Visual Analytics for Urban Data Analysis

Jan Burmeister, Jilin Liao, Jieqing Yang, Qingtian Wei, Kexin Wang

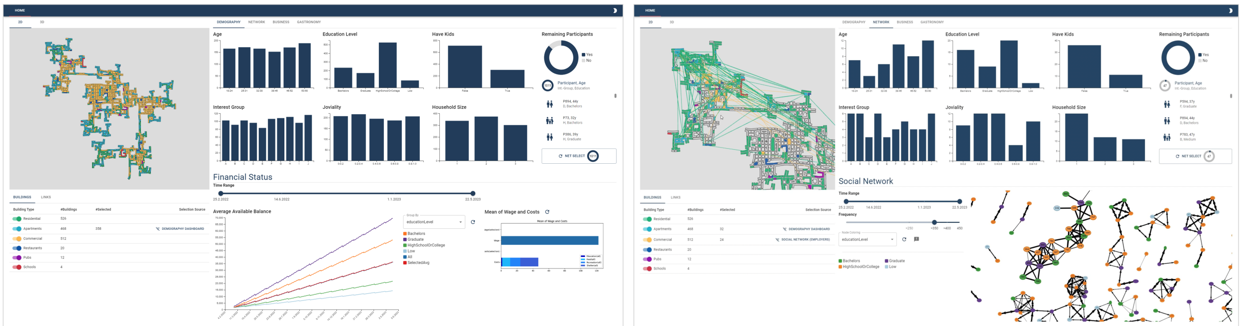

Fig. 1. The dashboard “Demography” (left) presents on overview of the composition of the whole population. In the fincancial section at the bottom, the analyst has split the data by education level to compare the aggregated balance over time. In the dashboard “Social Network” (right) the analyst has selected a group of participants from the node-link diagram, and checks for signficant traits in their value compositions. The group network is drawn on the map, revealing a spatial cluster in the upper left district of the city.

Abstract—The VAST 2022 Challenge 1 put us in place of an urban planning analyst for the fictitious city Engagement, Ohio. Given a large, interlinked dataset of its citizens, businesses and activities, we were tasked to assemble a summary of the state of Engagement to serve as basis for future investments. We present a visual analytics application to interactively explore this complex dataset using a combination of dashboards and a 2D/3D map. Our application uses interactive visualizations to iteratively build and analyze subgroups, sync their abstract data with spatial positions, and detect clusters and patterns in the data.

Index Terms—Visual Analytics, VAST Challenge

1 Introduction

The goal of the VAST Challenge 1 was to prepare a summary of the demographics, business, and social life of the fictitious city Engagement, Ohio. A collection of datasets provided descriptive information about a set of representative citizens, businesses, jobs, and buildings. Detailed logs captured the daily activities, interactions and financial situation of the participating citizens over a timespan of fourteen months. A special characteristic of the dataset is the strong linkage between the observed data entities due to various social and business relationships. Also notable is the linking between abstract and spatial data - every piece of information can be associated with a location or building. Location is a key trait in this context. Large cities commonly have well-known dedicated areas, each with their unique communities and distinct patterns of social behavior and activities. Knowledge about such local conditions is a key prerequisite in our use case of urban planning.

We are given the viewpoint of a data analyst within the city’s planning team, tasked to provide reports and insights on the current state of the city. Our job is to to provide a solid data foundation to support decision makers how to direct investments for future growth of the city. We approached this task with a set of interactive dashboards that are paired and linked with a 2D/3D map of the city. We present a visual analytics web application that provides the tools to explore this large dataset of linked entities, investigate and drill down into interesting subgroups, correlate patterns within the abstract data with patterns on the city map, and quickly check hypotheses that may come up in the team’s analysis process.

