

# ClusBridges: Clustering Life Routines and Data

William C. Fei\*, Yawen Lu\*\*, Hao Wang\*\*, Xingyu Jiang\*\*, Tianyi Zhang\*\*\*, Zhenyu Qian\*\*, Yingjie Chen\*\*  
West Lafayette High School, USA\*, Purdue University, USA\*\*, UNISOC Spreadtrum Communications, Inc\*\*\*

Email address: {feiw, lu976, wang5329, jiang718, qianz, victorchen}@purdue.edu; skyyyezhang@outlook.com

**Abstract**—ClusBridges allows analysis and visualization of city traffic, trends, and patterns at different levels from individuals to clusters and regions. The system clusters similar individual life routines to provide insight into the different lifestyles that people tend to live throughout the city. It primarily uses bridges, also known as arcs, to present traffic information between points through arc length, curvature, width, and color, which helps give its name “ClusBridges”.

**Index Terms**— Visual Analytics, City Management, Artificial Intelligence

## 1 INTRODUCTION

Monitoring and understanding people’s patterns of daily life are crucial to urban planning and development. At the same time, cluster analysis and cluster-based visualization turned out to be highly effective in a system design [1]. Based on the data collected from representative residents, including the location and time they travelled, duration of travel, and expenses at places, we designed ClusBridges, as depicted in **Fig. 1**, to draw inferences about city regions, traffic bottlenecks, life patterns, and individual life routines to help analyst understand different levels of patterns across the city. The system analyses and visualizes the life patterns of various aggregation levels in a city in a concise, efficient, and user-friendly manner. The system primarily uses arcs (which can also be called bridges) to present information through arc length, curvature, width, dash, and color. ClusBridges can display information about traffic of various points at different times, view individual routines and cluster similar life routines, and view region trends and metrics.

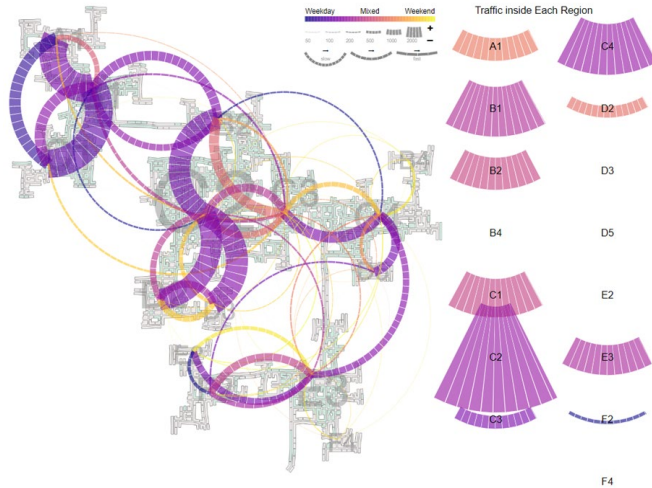


Figure 1. Traffic from Regions to Regions.

## 2 DATA PREPROCESSING AND ANALYSIS

Directly obtaining data at city level is not as straightforward as it lends itself to be. Instead, inferences from data must be made, along with data processing to make data more usable and understandable. We transformed the original datasets, which consisted of logging where and when participants traveled from location to location, into more meaningful and usable data. First, we sorted data by weekday/weekend times, months, and hours. We also then identified the actions that participants were doing when travelling (e.g., going out to eat, commuting to work). After completing basic sorting and categorizing, we then used K-means to cluster the logs and U-Map to

further verified the rationality of the clusters. We also sorted location information by sorting locations into “regions”.

## 3 SYSTEM ARCHITECTURE

ClusBridges consists of six modules. The first module is dedicated to displaying and interpreting traffic data between specific predetermined city regions, the second module to displaying and interpreting traffic data between specific predetermined city regions and specific locations, the third to viewing clusters of life routines, the fourth to individual life routines, the fifth to region trends over time, and the sixth to location metrics. There are options to sort by hours, weekends/weekdays, months in modules with a temporal factor. Also the user can filter routines involved with different types of locations and activities.

