JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 40, 265-282 (2024)
DOI: 10.6688/JISE.202403_40(2).0005

Visual Exploration Framework of Intersecting Users and Activities between Issues in Social Media

CHEN-CHI HU, CHUN-MIN LIAO AND MING-TE CHI+
Department of Computer Science
National Chengchi University
Taipei, 116 Taiwan
E-mail: {102753501; 104753040}@nccu.edu.tw; mtchi@cs.nccu.edu.tw

Analyzing the intersecting elements of hierarchical datasets is a critical task. We enhance the visual expression of hierarchical data and the visual exploration of the set/element attribute referring to the related visualizations. Therefore, we propose the visualization to analyze users' activities from the social media dataset and further explore the key users and their behaviors. We present the intersecting elements to achieve reliable scalability through a hierarchical visualization according to the attributes in the dataset. Moreover, we adopt the multi-view visualization to explore the posts and user profiles between issues to achieve an intuitive visual layout and integrate user interaction into hierarchical data. Furthermore, we compare user activities between issues, integrate the interacted attributes through the intersecting activities, and highlight the critical information of posts and user profiles. In our approach, we analyze the intersecting elements in issues of the multiple hierarchical sets from the social media and identify the impact events from the activities of user interactions.

Keywords: visual analytics, set visualization, glyph visualization, social media data, hierarchical data

1. INTRODUCTION

Data visualization presents data through interactive charts, graphs, and maps so that users can comprehend the data, resolve problems, and gain insights from the visualization. Given that the data integrate an abstract in data visualization, we create geometry within the encoding of transforming data into information (understanding, sense-making, and insight), thereby making data comprehensible to people. As the Internet evolves, massive related data have been presented in different formats and stored in databases since the 1990s. Social media and Internet forums are derivative to the first place where people obtain news, discuss and communicate information, trace trends, or surf something interesting. Our research colleague in the communication academy desires to examine the influence of speeches that inspire the people (opinion leaders) on the Internet. Hence, we implement a computer-aided system through data visualization to analyze the articles and find the opinion leaders on the Internet.

In social media, users browse the messages and express their opinions or reactions in different social groups. Taking the gossip board on PTT (one of the famous bulletin board systems (BBS) in Taiwan) as an example, the user comments through texts and expresses reactions on Issue 1 (politician's regular activities) and does so on Issue 2 (war of words

---

Received September 1, 2022; revised February 23 & April 24, 2023; accepted May 3, 2023.
Communicated by Chun-Rong Huang.
+ Corresponding author.
* This work was supported by The Ministry of Science and Technology, Taiwan, under GRANT No. MOST 111-2221-E-004-007.

between politicians). Therefore, the users with the same social pattern have at least an intersecting activity and linkage trajectory in multiple issues. Although the linkage trajectories of many users on social media are complicated, we have a strong interest in analyzing users' activities on social media and searching for active users (opinion leaders) who focus on issues through the way of sets and their intersections. PTT is the most popular BBS, with generally 130,000 peak users daily in Taiwan, similar to Reddit in the U.S. However, the cruder of the website is a collection of entries submitted by its registered users who submit a post about news, rate the posts, and join the discussion. When posts are submitted to the board, the users vote for/against the posts through their comments or reactions (push/neutral/boo).

Social media and Internet forums are hierarchical organizations. However, the hierarchical relation is a particular case of set relation. We construct the user-centered data inspired by set visualization techniques, multiple views design, and interaction across views. The appropriate literature decides the glyph design [1]. Thus, we evaluate the system to know the user's subject using experience, discuss task support, and compare it to other related works by a user study. Our contributions list as follows:

A. Propose an interactive framework for analyzing the intersecting activities with the corresponding users in multiple posts on social media communities.
B. Visualize the intersecting elements from the users with their corresponding activities on social media and provide the glyphs with scalability for data analysis.
C. Explore intersecting posts from the active user (opinion leaders) and trace the trajectories of the impact activities from the information implied with the glyphs (users).

2. RELATED WORK

We adopt the point of view in "Visualization Zoo" [1] to search the users from overlapping activities of two sets in a hierarchical tree between issues and conclude our related approaches to set visualization and glyph visualization.

2.1 Set Visualization

In set visualization, visualizing sets in many possible relations between data is a crucial challenge. Our goal is to find the intersecting activities from two sets with the users as the elements and posts as sets. According to a systematic overview, different kinds of set relations [2, 3] and set visualization [4-6] classify as Euler and Venn diagrams, overlays techniques, node-link diagrams, matrix-based techniques, and aggregation-based techniques [7, 8]. Euler and Venn's diagrams represent the sets or subsets in the closed regions and remove the sets that enclosed a region with an issue of scalability [9]. Overlays techniques present the sets by region, line, and glyph with the definition at the position of elements by other attributes [3]. The node-link diagram maps the sets to the nodes and the set relations to the edges and shows the clusters of each element. However, it limits the scalability of crossing lines and lacks the set relations by elements versus sets [6]. The matrix-based technique lists data on the matrix by rows and columns and moderately scales the number of elements and sets but represents the least number of set relations [2]. Finally, aggregation-based techniques encode the elements in a group of sets by a single visual element [7]. These approaches have the advantage of choosing an appropriate strategy for a given problem. However, some methods only set goals to visualize a single set or aggregate the data based on sets [4, 10] without focusing on the intersections between sets or even the color of the intersecting set and define the relations between sets in the color map [6, 9, 11]. Therefore, we adopt these advantages of the aggregation technique in our approach ("SetView" and "ElementView"). In addition, we analyze the interaction of a single user for different events or the same event for different users in social media between intersections of sets.

Some data from social media include hierarchical relationships for visualization owing to the multiple interactions of reactions: publishes, replies, pushes, and boos. Therefore, we have many choices for hierarchical visualization, such as indented tree layout [12-15], Sunburst [16-18], Treemap [19-21], and nested circle layout [22-24]. Indented and hierarchical circular tree layouts have the advantage of continuous hierarchical relations, aggregate information by enclosing low-level ones, and the navigation and information retrieval facilitated through an elegant set of browsing interactions among them. The only disadvantage is the lack of features that emphasize specific regional data. From another aspect, Sunburst and Treemap improve the performance, and usage with more significant learning costs partially recoups over time. After evaluating the advice from experts and users, we choose Sunburst, which has a better ability to convey structure and hierarchy.

