VisForum: A Visual Analysis System for Exploring User Groups in Online Forums

SIWEI FU, YONG WANG, and YI YANG, Hong Kong University of Science and Technology
QINGQING BI, Nanyang Technological University
FANGZHOU GUO, Zhijiang University
HUAMIN QU, Hong Kong University of Science and Technology

3

User grouping in asynchronous online forums is a common phenomenon nowadays. People with similar backgrounds or shared interests like to get together in group discussions. As tens of thousands of archived conversational posts accumulate, challenges emerge for forum administrators and analysts to effectively explore user groups in large-volume threads and gain meaningful insights into the hierarchical discussions. Identifying and comparing groups in discussion threads are nontrivial, since the number of users and posts increases with time and noises may hamper the detection of user groups. Researchers in data mining fields have proposed a large body of algorithms to explore user grouping. However, the mining result is not intuitive to understand and difficult for users to explore the details. To address these issues, we present VisForum, a visual analytic system allowing people to interactively explore user groups in a forum. We work closely with two educators who have released courses in Massive Open Online Courses (MOOC) platforms to compile a list of design goals to guide our design. Then, we design and implement a multi-coordinated interface as well as several novel glyphs, i.e., group glyph, user glyph, and set glyph, with different granularities. Accordingly, we propose the group Detecting & Sorting Algorithm to reduce noises in a collection of posts, and employ the concept of "forum-index" for users to identify high-impact forum members. Two case studies using real-world datasets demonstrate the usefulness of the system and the effectiveness of novel glyph designs. Furthermore, we conduct an in-lab user study to present the usability of VisForum.

CCS Concepts: • **Human-centered computing → Visual analytics**;

Additional Key Words and Phrases: MOOC forum, glyph design, application

ACM Reference format:
Siwei Fu, Yong Wang, Yi Yang, Qingqing Bi, Fangzhou Guo, and Huamin Qu. 2018. VisForum: A Visual Analysis System for Exploring User Groups in Online Forums. *ACM Trans. Interact. Intell. Syst.* 8, 1, Article 3 (February 2018), 21 pages.
https://doi.org/10.1145/3162075

The reviewing of this article was managed by special issue associate editors Yu- Ru Lin and Nan Cao.
Authors' addresses: S. Fu, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Hong Kong; email: sfuaa@connect.ust.hk; Y. Wang (corresponding author), Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Hong Kong; email: ywangct@cse.ust.hk; Y. Yang, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Hong Kong; email: canna1102.yy@gmail.com; Q. Bi, Nanyang Technological University, 50 Nanyang Ave, Singapore; email: bqingqing@gmail.com; F. Guo, Zhijiang University, Hang Zhou, China; email: guofz1234@gmail.com; H. Qu, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong, Hong Kong; email: huamin@cse.ust.hk.

https://doi.org/10.1145/3162075

1 INTRODUCTION

Web 2.0 technologies boost online forums, bulletin boards, blogs, and news groups. People flood into these online platforms to seek information, share interests, and socialize with each other. Thousands of user-generated data have been archived in these platforms, which provide great opportunities for researchers to study user interactions and engagements in a computer-mediated environment.

User grouping or clustering is a common behavior in online forums. People with similar backgrounds or shared interests are more likely to gather together in small group discussions. This is not only an important sociological research topic on interpersonal relationships, but also closely related to information diffusion [38] and knowledge construction [6]. By identifying "cliques" or "sub-groups" of users, we can easily differentiate the communication styles of participants and the friendliness of a community. For example, newcomers may find it difficult to get involved in the community where discussions are raised among senior members. Moreover, the intensiveness and sequence of interactions also play important roles in interpersonal relationship building. Studies have found that discussions may terminate prematurely if participants fail to receive an immediate response from their peers [9, 23].

However, the dynamics of asynchronous forum data (AFD) and a large volume of archived posts pose great challenges for forum administrators and analysts to discover user groups and gain insights from participants' interactions. Most conversation groups in online forums are formed temporally. People can join or leave a discussion with low switch cost. Group membership changes over time, which is hard to be tracked. Furthermore, AFD is in a hierarchical tree structure. Nested post-reply relationships make group detection a difficult task, even in one thread.

Previous researchers have made efforts to visualize AFD from different perspectives, such as users' social networks or hierarchical post-reply relationships of threads, but few have focused on visualization of participants' grouping behaviors. Studies in the data mining field have developed a large body of algorithms and methods to explore this topic, but the results generated from these analyses are hard to interpret and not intuitive to laymen.

To address these issues, we designed and developed a novel system, VisForum, to visualize user groups in AFD. VisForum consists of two modules, one for data processing and the other for system interface. In the data processing module, we design the Group Sorting Algorithm to gather posts from the same group, as well as craft an adapter to pre-process data from multiple sources. Inspired by Jorge's work [18], we also propose the concept of "forum-index" of a forum member to represent his/her impact in a forum.

The visualization module has a Web-based interface, which allows interactive exploration of groups from different granularities. Following the guides of a set of well-defined design rationales and tasks, we create multi-coordinated views to assist analysts to efficiently explore groups. Specifically, we design a sortable List View to summarize all threads to facilitate quick identification of interesting ones. We create group glyph, set glyph, and user glyph to enable users to obtain group information from various levels of details. We further develop a scatter plot equipped with circular-based glyphs to visualize group evolution over time.

Our contribution can be summarized as follows:

— We design three novel glyphs, i.e., group glyph, set glyph, and user glyph, to depict multi-facets of group characteristics and inter- and intra-group relationships in lengthy hierarchical threads.
— We propose the Group Detecting & Sorting Algorithm to accelerate the analysis process and the concept of "forum-index" in the context of a forum to capture the impact of a user.

—We develop an integrated visual analytical system which supports interactive explorations of user grouping behaviors in different levels of details in an online discussion forum.

2 RELATED WORK

In this section, we review the related work that can be categorized into four types: existing visual systems designed for AFD, user grouping behavior detection, glyph design, and visualization of social interactions in social networks.

2.1 Visual Systems for AFD

Previous studies have explored AFD visualization from different perspectives. The closest piece to our approach is VISM [24], which visualizes the sequence of user interactions and subgroup formulation in a radial tree layout. By plotting users at different levels of the radial tree, the depth of the levels shows the complexity of each thread and the arrow between two nodes shows the direction of information flows. It also allows users to view the interactions based on different time intervals. As a visual interface designed for e-learning online forums, VISM uses the node colors to encode different roles of participants, such as red as teacher or moderator and blue as student. It adopts the concept of "convex-hull," using dotted outlines to surround a set of nodes within a thread to differentiate each discussion in the whole forum. Pascual et al. [30] have developed a similar interface to VISM [24], which has a concentric and nested radial-tree layout. By enabling users to drag a node to the center of the tree, it further highlights the specific debate among a group of participants. These approaches are designed to visualize a small group of users. However, they do not scale well for general online forums where thousands of users are involved.

Another approach of AFD visualization systems related to users' grouping behaviors is based on social network analysis. The commonly used visualization technique is node-link diagram. One of the examples is Conversation Map [35], which computes a set of social networks in a forum based on post-reply interactions. It uses centrality degree of users in a social network to differentiate the importance of users in a community and visualizes social and semantic networks with a node-link diagram. A recent piece of work by Hoque and Carenini [21] develops ConVis to support multi-faceted exploration of blog conversations. It adopts an "overview+detail" approach and designs *thread view*, *facet view*, and *conversation view* to provide information at different granularities. The thread view provides a visual summary of the overall discussions by using a horizontally stacked bar. The facet view and conversation view convey details about the relationship between topic and authors as well as the content of each discussion. To analyze domain-specific forums, Kwon et al. [25] presented a design study where they collaborate with experts in online health communities to investigate online health forums. Similarly, iForum was developed [14], which examined Massive Open Online Courses (MOOC) forum data to help course instructors understand learning behavior of students. Aforementioned approaches scale well to a large quantity of data. However, they focus either on general social networks or on individual users. None of them focus on user groups.

In retrospect, existing visual analytics of users' grouping behaviors in online forums are still limited and new designs and techniques are needed to address this issue.

2.2 User Grouping Behavior Detection

Detection of user grouping behaviors has been well studied in data mining fields. We can simply calculate users' communication frequency and number of common friends as indicators of user similarities or conceptualize the problem into well-established concepts in graph studies, such as co-occurrence, clustering, and dense subgraphs.

Co-occurrence is a commonly used concept in text mining to study the frequent occurrence of two terms from a text corpus alongside each other in a certain order. In social network analysis, the hypergraph-based method [48] and machine learning algorithm [27] have been designed to find co-occurrence correlations among entities. Furthermore, well-established algorithms, such as IsoMap [40], Locally Linear Embedding (LLE) [34], spectral clustering [29], Multidimensional scaling (MDS) [22], CODE [16], and D-CODE [36] have been developed for analysis of dynamic co-occurrence data.

