

Zhiguang Zhou · Xiangjun Zhu · Yuhua Liu · Qianqian Ren · Changbo Wang · Tianlong Gu

# VisUPI: visual analytics for University Personality Inventory data

Received: 30 December 2017 / Revised: 20 April 2018 / Accepted: 21 April 2018 / Published online: 9 July 2018  
© The Visualization Society of Japan 2018

**Abstract** It is a feasible way to investigate the mental health status of college students through University Personality Inventory (UPI). However, the analysis of original UPI data is a difficult and tedious task, because the traditional software always requires the users to select and reload slices of data by hand. Also, the limited constant functions provided by the software cannot meet the various requirements of domain experts. In this paper, we propose VisUPI, a specialized visualization framework for UPI datasets, aiming at an in-depth analysis of the mental health status of college students. A circular view is firstly designed to layout the questionnaires and visualize the answers of individuals. Then, a decision tree model is employed to classify the investigated students, and a radial hierarchy chart is designed to present the relationship of different groups of students. According to the prior knowledge of UPI, we restructure the network relationship of specified questionnaires, and use the force-directed model to layout the questionnaires, enabling users to better perceive the answers of those students with serious mental health problems. Furthermore, multi-dimensional scaling is used to visualize the dissimilarity between different questionnaires, and the statistical answers are presented through bar charts in detail. Finally, the effectiveness and scalability of VisUPI are demonstrated through case studies with the real-world datasets and the domain-expert interviews.

**Keywords** Visual analytics · UPI · Visual design · Mental health · Hierarchy visualization

## 1 Introduction

With the rapid popularization and development of higher education, more and more young people have the opportunity to enter universities for further study in China. According to the recent statistics, the amount of

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s12650-018-0499-x>) contains supplementary material, which is available to authorized users.

---

Z. Zhou · X. Zhu · Y. Liu  
Zhejiang University of Finance and Economics, Hangzhou, China  
E-mail: zhgzhou1983@163.com

Q. Ren  
Zhejiang University of Water Resources and Electric Power, Hangzhou, China

C. Wang  
East China Normal University, Shanghai, China

T. Gu (✉)  
Guilin University of Electronic Technology, Guilin, China  
E-mail: cctlgu@guet.edu.cn

college students is so large and increases rapidly year by year. The massive recruitment of Chinese universities brings new problems for the college students, such as academic stress, employment pressure, economical burden, affective interaction and so on, which easily affect their psychological health.

As the university is the reserve of talents, and the college students are the pillars of future society, the problem of mental health for college students receives more attentions from the governments and college managers. A potential feasible way is to survey the psychology health of the college students by means of questionnaire. Each college student should answer a list of specific questions shortly after enrollment, which responses the psychological status from different angles. A widely used questionnaire for Chinese universities is the University Personality Inventory (UPI), which is comprised of three parts, including the fundamental information, questionnaire and the additional questions. The psychologists are able to get insights into the mental conditions of college students from the answers of questionnaire, and take suitable measures for those with bad mental health status as soon as possible.

The current survey of UPI is transferred onto the Internet, which simplifies the tedious course of data collection. However, it is still difficult to quickly conclude the mental health status of the colleague students, especially when the amount of the students is so large, because the psychologists must export and analyze the questionnaire data based on the statistical software, such as Excel, Statistic Package for Social Science (SPSS), etc. The analysis ways of the software are limited, which cannot provide the psychologists with effective customized functions to meet some special user requirements. Also, the software always needs the users to select slices and reload data by hand, which is a tedious and time-consuming task, and largely reduces the efficiency of data analysis. Therefore, it is necessary to design a specialized system aiming at the analysis of UPI data, allowing the psychologists to quickly detect the colleague students with serious mental health problems.

In this paper, we propose a visual analytics system for a deep analysis of UPI data. Firstly, a circular view is designed to present the questionnaire data, in which the questions are depicted as points and distributed in a radial coordinate system. A Voronoi diagram is employed to partition the radial coordinate system, allowing users quickly perceive the answers of questionnaires. Then, a decision tree model is employed to predict the relationship between the fundamental information and groups of students with different mental health status, by means of which the psychologists can easily conclude the potential causes of mental health problems. A radial hierarchy chart is further designed to improve the perception of tree structure, allowing users to focus on a single group of students and perceive the distribution of different groups. In order to investigate the groups of students with serious mental health problems, a network is constructed based on the prior knowledge of UPI, and a force-directed layout scheme is applied to present the different levels of mental diseases. We also use the MDS to reveal the structured dissimilarity between different questionnaires, and achieve the detailed information of the UPI data based on the answer distributions and the statistics of the answers. Finally, the effectiveness and scalability of VisUPI are demonstrated through the case studies with the real-world datasets, and the domain-expert interviews.

The major contributions of our study are listed as follows:

- A circular view to present the UPI data for a college student, including the visual designs of 56 real questionnaires and 4 lie-detecting questionnaires, allowing the psychologists to quickly perceive the mental status of investigated individuals.
- A decision tree model and a radial hierarchy chart, to classify the students with different mental health status, allowing the users to conclude the relationship between the mental status and the fundamental information.
- A network construction of the specific questionnaires and a force-directed graph layout scheme, allowing users to get deeper insights into the student groups with serious mental health problems.
- A MDS view to present the difference between questionnaires, and a set of bar charts to reveal the answer distribution, allowing users to perceive the relationship between the questionnaires and categories of students.
- A visualization framework is provided to focus on the analysis of UPI data, in which the visual designs and data mining techniques are integrated to facilitate the psychologists to quickly understand the mental status of college students, and further provide decision-making basis for the diagnosis and treatment of the college students with mental health problems.

The rest of this paper is structured as follows: The related work is reviewed in Sect. 2. Section 3 presents the analysis tasks and the system overview. The visual designs and the visual analytics method are detailed

in Sect. 4. Case studies in addition to domain-expert interviews are discussed in Sect. 5 and finally our conclusions are drawn in Sect. 6.

## 2 Related work

In this section, we give a brief overview of the UPI analysis, the applications of visual analytics and the visualizations of hierarchy.

### 2.1 UPI analysis

A large amount of research works have been conducted to investigate the mental health status of college students by means of UPI analysis. Zhang (2007) take the UPI questionnaires to investigate the psychodiathesis of the 1106 students of Shandong Administration Institute, and conclude that the condition of mental and body health is generally fine, with some individuals requiring close attention, based on the statistical analysis by means of the software of Linear Structural Relations (LISREL). Aiming at the exploration of the mental health conditions of 2075 students in a college, Su et al. (2008) compare the UPI questionnaires in the last two years by means of the software of SPSS, and find the account of students who have mental health problems has increased with time, and the mental health education and consolation should be continuously enhanced. Chen et al. (2014) investigate the mental health status of medical students by means of UPI questionnaires and analyze the related influence factors including family, society and education based on the software of SPSS. Zhang et al. (2015) examine the dimensions of UPI to provide more information about the specific mental problems for students at risk, and conclude that the physical symptoms, cognitive symptoms, emotional vulnerability, social avoidance, and interpersonal sensitivity are significantly correlated with the symptom checklist. It can be seen that the related works of UPI analysis are comprised of two respective parts, including the collection of UPI questionnaires and the analysis by means of statistical software. The course of UPI analysis is fragmentary, and always requires the users to reload the data by hand, which largely reduces the efficiency of data analysis. In addition, the analysis of UPI is limited to the functions provided by software, which cannot satisfy some special requirements of various analysis tasks.