- *Jan Burmeister is with Fraunhofer IGD. E-mail: jan.burmeister@igd.fraunhofer.de*
- *Jilin Liao, Jieqing Yang, Qingtian Wei, Kexin Wang are with TU Darmstadt. E-mail: {jilin.liao | jieqing.yang | qingtian.wei | kexin.wang}@stud.tu-darmstadt.de*

2 Data Processing and Implementation

The amount of data in the activity logs was too extensive to be queried at interactive speed that we were aiming for. We therefore preprocessed and stored derived data, e.g. aggregated attributes like operating hours of businesses per weekday, or mappings between participants and their apartment building extracted from their activities. In the process we also noticed an anomaly - some participants dropping out of the data collection after the first month - that we included accordingly on our dashboard pages. Two webservices (Python, Node.js) manage the resulting datasets and process queries and statistics requested by the web application (React).

3 Application Design

The web application combines a map of the city with four dashboards, one for each topic of interest: Demography, social networks, business and gastronomy. The dashboards describe the composition of participants and businesses regarding abstract attributes (e.g. age or employment strategies). The analyst can select subgroups by placing filters on these traits, analyze how those subgroups exhibit differences and similarities compared to the full dataset, and display the entities on the map to inspect their distribution and associations in space.

3.1 City Map

The map draws all buildings of the city color-coded according to their category (apartments, pubs, etc.). The analyst can turn solid coloring on/off for each category and draw only neutral-colored outlines. Limiting the map this way is often useful to support visual detection of spatial clusters (e.g. business hubs), or to assess distributions of rare buildings, such as social venues or schools, across the city. Dashboards can set filters on each category to visualize their current entity subgroup.

Links between buildings are visualized by lines drawn on top of the map. The dashboards can add such link groups to visualize social relationships between people, and link people to venues they have visited (workplaces, restaurants, pubs).

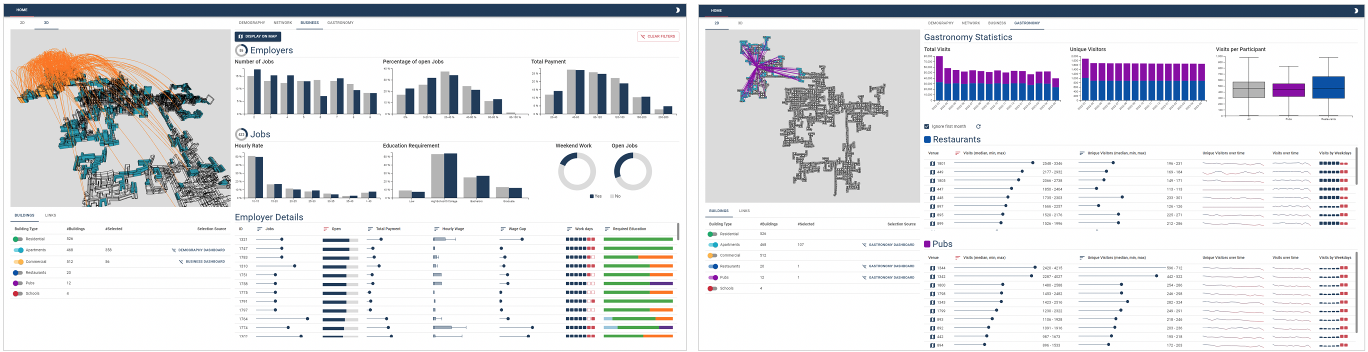

Fig. 2. The dashboard “Business” (left) shows the economic summary of businesses within a district that the analyst has selected by brushing on the map. We can see that attribute distributions are nearly identical to the whole dataset (blue bars represent the current selection, grey bars the full dataset). The map displays business buildings (orange) and the mostly nearby employee apartments (blue), linked with arcs in the 3D mode. The dashboard “Gastronomy” (right) displays the location and the guests’ apartments of a pub. We discovered that all but the largest pubs serve mainly their direct neighborhood, whereas the larger ones and restaurants attract guests from all over the city.

In the default 2D mode the map supports panning and zooming. In 3D the building heights correspond to the number of entities located there. Links are drawn as arcs above the buildings. The additional drawing space, paired with a movable camera (translate, rotate, zoom), is sometimes beneficial to sort out dense networks of links and discover outlier links that one might miss in the 2D version.

