Visual Analytics for Demographics, Social Networks and Business Base Pattern

Yuxiao Li, Xuexi Wang, Yue Wang, Ting Liu, Huiting Wang, Ziyue Lin, Siming Chen.
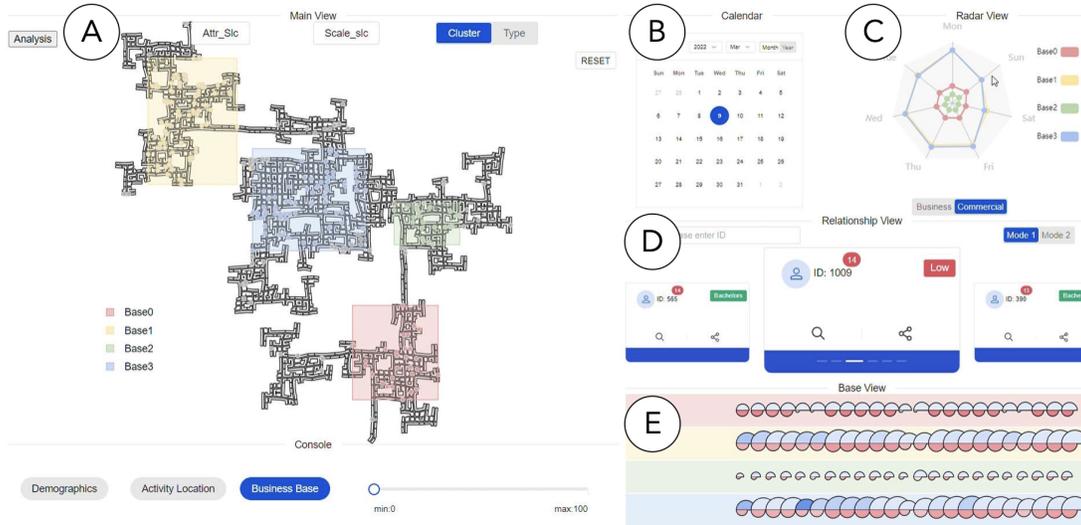
Fig. 1: System Interface: (A) Map View, (B) Calendar, (C) Radar View, (D) Relation View, (E) Base View

Abstract—We proposed an interactive visualization and analysis system for the VAST 2022 mini-challenge1 to analyze the basic demographic characteristics of the city as well as social characteristics by interactively selecting volunteer attributes, dates, etc. We can filter the volunteers' attributes (gender, education, etc.) according to our preferences and obtain visualization results by combining different attributes, or observe the social profiles among different groups of people by selecting dates, etc. and analyze them based on that. The exploration of social relationships is completed by ID search and k-means clustering. HDBSCAN clustering is used to complete the division of business bases and calculate the features that reflect both the work situation and business status of business bases based on the information of personnel's location distribution, account balance change, and salary level, respectively, based on which the business base model exploration is conducted. We will describe our design principles in detail in this paper.

---

1 INTRODUCTION

Mini-challenge 1 of VAST 2022 required a visual analysis of the basic profile of the city of Ohio. We were asked to analyze the demographic characteristics of the city, social relationship patterns, and business base patterns by giving volunteers attributes (gender, education, etc.), basic social behavior data, and some economic transaction data. The main challenge is to explore unknown patterns and perform deeper analyses based on the explored patterns. We implemented a visual analysis system that includes a map, a radar map, a calendar selector, a search and clustering view of social relationships, and a base view. Among them, the map supports the joint selection filtering and cross-sectional distribution display of multi-class features for demographics. The social relationship view supports the social relationship search for specific people and displays the k-means clustering visualization results of their related people. The calendar view, radar map, and base view work together to display the time-series characteristics of the base derived from HDBSCAN, including consumer traffic, work traffic, average consumption level, and average wage level of the business base. The five components of the visualization system coordinate with each other to support the exploration and discovery of demographic

characteristics, social relationship patterns, and business base pattern characteristics of the city.

2 VISUAL ANALYSIS SYSTEM

Our visual analysis system includes maps, radar charts, calendar selectors, search and clustering views of social relationships, and base views, in the following sections we will cover each of these in turn.

(a) People with engel coefficients greater than average and the radar view of the detailed attributes of volunteer 651

(b) A statistical analysis of people with children and their family size

Fig. 2: Demographics Mode in Map View

2.1 Map view

The map view is the basic part of our visual analysis system. The map view has three modes: Demographics, Activity Location and Business Base,each of which is described below.