### 2.2 Visual analytics applications

Visualization is able to convey information via graphical representations, which makes the complex data more accessible, understandable and usable, allowing the users to quickly analyze and reason about data and evidence (Liu et al. 2014). As a research hotspot in the field of visualization, visual analytics is especially concerned with coupling interactive visual representations with data mining techniques, which can be widely applied to analytic reasoning to support the sense-making process for various applications (Sacha et al. 2014). For example, Scholz and Lu (2014) propose a visual analytics tool to capture the dynamics of spatio-temporal evolution features for the urban activity patterns. Shi et al. (2016) propose a matrix-based visualization system aiming at a visualized forensic analysis on the unintelligible traffic datasets. Ferreira et al. (2013) propose a novel visual analytics model, allowing the domain experts to get insights into the mobility across cities by means of origin-destination queries. Wang et al. (2014) present a visual analysis system to explore sparse traffic trajectory data recorded by transportation cells. Ma et al. (2016) employ an Eulerian approach to analyze the crowd flows based on mobile phone data aiming at the exploration of human movement citywide. Zhou et al. (2015) introduce a visual analytic system named ENTVis allowing users to perceive the entropy-based traffic metrics and conduct an accurate traffic anomaly detection. Xia et al. (2017) propose a novel dimension relevance measure to indicate the cluster significance in the corresponding subspace and design a hyper-graph to visualize the internal structures of subspaces. He also proposes a novel dimensionality reduction method to present the distribution of pointwise LTS ( $x$  axis) and the variation of LTS in structures (the combination of  $x$  axis and  $y$  axis) (Xia et al. 2018). Zhao et al. (2014) present a visual analytics system called MVSec, enabling users to better understand the information flows in the condition of network security. Wang et al. (2017) design a visual interface to help users gauge utility loss while interactively and iteratively handling privacy issues of the original data, in which some widely known privacy models are integrated and compared under different use case scenarios. Zhang et al. (2014) design a visual analytics method to detect various issues of city utility services, and the visual aggregations

are further provided to transform numerous issues into legible events. Turkey et al. (2014) propose an interactive graphical model to present geographic variability of statistics of multiple attributes. Qu et al. (2007) present a visualization tool for the analysis of weather data, in which the enhanced parallel coordinates, pixel bar charts and weighted complete graphs are integrated to help domain experts investigate the laws and causes of air pollution problem in Hong Kong. Chen et al. (2016) propose an expressive visualization scheme for the multi-media NBA datasets, allowing the users to illustrate NBA games with multiple levels of details, such as season level, game level and session level. Yang et al. (2017) propose an integrated visual analytics framework to study the problem of blockwise brain network visual comparison based on the block information on the region of interest (ROI). In this paper, we also design a visual analytics framework for the psychologists to investigate the mental health status based on the UPI questionnaires.

### 2.3 Hierarchy visualization

A large number of tree visualizations have been proposed to present the hierarchical structures of data, which can be divided into different categories, such as the node-linked and space-filled. In the category of node-linked tree visualization, the relationship of nodes is represented by links. In order to optimize the visual perception of the hierarchical structures, various layout schemes are proposed, such as cone trees (Robertson 1991), hyperbolic geometry trees (Lamping et al. 2002), phyllotrees (Neumann et al. 2006) and point tree (Schulz et al. 2009). Although the node-linked tree is able to present an intuitive visualization of hierarchical structures in the data, it leaves the background space empty and takes up much space to distribute the nodes and links, which will be limited when it is applied to visualize a large-scale dataset (Landesberger et al. 2011). The space-filled tree visualizations are proposed to visualize the hierarchy, which make use of the spatial relations to reveal the hierarchical structures, such as adjacency and enclosure. In adjacent-based methods, the child nodes are distributed next to parent nodes in different layers, such as linear layers referenced in Landesberger et al. (2011) and sunburst layers referenced in Stasko and Zhang (2000). Different from the adjacent-based schemes, enclosure methods distribute the child nodes in the area of parent nodes. As a good example of enclosure method, treemaps display the child nodes by means of space-filled rectangles, which recursively subdivide the areas of parent nodes (Van Wijk and Huub 1999). As variances of the treemaps, Voronoi diagrams (Balzer et al. 2005) and bubble layouts (Bederson 2001) display the hierarchy by means of the other specified space-filled diagrams. The visualization of hierarchy has received much attention in the information visualization community, and the state of art for the display of trees is referenced in Landesberger et al. (2011). In this paper, a decision-making tree is constructed to relate the basic information and student categories with different mental health problems, and an improved sunburst display of the decision-making tree is designed to help users quickly perceive the internal features of the UPI data.

## 3 Problem characterization

In this section, the sources and structure of UPI data are firstly introduced. Then, a list of analytical tasks are characterized after the detailed discussions with domain experts. The pipeline of our visualization system is further presented, aiming at the completion of the specified data analysis tasks.

### 3.1 Data abstraction

In 1966, the UPI questionnaires were produced by the members of Japanese University Health Management Association, including the experienced Psychological consultants and psychiatrists. The UPI was brought into China by Professor Fan in 1991, which was further revised and used to discover the college students with mental health problems.

The UPI data is composed of three parts. The first part is the basic information of the investigated students, including the gender, family, parents, etc. The second part is the questionnaires, a total of 60 subjects, 56 of which can reflect the mental health status of the students from different perspectives, such as anxiety, distress, depression, interpersonal sensitivity and conflicts. The other 4 questionnaires are the pseudo measurement subjects, to identify the reliability of answers. The third part is the auxiliary problems,

which helps the psychological doctors to further evaluate and diagnose the mental health status of the investigated students.

According to the prior knowledge of UPI, a number of rules are defined to help psychologists quickly identify students with various mental health status. All investigated students can be classified into three categories based on the feedbacks of UPI. The first category is comprised of the students with serious mental health problems, including those the total scores of whom are larger than 25, or those the 25th questionnaire of whom is chosen as yes, or those more than 2 auxiliary problems of whom are answered as yes, or those putting forward specified requirements. The second category is comprised of the students with minor mental health symptoms, such as those the total scores of whom are between 20 and 25, or those one of the 8th, 16th, and 26th questionnaires of whom is answered as yes, or those one of the auxiliary problems of whom is answered as yes. The rest other than the above two categories of students are classified into the third category, a group of students without mental health problems.

In this paper, about 800 students from a common college of Zhejiang province in China are investigated by means of UPI. The basic information, the answers of questionnaires, and the categories of students are collected for psychiatrists to evaluate and diagnose the mental health status of the students. As usual, the psychiatrists should import the UPI data into some statistical software and conduct specific data analysis, which is a tedious and time-consuming task, often requiring the users to select slices and reload data by hand. Therefore, it is difficult enough for domain experts to perceive the mental health status of the investigated students and conclude the causes of mental health problems from various aspects of UPI data.

### 3.2 Task analysis

Aiming at a comprehensive and efficient analysis on the UPI data, we discussed with the domain experts in detail and characterized a list of analytical tasks as below.

**R.1 Individual visualization.** How to layout the UPI answers for an individual college student from different perspectives? How to present the category information based on the prior knowledge of UPI data? How to identify the credibility of UPI answers according to the specialized fake items?

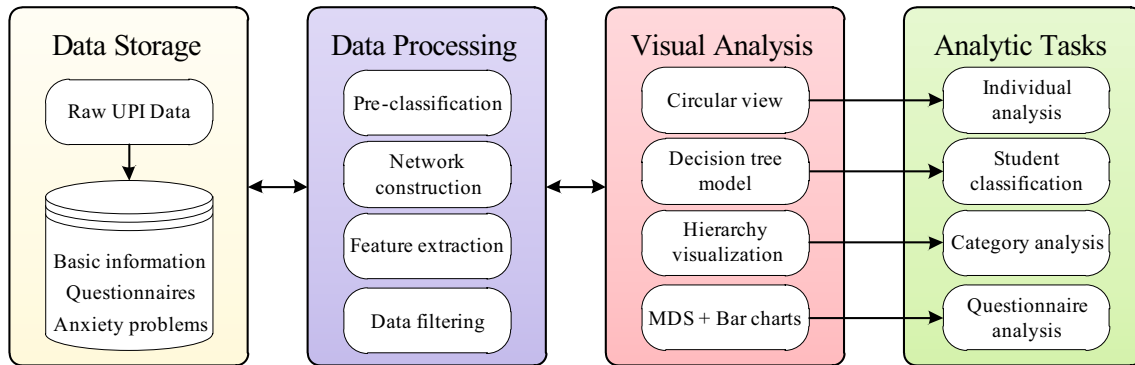
**R.2 Student classification.** Are the mental health problems related to the categories of college students based on the fundamental attributes? Which is the most relevant attribute that highly influences the mental health of students? How to construct a reasonable classification to help the domain experts quickly predict the mental health status of an investigated student?

**R.3 Category visualization.** How to provide a visual design for the student classification, allowing users to easily perceive and interact with? How to present the questionnaires for a category of interest? How to identify the causes of mental health problems for a category of interest?

**R.4 Questionnaire analysis.** How to provide convenient interactions for users to analyze the questionnaire of interest? How to measure the difference between questionnaires and visualize the dissimilarity of the questionnaires? How to present the detailed answer distribution for the questionnaire of interest?

