Ocean of Information: Fusing Aggregate & Individual Dynamics for Metropolitan Analysis

Mauro Martino, Francesco Calabrese, Giusy Di Lorenzo, Clio Andris, Liang Liu, Carlo Ratti
Senseable City Laboratory, Massachusetts Institute of Technology
E-mail: {mmartino,fcalabre,giusy,clio,liuliang,ratti}@mit.edu

ABSTRACT

In this paper, we propose a tool to explore human movement dynamics in a Metropolitan Area. By analyzing a mass of individual cell phone traces, we build a Human-City Interaction System for understanding urban mobility patterns at different user-controlled temporal and geographic scales. We solve the problems that are found in available tools for spatio-temporal analysis, by allowing seamless manipulability and introducing a simultaneous multi-scale visualization of individual and aggregate flows. Our tool is built to support the exploration and discovery of urban mobility patterns and the daily interactions of millions of people. Moreover, we implement an intelligent algorithm to evaluate the level of mobility homophily of people moving from place to place.

Author Keywords

Graph visualization, exploratory spatial data analysis, visual analysis, intelligent human information interaction, cell-phone data analysis.

ACM Classification Keywords

H.3.1 [Content Analysis and Indexing]: Abstracting methods; H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

1 INTRODUCTION

Collecting and analyzing massive amount of human mobility data has revealed interesting patterns in human dynamics [1], urban planning [10], and the spreading of viruses [11]. However, the tools needed to easily select and visualize massive mobility datasets and perform simple queries and selections to allow explanatory spatio-temporal data analysis are still lacking. The following tools and visualizations excel at displaying specific dimensions of the problem of spatio-temporal visual data analysis through using cell phone data.

Tools reference

MobiVis [5] (Figure 1.c) , MobileMiner [6] (Figure 1.d) and GeoTime [7] (Figure 1.e) are 3 tools that use data mining techniques for the analysis of mobile communication data. Developing tools from scratch with the required flexibility is difficult, time-consuming and requires skills not possessed by many engaged in geovisualization [8]. For this reason, tools have been developed that allow different interactive visualizations. MobiVis is a visual analytics system designed for exploring and discovering mobile data. The principal quality of this tool is to visualize complex social-spatial-temporal data and to filter the data with ontology graphs and interactive timecharts. MobiVis is suitable for both the expert and new user but is not compatible with other programs.

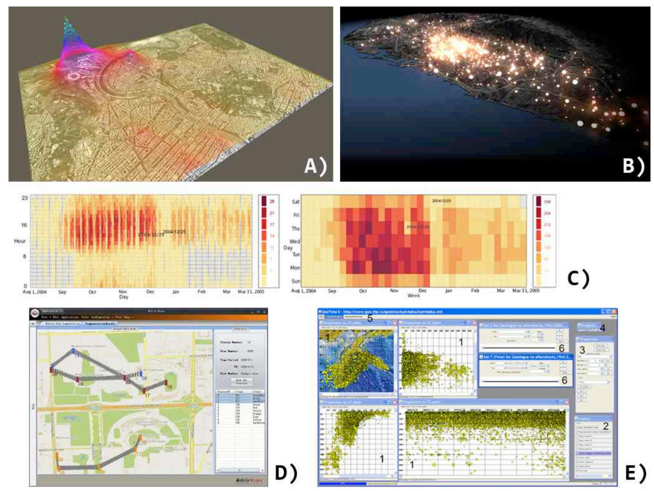


Fig. 1. Examples of visualizations of urban dynamics and visual analysis tools for exploring and understanding social-spatial-temporal mobile data.

MobileMiner is a tool to show a working data mining system on real mobile communication data. It is a good reference for understanding how to customize these types of programs for specific functions. In this case, data mining techniques are integrated into a mobile communication business solution. GeoTime is a tool for displaying and working with data over both space and time within a single, highly interactive 3D view. More importantly it operates with GIS and offers automated geo and temporal navigation. The principal concept is that individual frames of movement are translated into a continuous spatiotemporal representation.