2.2 Glyph Visualization

Given that the users have extraordinary attributes as soon as they overlap at the activities on social media, encoding the overlapping users as glyphs are intuitive to evaluate the multi-dimensional data. Therefore, we investigate the current set of visualization techniques and glyph evaluation. Glyph [25] is a point that encodes the shape, color, texture, size, orientation, aspect ratio, or curvature individually depending on the dimensions with at least one mark and the visual variables. In addition, the glyph has issues with mapping to the data of attributes [26], addressing the guidance on the usage [27], applications on interactive devices [28], layout [29], and visual exploration for the dimension reduction [30]. Therefore, we investigate line glyph, stripe glyph, clock glyph, and star glyph [31] under two data densities [32] or multiple data with quantitative values [33] or comparing with other factors between the data (*e.g.*, time series) [34]. We integrate these approaches and design our approach using the patterns of observed performance and the implications. These patterns encode glyphs nested inside the visualizations, such as hierarchies, social media, geographic data, or gathering many data points in one overview. According to the analysis of glyph types [27, 31, 32, 34] (*e.g.*, cyclic glyphs [35] or hierarchical glyphs [36]), the star glyph is a relatively compact data graphic that represents the setting, datasets, and tasks, and it is well studied for multi-dimensional data.

3. METHOD

3.1 System Overview and Data Processing

Our framework comprises three main components (Fig. 1): data crawling (from PTT) and preprocessing, server (database), and client (interface). Firstly, we crawl the attributes related to users and their activities from the PTT board (*e.g.*, gossiping, where on average, users create 2000 new posts per day) by utilizing an open-source PTT web crawler [37]. Each raw post consists of three hierarchical layers, *i.e.*, author, content, and counts of pushes (one of the reactions similar to "Like" on Facebook). Hence, the hierarchical structure comprises posts, comments, and reactions. Data preprocessing ensures that every data attribute has no null value in the post and eliminates any redundant values. The data attributes consist of post ID, post topic, user ID, board ID, reaction counts, and other essential attributes.

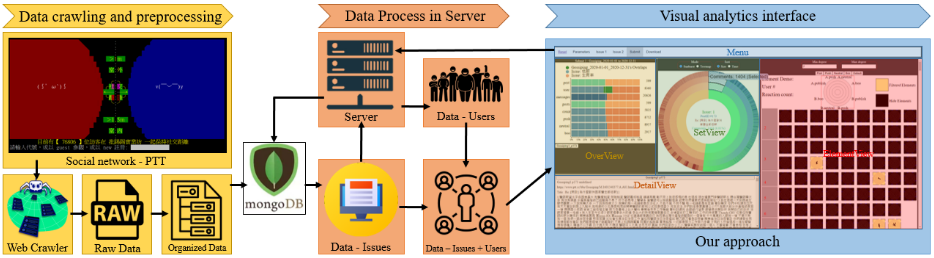

Fig. 1. System architecture and flow comprises three parts. The flow starts with data crawling and preprocessing, data process in the server, and then ends in the visual analytics interface.

After completing the data preprocessing, we store the data using MongoDB for server processing. The server handles client requests, retrieves the required data, and then constructs records that contain a list of users involved in each activity in those posts. Additionally, the server schedules merging the posts with the user list and then responds to the client. Furthermore, we incorporate *jieba* [38], a Chinese text segmentation tool that appends word columns to the records and extracts keywords from the post's content. This tool facilitates the creation of user-centered data while encountering the performance bottleneck. Consequently, we propose algorithm 1 to minimize the extra computing required for data transformation. In Algorithm 1, the user U_x takes action in every post $Post(P_i)$ from the query; the worst case is checking in $O(mn)$. When the user list $User(U_j)$ includes the U_x records, updates new action or creates a new user U_x in $User(U_j)$, the worst case is checking in Om). After iterating all posts and returning the user list $User(U_j)$ costs in $O(n * m^2)$. We found multiple users with overlapping behavior by taking action on multiple issues.

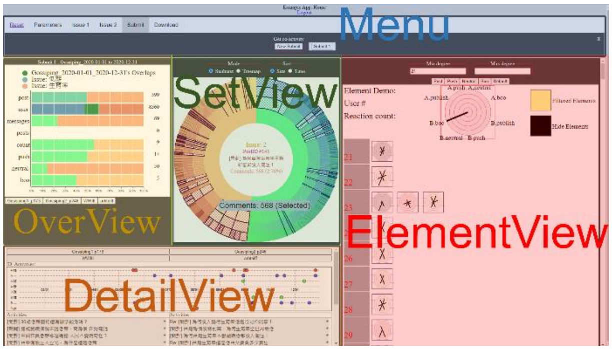

Fig. 2. Visual exploration framework consists of five views with different analysis functions.

Algorithm 1: Create active user list

Input: *Post*(P_i) is a list contain all query post data, P_i each contain their user activities log, *Post*(P_i) = [P_1, P_2, …, P_n].
Output: *User*(U_j) is a list collect all user has action in posts belong, *User*(U_i) = [U_1, U_2, …, U_m].

User(U_j) = [];
For i = 0; $i \leq n$; **do**
  While not end of user in P_i **do**
    if U_x find in *User*(U_j) **then**
      Update the U_x in U_j;
    end if
    Else
      Create U_x represent the user;
      Push U_x in *User*(U_j);
    end else
    next user;
  end while
end for
return *User*(U_j);

Alg. 1. Create an active user list by organizing the intersecting users.

The framework implemented by D3.js [39] includes multiple visualized views to handle different tasks, *Menu* on the top, *OverView* on the left, *SetView* in the middle, *ElementView* on the right, and *DetailView* on the bottom left (Fig. 2).

Menu (Fig. 2 (a)) provides tabs (Parameters, Issue 1, Issue 2, Submit, and Download) for the initial setting. The two tabs of Issues 1 and 2 load two records of querying posts. The Parameters tab sets the parameters for both queries: comment count (0–50), time interval, the keyword (*e.g.*, topic, content), active user ID, and other relative data (*e.g.*, full texts of the post). The Submit tab uploads the query items to the server, and the Download tab fetches the related content from the server.