### 3.3 System overview

Figure 1 presents an overview of the VisUPI system. After a set of data preprocessing operations on the original UPI data, three interfaces are provided for users to conduct an in-depth analysis, including the individual visualization, category visualization and questionnaire analysis. A circular view is designed to present the mental health status of individuals, by means of which end users can quickly perceive the categories of college students and their detailed UPI answers (R1). The decision tree model is employed to classify the students into different categories. A hierarchy chart is designed to indicate the relationship of students and a network visualization is presented, allowing the psychiatrists to conclude the major causes of the mental health problems (R2), and resort to effective treatment measures for the students with serious mental diseases (R3). In order to analyze the relationship of different questionnaires, a MDS view is provided to reveal the dissimilarity between questionnaires, and a stacked bar chart view is applied to present the detailed answer statistics information (R4). The above visualizations are highly related with the clues of interest, such as a questionnaire, a student or a category of students. The combination of all the views largely enables users to comprehensively analyze in-depth features of UPI data.



**Fig. 1** The workflow of the VisUPI

## 4 UPI visualization

This section describes a set of visualization schemes, allowing domain experts to get into deeper insights into the UPI data from different perspectives.

### 4.1 Individual visualization

Users are desired to comprehensively visualize the questionnaires and quickly perceive the mental status of an individual student in a synthetical view. Therefore, a circular view is designed for the individual visualization to layout the questionnaires, the top of which presents the questionnaires related to body problems, while the left bottom and right bottom of which, respectively, present the questionnaires related to anxiety and interpersonal sensitivity problems. In this view, the questionnaires are abstracted as points, and laid out uniformly in their corresponding sector regions. In order to guarantee the esthetics of the visualization of questionnaires, the abstracted points are distributed in the layers with different radii. Furthermore, the radii of the layers closed to the center are larger, while the radii of the layers far away from the center are smaller. The accounts of questionnaires located in the layer close to the center are smaller, while the accounts of questionnaires located far away from the center are larger, which makes the layout of questionnaires as uniform as possible.

According to the distribution of questionnaires, a Voronoi diagram is employed to segment the sector regions into different polygons. If the answer of a questionnaire is yes, the corresponding polygon is colored in red, or else the corresponding polygon is colored in fundamental colors, including light green, yellow and light red. The mental health status can be easily perceived by means of the distinct color mapping. Based on the prior knowledge of UPI, the students can be classified into three categories. For example, a student whose total score is larger than 25 is considered to have serious mental health problems. Therefore, the category that the student belongs to is automatically calculated, which is further used to map a meaningful color for the inner circle in the visual design of individuals. If the student belongs to a category with serious mental health problems, the inner circle is colored in red. Similarly, the inner circle is colored in blue when the student belongs to a category with minor mental health symptoms, and is colored in green for another category.

There are 4 fake questionnaires distributed outside of the inner circle, which are employed to measure the reliability of the UPI. The fake questionnaires and the normal questionnaires are intentionally compiled from opposite perspectives. If a student obtains higher score of normal questionnaires, while also obtains higher score of fake questionnaires, his questionnaires are considered as unbelievable items, and he will be required to conduct the UPI investigation once again. In order to help users quickly identify the unbelievable items from the UPI data, two additional circular bars are designed to surround the circular view of individual visualization. The outer bar colored in red is applied to quantify the score of normal questionnaires, and the inner bar colored in blue is applied to quantify the score of fake questionnaires. Therefore, the difference between two bars can be easily perceived and further used to change the color of the inner circle. If the difference is smaller than a user-specified threshold, the inner circle will be colored in gray to remind the users that the questionnaires are unbelievable.

Figure 2 presents the comparable questionnaire visualizations for three students. (a) Presents the questionnaire answers of a healthy student without mental health problems. A small number of red colors appears in the sectors of the circular view, and the inner circle is shaded in green based on the prior knowledge of UPI. (b) Presents the questionnaire answers of a student with serious mental health problems. A large number of polygons are shaded in red, especially the sector of anxiety. According to the prior knowledge of UPI, the inner circle is shaded in red. (c) Presents a set of unbelievable questionnaire answers. We can find that the fake questionnaires are answered with “no”, while the normal questionnaires are also answered with “no”. The difference of the circular bars surrounding the circular view is little, and the inner circle is shaded in gray, which further indicates that the answers of the student are unbelievable.

#### 4.2 Student classification

Based on the individual UPI analysis, the students can be classified into different groups, including those with serious mental health problems, those with minor mental health problems and those without mental health problems. Also, the college students belong to different categories according to the basic information, such as gender, family, parents, and so on. The domain experts are desired to find the relationship between the categories and the basic information, which, however, is hardly explored by means of traditional software.

In order to get deeper insights into the causes of mental health problems, a decision tree is employed to predict the relationship between the basic information and the different mental status. We make use of the traditional ID3 algorithm to construct the decision tree (Quinlan 1986), in which information gain is applied to measure the purities of branches, guiding a greedy course of data classification. The calculation of information gain is defined as follows:

$$IG(S|T) = Entropy(S) - \sum_{value(T)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{1}$$

where  $S$  is the raw dataset of UPI.  $value(T)$  is the values of attribute  $T$ . For example, man and male are the two values of attribute gender.  $|S_v|$  is the amount of the college students with value  $v$ .  $Entropy(S)$  is the entropy of the original dataset, which is defined as follows:

$$Entropy(S) = - \sum_{i=1}^n p_i \log_2 p_i \tag{2}$$

where  $n$  is the amount of all investigated students.  $p_i$  is the probability of the  $i$ th category of students from  $S$ .  $Entropy(S_v)$  is the entropy of the college students with value  $v$ , which is defined as follows:

$$Entropy(S_v) = - \sum_{i=1}^m p_i \log_2 p_i \tag{3}$$

where  $m$  is the amount of the students with value  $v$ .  $p_i$  is the probability of the  $i$ th category of students from  $S_v$ .

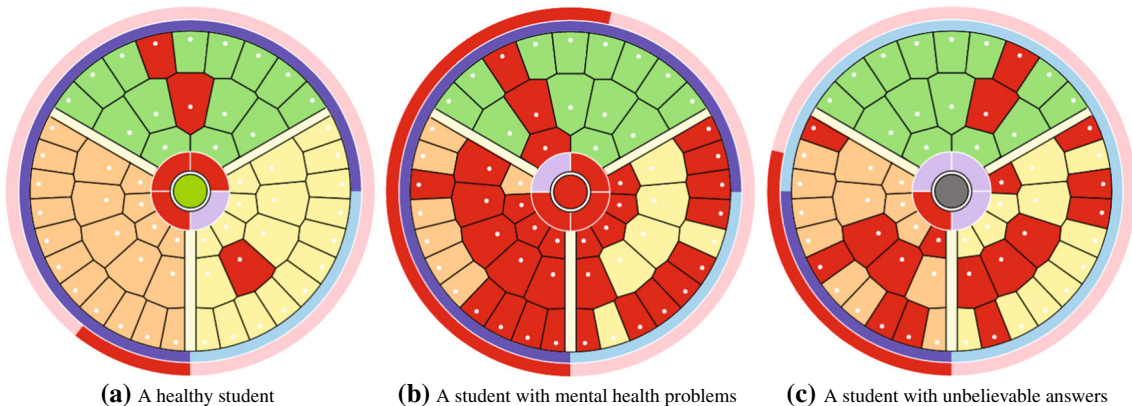


Fig. 2 The comparable questionnaire visualizations for three students of interest

In each step of classification, a prominent attribute with maximum IG value is selected to divide the leaf nodes. Therefore, the decision tree is greedily constructed in a top-down manner, and the prior attributes are considered as important factors, which might be highly related to mental health status.

### 4.3 Category visualization

Aiming at the presentation of the decision tree, a node-linked tree visualization is shown in Fig. 3a. Although the links are able to reveal the relationship of nodes, and the hierarchy of the decision tree can be visually obtained, plenty of backgrounds are left empty and some valuable information like the proportion distribution is missed in the node-linked diagram.