Visualization references

Real Time Rome [3] (Figure 1.a) and UrbanMobs [4] (Figure1.b) introduce a 3D perspective view and provide a sense of the collective emotions of a city. Aggregated data from cell phones is mapped onto the geography of the city during two special events over the summer of 2006: the World Cup finals match between Italy and France, and a Madonna concert. The visualizations show peaks in the volume of calls during stirring moments (such as Italy scoring a goal during the World Cup match) in a sense revealing the emotional signature of the city as well as where people are congregating. Similarly, Orange Labs and Faber Novel developed UrbanMobs as a tool to showcase popular emotion cartography through the analysis and visualization of citywide cellular network traffic activity. Both examples signal a shift in both the aesthetic qualities of visualizing dynamic urban data and in the methodology of positioning cell phones within urban space according to the location of cell phone towers that service those calls. This methodology may lose the detail of GPS data but in tapping into the network infrastructure of cell phones, it can harness vast amounts of data representing large swaths of the city, thus increasing the scale of representation.

In this paper, we propose the tool *Ocean of Information,* to explore the interactions between human and the city in a Metropolitan Area. By analyzing a mass of individual cell phone traces, we propose to build a Human-City Interaction system for understanding urban mobility patterns at both the individual and aggregate level, and furthermore, to work towards understanding large-scale dynamic human mobility patterns. Our most prominent innovations include:

- combining aggregated mobility patterns and individual traces in real time;
- building interactive tools for discovering patterns in large scale mobility data sets;
- visualizing massive dynamic datasets in both spatial and temporal scale.
- automatic detection of homophily between mobility patterns of people's traces.

Ocean of Information introduces improvements compared with similar tools, including temporal window-linked timelines and map manipulations. The sequential presentation of data, the 3D point of view and the ability to represent small multiple designs [9] increase the capacity of our tool to compare data and detect patterns in relation to the other reviewed tools.

3 DATASET

The dataset used in this project consists of cellular phone location data anonymously collected by AirSage[1] for close to one million cellphone users of one telecom operator in eastern Massachusetts, USA. This aggregated information is used to model, evaluate and analyze the location, movement and flow of people in the city. To guarantee anonymity, each user is identified with an encrypted unique identification number (ID). Moreover, the ID is reset every day in order to avoid the possibility of tracking people over a long period of time. The database for each user contains a measure of their geographic location in latitude/longitude, for each time they connect to the cell network.

Since, the location measurements collected for every user are often noisy and inconsistently sampled, we processed the raw data to extract a set of meaningful places and trips between those places. We define a *trajectory* as a sequence of chronological location points for each user.

A sub-trajectory is obtained by segmenting the trajectory with a spatial threshold ΔS, thus

$$Traj_i = \{p_1 \rightarrow p_2 \rightarrow \ldots \rightarrow p_n\}$$

where $p_i \in P, t_{i+1} > t_i$

and $distance(p_i, p_{i+1}) < \Delta S, \quad \forall i = 1, \ldots n$.

The segmentation aims at removing spatial gaps between two recorded points (p_i, p_{i+1}) of more than ΔS into a time interval ΔT. If a gap is found, p_i becomes the end point of the last sub-trajectory, and p_{i+1} becomes the starting point of the new sub-trajectory. Once sub-trajectories are detected, we first resample with a constant sampling time T_c and then apply to them a low pass filter in order to eliminate some measurement noise contained in the data.

For each sub-trajectory we determine the time at which the user stops traveling, and call the location stop S. S_t is a geographic region where the user stayed over a certain time interval. A stop can occur when the user remains in a certain geospatial region for a period. The extraction of a stop depends on two parameters: time distance threshold (T_{th}) and a spatial distance threshold (S_{th}). Therefore, a single stop s can be regarded as a virtual location characterized by a group of consecutive location points

$$P = \{p_s \rightarrow p_{s+1} \rightarrow \ldots \rightarrow p_n\}$$

where:

$$\max(distance(p_i, p_j)) < S_{th} \ldots \forall s \leq i, j \leq n$$

$$t_n - t_s > T_{th}$$

Once the stops have been detected, we want to identify the user's landmarks, and travel path throughout the day. To detect these landmarks, we group nearby stops and create a grid of 200 by 200 meter cells. The pixel size can be manipulated to best reflect the scope of activity in the area.

4 TOOL DESIGN

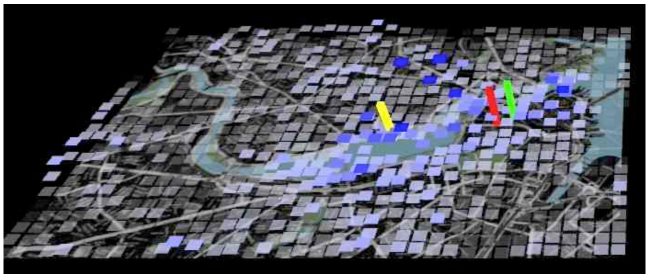
In light of the above discussion, we identified the following specific design requirements to guide the design and implementation of this tool.

