Noname manuscript No.
(will be inserted by the editor)

Unfolding the City: Exploring Urban Spatial Attractiveness with Individual Social Characteristics

Qi Wang · Min Lu · Yang Yue · Qingquan Li

Received: date / Accepted: date

Abstract **Keywords** Visual analytics · spatial attractiveness · individual characteristics · city

1 Introduction

bgm

past, trajectory, not social characteristics

in this paper

novelty

2 Related work

3 Online Census

The advent of mobile sensing techniques and social media applications makes it possible to collect spatial data from the social media source. Complementary to the conventional census, it brings the benefit of larger sampling frequency and a broader range in terms of space and time. It is possible to reach a wide range of individuals and collect the movement in human inactive time, such as the mid-night.

Q. Wang · Q. Li(✉)
the State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China, and also ShenZhen Key Laboratory of Spatial Smart Sensing and Services, ShenZhen University, Shenzhen 518060, China
Tel.: +86-13297983047
E-mail: wangqi@whu.edu.cn, liqq@szu.edu.cn

M. Lu · Y. Yue
School of Architecture and Urban Planning, and also ShenZhen Key Laboratory of Spatial Smart Sensing and Services, ShenZhen University, Shenzhen 518060, China
E-mail: minlu@szu.edu.cn, yueyang@szu.edu.cn

Fig. 1 Census Interface: (a) personal characteristics collecting page; (b) trips collecting page; (c) credit system page

In this work, we perform the census survey in Shenzhen, which is one of the most modern metropolia in China. The experiment is deployed on Wechat, a widely used social media application. Figure 1 shows the data collecting interfaces. Each individual hands in his or her personal characteristics. For privacy issue, all detailed personal information are desensitized to categorical levels (Figure 1(a)).

1. Gender
- 1-1 Female
- 1-2 Male

2. Age
- 2-1 <= 18: whose age is under 18
- 2-2 19-40: whose age from 19 to 40
- 2-3 41-60: whose age from 41 to 60
- 2-4 > 60: whose age is over 60

3. Residence
- 3-1 Yes: who has registered permanent residence
- 3-2 No: who doesn't has registered permanent residence

4. Education
- 4-1 Middel School: who highest education is or under middle school
- 4-2 High School: who highest education is high school
- 4-3 Undergraduate: who recieves a bachelor degree
- 4-4 Graduate: who recieves a Master or Ph.D degree

5. Estate
- 5-1 Yes: who has real estate
- 5-2 No: who doesn't has real estate

6. Income
- 6-1 <= 100k: whose annual pay is under 100k
- 6-2 100-200k: whose annual pay is from 100k to 200k
- 6-3 200-300k: whose annual pay is from 200k to 300k
- 6-4 300-500k: whose annual pay is from 300k to 500k
- 6-5 > 500k: whose annual pay is above 500k

7. Car
- 7-1 Yes: who at least has one car
- 7-2 No: who has no car

8. Job
- 8-1 Manager: whose is responsible for administering in a company
- 8-2 Technician: whoe does practical work in a laboratory or company
- 8-3 Businessman: who works in business or commerce
- 8-4 Servant: who performs duties for others
- 8-5 Workman: who works inlabour :
- 8-6 Government Officer: who works for government
- 8-7 Student: who is learner in school
- 8-8 Retiree: who retired
- 8-9 Other

Fig. 2 Profile of Individual: eight individual characteristics enrich the analysis of mobility patterns

Individual Characteristics Figure 2 lists the *eight domains*, including social, economic and demographic aspects, to give a generalized description of the individual characteristics. The profile will serve as the ingredients for the analysis of mobility patterns over diverse individuals.

Traveling Trips Besides to those individual characteristics, each individual can upload dynamic traveling trips (Figure 1(b)). Each trip requires the information of *start/end location*, *start/end time*, *traveling purpose*. To encourage the trip uploading, a credit system retains the contribution of individuals on trips and rewards the volunteers with the top credits (Figure 1(c)).

3.1 Basic Statistics of Data

Over the releasing time period from 2015-11 to 2016-01, 21435 individuals (48% females and 52% males) were reached and 229155 trips are collected. Each volunteer contributes 11 trips average.

Our case-study data is confined to a small proportion of Wechat users who opt to contribute their information and trips. Considering the caveat that self-selecting individuals are most unlikely to represent any clearly defined population (**?**), we performed a series of preliminary statistics to check whether it is rich enough to represent a wide range of the population in the city.

