View shows cosine distances between threads.

(b) Histogram View shows the distribution of measures of discussion threads.

Fig. 6.

(a) Similarity View & (b) Histogram View support exploration of threads according to similarity. The two views, as well as all the other views, are connected through linking and brushing. Users can filter threads in both views by specifying the region of interest.

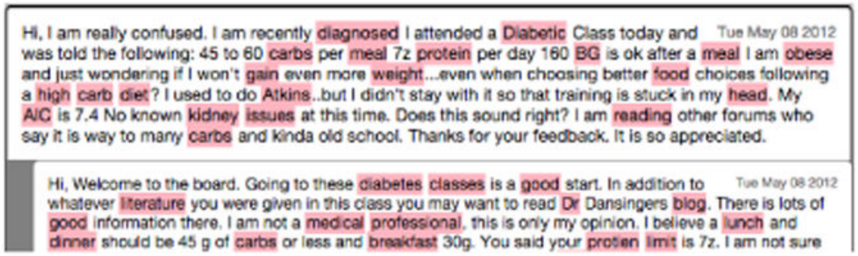


Fig. 7.

Text View shows contents of threads. Replies are indented. Text segments are marked and highlighted by medical keywords.

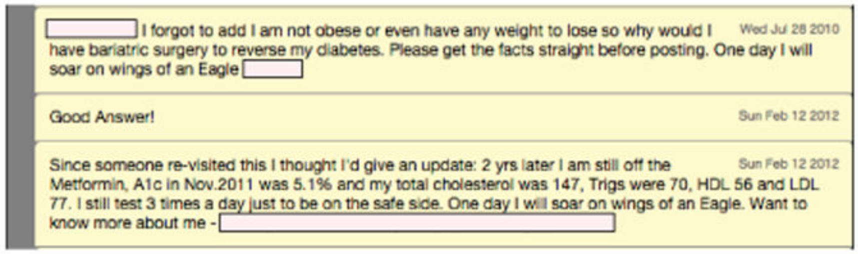


Fig. 8.

The captured text shows that a late comer left a reply 1.5 years after the last reply, which alerted the questioner to update the status.

Table 1

Design Recommendations for Visual Designs of VisOHC

Tasks	Recommendation for design
T1. Discover and compare characteristics of individual threads	
T1.1. Understand conversation patterns	Replier Mark: Make the vertical spread sufficient to show distinctive patterns Replier Circle & Beam: Avoid visual clutter by adjusting opacity Replier Timeline: Animate the transition; avoid occlusion by squeezing
T1.2. Understand contents of threads	Word Cloud: Differentiate the importance of keywords within threads and in overall threads. Text View: Enable users to search and mark interesting entities
T1.3. Understand correlations between diverse measures	Thread View: Select and filter to compare multiple threads Similarity View: Show clustering results
T2. Scale up and drill down	
T2.1. Find similar threads	Similarity View: Compute automatic suggestion for similarities Histogram View: Show correlations between measures
T2.2. Drill down to details	Thread View: Enable visual query to match conversation patterns Text View: Collapse similar text and show representative keywords
T3. Confirm findings in text	Text View: Annotate findings to keep analysis notes