Clustering and *dense subgraphs* are two other thoroughly-studied concepts about social groups in the theory of graphs and networks. In standard graph theory, a cluster is defined as a maximal complete subgraph of a graph *G*. Previous research in data mining fields has developed a large number of algorithms to identify tightly-connected clusters such as in the following: [13, 15, 20, 28, 38]. Dense subgraphs, on the other hand, have been defined as *cliques*, where every vertex connects to every other vertex, or some relaxation of the ideal clique [26]. Lee et al. [26] differentiates clustering and dense subgraphs, and concludes that "not all clustering algorithms are based on density, and not all types of dense components can be discovered with clustering algorithms." He further mentions that clustering methods are not suitable for absolute density criterion and cannot handle overlapping clusters.

Based on a thorough review of the above concepts, we decide to adopt the concept of dense components to define user groups and exploit exact enumeration algorithms to calculate dense graphs.

2.3 Glyph-based Visualization Design

Glyphs can convey a large amount of information in a single shape or symbol [44]. According to Borgo et al. [5], *glyph* is defined as "a small visual object that can be used independently and constructively to depict attributes of a data record or the composition of a set of data records." Glyph has been widely used in the visualization of various data. For example, Dunne and Shneiderman [12] applied glyphs in the simplification of large graph visualization and used glyphs to improve the graph readability. The recent work by Duffy et al. [11] introduced a glyph-based visualization for semen analysis and encoded 20 variables through different visual channels of the glyphs. Previous research holds a large body of works on glyph-based visualization. Many surveys such as those by Borgo et al. [5], Ropinski et al. [33], and Ward [44, 45] have attempted to summarize the latest state-of-the-art work on this topic. Ward [44] classifies the techniques of glyph positioning into data-driven and structure-driven approaches and presents a taxonomy of glyph placement strategies. In another work, Ward [45] proposes a technical framework of design principles and practices for multivariate data glyphs. They also elaborate methods and biases of visual mapping. The authors in Rapinski et al. [33] discuss the glyph-based visual techniques exploited for spatial multivariate medical data. The work by Borgo et al. [5] further clarifies the concept of glyph and describes a thorough glyph design space, guidelines and techniques, as well as applications in different contexts.

To sum up, a well-designed glyph can stimulate the human perceptual system and facilitate efficient information processing. Therefore, we also utilize glyph-based visualization to present synchronous forum data, which is multi-dimensional in nature. Many generic glyphs also exist, such as face glyphs, box glyphs, and star glyphs, but the information they convey is always limited to the data and design. Therefore, in the context of AFD, we carefully designed glyphs to encode multivariate user profile and group information.

2.4 Interaction Visualization in Social Networks

Visual analytics approaches have been extensively applied in the analysis of social interactions between users in social networks. Prior work mainly focuses on either analyzing the interaction structures between users and their temporal evolution from a local perspective or exploring the information diffusion in social networks. For example, Wu et al. proposed egoSlider [47], a visualization system to support evolutionary analysis of egocentric networks at three different levels. Similarly, EgoLines [49] used a "subway map" metaphor to support the egocentric analysis of dynamic networks, where the social interactions between the ego node and its surrounding nodes can be conveniently explored. Shi et al. [37] presented a 1.5D visualization design to display and analyze dynamic ego-networks, encoding the temporal information through positioning the non-focus nodes along a trend glyph. Cao et al. [7] visualized the egocentric social interactions of both interaction initiators and responders. Apart from an egocentric perspective, some research also studied the community consisting of users with more frequent interactions between them and its temporal development. Reda et al. [32] used X-axis to represent time and Y-axis to position individuals in the appropriate community threads, visualizing the temporal evolution details of communities. A similar approach is also presented in Vehlow et al. [41], where the flow metaphor and node-link diagram are combined to show the community structure and its evolution. Other researchers paid more attention to the interaction result in social networks, i.e., the information diffusion between users. For instance, Whisper [8] and D-Map [10] support the exploration of information propagation among individual users. Wang et al. introduced a visualization system to analyze how information flows across multiple social groups [43].

Different from prior work, this article focuses on exploring the user interactions and the corresponding user groups in the replying threads of online forums.

3 GROUP DEFINITION AND ALGORITHM

A thread consists of a root post and multiple replies, which are archived in sequence by published time and nest in a hierarchical structure. Before giving the definition of group, we first introduce several concepts that will be referred to in later discussions.

Following Wang's definition [42], let $T = \{t_0, t_1, \ldots, t_{n-1}\}$ be a set of threads in a particular asynchronous online forum. Each thread t_i initializes from a ***root post*** p_0, and consists of m posts $\{p_0, p_1, \ldots, p_{m-1}\}$ in total including the root post. Each post p_i has a unique *post_id* and is created by a user u_i with a unique *user_id*. The post p_i is said to be the ***parent post*** of p_j if and only if p_j is an immediate follow-up post of p_i. Each path originating from the root post to the leaf post forms a ***reply sequence*** s_i (see Figure 1). In one thread, although all posts correspond to the same topic, each reply sequence can be taken as a self-consistent sub-threaded discussion [42]. Therefore, we define conversation group within the scope of a reply sequence rather than the whole thread. A reply sequence can be formed as a directed graph G if we represent each user as a node. Below, we define conversation group in two scenarios.

- **Two-member Group**
  Two forum members, $user_a$ and $user_b$ $(a \neq b)$, form a conversation group if and only if in one reply sequence, $user_a$ and $user_b$ have replied to each other at least once.
- **N-member Group ($N \geq 2$)**
  If more than two forum members discuss in a reply sequence, the ideal scenario is that each two members have replied to each other. However, a maximally complete graph (or subgraph) restricts indirect connections among users. To avoid this downside of cliques, we take the form of relaxations of the pure clique measure.

  Given a reply sequence represented as a directed graph $G(V, E)$, we construct a corresponding undirected graph $G'(V, E')$. If $|E(\{u, v\})| \geq 2$, where $\{u, v\} \in V$, we add an edge connecting u, v in G'. We iterate all node pairs in G until all potential edges are added to G'.

  Inspired by Lee et al. [26], we define the group(s) in G' as the maximal degree-based ***quasi-cliques*** with size at least two and degree of each vertex at least $\gamma(|V| - 1)$. We set γ

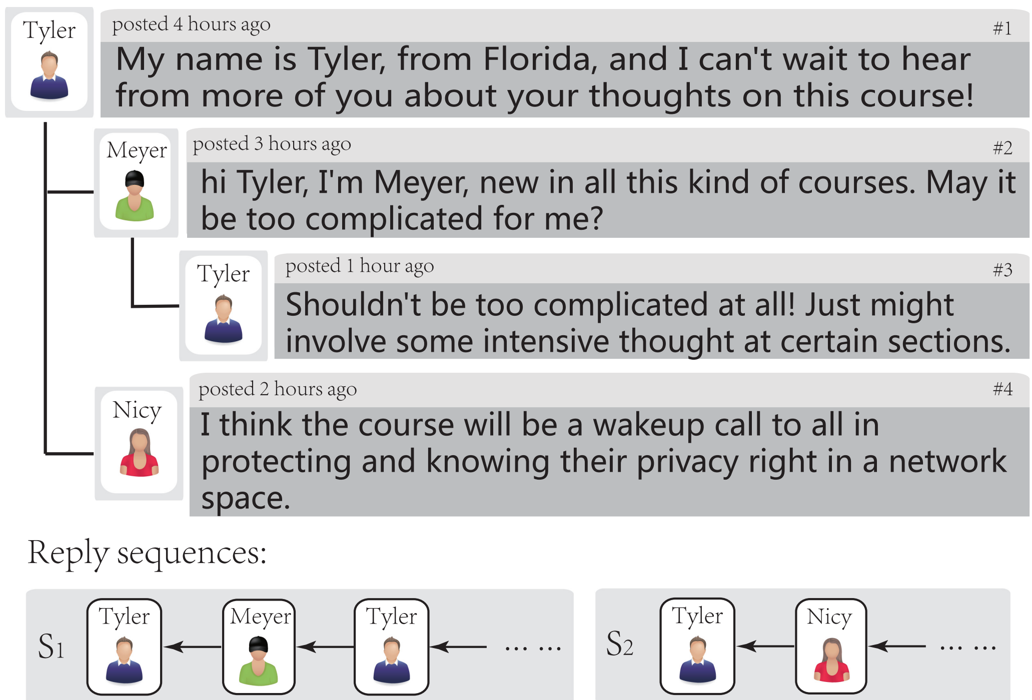

Fig. 1. A real thread sequence depicted as tree structure due to reply relations. Two reply sequences exist in this thread: one is Tyler ← Meyer ← Tyler, and the other is Tyler ← Nicy.

as 0.5, which means all group members have conversations with at least half of the members in the group. Our system keeps the flexibility to adjust γ based on different group definitions. For example, if a user group is defined as a maximally complete graph, γ should be set to 1.