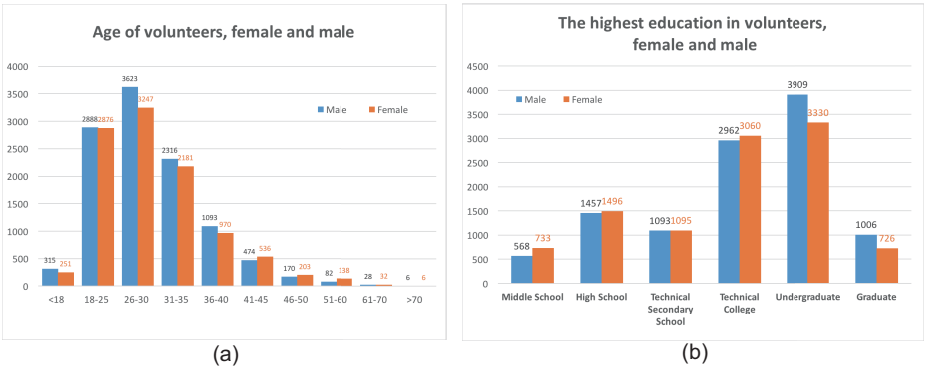

Fig. 3 Age and Education Distribution: (a) age; (b) education

Figure 3 gives the distribution of age and education over the population. It shows that samples cover a wide range of ages, dominating between 18 to 45. There is also a few records pertaining to individuals below the age of 18 or above 70. The distribution follows the fact that Shenzhen is a city where the majority is young people. According to the 2015 Annual Census Statistics report[1], people aging 15-64 occupy 83.23% and the median age is 31.5. Figure 3 gives the distribution of education levels, ranging from low to high. The technical college and university dominate the samples at the 61% occupancy rate.

Figure 4(a) shows the job types of sampled individuals, who are servants, workers, officers, businessmen and so on. The covering of jobs is pretty wide. Figure 4(b) gives the radar

[1] http://www.sztj.gov.cn/xxgk/tjsj/pcgb/201606/t20160614_3697000.htm

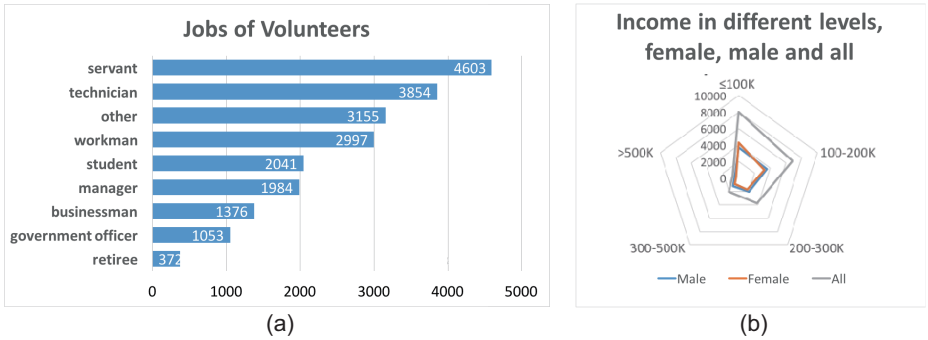

Fig. 4 Job and Income Distribution: (a) job; (b) income

diagram of the annual pay. The majority get paid below 200 000. Individuals with higher salary are also reached in our census.

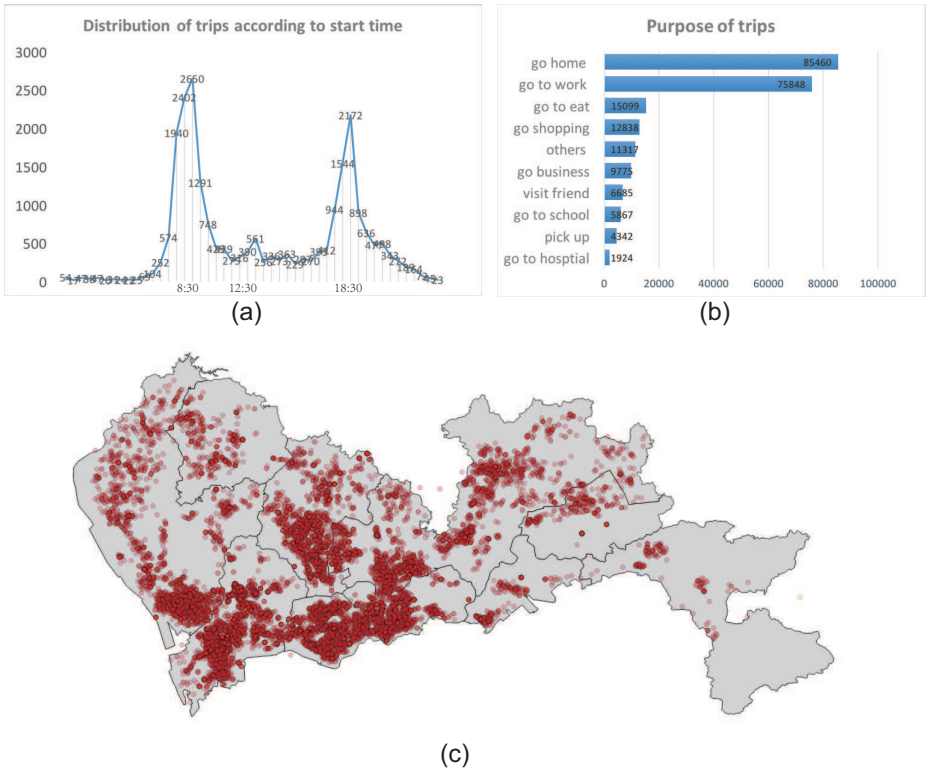

Fig. 5 Statistics of Trips: (a) start time of trips; (b) purpose of trips; (c) distribution of origins/destinations