### 3.1 Region Trends and Location Metrics

We displayed the number of visits, average travel time in mins to every location, and the visit distributions of each type of locations (apartments, employers, pubs, restaurants, school, etc). Based on the observations on the typology of the road system, types of locations, no. of visits and minutes used per travel, we divide the entire city in 15 specific zones to increase usability of data and improve overall clearness of analysis. The data was visualized on Engagement, Ohio’s map, as shown in **Fig. 2**.

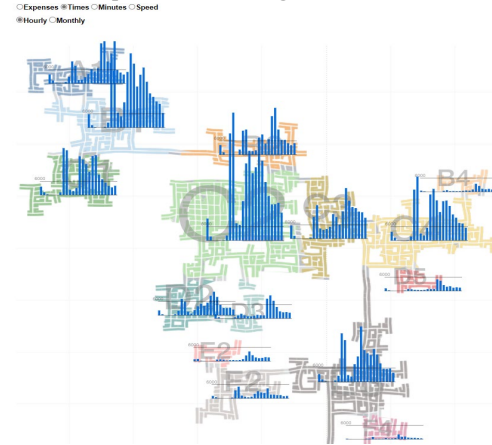


Figure 2: The entire city is divided into 15 regions. Various trends were visualized to show the monthly or hourly trends of all the regions.

Regional trends were visualized after sorting data by hour and types of locations. Region trends can be viewed either hourly or monthly, and trends are available for the number of visits, travel speed to an area, travel time to an area, and average expenses spent in an area.

Overall traffics from region to region, or from region to location, was visualized through the use of arcs. The color of the arc indicates how relatively frequency that path is taken on weekends vs weekdays. A yellow arc means the traffic mainly happened during weekend, a blue arc means the traffic mainly happened during weekdays, while a color in the middle means the traffic is mixed. The thickness of an arc indicates how many times that specific path is taken in the requested time frame. The curvature/dash size of an arc indicates the average speed of the path, which is calculated by taking the average travel time between the two regions and dividing it by their Manhattan distance. The slower the speed, the shorter the dash and the bigger of the curvature.

Since traffic is bidirectional, traffic going from point A to point B is always represented by the arc on the right connecting point A and point B. Overall travel from a specific location to a region was visualized similarly. The only difference was that each location also had a small drawing indicating what kind of location it was (e.g. pub, restaurant, school, employer). In addition, total traffic to each region was visualized similarly through the use of an arc (shown right in Figure 1). Locations activities were visualized through the use of line and bar graphs. As shown in the right side of Fig. 3, the bar graphs represent total activity throughout the months, the green line graph represents total activity throughout the hours of the day on weekends, the orange line graph represents total activity throughout the hours of the day on weekdays. The 24-hour duration chart shows the average duration of stays throughout the hours of the day. Mouse over an location will zoom-in and display the activity chart in its local position on the map, which help the analyst to make sense of the details of activities happened in the picked location. These modules allow users to sort by weekends, weekdays, months, hours, and location types.