Based on the above definition, the task of group detection is equivalent to identifying all dense components in a reply sequence.

After group detection, a reply sequence is transformed to a set of groups and a number of posts that do not belong to any groups, such as RED in Figure 2. Then, we sort groups and non-group posts based on published time. In particular, we consider average published time of all posts as the published time of a group. To order reply sequences, we apply similar strategies. That is, we calculate average published time of all posts in a reply sequence and sort reply sequences based on this value. The output of the algorithm is a sequence of groups containing posts with reply relationships and non-group posts.

The complexity of our algorithm is $O(m \times n^2)$, where m is the number of reply sequences in a thread and n is the maximum number of posts in a reply sequence. Although the group detection and reordering algorithm is efficient and able to preserve temporal information of user groups, we notice some limitations. First, as we focus on user groups rather than individual posts, the temporal and structural information of original posts is lost after reordering. Second, the thread root is repeated in each reply sequence, which may cause redundancy issues. As described in Section 4.3, this issue is alleviated after applying space-efficient visual representation for user groups.

Detailed description of the Group Detection & Sorting Algorithm are shown in Algorithm 1. An intuitive demonstration of the algorithm is depicted in Figure 2.

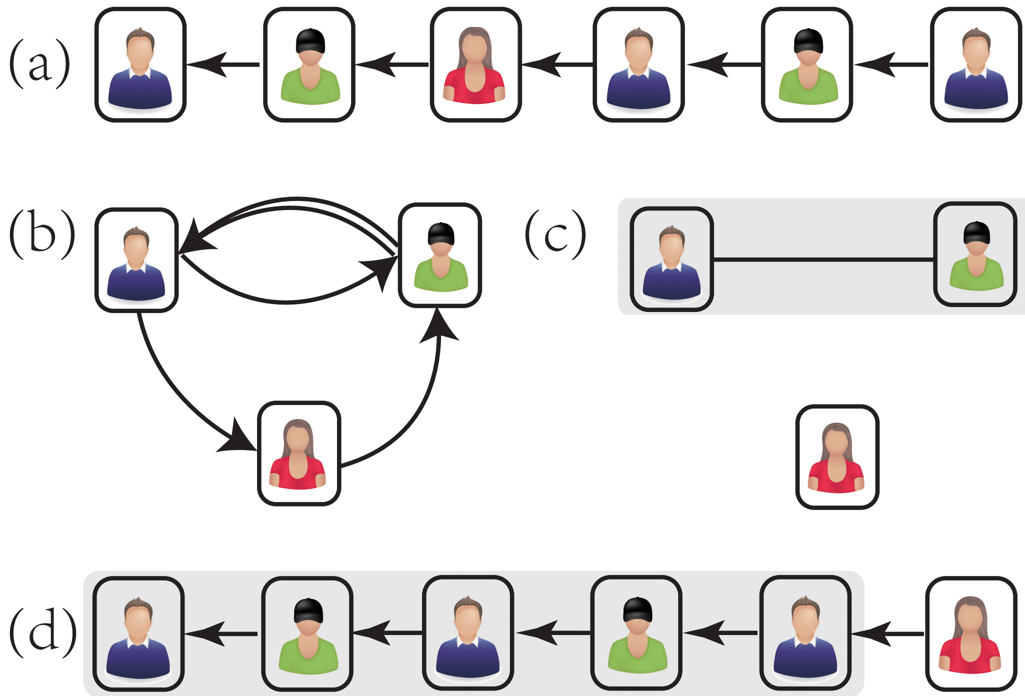

Fig. 2. (a) The original reply sequence. (b) We first transform the reply sequence to a directed graph G. (c) Then, we convert G to an undirected graph G' and detect dense components in G' (BLUE and GREEN). (d) Finally, we sort a sequence of groups and non-group posts according to temporal information.

ALGORITHM 1:

Input: A sequence of posts $P = \{p_0, p_1, p_2, \ldots, p_n\}$
Output: A reordered sequence *out*
Get reply sequences $S = \{s_0, s_1, \ldots, s_n\}$ from P
for *each* s_i *in* S **do**
    Transform s_i to a directed graph G
    Convert G to an undirected graph G'
    Detect dense components in G'
    Sort groups and non-group posts in s_i'
end
return *out*

4 SYSTEM DESIGN

This section includes three parts: descriptions of design rationales (4.1), an overview of system architecture (4.2), and detailed explanations of each visual design (4.3).

4.1 Design Goals

This project originated from domain experts' requirements. We work closely with two educators, who have released courses on Coursera and Edx. They are not co-authors of this article, but they were both interested in students' grouping behaviors in MOOC forums. Based on several sessions of interviews with these domain experts as well as literature reviews from related fields, we compiled a list of design goals to guide our visual design:

(1) *Exploring conversation groups from multiple facets.* Thread-by-thread is a traditional method of viewing AFD. However, both domain experts proposed that it was inefficient to switch between threads to check the posts by a specific conversation group. Additionally, it is not an easy task to explore a particular group within a lengthy thread when noises break the continuity of a discussion. Thus, they suggested that exploration of conversation groups from multiple facets should be supported in the visual system. Specifically, they required that the system could be able to explore groups when they raised interests in particular users as well as within a particular thread.
(2) *Unfolding the dynamics of the forum.* Forums evolve over time. However, exploring groups during specific time periods is a tedious work for the current forum system. Both domain experts suggested that an overview of the dynamics of a forum is essential for them to begin their exploration and analysis.
(3) *Identifying interesting threads with efficiency.* Selecting interesting threads from thousands of threads is a difficult task. Traditionally, users have to go through the contents of a thread to check the topic, length, sentiment, and participants, which is lack of efficiency. Therefore, the system should provide summarized information of threads and support sorting and filtering functionalities.
(4) *Indicating details of groups.* A group is a multi-dimensional entity. Group size, activeness, and sentiment are all critical factors for the differentiation of groups. Also, each individual has multiple attributes, such as post distributions and interaction patterns. In the interview, experts were especially interested in identifying "influential" forum members who can be assigned as potential "Community Teaching Assistants." The system should allow experts to easily explore details of each group with a succinct presentation.
(5) *Comparing groups locally and globally.* Most conversation groups in online forums are formed temporarily. People engage in a conversation to agree with or argue against a specific topic. On one hand, multiple groups may share the same forum members in one thread. On the other hand, seeing the forum in a bird's-eye view, group evolution allows us to view the development of interpersonal relationships and the changes of individuals' involvement in the community. To gain insights into the relationship among groups within one thread and across multiple threads, the system should support the comparison of groups both locally and globally.
(6) *Integrating semantic and sentimental analysis.* Texts convey meanings as well as emotions. During the interviews, both experts mentioned they were interested in the textual contents of a forum and the discussion atmosphere. They'd like to explore raw discussions and members' attitudes toward a topic or a person. Thus, our system was required to integrate semantic and sentimental analysis of conversations.

4.2 System Overview

Figure 4 demonstrates the system architecture of VisForum. It consists of two main components: a time-efficient data processing module and a user-friendly interface. The system workflow starts from the data processing component where we build adapters for various forum data to clean and format the data into predefined structures. The data is then stored in the MongoDB[1] and indexed for efficient queries. Furthermore, we execute Group Detecting & Sorting Algorithm (Algorithm 1)

[1] https://www.mongodb.org/.

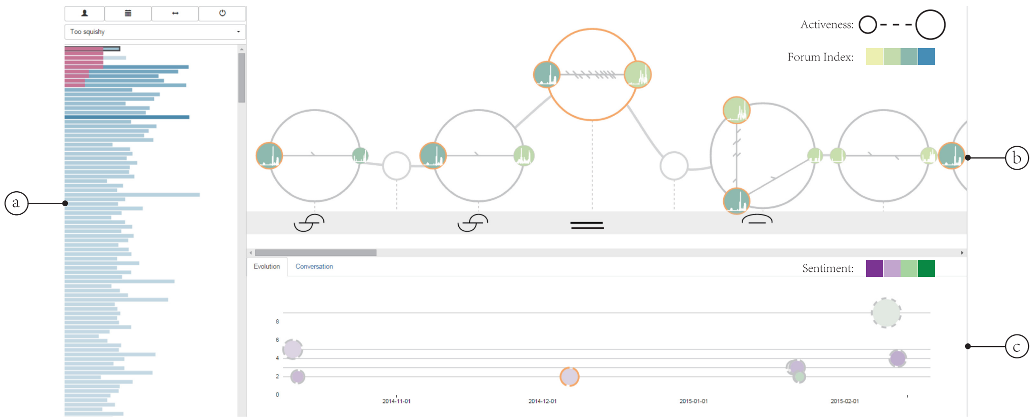

Fig. 3. VisForum is a multiple-view visual analysis system intended for exploring, comparing, and tracking conversation groups in asynchronous online forums. The picture shows the interface of VisForum with real forum dataset, The MOOC forum dataset, plugged in (a) is List View, (b) is Group Sequence View, and (c) is Group Evolution View.