2.1.1 Demographics Mode

In Fig. 2a we can see that the volunteers are scattered as dots on the map, and their position on the map is determined using the location of their residence. The type of street is mapped to the color, blue for commercial and green for. Initially, all volunteers are uniformly represented as small blue circles of the same size. By clicking on the attr selection button, specific attributes can be filtered and different attributes can be combined. Visualization of a specific attribute can be performed by clicking on the scale selection button, which maps the values of the selected attribute from small to large to the color and observes the distribution of different values of the attribute over geographical locations. After selecting all the attributes you want to observe, you can click the Analysis button to observe the proportion of values taken by all volunteers with respect to the selected attribute, as in Fig. 2b. Since volunteers have many attributes, we designed an additional auxiliary tooltip that, when the mouse hovers over a small circle, a radar plot of the volunteer's base attributes appears for auxiliary analysis. In addition, we included the Engel coefficient as an additional attribute of the volunteers, because the Engel coefficient is an important indicator of the standard of living [1]. The formula for calculating Engel's coefficient is as follows:

$$Engel's\ coefficient = \frac{Total\ food\ expenditure}{Total\ personal\ consumption\ expenditures}$$

2.1.2 Activity Location Mode

This mode displays the location of the experimenter on the map on a non-commuting day according to the selected date, as shown in Fig. 3b. In this mode, clicking on the semicircle in base view makes the map view show the distribution of people's location on corresponding day.

2.1.3 Business Base Mode

The geographic coordinates of all the Restaurants, Pubs and Employers in the town are clustered by HDBSCAN and adjusted to the city map structure to get the 4 main business bases, as shown in the figure below.

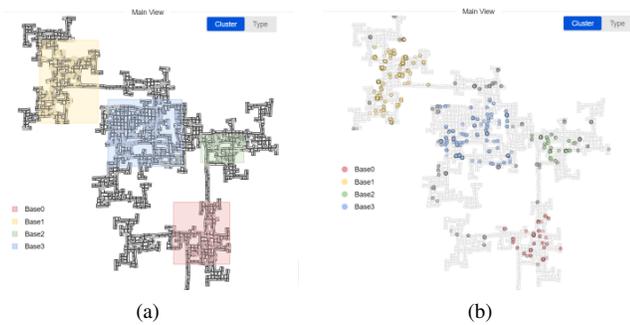(a)   (b)

Fig. 3: Business Base Mode and Activity Location Mode in Map View

Fig. 3a shows the clustering results, where the gray points are those clustered as -1 class. The four rectangles in the figure are the final four selected commercial bases.

2.2 Radar View

To analyze the patterns present in the commercial bases, we mainly considered the work and consumption situations of the volunteers in these bases. For the work situation, we consider for each site the number of people working in the site on a daily basis and the average hourly wage of all. For consumption, we considered for each base the number of people who spend money in the base per day, other than

transportation, and the average spending level per unit. Radar View calculates the average distribution of the two headcount characteristics with respect to the day of the week for the 30 days before and after the selected date, which helps us to explore the distribution characteristics of the headcount characteristics with respect to the day of the week, as shown in Fig. 1(C)

2.3 Relationship View
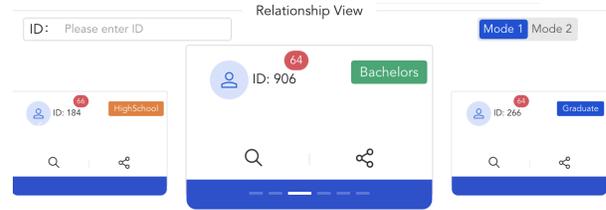
(a) The initial state of relationship view and 6 recommended candidates

(b) The social relationships of candidate 730 and specific social activity with candidate 390

Fig. 4: Relationship View

In the relationship view , the initial state(Fig. 4a) shows the six most socially active people per day, based on the date selected by the calendar. Each person's business card shows basic information about the individual and details are displayed when hovering over the magnifying glass button. We can also enter in the input box in the top left corner the social connections of any character we want to find for that day, but the ID must be legitimate. By clicking on the explore button on the right hand side of the card, we can access mode 2(Fig. 4b) to explore the person's social network. On the left is a force map, where each node is an ID and each edge represents a social relationship between two people. Each edge can be categorised by selecting the k-value for k-means clustering, making it easy to view social patterns, as well as selecting the specific category of social patterns you want to view from the drop-down box. A node or edge can be selected from the force guide to further select the desired information. On the right is a view of the relationship between the two people on the selected edge, in which you can see the time, distance, duration and some basic information about the social relationship that occurred between the two people after matching.

2.4 Base View

Base View shows the daily characteristics of each base for the 30 days before and after the date selected in the calendar. The red semicircle indicates the work situation, the radius size indicates the total number of people working in the base on that day, and the color shade indicates the average hourly wage of all people working in the base on that day. The blue semicircle represents the consumption situation, and the radius size represents the total number of trips made in the base on that day, except for transportation, and the color shade represents the average consumption level per order, as shown in Fig. 1(E).

REFERENCES

[1] F. Shujuan. A study of the influencing factors of the engel coefficient based on multi-factor analysis of the economics and ols. *Journal of Shanghai Business School*, 2012.