In the *OverView* (Figs. 2 (b) and 4), we provide information for comparing datasets through a stacked bar chart and tooltip demonstrating the statistical information of elements and datasets. The bar chart is divided into the left section (Issue 1) and the right section (Issue 2). The color blue represents Issue 1, orange represents Issue 2, and the green represents the intersection users of Issues 1 and 2. For example, the attribute indexes of PTT data (Figs. 2 (b) and 4), such as post-count, active user counts, and activity-count (*e.g.*, reply, reactions), summarize specific elements for the process of *SetView* or *ElementView*.

SetView visualizes the relations, attributes, and particular sets, which analyze the hierarchical relation from posts. In some exceptional cases, the hierarchical relations visualize the space-filling technique, either Treemap [19-21] or Sunburst [16-18] (Figs. 2 (c) and 5). The space-filling technique views the proportion of reactions in their category area. Large areas in Treemap or border arc in Sunburst indicate that the post joined many users and activities to the corresponding issue. Sunburst has the center space to arrange more information (*e.g.*, text), which has an advantage over Treemap, and the users prefer Sunburst (questionnaire survey). Color encodes two query sets by orange-red and green-blue in *SetView*. Moreover, luminance encodes the hierarchical relations of issue, post, comment, and reactions, and the types of reactions (push, neutral, and boo) encoded with color in PTT.

The matching element in *ElementView* refers to the users who have participated in the selected posts' activities. As shown in Figs. 2 (d) and 6, *ElementView* displays overlapping glyphs that represent these users. By default, all matching elements are scaled into an overview in one area, and users can zoom and pan to navigate through the glyphs for more information. *ElementView* allows users to filter elements based on their activity type and interval degrees. For instance, users can filter PTT activities by five types: Default (All types), Publish, Push, Neutral, and Boo. If users select the "Push" button, the glyphs will represent users with at least one post in both query sets. This function helps focus on the overlap of users with specific activities.

The glyph alignment in Powerset [25] inspired us. The elements are sorted by degrees of the numbers rectangle (2~*max*); if the glyph counts *n* over line limit *m*, it is divided by 2. If the value *m*/2 is over the limit, then *n* is divided by 3, and so on. This way, the glyph alignment allocates the whole dataset in a view. Of course, the lower degree of elements is numerous, but the higher degree is more critical than, the lower one. Hence, we omit the lower but keep, the higher degree of elements. *DetailView* shows the interactive information by hovering the sets in *SetView* or glyphs in *ElementView* (Figs. 2 (e) and 3). For example, the attributes of the selected post include Post ID, URL, title, content, and user reactions (*e.g.*, push, neutral, boo). The contents of selected users include the timeline and the list of activities. In the timeline, every dot denotes an action. The contents classify into two parts to show the active users in the set, and each action type owns a horizontal line to indicate the different types of actions.

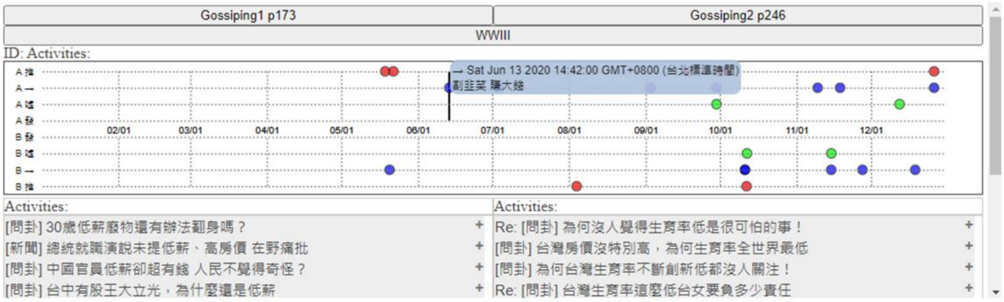

Fig. 3. DetailView shows the user activities from PTT. The red node shows the reaction "Push," the blue one is "Reply," and the green one is "Boo."

3.2 Visual Design

We derived the following requirements for the task analysis from our visual framework.

R1: Analyze the intersecting users between posts on PTT.
R2: Analyze the intersecting activities between users on PTT.
R3: Analyze the activities (posts) of the specific users at different periods.
R4: Analyze the overlapping activities of the users and determine the relationship.
R5: Trace more detailed information in viewing the related posts and the users' profiles.

To analyze the intersecting activities in social media, we identify the user behavior and further describe research problems related to the requirements as follows:

Prob(1): How to visualize the activities (posts) on PTT between specific users at different times (R1, R3, and R4)? Two design decisions are available for this problem: determine (1) the intersecting activities between two identical issues and (2) the opinion leaders to each issue. Sec. 3.3 discusses the approach to the intersecting activities and the corresponding opinion leaders.

Prob(2): How to visualize the users interacting on PTT at different times and determine the hierarchy among users (R1, R2, R3, and R4)? This problem extends further from the previous problem to determine users' hierarchy to the corresponding opinion leaders. Therefore, we will discuss the complex hierarchies in Sec. 3.4.

Prob(3): How to visualize the overlapping activities and trace the target posts or users (R1, R3, R4, and R5)? This problem is also an extension of Problem 2. We also analyze the reactions and comments to each issue and determine the hierarchy of the users. Finally, we will discuss the overlapping activities and information tracing in Sec. 3.5.

We compile the summarized design goals to meet the requirements and support the analytical tasks of the communication college experts and the insights from the desired data:

G1. Determine the overlapping activities from two identical issues among users – exploring the overlapping activities by the users between two issues and tracing whether the activities are promoted to the other users by the opinion leaders are keys to the relation of two issues.

G2. Explore the hierarchies of the users to the related post – exploring the posts from PTT in the original hierarchy and the user lists and observing whether the activities on both issues are relative at different times.

G3. View the information of posts and the profiles of the interesting users – specific users who are highly active in the posts are the evidence. The related experts confirm these candidate users and posts from the detailed information.

We propose our approach to visually analyze the intersecting activities from the target users for the experts based on these design goals. This approach follows a hierarchical exploration scheme (G2) to facilitate practical visual analysis.

Set Visualization (G1): This type of visualization performs the intersecting activities between two identical posts. However, the disadvantage is losing detailed information, such as the hierarchies of the users and the other reactions from the users.

Hierarchical Visualization (G2): This type of visualization shows the hierarchies of the users to the related posts, but it is also functional for the hierarchies without detailed information, such as the reactions from the users.