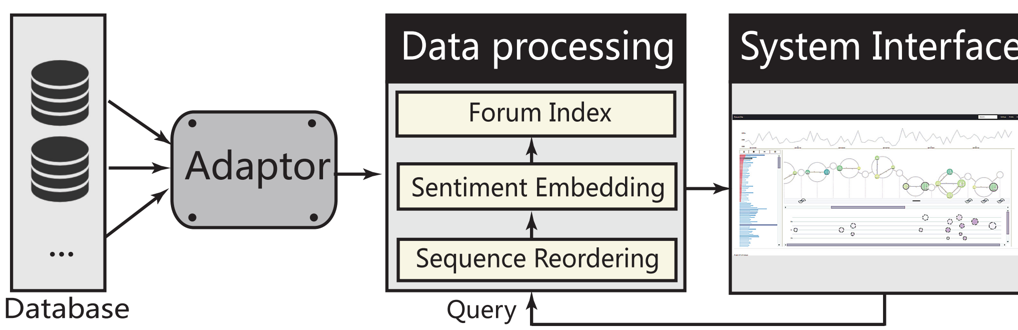

Fig. 4. Workflow of VisForum system architecture. The adapter collects various forum data from the database and unifies them into the predefined structure. Then, all data is passed to the data processing module, which interacts with the system interface.

for each thread to cluster posts of the same group, and conduct sentiment analysis for each group in the thread in real time. The system interface consisting of four coordinated views is closely coupled with the data processing module. The interface will be presented in details in the next section.

4.3 System Interface

Following the design rationales and multi-coordinated view design principles proposed by Baldonado et al. [3], we designed and implemented the VisForum interface with four coordinated views (1): a Time Variant View, a List View (see Figure 3(a)), a Group Sequence View (see Figure 3(b)), and a Group Evolution View (see Figure 3(c)). The Group Evolution View and the Group Sequence View are core analytical modules. The Time Variant View allows analysts to examine the forum at any time period. The List View is designed to assist users' exploration of grouping behavior from threads perspective.

4.3.1 Group Sequence View. The Group Sequence View is composed of two parts. One is a sequential demonstration of groups in one thread Design goal (4), and the other is about the set relationship comparison between two groups Design goal (5).

Table 1. A Summary of the Advantages and Drawbacks of the Edge Glyph

Advantages	Drawbacks
1. Present bi-directional relationships in a space-efficient manner.	1. Temporal information is hard to compare among different edge glyphs.
2. Display the number of replies explicitly.	2. Visual clutter may emerge when edge glyphs cross together.
3. Show distribution of the communication.	3. The order of replies is not clear.

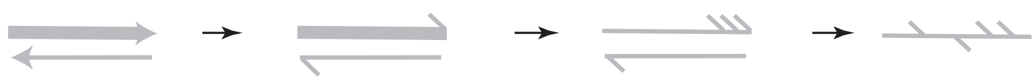

Fig. 5. This picture demonstrates the way of designing edge glyphs. (a) Original two directed lines; (b) halve the arrow; (c) encode the number of replies to the arrow number instead of line width; (d) combine two lines and distribute the arrows due to post time.

We design *group glyph, user glyph, and set glyph* in the Group Sequence View to allow analysts to dive into details of each group Design goal (4) and relations among groups Design goal (5).

Group Glyph

Regular expression is effective in matching long strings with specific patterns by which long strings can be abbreviated to a concise form. Inspired by this idea, we create glyphs as patterns to represent a group of posts in a short way. By using group glyph, we show a long sequence of posts (Figure 2(d)) in a space-efficient presentation. Meanwhile, for consecutive non-group posts, they are regarded as noises and are simplified by using a blank circle.

Group glyph is designed to address users' requests on group details (4). A group glyph (see Figure 6(b)) has an outer circle with user glyphs and inner bumped edges. Group size is depicted by the number of user glyphs and group. The vertical positions of group glyphs represent the relative frequencies of group occurrence. The higher a group locates, the more frequently it appears in the forum.

Before coming up with the bumped edges, we used two directed lines with line width mapping reply frequencies (see Figure 6(c)). However, after importing different datasets into our system, we found this design had a severe visual clutter problem. To be precise, for a high-degree user in the group glyph, several in-degree arrows and out-degree arrows may cluster together. To overcome this shortcoming, we combine the lines and place the half arrows from each line on both sides of the edge (Figure 5). Compared with two directed lines, bumped edges alleviate visual clutter near the user glyph while preserving reply information between two users.

The number of bumps represents the quantity of replies, and the position of bumps indicates the time of the communications. From the edge glyph, we are able to answer questions like, "How many replies from one user to another?" and "What is the distribution of the communication between two users?" However, the temporal information of edge bumps is hard to perceive and compare because edge glyphs between different user glyphs are different at scale. Further, when two edge glyphs cross together, bumps may cause visual clutters if they are posited close to the crossing. A discussion of advantages and drawbacks of edge glyphs are presented in Table 1.

User Glyph

User glyph is designed as a filled circle with spark line inside (see Figure 6(a)). Circle radius is mapped to the number of posts created by a user, which reflects his or her activeness in the whole

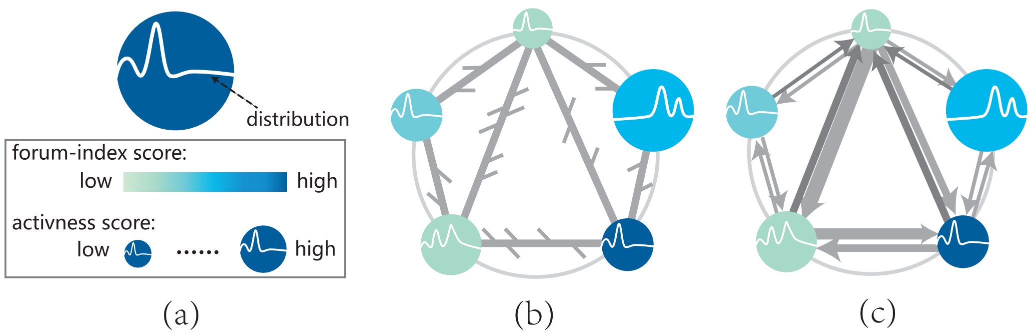

Fig. 6. (a) Demonstrates that the user glyph presents three key attributes: post distribution, activeness, and Forum-index score; (b) shows the group glyph equipped with bumped edges while (c) displays an alternative design.

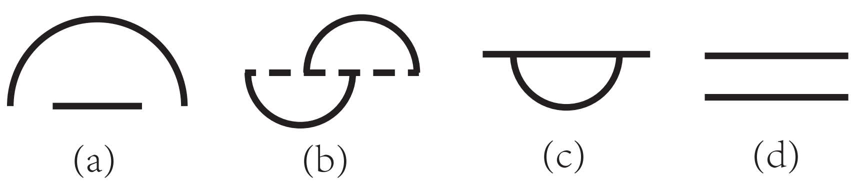

Fig. 7. Four glyphs demonstrate four set relations: (a) the compared group is the **superset** of the baseline group; (b) two groups **have common forum members** but one does not contain the other; (c) the compared group is the **subset** of the baseline group; (d) two groups are **equal** in terms of forum members.

forum. The bigger the radius is, the more active the user is. The spark line inside each circle presents the user's post distribution starting from the registration date until now, which shows the activeness of a user over time. The background color of the glyph is mapped to the "social influence" of the user. We use a set of sequential colors chosen from ColorBrewer [17] and highlight influential users by darker background color. From visual cues, experts are able to identify the groups containing "influential and active" users easily.

"Social influence" of a member in the forum indicates the ability to attract other members into discussions. Inspired by Jorge's work [18] in H-index, which aims to measure both productivity and citation impact of the published body of work of a scientist or scholar, we proposed the concept of "*Forum-index*" to reflect his/her "impact" in a forum.

Forum-index Definition

A member has index x if x of N threads he/she participated in has at least x posts replying to him/her, and the other $(N - x)$ threads have no more than x replies to him/her.

A member's "*Forum-index*" captures not only his/her activeness in discussions, but also the frequency of interactive replies with other members. Hence, this helps us identify influential forum members.

Set Glyph

In order to compare groups by group members within a thread Design goal (5), we develop four intuitive set glyphs. Figure 7 demonstrates four glyphs representing four types of set relationships: superset, subset, intersection, and equal. The superset glyph originates from the cap metaphor, which indicates that the compared set covers the baseline set. On the contrary, the subset glyph is the inverse of a cap. Intersection is presented as two caps overlapping with each other while the equal relation is presented as the equal notation.