In order to help users better perceive the proportion distribution of each categories of students, a space-filled radial hierarchy diagram is employed to visualize the hierarchy of decision tree, as shown in Fig. 3b. The rings close to the center present the higher level leaf nodes, while the rings far away from the center show the lower level leaf nodes. The angles of different areas are applied to map the amounts of students in the corresponding categories. The colors of areas are mapped according to the proportion of different categories of the corresponding leaf nodes, which is calculated as follows:

$$\text{Hue} = \alpha * p_1 * \text{Hue}_1 + \beta * p_2 * \text{Hue}_2 + \gamma * p_3 * \text{Hue}_3 \tag{4}$$

where Hue is the hue of current leaf node, which is weighted by the hue values of different categories, the proportions of which are  $p_1, p_2$  and  $p_3$  in the current leaf node.  $\text{Hue}_1, \text{Hue}_2$  and  $\text{Hue}_3$  are the normalized hue values of the specified colors, including blue, yellow and red. Therefore, the distribution of different categories of students can be easily perceived based on the normalized Hue. For example, the color of the leaf node is mapped close to red, when the category of students with serious mental problems occupies a large part of the leaf node.

In addition to the angles and colors, we further improve the sunburst diagram with an internal pie chart and a logarithmic operation. Although the radial diagram is able to show the hierarchy of decision tree, it cannot present the distribution of three categories of students, especially when the leaf node is not pure and would not be further divided. In order to accurately present the composition of each leaf node, a pie chart is designed to reveal the proportions of different categories and laid out in the center of the sunburst diagram. By default, the pie chart shows the proportions of different categories of all investigated students. When a leaf node is interactively specified by users, the pie chart will be updated with the corresponding proportions of the leaf node. Figure 3b shows the pie chart for a specified category of interest. Another limitation of the radial hierarchy diagram is that it is always difficult to identify the detailed distribution of a leaf area due to the huge difference between the accounts of categories. Therefore, we reduce the original accounts of categories into logarithmic values, which will also retain the difference between categories at the expense of accuracy. Figure 3c shows the logarithmic distribution of radial hierarchy diagram, allowing users to better perceive the distribution of categories. The actual proportion values can be obtained by means of the interactive suspended prompts.

Based on an overview of the decision tree, we also provide a set of visual designs to focus on a leaf node of interest. In addition to the individual visualization, a gradual color mapping is applied to present the

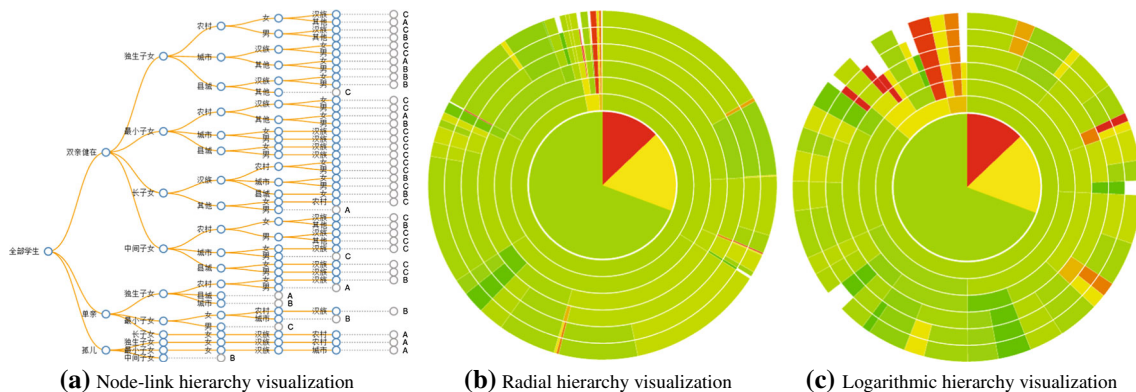


Fig. 3 The comparable hierarchy visualizations for a decision tree

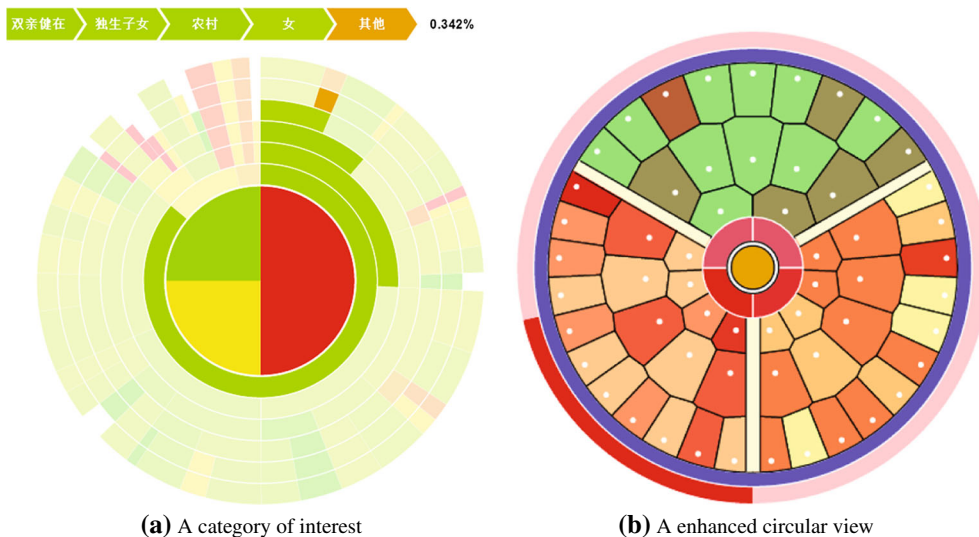


answer distribution. If a questionnaire is answered as yes by most of the students, the color of the corresponding polygon is close to red, or else the color is close to fundamental colors. Figure 4 shows an enhanced visualization of the sunburst diagram for a category of interest, by means of which the users can easily perceive the overall mental status of the investigated students.

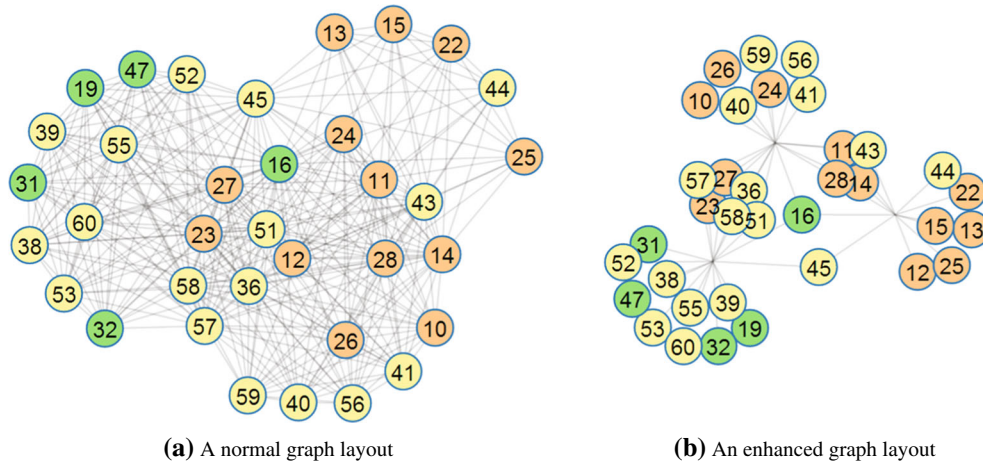
In order to get deeper insights into the categories with serious mental health problems, a set of questionnaires revealing the mental status of investigated students from various perspectives are specified according to the prior knowledge of UPI. The 16th, 19th, 23th, 27th, 31th, 32th, 36th, 38th, 39th, 45th, 47th, 51th, 52th, 53th, 55th, 57th, 58th, 60th questionnaires are considered highly related to neuroses. The 11th, 12th, 12th, 14th, 15th, 16th, 22th, 25th, 28th, 43th, 44th, 45th questionnaires are highly related to depression. The 10th, 11th, 14th, 16th, 23th, 24th, 26th, 27th, 28th, 36th, 40th, 41th, 43th, 51th, 56th, 57th, 58th, 59th questionnaires are related to schizophrenia. Therefore, the categories with serious mental health problems can be further divided into another three categories, including those tend to neuroses, depression and schizophrenia. Due to the irregular distribution of the specified questionnaires, it is difficult to perceive the mental status of an individual or a category of interest. To visually present the specified questionnaires as well as their relationship, a questionnaire network is constructed with each questionnaire abstracted as a node. The questionnaires in a category are linked to each other through edges. Figure 5a shows the layout of three categories of specified questionnaires. In order to better visualize the relationship of the questionnaires and further help the user quickly identify the mental status of the individual or group of interest, we update the questionnaire network with three virtual central nodes. In each questionnaire category, questionnaires are all linked to the corresponding central node, without linking to each other. The force-directed layout of three questionnaire categories is able to better present the specified questionnaires as shown in Figure 5b, which largely supports the visual analytics of various mental health problems.