Glyph Visualization (G3): The glyph visualization shows detailed information (*e.g.*, the reactions from the users and the reactions for analyzing a single user).

3.3 Intersections in Set Visualization

Given a universal set E, we define system F consisting of n sets contained in E as follows:

$$F = \{S_1, S_2, \ldots, S_n\}: S_{1\le n} \subset E. \tag{1}$$

In the context of PTT, we can think of F as the overall set of all discussion boards, and S_i as a subset of F that represents a particular user's participation in some of the discussion boards. Each set S_i has intersections in F that correspond to several sets X as $X \subseteq F$. The number of these intersections is equal to the cardinality of $P(F) = \{X: X \subseteq F\}$, and the number of power sets F is equal to 2^n. $P(F)$ represents intersections that differentiate two types of intersections with full and exclusive intersections. Given sets $X = \{S_{i_1}, S_{i_2}, \ldots, S_{i_n}\} \subseteq E$ in the system F:

In a full intersection, all elements belong to the sets regardless of the other set elements:

$$I(X) = \cap(X) = S_{i_1} \cap S_{i_2} \cap \ldots \cap S_{i_k}. \tag{2}$$

This definition has not clarified the elements that belong to X or $F\backslash X$ (elements belong to F and do not belong to X). Moreover, the full intersection of the set $I(X)$: $X \in P(F)$ does not constitute a partition of the universal set of elements E. Considering that $I(X_1 \subseteq F)$ and $I(X_2 \subseteq F)$ might overlap, exclusive intersection contains all elements that belong to all sets $S_{i_1}, S_{i_2}, \ldots, S_{i_k}$ and exclude the other sets in F.

$$I_{ex}(X) = \cap(X) \backslash (F\backslash X). \tag{3}$$

The set of all intersections in $I_{ex}(X)$: $X \in P(F)$ properly partitions the universal set with all elements E. Elements in a particular intersection only belong to the regions of the Venn diagram of the set system; this helps detect differences between groups of elements based on all elements.

Degree E denotes the number of sets that belong to F. All elements $e \in I_{ex}(X)$ in an exclusive intersection have the same degree $|X|$:

$$Degree(I_{ex}(X)) = |X|. \tag{4}$$

The mechanism for exploring the intersections between two sets involves identifying active users who have participated in both issues through multiple posts. This is represented in Fig. 4 by the deep green portion of the bar chart.

3.4 Hierarchical Visualization

The data in this study have a hierarchical relationship, and user activities are included in the set of data content according to attributes in social media. For example, a user in PTT may publish a post or participate in the opinions in the post. The user belongs to the activity participant set of this post. The discussions in each post form a set, and the users who participate in the discussion are the elements of the set. For example, a user x commented on the post P in Issue a, x belongs to the set of post P and the set of Issue a, and Issue a contains the post P. Therefore, the hierarchical relationship transforms into the relationship between sets and elements. Query users who are active in two issues (PTT) visualize the layout of the intersections by expressing hierarchical relationships, set and element relationships, and one of the most critical tasks.

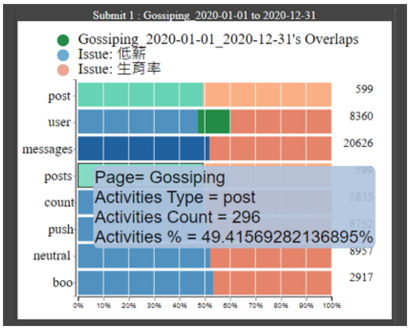

Fig. 4. *OverView* of PTT example (Low salary vs. Low fertile rate).

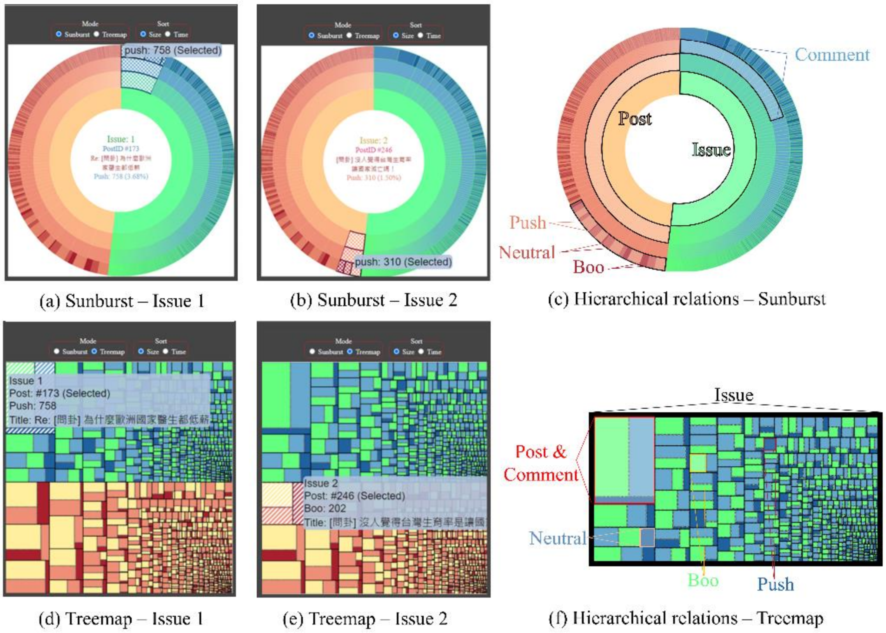

Fig. 5. Showcases the visual exploration provided by *Setview* using two space-filling techniques – Sunburst and Treemap – to address different issues (Low salary vs. Low fertile rate).

3.4 Hierarchical Visualization

The data in this study have a hierarchical relationship, and user activities are included in the set of data content according to attributes in social media. For example, a user in PTT may publish a post or participate in the opinions in the post. The user belongs to the activity participant set of this post. The discussions in each post form a set, and the users who participate in the discussion are the elements of the set. For example, a user x commented on the post P in Issue a, x belongs to the set of post P and the set of Issue a, and Issue a contains the post P. Therefore, the hierarchical relationship transforms into the relationship between sets and elements. Query users who are active in two issues (PTT) visualize the layout of the intersections by expressing hierarchical relationships, set and element relationships, and one of the most critical tasks.