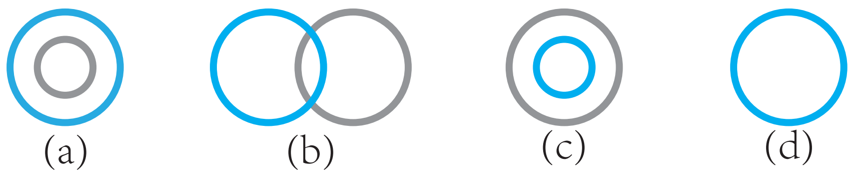

Fig. 8. The design of set glyphs building on Venn diagrams. The baseline group is in gray while the compared group is in blue. (a) The compared group is the **superset** of the baseline group; (b) two groups **intersect** together; (c) the compared group is the **subset** of the baseline group; (d) two groups are **equal** in terms of forum members.

When a group is selected in the Group Sequence View, this group is considered as the baseline group. All groups are compared with this group in terms of set relations. Set glyphs are drawn below each group glyph. However, if one group does not have common group members with the baseline group, we do not draw any glyphs below that group.

During the study, we first proposed set glyphs based on Venn diagrams, which is a common approach to present set relationships. As shown in Figure 8, we use color channel to distinguish the baseline group and the compared group. Compared with a previous design (Figure 7), the additional visual encoding requires more mental effort for end users. In particular, Figure 8(a) and (c) are opposite in meanings but the same in topology, which is not intuitive and misleading. Finally, we proposed the design shown in Figure 7. Although the design requires learning, the learning curve is shallow and users are able to distinguish each glyph by topology only without remembering visual encodings. We plan to conduct an in-depth user study to further confirm these observations in future research.

4.3.2 Group Evolution View. The evolution of one group means that other members join or leave the group, which is denoted as g_i. The Group Evolution View is designed to demonstrate the changes of a specific group across time (5). This view is displayed after a conversation group is selected in the Group Sequence View.

The Group Evolution View is presented as a scatter plot equipped with a simplified group glyph depicting each point. Given a selected group g_i, the x-axis is the timeline ranging from the earliest occurrence time of g_j, or its evolution, to the latest occurrence time of g_j, where g_j means g_i and its evolution. The y-axis depicts the number of participants in the group. Groups with more participants are positioned higher vertically. We select a number of critical features for a conversation group, and design a circular-based glyph to encode these features. More specifically, as shown in Figure 9, the radius of the segmented circle encodes the number of posts that the group created. A larger circle means that the group has more heated discussions. Moreover, we encode the number of group members again using the segment number in each circle because in the interviews with domain experts, they suggested to map the number of participants directly to each circle so that they do not need to read the value in the y-axis when examining a change of the group.

We also use the Natural Language Processing APIs[2] and NLTK [4] to obtain the semantics of each conversation group and map the results to the node background color Design goal (6). We choose two diverging colors from ColorBrewer [17] to encode the distribution of semantics ranging from negative (purple) to positive (brown). Neutral is presented as white.

The Group Evolution View works accordingly with the List View and the Group Sequence View Design goal (1). When a group g_i is selected in the Group Evolution View, threads containing g_i

[2]Natural Language Processing APIs: http://text-processing.com/.

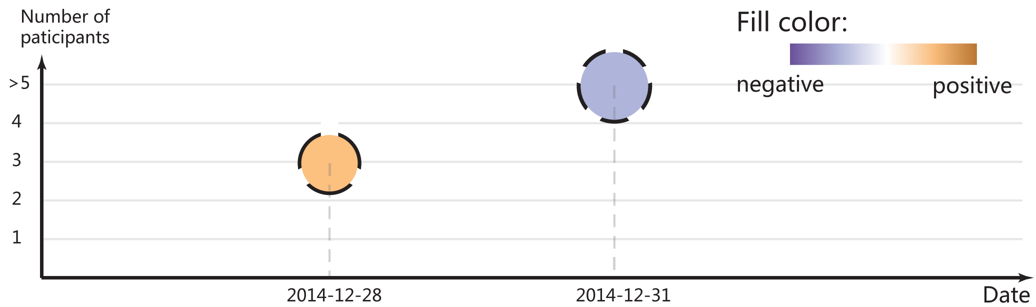

Fig. 9. This figure demonstrates the Group Evolution View. Each segmented circle represents a group. The sentiment of the group discussion is presented using color.

and its evolution are highlighted in the List View. Also, threads containing g_i will be displayed in the Group Sequence View. From the iteration of exploring thread→ group→ thread, users are able to go through multiple heated group discussions or explore the evolution of one group.

4.3.3 Other Coordinated Views. For the integrity of the system, the interface also contains the Time Variant View and the List View (Figure 3(a)). The Time Variant View displays the dynamic activeness of the forum using a line chart (2). It also acts as a filter, which means by brushing a date range, other views change accordingly. The List View is designed to summarize and filter threads effectively based on various criteria(3).

5 CASE STUDIES

We demonstrate the usability of our system with two distinct AFD datasets, one is a course forum released in the MOOC platform Edx[3] and the other is web-scraped data from The Lord of the Rings Online^{TM} forum (we use “Lotro” as abbreviation). Data from Edx contains 13,289 posts distributed in 1,978 threads. There are 2,712 forum members active in this three-month course. Meanwhile, Lotro has 98,113 posts created by 6,117 users during 2013-10-16 and 2015-03-14.

5.1 Case 1: Course Discussion Forum in MOOC

We first invited an instructor who had released a course in Edx to use our system. His course started at 2014-06-18 and ended at 2014-09-08 with 2,712 registered students. After loading the data into VisForum, the Time Variant View first caught the eyes of the instructor, as shown in Figure 10. He noticed that at the end of the course, group discussions disappeared in about three days and became extremely heated after that. This reminded the expert that the forum had been shut down during the final exam and reopened three days later. Students were eager to know of the exam. Therefore, they flooded into the forum and engaged in conversations frequently when the forum was reopened. An expert brushed the “peak” of September in Time Variant View and picked up one of the longest threads from the List View Design goal (3). The Group Sequence View immediately updated, which allowed him to choose a group with two engagements, among which one of them was a highly influential user Design goal (4). He checked discussions in conversation view after finding that the conversation was negative in the Group Evolution View Design goal (6). The posts, such as “I’m glad I wasn’t the only one that thought the multiple choice questions were the most difficult part of the exam. (*Posted on 2014-09-02 by user 1*)” and “Will the answers to the final exam questions be released? (*Posted on 2014-09-02 by user 2*)” confirmed his assumption and demonstrated the efficiency of our system. However, after checking more negative posts, the

[3]Edx: www.edx.org, one of the most popular MOOC platforms.

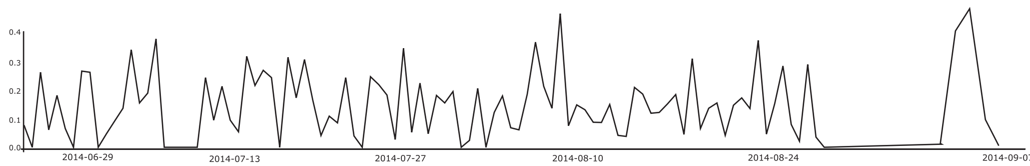

Fig. 10. This figure demonstrates the Time Variant View after the MOOC forum data is loaded.

instructor pointed out that a number of problem discussions are marked as negative by mistake. Due to the complexity and variation of human languages, it is challenging to identify sentiment of sentences according to semantic meanings. We plan to resolve this issue when more advanced analytical techniques are developed.

Then the expert decided to check the performance of the "Community Teaching Assistant (CTA)." CTAs are forum moderators who help instructors answer questions related to course structure, class schedule, and teaching materials. Response time and reply contents of CTAs are critical for community building and also to influence students' assessment of the course. Traditionally, it is difficult to evaluate the CTAs' performance in online forums because of two reasons: (1) Identifying all conversations in which CTAs are involved is time consuming. (2) Tracking conversations is a tedious task in lengthy threads. By using VisForum, the instructor selected a specific group with active influential users in it (4), as shown in Figure 3(b). He confirmed that one of the users in the group is a CTA by user name. He clicked the group and observed that this CTA has joined in other groups in this thread from set glyphs (5), as shown in Figure 3(a). He greatly appreciated the set glyph design because it clearly showed the interactive patterns of the CTA with other students. Furthermore, post distribution in user glyph demonstrated that this CTA had been actively discussing with other students (Figure 3(b)). He also found the highlighting and color-encoded sentiment in the Group Evolution View (Figure 3(c)) useful because they saved the instructor's time to check TAs' occurrence in different groups and their communication styles.