#### 4.4 Questionnaire analysis

The prior knowledge largely navigates the exploration of UPI data. For example, the investigated students can be classified into different categories, and the questionnaires are specified to get insights into the mental status of students from different perspectives. In addition to the prior knowledge of UPI, the psychiatrists are also desired to find the dissimilarity between the questionnaires and explore the detailed information of questionnaires. Therefore, a commonly-used multi-dimensional scaling (MDS) method is applied to present the difference between questionnaires, with the dissimilarity matrix  $D$  evaluated as follows:



**Fig. 4** The enhanced visualization of the sunburst diagram for a category of interest, **a** presents a category selected by experts and **b** shows the answer distributions for the category, in which most students in the category are troubled by mental health problems, especially for the sectors of anxiety and interpersonal sensitivity



**Fig. 5** The comparative visualizations for a questionnaire network based on a set of specified questionnaires. **a** The questionnaires belonging to a same group are linked to each other. **b** The questionnaires belonging to a same group are linked to a virtual node, without linking to each other. Although the layout results enables users to find the groups of specified questionnaires, there are still some questionnaires misleading the visual perception of questionnaire groups, such as 24th item in (a)

$$D = \begin{pmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,n} \\ d_{2,1} & d_{2,2} & \dots & d_{2,n} \\ \dots & \dots & \dots & \dots \\ d_{n,1} & d_{n,2} & \dots & d_{n,n} \end{pmatrix} \quad (5)$$

where  $n$  is the amount of questionnaires, which is comprised of 60 questionnaires in the proposed system, including 56 normal questionnaires and 4 fake questionnaires.  $d_{i,j}$  is the distance between questionnaire  $i$  and questionnaire  $j$ , which is calculated as follows:

$$d_{i,j} = \frac{M - m_{i,j}}{M}$$

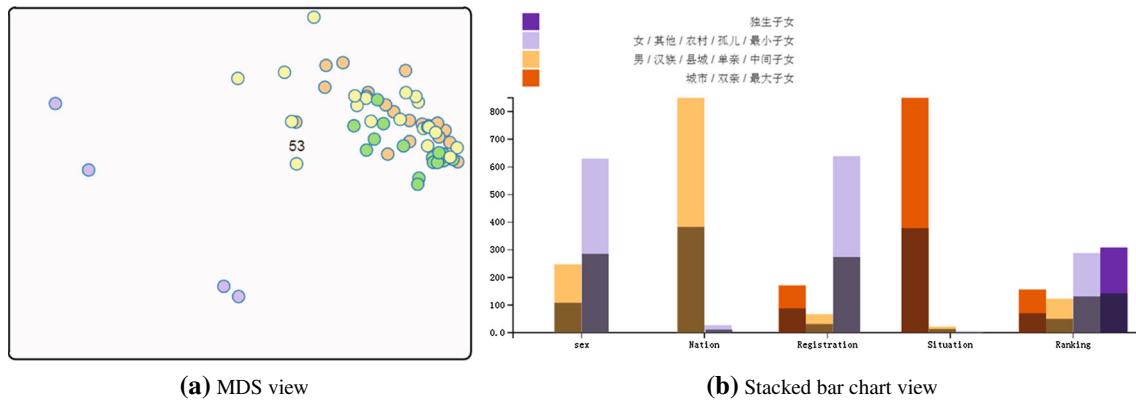
where  $M$  is the total amount of the investigated students.  $m_{i,j}$  is the amount of students who deal with the questionnaire  $i$  and questionnaire  $j$  with the same answers.

In order to visually present the dissimilarity between questionnaires, MDS is applied to project the questionnaires into 2D coordinates, with an objective function defined as follows:

$$\min_{x_1, \dots, x_n} \sum_{i < j} (\|x_i - x_j\| - d_{i,j})^2$$

where  $x_1, \dots, x_n$  are the optimal 2D vectors of the  $n$  questionnaires, obtained by means of an eigenvalue decomposition method. Figure 6a presents the MDS results, in which each point represents an original questionnaire, and the distance between different points largely indicates the dissimilarity between the questionnaires.

Given the dissimilarity between different questionnaires, we provide an interaction for users to focus on a questionnaire of interest, and then a stacked bar chart diagram is designed to present the detailed distribution of answers across different categories. Figure 6b shows the answer distribution of the questionnaire of interest. The different colors corresponds to various kinds of fundamental information, such as sex, nation, registration, et al. The heights of colorful bar charts presents the distribution of all students, while the heights of black bar charts indicates the properties of students who answer the questionnaire as “yes”. Therefore, the detailed answer distribution of questionnaires can be easily captured by means of the stacked bar charts, which further enable the experts to get deeper insights into the UPI dataset.



**Fig. 6** The visualization results of questionnaire analysis. **a** Presents the difference between questionnaires. **b** Presents the detailed answer distributions of questionnaires

### 5 Experimental results

The entire system was developed using D3.js. We tested the proposed framework and visualization techniques on a computer with 3.60 GHz Intel(R) Core(TM) i5-3470 CPU and 8GB memory. The interface of VisUPI is shown in Fig. 7. In this section, the usability of our system is evaluated by means of case study, expert interview and discussion.

#### 5.1 Case study

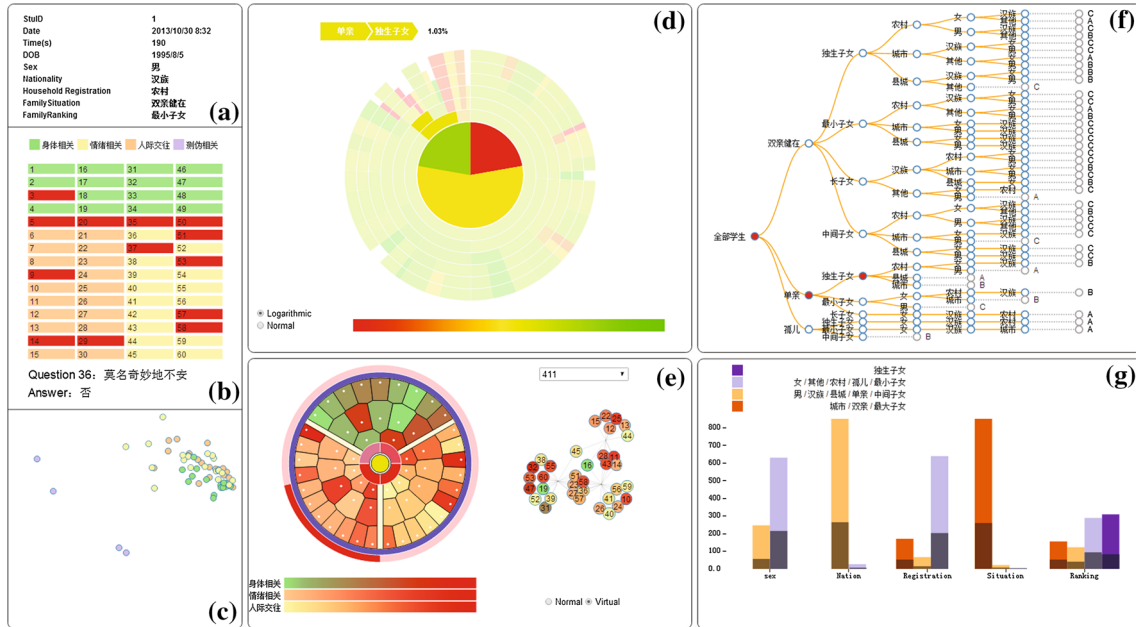
We conducted the case studies with the domain experts who are familiar with VisUPI, including the visual designs and the UPI datasets.

##### 5.1.1 Mental health status exploration

In this scenario, the experts used the VisUPI to explore the mental health status of the college students.

*Exploring individual questionnaires* The experts sought to quickly perceive the mental health status of an individual student (R1). To this end, they interacted with a pull-down menu, and selected a student of interest. The original list of questionnaires is shown in Fig. 8a, through which the experts achieved the answers of questionnaires. Figure 8b presented the circular design. The experts easily concluded that the answers of the student were unbelievable, since the difference of the two circular bars was small enough, and the color of inner circle was gray. Then, the experts selected another student of interest, and the corresponding circular view is shown in Fig. 8c. The inner circle was shaded in red, which indicated that the student belonged to a category with serious mental health problems. In addition, a large amount of polygons in the left bottom of circular view were shaded in red, which indicated that the mental health problems were largely resulted from anxiety factors. In order to further identify the severity of mental health problems for the student, the experts observed the force-directed layout of the specified questionnaires, to analyze the levels of mental symptoms from different perspectives. As shown in Fig. 8d, most of the answers were distributed in the community of neuroses, indicating that the student was affected with serious neurological symptoms, and required the psychologists to conduct corresponding treatments.