To visualize the layout of intersections between active users in two issues (PTT), we express the hierarchical relationships, set and element relationships, one of the most critical tasks. Based on the aforementioned data characteristics, we adopt Sunburst [16-18] and Treemap [19-21] for visually analyzing the target user from hierarchies. In Sunburst mode (Figs. 5 (a)-(c)), we implement issues, posts, comments, and reactions from the inside out and sort the posts by size/time clockwise. Sunburst provides a circular hierarchical layout that allows users to understand the relationships between sets and elements easily. Similarly, we implement the reactions of a post in Treemap (Figs. 5 (d)-(f)) and sort the posts by size/time from top to bottom and left to right. Treemap provides a rectangular hierarchical layout that can display a large amount of data and provide a clear overview of the data.

3.5 Glyph Visualization

Multi-dimensional data visualization encodes every dimension in the graph after dimensional reduction and maintains the critical attributes. In our approach, we adopt glyph evaluation [27, 32] for multi-dimensional data visualization through star glyph [31]. Although linear glyphs (bar or line charts) are more accurate at value judging and trend detection, the star glyph (circular layout) can compare attributes among different glyphs while precisely querying specific attributes. We design star glyph of whisker (only axis) with grid on the axis for easier detection of similar data (Fig. 6); providing reference to improve the correctness of similarity judgments only for star glyph with contours. Contours enhance the detection of the similarity to the other types, for example, shapes are distracting by verifying data similarity. Every attribute (action type) is significantly different in quantities at the same scale level, so we scale the attributes on the axis. The greatest attribute of all glyphs normalizes to the full length of the axis for comparing values of all showing glyphs. The glyph axis splits to the upper and bottom sides and directs reversely (Fig. 7) to compare two attributes. For example, in the attributes of PTT Issues 1 and 2 in Fig. 7, the intersecting attributes are in the reverse direction. Still, the label and value only show clear viewing during the mouse hover. The left glyph indicates that the reaction of this user on Issue 1 is 17 times without other actions, but the reactions of this user on Issue 2 are 27 times, with 84 times comments.

3.6 Interactions in Multi-view Visualization

"Overview first, zoom and filter, then details-on-demand." B. Shneiderman proposed on Visual Information Seeking Mantra [40]. This approach inspired us to design detailed interactions on demand. The system shows tooltips while navigating *OverView*, *SetView*, and *ElementView* and provides interactions on sets and elements with the information on *DetailView*. *ElementView* is full of glyphs that use zoom and pan to assist in navigating glyphs to the detailed information or all glyphs in a glance. Moreover, the demonstration grid in the *ElementView* shows the zooming glyph with detailed information by hovering the mouse over the target glyph in the glyph list. Selection is an interaction for multiple views while surfing a post in *SetView* then highlighting the corresponding user in a glyph

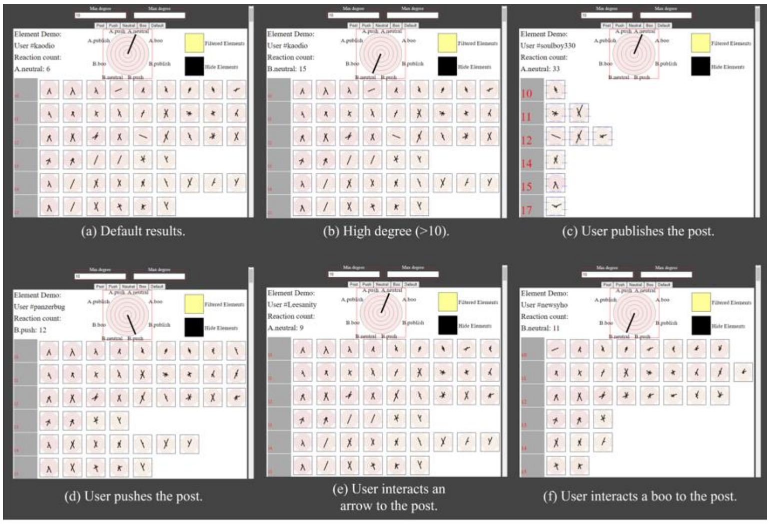

Fig. 6. ElementView of PTT. The layout interacts with different conditions: (a) default results, (b) high degree (>10), (c) user publishes the post, (d) user pushes the post, (e) user interacts an arrow to the post, and (f) user interacts a boo to the post.

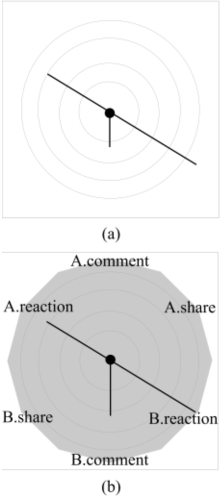

Fig. 7. Glyph design; (a) default; (b) selected glyph.

or selecting a user in *ElementView*, and then highlighting (grid pattern) the post in *SetView*. Both interactions list the detailed information on *DetailView*. Furthermore, the highlight color indicates the number of match conditions encountered in multiple selections. For example, *ElementView* shows users in yellow if they are active in both issues, gray when joining one of the posts, and black when not in any posts. Moreover, *SetView* shows sets in the filling pattern when the users are active in the post, and it performs in the original color if no users join.

4. ANALYSIS

4.1 Case Analysis

We analyzed the social issues of low salary [LS] and lower fertility rates [LFR] in 2020 as examples with the help of visualization experts. Our first step was to analyze PTT posts related to [LS] and [LFR] (Figs. 4 and 8 *OverView*). The results showed that [LFR] had more post counts but fewer active users than [LS] ([LFR]-362 vs. [LS]-296 post-count and [LFR]-5,073 vs. [LS]-5,333 active users). Moreover, the proportion of reactions (push/boo) in [LFR] (4,466/1,559) was also less than in [LS] (4,930/1,582). Using *SetView*, we gained insights into user activity and post topics related to [LS] and [LFR]. We found that users were more active in [LS] and discussed "*Why do medical doctors get paid few in Europe*?" (Fig. 8 on the bottom left of *DetailView*). On the other hand, users discussed [LFR] and topics related to "*Does anyone think that the lower fertility rate in Taiwan is destroying the country*?" (Fig. 8 on the bottom right of *DetailView*).