Furthermore, instructors pointed out that VisForum is useful in discovering possible CTAs for their newly released courses. CTAs are not pre-assigned when the course starts; some students become CTAs if they are active and ready to help others. Possible CTAs are distinguished from other students by node size and color in our system. When a new course has been released for two weeks, instructors can detect the possible CTAs by two-week discussions using VisForum.

5.2 Case 2: Comparison with Other Visual Representations

Node-link diagram is one of the most intuitive design choices to depict social networks. In the early prototype, we have adopted node-link diagrams to show the interactions between forum users in a thread. Figure 12 displays the replying relationships among members in the Lotro forum, where the forum members are presented as small circles and force-directed layout is used to draw the graph. The circle radius encodes member activeness and the color indicates the impact of that forum member. The link width represents the mutual reply frequency between two forum members. Node-link diagram is able to present the overall interactions between members of a long thread. For example, from Figure 12, we can easily observe that a high impact node *A* has closely interacted with *B*. Some prior techniques including edge bundling [2, 19], node clustering [1, 46], and sampling [31] can be used to reduce the visual clutter of node-link diagram and further increase the readability. However, it cannot demonstrate the detailed reply sequence in a forum thread. For example, three inter-connected nodes in the diagram do not mean they are in the same reply sequence.

Our Group Sequence View, as shown in Figure 11, visualizes the same thread as Figure 12. Since the view is too long to fit in one screen, we only demonstrate partial glyphs in the view. From part

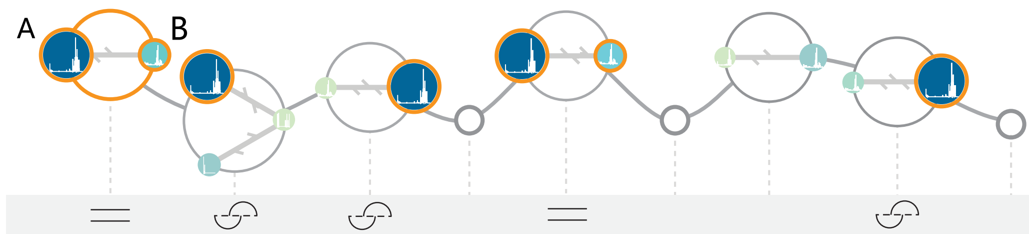

Fig. 11. This figure demonstrates the Group Sequence View for one thread. Selected group is highlighted in orange. Set relations among this group to all others are presented using set glyphs below.

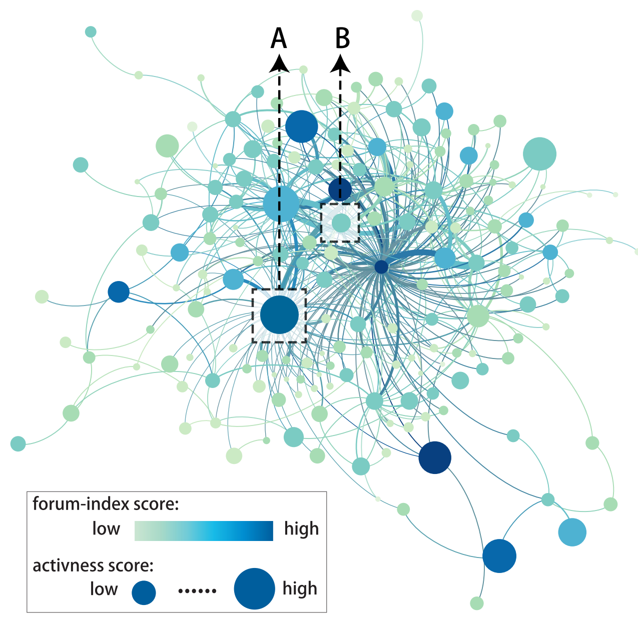

Fig. 12. The node-link diagram depicts the connections among members of one thread due to the reply relations. The two highlighted nodes, i.e., A and B, closely interact with each other.

of the thread, we can see that the user *A* has conversations with user *B* two times, as indicated by set glyphs. The two conversations appear in different parts of the thread. Also, from the y-position of this group, we can learn that these two members have joined in the same conversation groups multiple times. From the bumped edge in the group glyph we can learn the reply circles between these two users in detail (4).

In conclusion, the Group Sequence View reveals more detailed relations between the two members. However, it fails to demonstrate aggregated interactions between any two members in one thread.

6 USER STUDY

In this section, we focus on systematically evaluating the usability and effectiveness of VisForum. We first introduce the study setup, then analyze the quantitative results of the study. Finally, we summarize and discuss general feedback from participants.

6.1 Participants and Apparatus

We recruited 10 participants (age range of 23 to 30) for the study. Eight out of ten are students with a background in computer science. Most of them have basic understanding of visualization by using tools such as Microsoft Excel. Notably, two of them have experience in visual analytics. All participants had explored online forums of various kinds. Prior to the study, we checked the participants for color blindness. The experiments were conducted on a MacBook Pro with 2.5GHz Intel Core i7, equipped with 16GB of RAM.

6.2 Procedure

After obtaining the consents from participants, we first introduced VisForum with detailed information on the research context, design purpose, system architecture, and functionalities. To help participants familiarize with the tasks and visual designs, we provided a training session and presented sample questions covering major visual designs and tasks. In the end of the training session, a set of tests were used to confirm that users can understand and recognize the visual encodings correctly. Overall, the introduction and training session took about 20 minutes per subject.

The study was performed using the "think-aloud" method, which encouraged users to say whatever they were looking at, thinking, doing, and feeling when performing tasks. We also marked down the difficulties they had encountered and explained the specifications to their concerns. The completion times for each task were recorded to approximate the potential for improvements in the interface design.

At the end of the study, we gave each participant a questionnaire with six questions, which evaluated their perceived usability of VisForum on a 7-point Likert scale ranging from (1) totally disagree to (7) totally agree. The measurement items covered topics such as helpfulness, intuitiveness, and usefulness of each view. We also interviewed these participants after the experiment with open questions to collect their suggestions for improvement.

6.3 Data

To avoid the learning effect, the data used in the training session is different from the data for the real user study. In the training session, we guide the participants to explore VisForum using MOOC forum data, the same dataset used in Section 5.1. While in the real user study, we employ the "Lotro" dataset, which is the same as that in Section 5.2.

6.4 Tasks

We designed six tasks that the participants had to perform using VisForum. The tasks cover important features of the system, including sorting threads, identifying groups, viewing group and user details, comparing group relationships, and monitoring group evolution. Detailed task descriptions can be found in Table 2. The order of these tasks are designed to follow the well-known information-seeking mantra [39], where the tasks focusing on the overall understanding of the forum data are put before the tasks related to detailed exploration.

6.5 Results and Discussions

In general, all participants were able to perform the tasks and achieve the goals within a short period of time. The completion time of all tasks is shown in Figure 13.

After sorting the six box and whisker diagrams by median number, the first two box and whisker diagrams have the shortest median completion time. That is, half users are able to finish T3 and T6 in a short period of time. That may be because users are able to finish T3 and T6 with one visual channel and little interaction. The long task completion time for T1, however, is unexpected

Table 2. User Tasks for the User Study

ID	View to Test	Tasks
T1	List View	Find threads in the List View with the most posts, earliest creation time, and most user groups.
T2	Thread View	Identify the user group that occurs most frequently within the selected tread of T1.
T3	Group Sequence View - Group Glyph	Find the most active user in the selected user group of T2.
T4	Group Sequence View - Edge Glyph	Describe the interaction patterns of the user you found in T3.
T5	Group Sequence View - Set Glyph	Explore the relationship between the selected user group and other groups in the same thread.
T6	Group Evolution View	In the group evolution view, select one glyph with the maximum number of group members.

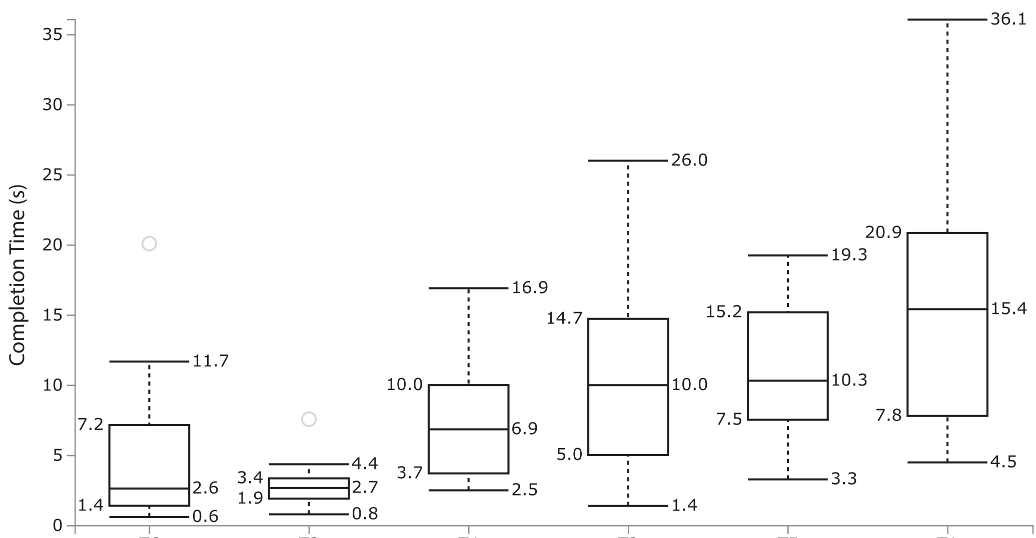

Fig. 13. The statistics of completion time of different tasks. Six box and whisker diagrams are sorted by median number.

because the average Likert score of T1 is the highest, as shown in Figure 14, which means most users agree that the List View can support T1 well. The reasons of such mismatch may be twofold. First, this task requires more interactions than other tasks. Second, some users complain that they are not able to recall the sorting strategy by using icons on the buttons. They suggest we add labels on the buttons.