*Category analysis and visualization* The experts sought to observe the global distribution of investigated students according to their mental diagnosis and fundamental information (R2). To this end, they first loaded an original dataset into the system, including the fundamental information of the students and their UPI answers. A decision tree model was conducted to classify the students with different mental health status, and construct a hierarchical structure for the different levels of categories. From the hierarchy shown in Fig. 9a, the experts identified that the mental health problems were largely related to the completeness of family. According to the hierarchy visualization (R3), the experts easily concluded that the category in which the students are living in the families with single parent, ranking elder, female, Han nationality and



**Fig. 7** The system interface of VisUPI. The basic information of the investigated student is presented in (a). The 56 insightful questionnaires and 4 fake questionnaires are listed in (b). The MDS view is presented in (c) to reveal the difference between questionnaires. A node-link diagram in (f) and a radial hierarchy view in (d) are visualized accordantly to depict the decision tree model of student classification. The circular design and network visualization are presented in (e) and the stacked bar charts are shown in (g)

born in the countryside, always presented poor mental health status. From the radial hierarchy chart shown in Fig. 9b, the experts caught the distribution of different categories according the sizes of angles. Based on the synthetic colors, the experts obtained the overview of mental health status, and captured the purities of different categories. The logarithmic scheme was also used to enhance the visual perception of radial hierarchy diagram, which largely helped the experts to perceive and interact with the categories of interest. In addition, the experts sought to conduct an in-depth analysis of the category. They selected the corresponding region, the inner pie chart of sunburst diagram presented the proportions of students with different mental health status. In the meanwhile, the circular view and the network view showed the answer distribution as shown in Fig. 9c, d, in which the colors of regions were mapped gradually between the fundamental colors to red. It can be seen that the mental health status of the category was poor enough, which was largely caused by anxiety and interpersonal communication problems.

5.1.2 Questionnaire analysis

This case is aimed to demonstrate the effectiveness of the analysis of questionnaires.

*Dissimilarity detection* The experts sought to find whether the questionnaires presented similar features according to the answers of college students (R4). Figure 10a presented the questionnaire list for a student of interest. When the experts clicked a questionnaire, the question and corresponding answer were detailed at the bottom of questionnaire list. Based on the answers of investigated students, the dissimilarity matrix was constructed and the difference between questionnaires were conveyed through the distance of points in the MDS view as shown in Fig. 10b. The experts found that four fake questionnaires were distributed far away from other questionnaires. The questionnaires related to body, anxiety and interpersonal sensitivity problems were distributed within each other, without showing remarkable clustering features. A large number of questionnaires were distributed close to each other, which meant that they shared similar answers. As one of these questionnaires, the description of 55th questionnaire was “Do you feel yourself smell strange?”. It was an abnormal questionnaire that most of the students answered it as “no”, so that the corresponding MDS projected point was close to other similar questionnaires.

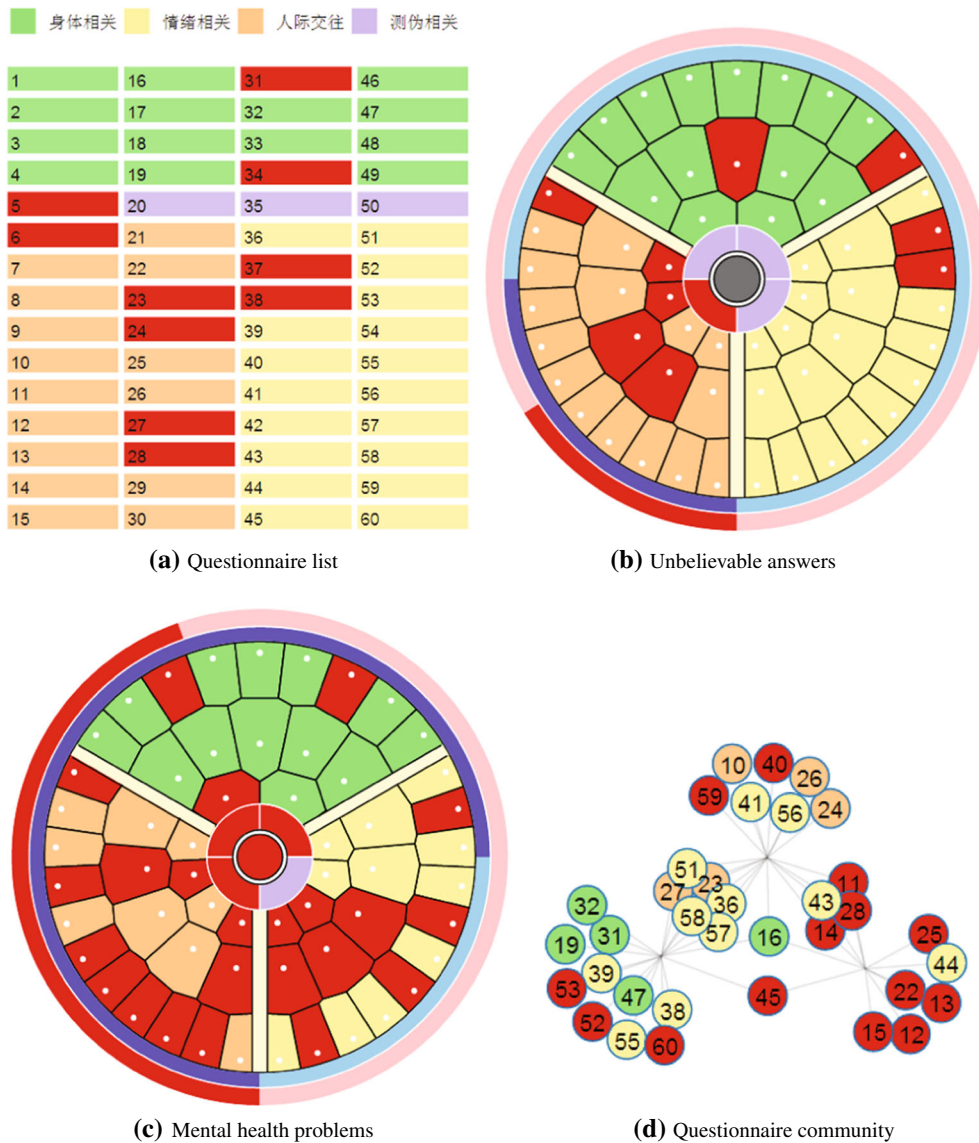


Fig. 8 The visualizations for the exploration of individual questionnaires

*Answer distribution* The experts sought to explore the answer distributions of the questionnaires. After the original UPI data were loaded into the system, traditional bar charts were presented to indicate the distribution of investigated students from different perspectives. When the experts selected the questionnaire of interest, the stacked bar charts were constructed to reveal the proportion of students whose answers were yes for the questionnaire. Figure 11a shows the answer distribution of the 55th questionnaire. The experts found that few students answered this questionnaire as “yes”, and the black stacked bars were lower enough in each multiple attribute bar charts. As a difference, Fig. 11b shows the answer distribution of 20th questionnaire, which was a fake item depicted as “Do you always be full of life?”. Most of the students answered this questionnaire as “yes”, and the black stacked bars were higher enough.