The visualizations should be interactive but simple, allowing the user to explore the evolution of the event in space and time. Same maps are not meant to be visual analytics tools, so they should not provide quantitative details on the data, but other maps provide quantitative details. Inspired by Norman's work [2], we use color, saturation, and luminosity to improve the aesthetic and emotional impact of the project.

Digital Skin

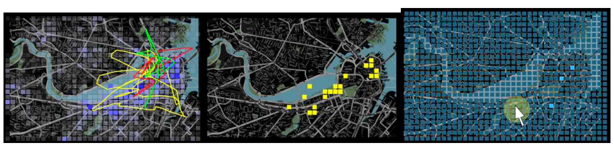
---

1 www.airsage.com



The patch graph movement represents the aggregate data; the color of the patches changes in relation to the number of people in each area (projected as a patch onto the floor). The data are displayed dynamically over time and in geographic space in order to represent flows of activity (i.e. pedestrians and cars). The visualizations attempt to represent activity in almost real-time with a 5 minute delay. The patches should create a "*digital skin*" of urban spaces, with a simple and clear visual language system. They should convey the social dynamics of a crowd, which are real phenomena, using informational data. The user should perceive the former, not the latter. They should establish a strong relationship between the crowd and the urban landscape. For this reason, we considered the spatial representation of the information as a given, reflecting the geographic nature of the data.

Multiview and Interactive Areas



Multiview is the basis for explanatory data visualization and analysis, since it helps the user search, locate and find new information with no prior knowledge. We selected multiviews to help make sense of the large volume of data. Each view is linked to the others using the CMV (coordinate multiview view) technique and changed automatically while interacting and querying the data.

Moreover, we added an interactive area to query information about different locations in the city at the same time.

5 INTELLIGENT INTERACTION

As described above, the tool allows visualizing the massive amount of information contained in the close to 1 million traces, by combining aggregate behaviors and individual traces. However, to allow easily detecting similar mobility patterns in the data, we further introduce intelligent components to measure how much users move similarly in the environment. This measure is then associated to a particular place and time interval, and allows for creating maps that illustrate at a give time which areas of the city are aggregating more similar people.

5.1 MOBILITY HOMOPHILY

Given an area C, and a time interval T, let us consider all traces that cross this area in the time interval T. For each of those traces, we can determine the location and time of the stop preceding time T, and the location and time of the stop following that.

If we group together those locations and times for all traces, we can derive the following sets:

$$before(C,T) = \{(B_1,Tb_1,nb_1),\ldots,(B_n,Tb_n,nb_n)\}$$

$$after(C,T) = \{(A_1,Ta_1,na_1),\ldots,(A_m,Ta_m,na_m)\}$$

where the tripe (B_1,Tb_1,nb_1) represents nb_1 traces who have passed through location B_1 at time Tb_1. For every trace t will then exist two indexes i and j so that $t \in (B_i,Tb_i,nb_i)$ and $t \in (A_j,Ta_j,na_j)$.

We can then define the folliwing mobility homophily index as follows

$$h(C,T) = \frac{1}{n\,m}\sum_i\sum_j \frac{\#\text{traces in } (B_i,Tb_i,nb_i) \text{ and } (A_j,Ta_j,na_j)}{\min\{nb_i,na_j\}}$$

The index ranges from 0 to 1 and represents how much a given area, at a given time, brings together people that behave similarly from a mobility point of view. Repeating this computation for every area in the city, and consecutive time intervals, we can create a map showing at a give time which areas are aggregating more similar people.

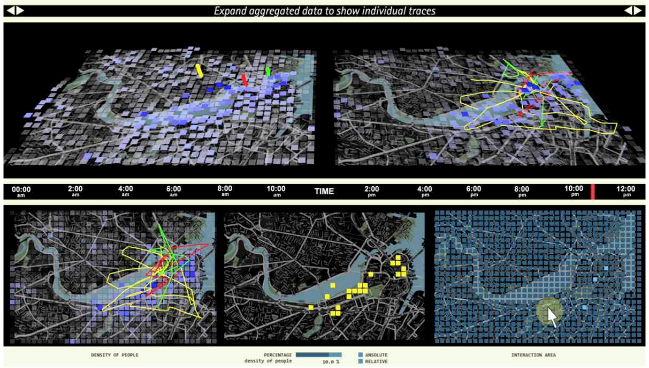
5.2 PANELS

The tool is designed to help answer questions regarding human movement dynamics in the city. To do this, we characterize each trace with succinct variables: distance, duration, speed, tortousity. The following view is of downtown Boston. Aggregate-data interaction allows for the selection and visualization of grouped individual traces for a population density distribution over space and time. The density of users in an area is evaluated by summing the number of users that fall into the area at a given time. Individual-data interaction deals with the selection and visualization of individual traces based on geographic location and time.