Data analysis of the questionnaire also shows positive results of the system's usability (Figure 14). The average ratings of each functionality exceed five points, which means participants

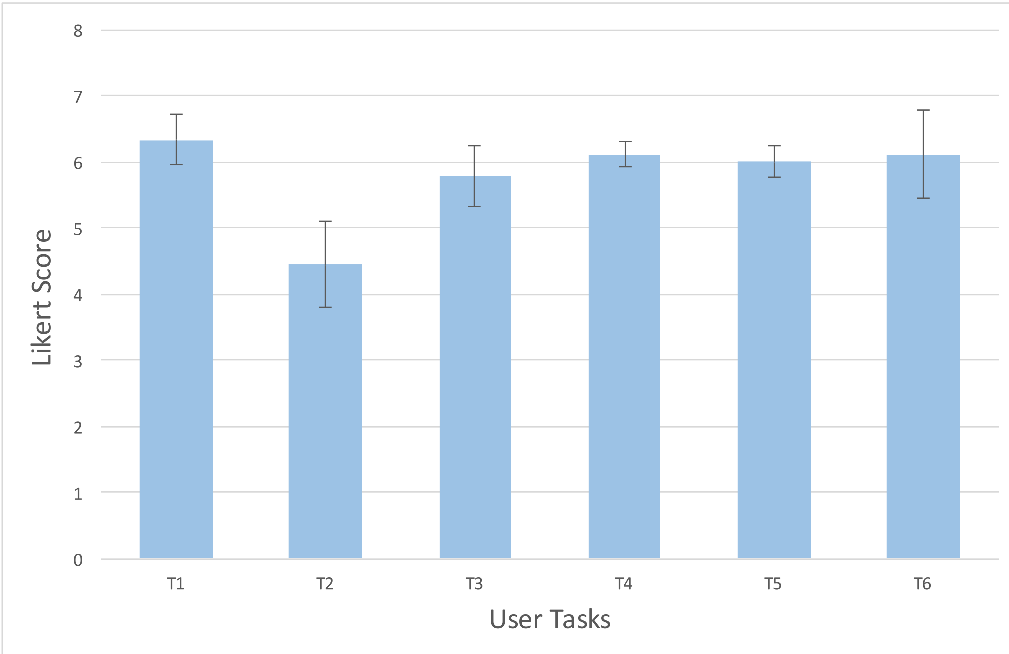

Fig. 14. Results of 7-point Likert scale of six user tasks.

perceive that these techniques are clear, easily understood, and helpful. However, we note that our system does not support T2 well, as indicated by the lowest average Likert score for T2. When identifying user groups with the most occurrence, users have to scroll left and right to find the one with the highest value on the Y-axis, which is attention- and time-consuming. One participant comments, "*It is not easy to compare the height of group glyphs, which are not shown in the same screen space (given a lengthy thread). Maybe displaying the value of Y-axis for each group glyph is helpful.*" In future research, we plan to design a scalable overview of the Thread View so that users are able to compare all group glyphs at first glance. And VisForum should allow users to dig into individual group glyphs by zooming and panning.

6.6 General Feedback

All participants showed great interest in using VisForum, and provided positive feedbacks during the interview sessions in three aspects. First, the topic is interesting and our system is able to meet various requirements of group identification and exploration. Second, various sorting strategies in the List View help them find threads of interest. Third, the edge glyph is intuitive and informative, and it helps participants understand and explore reply interactions between users.

Further, our participants provided valuable feedback to further improve the usability and usefulness of VisForum. First, 6 out of 10 participants reported that the Group Sequence View should allow the comparison of all user groups at the same time. As discussed before, we plan to provide a scalable overview to address this issue. Second, three participants noted that the edge glyph fails to show the replying order. Highlighting the first reply may solve this issue. Third, to improve the usability of VisForum, two participants suggest VisForum augment legends and axis of each view to reduce their memory load in visual comparison and exploration.

7 DISCUSSION AND CONCLUSION

We proposed various glyphs and coordinated algorithms for thread compression in order to facilitate group exploration in a lengthy thread. However, our design poses new scalability issues. For example, (1) in the Group Sequence View, if a thread contains more than 30 groups, users find it difficult to remember the position of the selected group and they have to scroll the sequence back and forth. This is partially because the design of group glyph and user glyph in the Group Sequence View is not space efficient. We plan to design a compact layout for the group glyph and user glyph to better utilize the screen space and ease the interaction. (2) If a group has more than five members, edges with bumps crossing together may pose a burden in differentiating reply frequencies depicted as bumps. However, this can be solved by Divide and Conquer. We can separate a big group into several smaller ones. (3) Furthermore, in the Group Evolution View, if multiple groups of similar size are formed at the same time, nodes in the Group Evolution View will overlap. This problem can be simply addressed by using collision detection algorithms. We also noted that proportional symbols are used both in the Group Evolution View and the Group Sequence View with different meanings, which may confuse end users. A careful design in the Group Evolution View can mitigate this issue.

It is also worth noting that although novel glyphs are designed in the context of MOOC forum analysis, many of them can be applied to a border application. For example, the design of Edge Glyph can be applied for visualizing general directed graphs to show edge strength in a quantitative presentation. However, it is not clear how the Edge Glyph is applied to a dense directed graph with edge crossings. Further user study is required to evaluate the scalability of the design.

In this article, we have presented VisForum, a novel visual analysis system for interactively exploring, comparing, and tracking conversation groups in asynchronous online forums. VisForum integrates four coordinated views with a set of novel designs to help users explore groups from multiple view perspectives interactively. The Group Detecting & Sorting Algorithm and the concept of "Forum-index" are proposed to assist the analysis process. Two case studies and one user study have demonstrated the usefulness and usability of the system. In particular, the Group Sequence View plays a primary role in user-based group comparison, and the interaction across multiple views enables users to quickly explore various related groups and obtain an overview of the whole forum in terms of groups.

We have found promising directions for future research. First, the current prototype supports the presentation of user groups in a single thread in the Group Sequence View. It will be interesting to investigate the same user groups in multiple threads at the same time. Further, the system can support semantic clustering of threads to ease the interaction. Second, we plan to integrate more data mining techniques in order to gain insights into the discussions. Third, although the system has originally been designed for expert users in MOOC forum analysis, case studies and user studies demonstrate that our method can be generalized to other forums. We plan to widely extend its usage scenario in the near future.

REFERENCES

[1] David Auber, Yves Chiricota, Fabien Jourdan, and Guy Melançon. 2003. Multiscale visualization of small world networks. In *Proceedings of the 9th Annual IEEE Conference on Information Visualization (INFOVIS'03)*. IEEE Computer Society, Washington, D.C., 75–81.

[2] B. Bach, N. H. Riche, C. Hurter, K. Marriott, and T. Dwyer. 2017. Towards unambiguous edge bundling: Investigating confluent drawings for network visualization. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 541–550. DOI: http://dx.doi.org/10.1109/TVCG.2016.2598958

[3] Michelle Q. Wang Baldonado, Allison Woodruff, and Allan Kuchinsky. 2000. Guidelines for using multiple views in information visualization. In *Advanced Visual Interfaces*. 110–119.

[4] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*. "O'Reilly Media, Inc."

[5] Rita Borgo, Johannes Kehrer, David H. Chung, Eamonn Maguire, Robert S. Laramee, Helwig Hauser, Matthew Ward, and Min Chen. 2013. Glyph-based visualization: Foundations, design guidelines, techniques, and applications. *Eurographics State of the Art Reports* (2013), 39–63.

[6] Tharrenos Bratitsis and Angelique Dimitracopoulou. 2006. Monitoring and analyzing group interactions in asynchronous discussions with the DIAS system. In *Groupware: Design, Implementation, and Use*. Springer, 54–61.