### 5.2 Expert interview

In order to further evaluate the effectiveness of VisUPI, we present our system interface and user studies to domain experts and collected their feedback based on one-on-one interviews. One of them is working in a psychological institution (Expert A), whose interests include the mental health of college students. Another

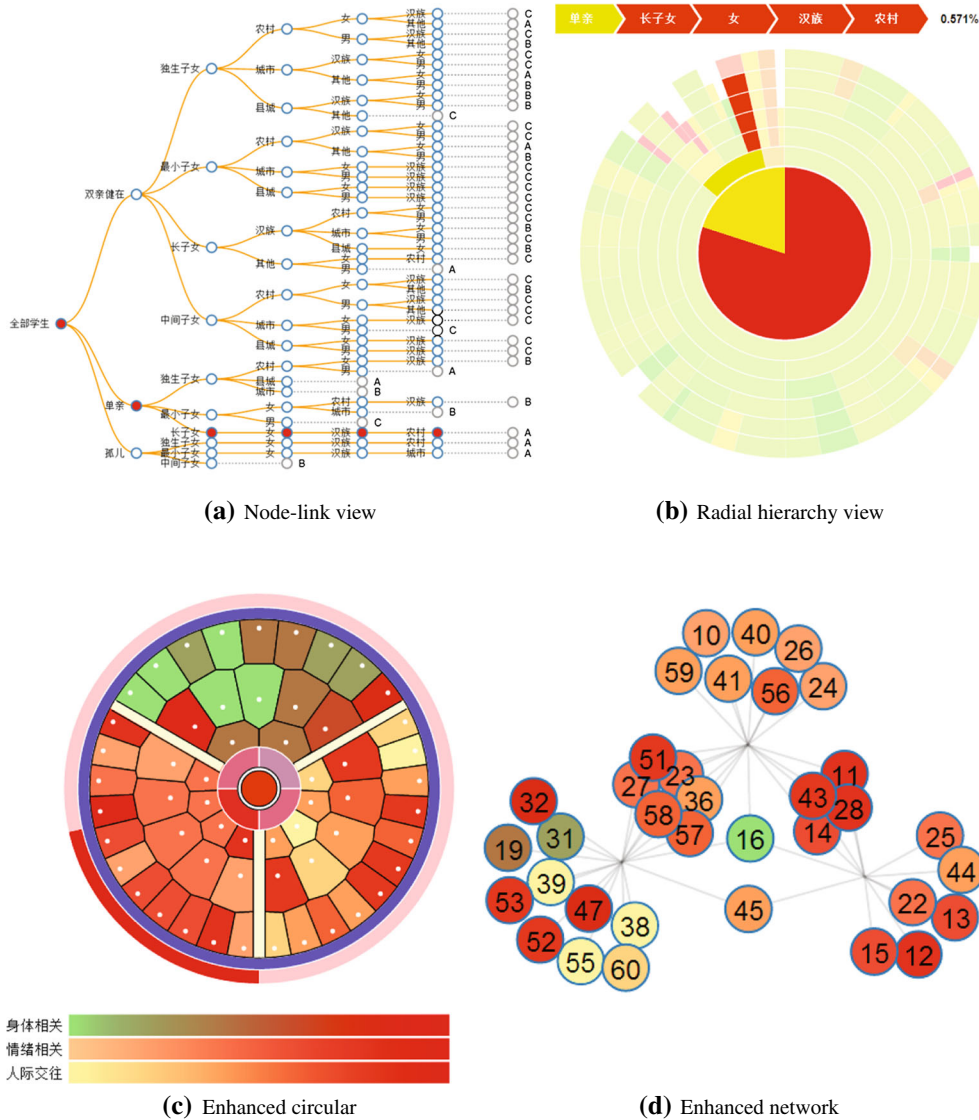


Fig. 9 The visualizations for category analysis

is a professor from a university (Expert B), whose interests include statistical analysis and data mining. He is familiar with various kinds of statistical software and also used SPSS to conclude from massive questionnaire datasets.

5.2.1 Visual design and interactions

Two experts both confirmed that VisUPI was well designed and the interface was quite user-friendly. The visual designs of the decision tree model and questionnaires received high praises from the users, especially those psychologists with much domain knowledge. They believed that the system could be easily applied by users for the exploration of UPI data. Expert A commended that “It is complex and time-consuming to extract valuable information from the UPI dataset. This system is able to integrate decision tree model and visual designs, allowing users to quickly classify the students and perceive the answers of questionnaires.” Expert B praised the user interactions of our system and commended that “The visualization windows are well associated based on a set of meaningful relations, such as questionnaire ID, student ID and categories, enabling users to easily interact with the system.”

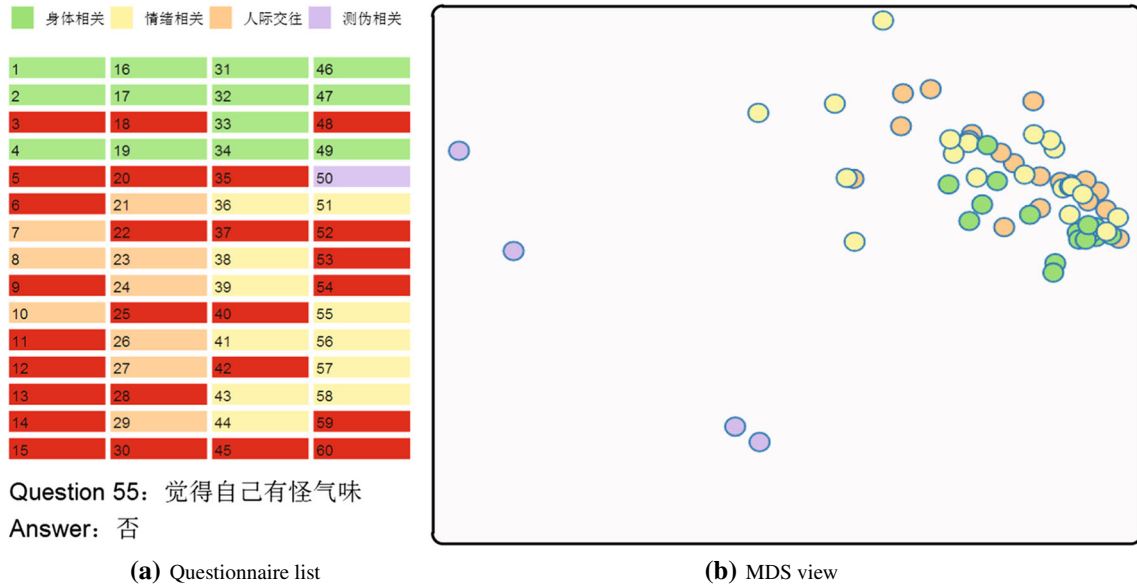


Fig. 10 Dissimilarity visualization for questionnaire analysis

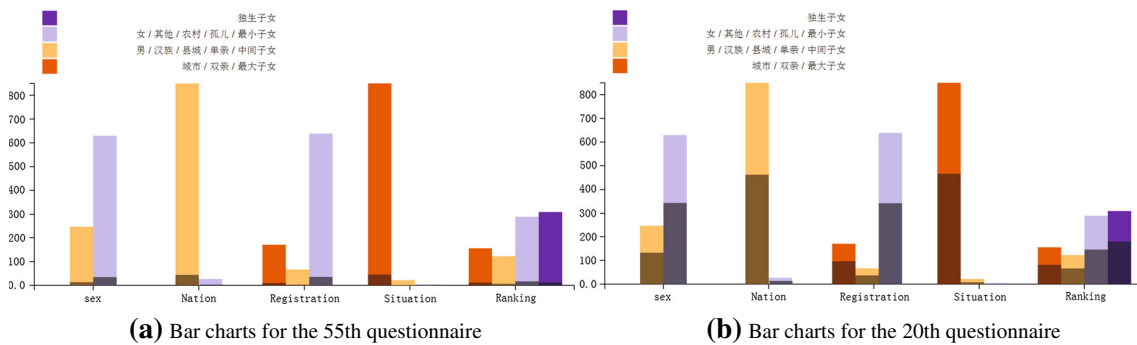


Fig. 11 The comparable results of answer distribution for the questionnaire analysis

### 5.2.2 Usability and improvements

The experts also appreciated the functions provided by VisUPI. They all agreed that the system enabled users to visualize the distribution of answers, perceive the questionnaires of students and analyze the potential causes of mental health problems. In addition, the experts also provided some valuable comments for the improvements of our system. Expert A mentioned that “Apart from the objective questionnaires, there are still some subjective items. The answers of subjective questionnaires are recorded by texts, which will be good supplements for the diagnosis of the mental health problems of students.” Expert B mentioned that “The functions provided in current system are dependent on the prior knowledge of UPI. In fact, the usability of UPI for the diagnosis of mental health is uncertain, and the questionnaires should be updated in accordance with times. It is responsible for the system to find the improper questionnaires and further helps the psychologists refine all aspects of the UPI.”

### 5.3 Discussion

The effectiveness of VisUPI is well demonstrated by case study and expert interviews, while there is still some space for improvement. A decision tree model is employed to relate the basic information and student groups with different mental health status. Nevertheless, the kinds of basic information is limited, which will easily generate a set of impure leaf nodes. In the cases, the classification of students will be uncertain, which

will disturb the accuracy of the diagnosis for the students with different mental health status. A possible solution for the problem is to gather more information in the course of UPI data collection, which might be highly related to the mental health problems. Therefore, we can construct the decision tree model with more levels, to enable the accurate classification of students.