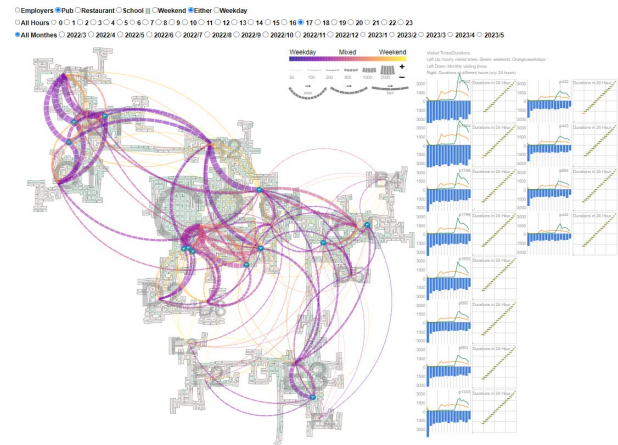


Figure 3. Traffic from Regions to Locations Module

### 3.2 Routine Clusters

With the detailed daily routines for each individual provided in the database, we clustered all travel logs (check-in type: weekend or weekday, check-in time, stay time, source location, destination location, etc) into different life routines. Firstly, we split all the logs according to specific location types (schools, pubs, restaurants, employers, apartments, etc). For each type of location, we clustered the captured logs based on the time stamps and location information. K-Means++ initialization is employed to obtain K initial cluster centers and treat them as one particle. Iterating this process for the provided number of rounds will obtain the desired number of particles. Secondly, the elbow method is applied to help to choose the optimum value of ‘K’ (number of clusters) by fitting the model with a range of values of ‘K’. The optimal K value is determined when the cost function does not decrease obviously anymore. Finally, the coarse-level clustered data extracted in the above steps

for each type of location are fused together to cluster all given logs. The excellent clustered groups, named routine clusters, are generated as Fig. 4. Activity clusters are visualized as horizontal bars, with different colors to show the type of activity. Participants that have similar activities are shown as vertical bars.

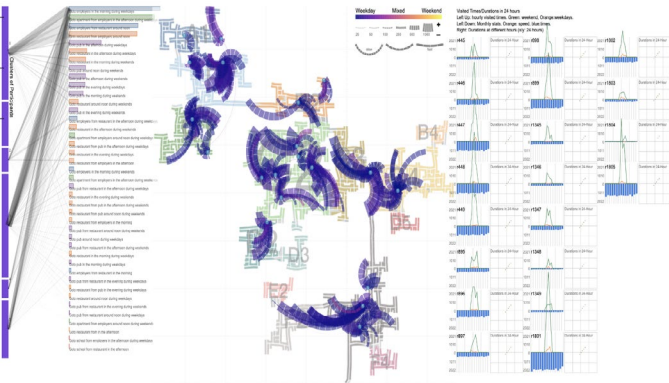


Figure 4. Clusters of participants and routine clusters (e.g., people in group 1 go to restaurants for lunch).

### 3.3 Individual Routines

Similarly to how routine clusters were visualized, individual life routines were visualized through commonly observed patterns, such as going to a restaurant from an employer location around noon. The frequency of such behaviors is shown in the bar graph on the left in Fig. 5. Similar to how traffic was visualized, individual travel was visualized in the same manner as arcs to show their daily destinations including their employers, which pubs/restaurants they preferred and at which day/time they end to visit pubs. Activity graphs show their detailed visits and durations of stay at each of the locations.

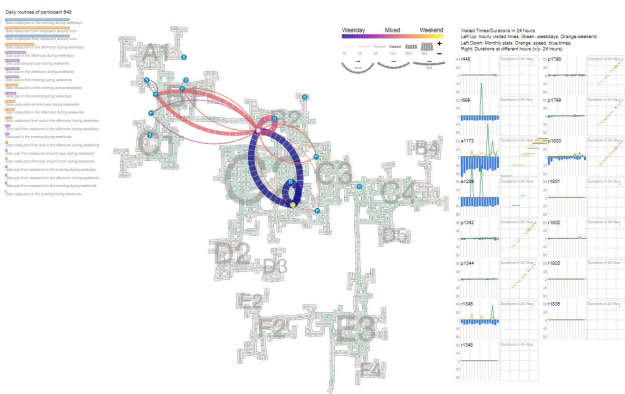


Figure 5. Individual's routines.

## 4 CONCLUSION

ClusBridges visualizes life routines and traffic from individuals to city scales through clustering, individuals' routines are aggregated at different levels. For each life routine group, we provided a detailed description of travel destinations and statistics about travel time and stay duration. It can visualize traffic between regions, traffic between locations and regions, routine clusters, individual routines, and region trends and location metrics. The analyst can gain insights into the life routines of individuals and the whole city, and make corresponding development plan for the city of Engagement.

## REFERENCES

[1] Van Wijk, Jarke J., and Edward R. Van Selow. "Cluster and calendar based visualization of time series data." Proceedings 1999 IEEE Symposium on Information Visualization (InfoVis'99). IEEE, 1999.