ElementView helped us filter out random users and target active users who had participated in the issues. We found that most active users joined posts more than ten times (Fig. 6 (b)). Users were only allowed to choose a reaction once and had to use an arrow to fill in the remaining text if they exceeded the one-line length. Our framework provides a function that allows users to click the reaction button to trigger the reactions in both sets. The case study revealed that most users used the ‘push’ reaction and replied to the posts more often than the ‘boo’ reaction. We also found that some topics in PTT were hosted by a small group of users, such as a user named “WWIII,” an active author discussing and reacting to [LS] posts. Many color dots were plotted on the timeline in *DetailView* (Figs. 3 and 8) to indicate the intersecting activities of users like “WWIII” and ‘arbteiff.’ *OverView* selected post-ID 173 vs. 246 from user “WWIII” and *DetailView* confirmed that the user had reacted to both posts (Fig. 8).

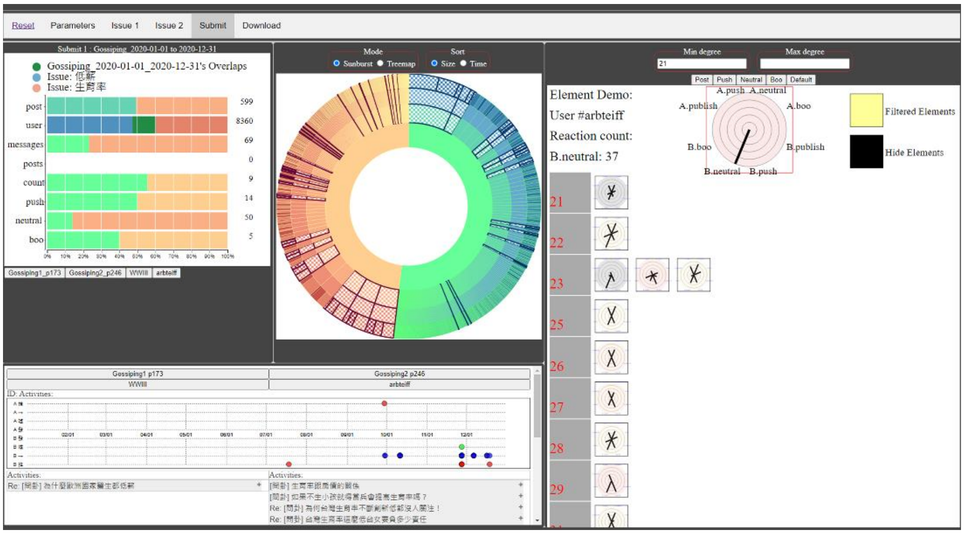

Fig. 8. The interactive results of the two intersecting users (WWIII and arbteiff) between the posts (p173 and p246) with the issues ([LS] and [LFR]) in the PTT gossiping board.

Fig. 9. The interactive results of the two intersecting users (kutkin and kcclasaki) between the posts (p164 and p516) in with the issues ([salmon] and [changing name]) in the PTT gossiping board.

To provide another case study, we analyzed the issues of [salmon] and [changing name] on PTT between 2021-2022. In *OverView* (Fig. 9), we observed that there were more posts related to [salmon] than [changing name], but the [salmon] had more active users based on the post-count ([salmon]-1,000 vs. [changing name]-974), active users ([salmon]-13,721 vs. [changing name]-13,255) and reaction (push/boo) proportions, which were [salmon](106/4) vs. [changing name](70/1). *SetView* revealed a more significant user activity in [salmon] and the post's topic: "*Laugh at Taiwanese beggars! Korean media report* '*Salmon Rebellion*'." (Fig. 9 *SetView* top right). Moreover, *SetView* revealed a higher activity level in [changing name] in "*Shame! Foreign media report that the Taiwanese changed the name because of the free sushi.*" (Fig. 9 *SetView* bottom left).

To target the active users, *ElementView* (Fig. 9) revealed that most users who participated in the discussions joined posts with more than 100 replies. This filtering helps to exclude users who randomly participated in the issue. In addition, we identified some topics hosted by a small group of users. For instance, a user named "kutkin" was an active author who discussed and reacted to [changing name] posts, resulting in many color dots plotted on the timeline in *DetailView* (Fig. 9). *OverView* selected post-ID 164 vs. 516 from user "kutkin" and *DetailView* confirmed that the user had reacted to both posts. The intersecting activities were not limited to "kutkin" but also applied to other users, such as "kcclasaki," as shown in our approach (Fig. 9).

4.2 User Study

We conducted a 60-minute interview with three domain experts, including two computer science professors (Experts A and B) and one communication college professor (Expert C). All of them are well-versed in social media visualization and analysis. We demonstrated our system during the interview and solicited their feedback on usability, visual design, and human-computer interaction. Additionally, they provided valuable suggestions for further optimization and improvement of our approach. Before the interview, we provided the experts with a guideline of the relevant views to ensure they were familiar with the exploration interfaces. Our approach impressed the experts, particularly in analyzing opinion leaders and identifying their intersecting activities. Experts A and B suggested improvements, such as providing proper exploration to guide users to observe specific users from issues, posts, comments, and reactions. This way, users can complete the task of reaching the commenters (the users who participated in posts corresponding to the issues) and listing their activities. The previous exploration also allowed users to compare the opinion leaders (commenters with dominant reactions between the posts) and reach out about their interesting activities. Regarding system usability, Expert C expressed great interest in using the dashboard view to support the exploration and analysis of intersecting users and activities between issues in social media. Expert C also suggested improvements, such as highlighting opinion leaders in SetView and using interactions to verify their value from ElementView. Based on the experts' feedback, we provided them with a comprehensive approach to exploring intersecting users and activities between issues in social media.

We also conducted some students of nine males and two females aged 20–30 majoring in computer science and one female aged 25–30 majoring in journalism. Moreover, the frequency of browsing PTT is evenly distributed many times a day, once a month, and even never. Therefore, the user study consists of two parts in our approach. The first part requests the users to find an issue in the discussion from PTT after experiencing our framework. At first, we instruct the users to explore the issue in the gossiping board on PTT, which they are concerned about, or randomly pick an issue in 2020. Then, they answer the questionnaire after experiencing our framework. The results of the questionnaires indicate that users determine the counts of reactions, counts of activities, active users, the more reactions to the specific users, existence/absence of similar users and the minor difference between them, the inspiration after experiencing the system, and the feedback to our approach.