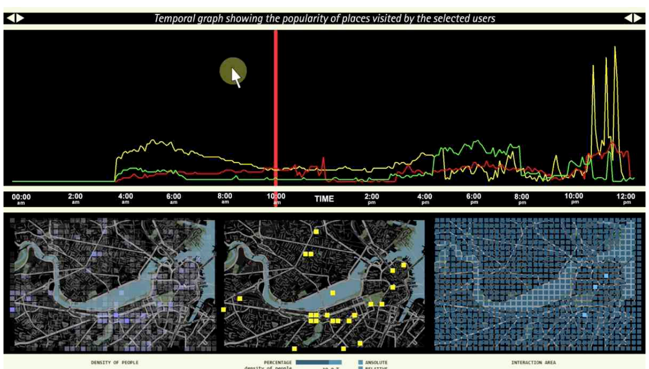
The following panels illustrate different scenes of the visualization.

Panel 1: 3D visualization of density distribution of people over time. The user is able to adjust the viewpoint of the 3D scene. Changing the viewpoint permits the user to view and

zoom around the scene. For example, the user could view an alignment of two different lines, or two objects of the same shading overlapping from a particular view.



Panel 2: A large scale rendering of downtown traces. This image allows the user to examine tortousity and envision activity space by the spread of each individual trace. The user manipulates the view by interactively selecting specific areas (wherein all the traces that pass through that geographic area are highlighted). The user can also interactively select individual traces with traditional union and intersection functions. The temporal frame in use adds an extra element, as when the trajectories are selected at time t, they will remain in place at time t+x for a smooth comparative analysis.



Panel 3: The places where Mobility Homophily is high are shown on the map (bottom center), together with a temporal graph showing the popularity of places visited by the users passing the selected areas (top).

6. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a tool to explore human movement dynamics in a Metropolitan Area. By analyzing a mass of individual cell phone traces, we have built a Human-City Interaction System for understanding urban mobility patterns at malleable temporal and geographic scales.

The tool enables the incorporation of different advanced data analysis methods, and provides a unified interface for performing sophisticated analytic tasks. To test the potential of the tool associated to the used dataset, we have implemented an intelligent algorithm to evaluate the level of mobility homophily of people moving from places to places.

ACKNOWLEDGEMENTS

The work described in this paper has been supported by AirSage.

REFERENCES

1. González, M. C., Hidalgo, C. A., and Barabási, A. L. (2008). Understanding individual human mobility patterns. Nature 453, pp. 779-782.

2. D. Norman. Emotional Design: Why we love (or hate) everyday things. Basic Books, 2004.

3. J. Reades, F. Calabrese, A. Sevtsuk, and C. Ratti. Cellular census: Explorations in urban data collection. IEEE Pervasive Computing Magazine, 6(3):30–38, 2007.

4. Orange Labs and Faber Nove. UrbanMobs, http://www.fabernovel.com, Revised in 30 August 2009.

5. Z. Shen, K. Ma. MobiVis - A Visualization System for Exploring Mobile Data. Visualization & Interface Design Innovation (VIDi), 2007.

6. T. Wang, B. Yang, J. Gao. MobileMiner: A Real World Case Study of Data Mining in Mobile Communication. SIGMOD 2009.

7. T. Kapler and W. Wright. GeoTime Information Visualization. InfoVis 2004.

8. J.A. Dykes, "Exploring spatial data representation with dynamic graphics", Computers and Geosciences 23 (4), pp. 345-370, 1997.

9. Tufte, E., Envisioning Information, Graphics Press, Cheshire, CT, 1990.

10. Reades J, Calabrese F, Ratti C, 2009, "Eigenplaces: analysing cities using the space – time structure of the mobile phone network" Environment and Planning B: Planning and Design 36(5) 824 – 836

11. P. Wang, M.C. González, C.A. Hidalgo and A.-L.Barabási Understanding the spreading patterns of mobile phones viruses, Science 324, 1071-1076 (2009).