[7] Nan Cao, Yu-Ru Lin, Fan Du, and Dashun Wang. 2016. Episogram: Visual summarization of egocentric social interactions. *IEEE Computer Graphics and Applications* 36, 5 (2016), 72–81.

[8] Nan Cao, Yu-Ru Lin, Xiaohua Sun, David Lazer, Shixia Liu, and Huamin Qu. 2012. Whisper: Tracing the spatiotemporal process of information diffusion in real time. *IEEE Transactions on Visualization and Computer Graphics* 18, 12 (2012), 2649–2658.

[9] J. C. C. Chan, Khe Foon Hew, and Wing Sum Cheung. 2009. Asynchronous online discussion thread development: Examining growth patterns and peer-facilitation techniques. *Journal of Computer Assisted Learning* 25, 5 (2009), 438–452.

[10] Siming Chen, Shuai Chen, Zhenhuang Wang, Jie Liang, Xiaoru Yuan, Nan Cao, and Yadong Wu. 2016. D-map: Visual analysis of ego-centric information diffusion patterns in social media. In *Proceedings of the 2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 41–50.

[11] Brian Duffy, Jeyarajan Thiyagalingam, Simon Walton, David J. Smith, Anne Trefethen, Jackson C. Kirkman-Brown, Eamonn A. Gaffney, and Min Chen. 2015. Glyph-based video visualization for semen analysis. *IEEE Transactions on Visualization and Computer Graphics* 21, 8 (2015), 980–993.

[12] Cody Dunne and Ben Shneiderman. 2013. Motif simplification: Improving network visualization readability with fan, connector, and clique glyphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3247–3256.

[13] Gary William Flake, Robert E. Tarjan, and Kostas Tsioutsiouliklis. 2004. Graph clustering and minimum cut trees. *Internet Mathematics* 1, 4 (2004), 385–408.

[14] S. Fu, J. Zhao, W. Cui, and H. Qu. 2017. Visual analysis of MOOC forums with iForum. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (Jan. 2017), 201–210. DOI: http://dx.doi.org/10.1109/TVCG.2016.2598444

[15] Michelle Girvan and Mark E. J. Newman. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99, 12 (2002), 7821–7826.

[16] Amir Globerson, Gal Chechik, Fernando Pereira, et al. 2007. Euclidean embedding of co-occurrence data. *Journal of Machine Learning Research* 8 (2007), 2265–2295.

[17] Mark Harrower and Cynthia A. Brewer. 2003. ColorBrewer. org: An online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37.

[18] Jorge E. Hirsch. 2005. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America* 102, 46 (2005), 16569–16572.

[19] D. Holten. 2006. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics* 12, 5 (Sept. 2006), 741–748. DOI: http://dx.doi.org/10.1109/TVCG.2006.147

[20] John Hopcroft, Omar Khan, Brian Kulis, and Bart Selman. 2003. Natural communities in large linked networks. In *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 541–546.

[21] Enamul Hoque and Giuseppe Carenini. 2014. ConVis: A visual text analytic system for exploring blog conversations. In *Computer Graphics Forum*, Vol. 33. Wiley Online Library, 221–230.

[22] Borg Ingwer and Patrick J. F. Groenen. 2005. *Modern Multidimensional Scaling: Theory and Applications*. Springer.

[23] Allan C. Jeong. 2005. The combined effects of response time and message content on growth patterns of discussion threads in computer-supported collaborative argumentation. *International Journal of E-Learning & Distance Education* 19, 1 (2005).

[24] Sujana Jyothi, Claire McAvinia, and John Keating. 2012. A visualisation tool to aid exploration of students' interactions in asynchronous online communication. *Computers & Education* 58, 1 (2012), 30–42.

[25] B. C. Kwon, S. Kim, S. Lee, J. Choo, J. Huh, and J. S. Yi. 2016. VisOHC: Designing visual analytics for online health communities. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 71–80. DOI: http://dx.doi.org/10.1109/TVCG.2015.2467555

[26] Victor E. Lee, Ning Ruan, Ruoming Jin, and Charu Aggarwal. 2010. A survey of algorithms for dense subgraph discovery. In *Managing and Mining Graph Data*. Springer, 303–336.

[27] Nan Li and Desheng Dash Wu. 2010. Using text mining and sentiment analysis for online forums hotspot detection and forecast. *Decision Support Systems* 48, 2 (2010), 354–368.

[28] R. Duncan Luce. 1950. Connectivity and generalized cliques in sociometric group structure. *Psychometrika* 15, 2 (1950), 169–190.

[29] Andrew Y. Ng, Michael I. Jordan, Yair Weiss, et al. 2002. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems* 2 (2002), 849–856.

[30] Victor Pascual-Cid and Andreas Kaltenbrunner. 2009. Exploring asynchronous online discussions through hierarchical visualisation. In *Proceedings of the 13th International Conference on Information Visualisation, 2009*. IEEE, 191–196.
[31] D. Rafiei. 2005. Effectively visualizing large networks through sampling. In *VIS 05. IEEE Visualization, 2005*. 375–382. DOI: http://dx.doi.org/10.1109/VISUAL.2005.1532819
[32] Khairi Reda, Chayant Tantipathananandh, Andrew Johnson, Jason Leigh, and Tanya Berger-Wolf. 2011. Visualizing the evolution of community structures in dynamic social networks. In *Computer Graphics Forum*, Vol. 30. Wiley Online Library, 1061–1070.
[33] Timo Ropinski, Steffen Oeltze, and Bernhard Preim. 2011. Survey of glyph-based visualization techniques for spatial multivariate medical data. *Computers & Graphics* 35, 2 (2011), 392–401.
[34] Sam T. Roweis and Lawrence K. Saul. 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* 290, 5500 (2000), 2323–2326.
[35] Warren Sack. 2000. Discourse diagrams: Interface design for very large-scale conversations. In *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000*. IEEE.
[36] Purnamrita Sarkar, Sajid M. Siddiqi, and Geoffrey J. Gordon. 2007. A latent space approach to dynamic embedding of co-occurrence data. In *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics (AI-STATS'07)*.
[37] Lei Shi, Chen Wang, Zhen Wen, Huamin Qu, Chuang Lin, and Qi Liao. 2015. 1.5D egocentric dynamic network visualization. *IEEE Transactions on Visualization and Computer Graphics* 21, 5 (2015), 624–637.
[38] Xiaolin Shi, Jun Zhu, Rui Cai, and Lei Zhang. 2009. User grouping behavior in online forums. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 777–786.
[39] Ben Shneiderman. 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of IEEE Symposium on Visual Languages, 1996*. IEEE, 336–343.
[40] Joshua B. Tenenbaum, Vin De Silva, and John C. Langford. 2000. A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 5500 (2000), 2319–2323.
[41] Corinna Vehlow, Fabian Beck, Patrick Auwärter, and Daniel Weiskopf. 2015. Visualizing the evolution of communities in dynamic graphs. In *Computer Graphics Forum*, Vol. 34. Wiley Online Library, 277–288.
[42] Hongning Wang, Chi Wang, ChengXiang Zhai, and Jiawei Han. 2011. Learning online discussion structures by conditional random fields. In *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 435–444.
[43] Xiting Wang, Shixia Liu, Yang Chen, Tai-Quan Peng, Jing Su, Jing Yang, and Baining Guo. 2016. How ideas flow across multiple social groups. In *Proceedings of the 2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*. IEEE, 51–60.
[44] Matthew O. Ward. 2002. A taxonomy of glyph placement strategies for multidimensional data visualization. *Information Visualization* 1, 3–4 (2002), 194–210.
[45] Matthew O. Ward. 2008. Multivariate data glyphs: Principles and practice. In *Handbook of Data Visualization*. Springer, 179–198.
[46] Martin Wattenberg. 2006. Visual exploration of multivariate graphs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06)*. ACM, New York, NY, 811–819. DOI: http://dx.doi.org/10.1145/1124772.1124891
[47] Yanhong Wu, Naveen Pitipornvivat, Jian Zhao, Sixiao Yang, Guowei Huang, and Huamin Qu. 2016. egoslider: Visual analysis of egocentric network evolution. *IEEE Transactions on Visualization and Computer Graphics* 22, 1 (2016), 260–269.
[48] Ruliang Xiao, Youcong Ni, Xin Du, and Ping Gong. 2010. Towards a correlation cooccurrence model generating approach to folksonomy. In *Proceedings of the 2010 International Conference on Web Information Systems and Mining (WISM)*, Vol. 2. IEEE, 399–403.
[49] Jian Zhao, Michael Glueck, Fanny Chevalier, Yanhong Wu, and Azam Khan. 2016. Egocentric analysis of dynamic networks with egolines. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 5003–5014.

Received January 2017; revised October 2017; accepted November 2017