The second user study requests the users to mark our approach after experiencing it in the five-point Likert scale survey (Fig. 10). The users agree with our framework in general statistics. For example, *Easily comprehend in OverView*, which has the highest mark of **4.42**. Apart from this, *DetailView* provides functional tabs, *DetailView* provides a timeline for activities, and Easy to find the concerned posts by chosen users with a higher mark of **4.25**. These statistics point out that the advantage of our approach is that the integration of multiple views assists users in comprehending the data. On the contrary, the weakness of our approach is the lack of hints for interactions. Thus, we improved our system with instructions and added tooltips when hovering over the icons.

4.3 Evaluation and Comparison

The functionality is less in our system than the set-related tasks [2]. However, we provide the most tasks related to the elements, hierarchical related to the tasks, simplified tasks related to the sets, and tasks associated with the attribute of elements except those directly related to the attributes. The scalability of sets is hundreds of sets in hierarchy relation from Treemap [19-21] or Sunburst [16-18] approach. The elements are limited to less than one thousand glyphs without losing the perception. Compared with Upset [4] and Powerset [10], more sets are combined with the hierarchy relation in our approach, but a few elements are combined with attributes' detailed information.

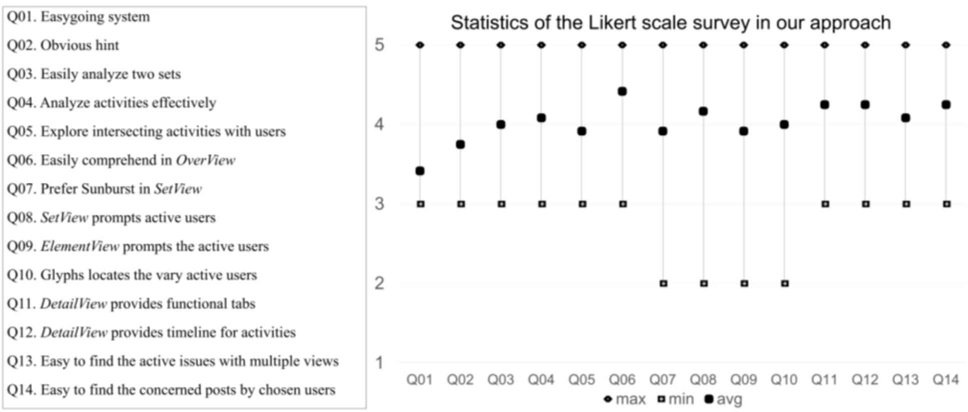

Fig. 10. The questionnaire list and the box and whisker plot after experiencing our approach.

4.4 Result and Discussion

After experiencing our framework, users highly agreed with comparing two sets of data. Visualizations can be performed in each view once the two sets of issues and the intersections of user activities have been explored. After the user study, the users prefer Sunburst [16-18] over Treemap [19-21] in *SetView* and Glyphs in *ElementView*, which has several active users. However, our approach has some limitations. First, the glyph function is not effective in the case of significant intersections and is time-consuming. A circular layout could further improve the layout with a higher degree of elements at the center space. Moreover, *ElementView* shows the intersecting elements with the details of the elements belonging to specific sets in the demonstration grid. The users have no idea how to determine users' intersecting behavior but only explore two sets. Thus, we provide a comparison with multiple sets in the demonstration grid. In addition, the time interval after the submitted query lacks interactions. A time interval could improve interaction in future work.

5. CONCLUSION AND FUTURE WORK

We demonstrate a multiple-view design and interactions to navigate users' activities of hierarchical data in social media (PTT) and analyze the intersecting users in a post according to different query conditions between sets. However, set visualization is more inspiring than hierarchical visualization. In our approach, sets of posts and elements of the activities are performed by space filling. Every view has functionality and interacts with one another. For example, *OverView* shows the overall statistic, *SetView* shows the hierarchical set relation by users' activities, and *ElementView* shows the elements intersecting two sets with the attributes. Finally, *Detailview* shows the detailed information of the corresponding set or element. Our design meets our goals. Thus, related researchers can intuitively explore the interesting posts and users, compare users' activities in time series on both issues, observe the intersections among user activities, and analyze the detailed information of posts and users' profiles. We also provide a demonstration grid for the scalability of glyphs while surfing the sets to show the elements of the different sets. Future work can analyze more than two sets of intersections and intersecting timelines for set visualization.

REFERENCES

1. J. Heer, M. Bostock, and V. Ogievetsky, "A tour through the visualization zoo," *Queue*, Vol. 8, 2010, p. 20.
2. B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers, "Visualizing sets and set-typed data: State-of-the-art and future challenges," in *EuroVis* (*STARs*), R. Borgo, R. Maciejewski, and I. Viola, (eds.), The Eurographics Association, 2014.
3. C. Vehlow, F. Beck, and D. Weiskopf, "The state of the art in visualizing group structures in graphs," in *EuroVis* (*STARs*), 2015, pp. 21-40.

4. A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, "Upset: visualization of intersecting sets," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, 2014, pp. 1983-1992.
5. J. R. Conway, A. Lex, and N. Gehlenborg, "Upsetr: an r package for the visualization of intersecting sets and their properties," *Bioinformatics*, Vol. 33, 2017, pp. 2938-2940.
6. B. Alsallakh, L. Micallef, W. Aigner, H. Hauser, S. Miksch, and P. Rodgers, "The state-of-the-art of set visualization," in *Computer Graphics Forum*, Vol. 35, 2016, pp. 234-260.
7. S. Luz and M. Masoodian, "A comparison of linear and mosaic diagrams for set visualization," *Information Visualization*, Vol. 18, 2019, pp. 297-310.
8. H. Chung, S. Nandhakumar, and S. Yang, "Gridset: Visualizing individual elements and attributes for analysis of set-typed data," *IEEE Transactions on Visualization and Computer Graphics*, Vol. ---, 2020, pp. -- --.
9. J.-B. Lamy, H. Berthelot, C. Capron, and M. Favre, "Rainbow boxes: a new technique for overlapping set visualization and two applications in the biomedical domain," *Journal of Visual Languages & Computing*, Vol. 43, 2017, pp. 71-82.
10. B. Alsallakh and L. Ren, "Powerset: A comprehensive visualization of set intersections," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, 2017, pp. 361-370.
11. J.-B. Lamy and R. Tsopra, "Rainbio: Proportional visualization of large sets in biology," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 26, 2019, pp. 3285-3298.
12. N. Kong, J. Heer, and M. Agrawala, "Perceptual guidelines for creating rectangular treemaps," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 16, 2010, pp. 990-998.
13. M. Sondag, B. Speckmann, and K. Verbeek, "Stable treemaps via local moves," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, 2017, pp. 729-738.
14. W. Köpp and T. Weinkauf, "Temporal treemaps: Static visualization of evolving trees," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25, 2018, pp. 534-543.
15. M. Sondag, W. Meulemans, C. Schulz, K. Verbeek, D. Weiskopf, and B. Speckmann, "Uncertainty treemaps," in *Proceedings of IEEE Pacific Visualization Symposium*, 2020, pp. 111-120.
16. J. Stasko and E. Zhang, "Focus$^+$ context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations," in *Proceedings of IEEE Symposium on Information Visualization*, 2000, pp. 57-65.
17. J. Stasko, R. Catrambone, M. Guzdial, and K. McDonald, "An evaluation of space-filling information visualizations for depicting hierarchical structures," *International Journal of Human-Computer Studies*, Vol. 53, 2000, pp. 663-694.
18. M. F. Rey and C. M. D. S. Freitas, "A visualization-based approach for the taxonomy browser interface," in *Proceedings of the XVI Brazilian Symposium on Human Factors in Computing Systems*, 2017, pp. 1-10.
19. H.-J. Schulz, "Treevis.net: A tree visualization reference," *IEEE Computer Graphics and Applications*, Vol. 31, 2011, pp. 11-15.