3.2 Dashboards

Demography: This page summarizes the overall composition of Engagement’s population. In the upper part, bar charts visualize the distributions of age, level of education, household size, having kids, interest group and joviality. The analyst can select specific subgroups by brushing on the charts. The selection can be synced with the map by placing a filter on apartment buildings. A donut chart visualizes the fraction of anomaly participants. The analyst can add a filter for these two subsets, to either investigate these participants in detail, or remove them from further analysis on this page.

The lower section of this dashboard visualizes the financial status of the currently selected group. A line chart displays the average balance of the participants over time. The analyst can focus on a time range, and also split the group by the attributes mentioned above. Each group is then displayed with its own line, enabling comparisons between the average balance of e.g. different education levels or household sizes. The chart is complemented with a pair of stacked bar charts that present a breakdown of the aggregated average income and spendings.

Social Network: This page focuses on the social relationships between the participants. The upper part is shared with the demography page, and can be used to define a subgroup of participants. The lower part hosts a node-link diagram as a visual representation of the city’s social networks. Nodes map to participants, edges to weighted connections between them. The width of the drawn edges correspond to the frequency of encounters, i.e. the strength of the relationship. The analyst can define a time range and minimum frequency of encounters to build the network. By tuning these controls, the analyst can filter only for those strong connections and prevent overloading the network into an uninformative hairball of links. A force-based positioning of the nodes separates groups, so strongly connected participants within the network become distinguishable. The analyst can drill down into individual groups by multiple means: First, the network nodes can be colored by the demographic attributes, which might give an indication if a group consists of people with e.g. the same interests or family status. While this provides an overview for the whole network, the analyst can also select a connected component on the network and set those participants as selection for this dashboard. The analyst can now check the distribution charts for distinctive features that characterize this group. The node-link structure also transfers to the map: The participants’ apartment buildings are filtered, and lines are drawn between the buildings corresponding to the edges in the node-link diagram.

Business: This page summarizes the business side and the current job market of the city. It links statistics of several data entities within one page: Employers, jobs and employed participants.

The top row of bar charts describes the employers of the city by how many jobs they offer, what percentage of them are currently open, and how much they pay combined over all offered jobs. The second row of distributions deals with individual jobs. It characterizes Engagement’s job market by statistics of hourly rate, required education, whether working hours occur on weekends, and how many jobs are currently open. The data sections are linked. An analyst can e.g. select employers with a currently large percentage of open jobs to find out more about the nature of these companies by examining whether those are small or large companies, and what types of jobs they are looking for.

A sortable table at the bottom provides full details of the current selection of employers. The table can be sorted by employer attributes and derived statistics of the jobs they offer. This is useful to find outlier cases, examine potential correlations between attributes, and drill down into the details of a selected subset.

Each selection defines a subset of jobs, employers, and the participants employed there. The analyst can choose to display the buildings of employers and employees on the map, and also draw links between employees and their workplace to investigate if a specific selection of business is clustered in the same area, and whether employees tend to live close by their workplace or commute longer distances. The analyst can also select an area on the map via rectangle-selection, and select all employers in this area for this dashboard to investigate the composition of businesses within a particular district of the city.

Gastronomy: This page describes the health of Engagement’s pubs and restaurants. It might be indicative for both the social life and business side of the city, as spendings in social venues tend to be tied to the citizens’ financial health. We therefore dedicated several visualizations to identify such patterns. The stacked bars in the top row visualize the volume of visits, unique visitors, and visits per participants to restaurants and pubs over time.

All restaurants and pubs are listed in detail tables, sortable by aggregate statistics of their number of total visits and unique visitors. Sparklines depict how the venues are frequented over time. A set of boxes visualizes how visits are distributed on the weekdays. A venue can be visualized on the map as network of lines, connecting the location of the venue with the residences of its guests.

4 Conclusion

We have presented a visual analytics application to support the analyst and planning team of Engagement to gain an overview over the current state of the city. It provides the tools to interactively explore groups of interest, find outstanding entities, and correlate them with spatial information. Our team used exlusively this application to answer the questions of the VAST challenge 1 and compose our report.