MDS is used to quantify the difference between questionnaires, and the stacked bar charts are applied to present the distribution of answers from different perspectives. However, the complex features of questionnaires for a category of students are ignored in current system. The analysis of questionnaires for different categories of students will be helpful for domain experts to deduce the causes of mental health problems. A possible solution for the problem is to quantify the difference between questionnaires across various of groups and find the laws of answers for the students with different mental health problems.

The diagnosis of different mental health problems is based on the answers of 56 objective questionnaires, according to the prior knowledge of UPI. However, the current questionnaires are limited to insight into the mental health status of students. Various kinds of questionnaires are required to reveal the mental conditions of students, such as the subjective questionnaires. Therefore, the system should provide a set of functions to support the processing and analysis of textual datasets. With a more comprehensive investigation of students, the VisUPI will play more important roles in the analysis of UPI, enabling the psychologists to perceive the mental health status of students more quickly and accurately.

## 6 Conclusion

In this paper, we propose a visual analytics system VisUPI for the in-depth exploration of original UPI data. A circular view is designed to visualize the questionnaires of individuals. The layout of questionnaires and the corresponding colors indicate the mental health statuses of students. We also make use of a decision tree model to classify the students based on their answers of questionnaires and fundamental information. A radial hierarchy diagram is designed to present the hierarchical structures of decision tree. When a group of students is focused, the circular view will present the answer distributions by means of gradual colors. In addition, a network is constructed based on the relationship of the specified questionnaires, and a force-directed graph layout method is applied to enhance the visual perception of specified questionnaires. MDS is employed to quantify the difference between questionnaires, and a stacked bar chart is designed to get insights into the detailed answer distributions for the questionnaires of interest. Furthermore, a set of user-friendly interactions are provided to enhance the usability of VisUPI, enabling the experts explore the UPI dataset according to their prior knowledge. At last, we demonstrate the effectiveness and scalability of VisUPI through case studies with the real-world datasets and the domain-expert interviews.

**Acknowledgements** This work was supported by NFS of China Project Nos. 61303133, U1501252, U1711263, the Natural Science Foundation of Zhejiang Province No. LY18F020024, the National Statistical Scientific Research Project No. 2015LD03 and the First Class Discipline of Zhejiang—A (Zhejiang University of Finance and Economics-Statistics).

## References

- Balzer M, Deussen O, Lewerentz C (2005) Voronoi treemaps for the visualization of software metrics. In: *Softvis*, pp 165–172
- Bederson BB (2001) Photomesa: a zoomable image browser using quantum treemaps and bubblemaps. In: *ACM symposium on user interface software and technology*, pp 71–80
- Chen W, Lao T, Xia J, Huang X, Zhu WB, Hu Guan H (2016) Gameflow: narrative visualization of NBA basketball games. *IEEE Trans Multimed* 18:2247–2256
- Chen D, Yueping LI, Wang X (2014) Study on the status of medical students' mental health and influencing factors. *Chin J Soc Med* 4:254–256
- Ferreira N, Poco J, Vo HT, Freire J, Silva CT (2013) Visual exploration of big spatio-temporal urban data: a study of New York City taxi trips. *IEEE Trans Vis Comput Graph* 19(12):2149
- Lamping J, Rao R, Pirolli P (2002) A focus + context technique based on hyperbolic geometry for visualizing large hierarchies. In: *Proceedings of the SIGCHI conference on human factors in computing systems*, pp 401–408
- Liu S, Cui W, Wu Y, Liu M (2014) A survey on information visualization: recent advances and challenges. *Vis Comput* 30(12):1373–1393
- Ma Y, Lin T, Cao Z, Li C, Wang F, Chen W (2016) Mobility viewer: an eulerian approach for studying urban crowd flow. *IEEE Trans Intell Transp Syst* 17(9):2627–2636
- Neumann P, Carpendale S, Agarwala A (2006) Phyllotrees: phyllotactic patterns for tree layout. In: *Eurovis 06 Proceedings of eurographics*, pp 59–66



- Qu H, Chan WY, Xu A, Chung KL, Lau KH, Guo P (2007) Visual analysis of the air pollution problem in hong kong. *IEEE Trans Vis Comput Graph* 13(6):1408–1415
- Quinlan JR (1986) Induction on decision tree. *Mach Learn* 1(1):81–106
- Robertson GG (1991) Cone trees: animated 3d visualizations of hierarchical information. In: Sigchi conference on human factors in computing systems, pp 189–194
- Sacha D, Stoffel A, Stoffel F, Kwon BC (2014) Knowledge generation model for visual analytics. *IEEE Trans Vis Comput Graph* 20(12):1604–1613
- Scholz RW, Lu Y (2014) Detection of dynamic activity patterns at a collective level from large-volume trajectory data. *Int J Geogr Inf Sci* 28(5):946–963
- Schulz HJ, Hadlak S, Schumann H (2009) Point-based tree representation: a new approach for large hierarchies. In: *IEEE pacific visualization symposium pacificvis*, pp 81–88
- Shi R, Yang M, Zhao Y, Zhou F, Huang W, Zhang S (2016) A matrix-based visualization system for network traffic forensics. *IEEE Syst J* 10(4):1350–1360
- Stasko J, Zhang E (2000) Focus + context display and navigation techniques for enhancing radial, space-filling hierarchy visualizations. In: *IEEE symposium on information visualization, 2000. InfoVis 2000*, pp 57–65
- Su Y (2008) Mental health condition of freshmen by UPI. *China J Health Psychol* 16:1133–1134
- Turkay C, Slingsby A, Hauser H, Wood J (2014) Attribute signatures: dynamic visual summaries for analyzing multivariate geographical data. *IEEE Trans Vis Comput Graph* 20(12):2033–2042
- Van Wijk JJ, Huub VDW (1999) Cushion treemaps: visualization of hierarchical information. In: *IEEE symposium on information visualization*, p 73
- Von Landesberger T, Kuijper A, Schreck T, Kohlhammer J, Van Wijk JJ, Fekete JD, Fellner DW (2011) Visual analysis of large graphs: state of the art and future research challenges. *Comput Graph Forum* 30(6):1719–1749
- Wang Z, Ye T, Lu M, Yuan X (2014) Visual exploration of sparse traffic trajectory data. *IEEE Trans Vis Comput Graph* 20(12):1813
- Wang X, Chou JK, Chen W, Guan H, Chen W, Lao T, Ma KL (2017) A utility-aware visual approach for anonymizing multi-attribute tabular data. *IEEE Trans Vis Comput Graph* PP99:1–1
- Xia J, Jiang G, Zhang YH, Li R, Chen W (2017) Visual subspace clustering based on dimension relevance. *J Vis Lang Comput* 41:79–88
- Xia J, Ye F, Chen W, Wang Y, Chen W, Ma Y, Tung AKH (2018) Ldsscanner: exploratory analysis of low-dimensional structures in high-dimensional datasets. *IEEE Trans Vis Comput Graph* 24(1):236–245
- Yang X, Lei S, Daianu M, Tong H, Liu Q, Thompson P (2017) Blockwise human brain network visual comparison using nodetrix representation. *IEEE Trans Vis Comput Graph* 23:181–190
- Zhang ML (2007) Effect on university personality inventory(upi)of 1087 college freshman. *Progr Mod Biomed* 7:1093–1095
- Zhang J, Yanli E, Ma J, Zhao Y, Xu B, Sun L, Chen J, Yuan X (2014) Visual analysis of public utility service problems in a metropolis. *IEEE Trans Vis Comput Graph* 20(12):1843–1852
- Zhang J, Lanza S, Zhang M, Su B (2015) Structure of the university personality inventory for chinese college students. *Psychol Rep* 116(3):821
- Zhao Y, Liang X, Fan X, Wang Y, Yang M, Zhou F (2014) Mvsec: multi-perspective and deductive visual analytics on heterogeneous network security data. *J Vis* 17(3):181–196
- Zhou F, Huang W, Zhao Y, Shi Y, Liang X, Fan X (2015) Entvis: a visual analytic tool for entropy-based network traffic anomaly detection. *IEEE Comput Graph Appl* 35(6):42–50. <https://doi.org/10.1109/MCG.2015.97>