20. K. Kucher and A. Kerren, "Text visualization techniques: Taxonomy, visual survey, and community insights," in P*roceedings of IEEE Pacific Visualization Symposium*, 2015, pp. 117-121.
21. Y. Ming, H. Qu, and E. Bertini, "Rulematrix: Visualizing and understanding classifiers with rules," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25, 2018, pp. 342-352.
22. R. Boardman, "Bubble trees the visualization of hierarchical information structures," in *Proceedings of ACM CHI Extended Abstracts on Human Factors in Computing Systems*, 2000, pp. 315-316.
23. J. Görtler, C. Schulz, D. Weiskopf, and O. Deussen, "Bubble treemaps for uncertainty visualization," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, 2017, pp. 719-728.
24. G. Li, Y. Zhang, Y. Dong, J. Liang, J. Zhang, J. Wang, M. J. McGuffin, and X. Yuan, "Barcodetree: Scalable comparison of multiple hierarchies," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 26, 2019, pp. 1022-1032.
25. R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, "Glyph-based visualization: Foundations, design guidelines, techniques, and applications," in *Eurographics* (*STARs*), 2013, pp. 39-63.
26. M. O. Ward, "Multivariate data glyphs: Principles and practice," *Handbook of Data Visualization*, 2008, pp. 179-198.
27. J. Fuchs, F. Fischer, F. Mansmann, E. Bertini, and P. Isenberg, "Evaluation of alternative glyph designs for time series data in a small multiple setting," in *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems*, 2013, pp. 3237-3246.
28. T. Blascheck, L. Besancon, A. Bezerianos, B. Lee, and P. Isenberg, "Glanceable visualization: Studies of data comparison performance on smartwatches," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 25, 2018, pp. 630-640.
29. T. Opach, S. Popelka, J. Dolezalova, and J. K. Rød, "Star and polyline glyphs in a grid plot and on a map display: which perform better?" *Cartography and Geographic Information Science*, Vol. 45, 2018, pp. 400-419.
30. D. Kammer, M. Keck, T. Gründer, A. Maasch, T. Thom, M. Kleinsteuber, and R. Groh, "Glyphboard: Visual exploration of high-dimensional data combining glyphs with dimensionality reduction," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 26, 2020, pp. 1661-1671.
31. J. Fuchs, P. Isenberg, A. Bezerianos, F. Fischer, and E. Bertini, "The influence of contour on similarity perception of star glyphs," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 20, 2014, pp. 2251-2260.
32. J. Fuchs, P. Isenberg, A. Bezerianos, and D. Keim, "A systematic review of experimental studies on data glyphs," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 23, 2017, pp. 1863-1879.
33. A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini, "Towards understanding human similarity perception in the analysis of large sets of scatter plots," in *Proceedings of ACM CHI Conference on Human Factors in Computing Systems*, 2016, pp. 3659-3669.
34. M. Gleicher, "Considerations for visualizing comparison," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 24, 2017, pp. 413-423.
35. J. Suschnigg, B. Mutlu, A. K. Fuchs, V. Sabol, S. Thalmann, and T. Schreck, "Exploration of anomalies in cyclic multivariate industrial time-series data for condition moi-

toring," in *Proceedings of EDBT/ICDT Workshops*, 2020, pp. 1-8.
36. D. Rees, R. S. Laramee, P. Brookes, T. D'Cruze, G. A. Smith, and A. Miah, "Agentvis: Visual analysis of agent behavior with hierarchical glyphs," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, 2021, pp. 3626-3643.
37. PTT-web-crawler@ONLINE, https://github.com/jwlin/ptt-web-crawler.
38. J. Sun, "jieba: Chinese word segmentation tool," https://github.com/fxsjy/jieba, 2012.
39. M. Bostock, V. Ogievetsky, and J. Heer, "D3 data-driven documents," *IEEE Transactions on Visualization and Computer Graphics*, Vol. 17, 2011, pp. 2301-2309.
40. B. Shneiderman and C. Plaisant, "Treemaps for space-constrained visualization of hierarchies," http://www.ifs.tuwien.ac.at/~silvia/wien/vu-infovis/articles/shneiderman_tree map-history_1998-2009.pdf, 1998.

Chen-Chi Hu (胡臻騏) is currently pursuing a Ph.D. degree in the Department of Computer Science, National Chengchi University. His research interests include computer graphics, stylized rendering, and information visualization.

Chun-Min Liao (廖春閔) received the MS degree from the Department of Computer Science, National Chengchi University. His research interests include computer graphics and information visualization.

Ming-Te Chi (紀明德) received the MS and Ph.D. degree from the Department of Computer Science and Information Engineering, National Cheng-Kung University, Taiwan, in 2003 and 2008, respectively. He is currently an Associate Professor of the Department of Computer Science, National Chengchi University. His research interests include computer graphics, stylized rendering, and information visualization.