Figure 5 shows the basic statistics related to trips. In Figure 5(a), the active traveling time (here, we take the start time as the representative of active time) follows the common knowledge of urban life. There are obvious morning and late afternoon peaks. Figure 5(b)

gives the counting of different traveling purposes. 95% trips are tagged with clear raveling purpose in the data. 33% are going home and 37% are going to work. Besides this kind of routine traveling, there are also substantial trips such as going shopping, going the hospital, etc. Figure 5(c) shows the spatial distribution of the origins and destinations. It is found that more dots are located in Futian and Nanshan districts, the city's heart than the surrounding areas. This is consistent to Batty's exposition of the focus of city networks and interaction patterns (**?**).

Because there is always inevitable bias inherent in fully representing the ground truth of the population, the preliminary statistical analysis shows positive sign of a relatively even sampling of the population.

4 Overview

By communicating with domain expert, we summarize the analytic tasks identified from the interviews of domain experts. Aimed at these tasks, we derive design considerations and introduce the workflow of the proposed visual analytics system.

4.1 Background

For understanding how to express spatial attractiveness, we have several interviewers with an expert in GIS who works on human mobility and urban informatics. The experts pointed out that the study of social-spatial interactions has become more important with big and open data evolved in. Instead of evaluated by static indicators, such as the characteristics of the spatial facilities, spatial attractiveness can be evaluated dynamically with human movement data. People have their own preferences when they go out for a specific purpose. It reflects the attractiveness of the place they selected. When the research scale is expanded into city-scale, whether spatial attractiveness is similar of the citizens with similar social characteristics. However, for GIS researchers, there is no efficient and effective methods which help to explore spatial attractiveness, citizens with multiple social characteristics, and their relationships. After several rounds of interviews with experts, we summarize the main analysis criteria below:

- *Multiple social characteristics*
- *Spatial attractiveness* In urban planning and GIS, the definitions of spatial attractiveness is based on gravity model[WQ: add refer]. It means the more distant the less attractiveness. Hence, both the quantity of people attracted and movement distances are two crucial concepts to be explored.
- *Attractiveness diversity* In urban planning, functional area is an important concept. When the area has more functions, for people here, their living requirements are more easily met, the attractiveness diversity of the area is higher. .....

4.2 Tasks

The exploration of core concepts above are resolved further with following tasks... to align the analysis of mobility patterns with individual characteristics. Before diving into the design of the system, introduce the couple of specifical tasks the system intended for.

- *Task 1: Identify the group with specific individual characteristics[WQ: What kind of citizens tend to be attracted?]*: to get groups of people with common or close attributes, to explore the correlation among individual characteristics.
- *Task 2: What is the correlation between space and citizens? (Spatial Attractiveness)* to understand... group of citizens are attracted to different places, e.g., to know the similarities and differences between different groups in their mobility patterns, to investigate the relationship between movement and individual characteristics.
- *Task 3: What is the detailed properties of attractive differences? ...* How many citizens are extracted, the origins (Where do the attracted citizens come from?) What is the attractive differences over groups?

4.3 Design Considerations

With these three tasks, we derive following design considerations:

- *Intuitive perception of an individual as an organic complex (C1)*: the system should make use of users' daily life experience in knowing people to provide intuitive visualization, instead of the lifeless representation by number. The visual design needs to help end-users to pick desirable ones from the mass.
- *Good overview of the multivariate individuals (C2)*: following the visual analytics manta by Shneiderman (**?**), it is very important to provide a good overview of all the individuals then the users know where to explore.
- *Effective multivariable cross-filter for individual characteristics (C3)*: there are eight domains to describe an individual. The system is supposed to provide a straight-forward way for easy filtering by the eight criteria.
- *Compact visualization of mobility patterns in the constraint of spatial space (C4)*: the analysis of mobility patterns not only includes the conventional spatial and temporal dimensions, but also other abstract dimensions, e.g., travel purpose, visiting frequency, etc. The system should handle a compact layout to support the easy correlation between spatial and abstract information.
- *Flexible interactions to explore the mobility patterns either within one group or between groups (C5)*: to support the comparison among groups, *Task 3*, the system should maintain flexible interactions which allow the end-users to explore freely.

5 Visual Design

With those design considerations described before, we develop a visual analytics system as Figure 6 shows. It composes of individual panel (the left part) for individuals (*Task 1*) and spatial panel (the right part) for the mobility pattern (*Task 2 and 3*). In the individual panel, interactive infographics, t-SNE visualization, and data-driven social profiles support users to narrow the scope down to groups of individuals with interested characteristics. With the chosen group of individuals, spatial attractiveness is visualized and explored in a map view with the [WQ: no name] glyph we designed.

5.1 Social Profile Visualization

To express the multiple dimensional social characteristics, we use three scales: bar glyph for all citizens, octagon glyph for a group of persons, and cartoon glyph for one person.

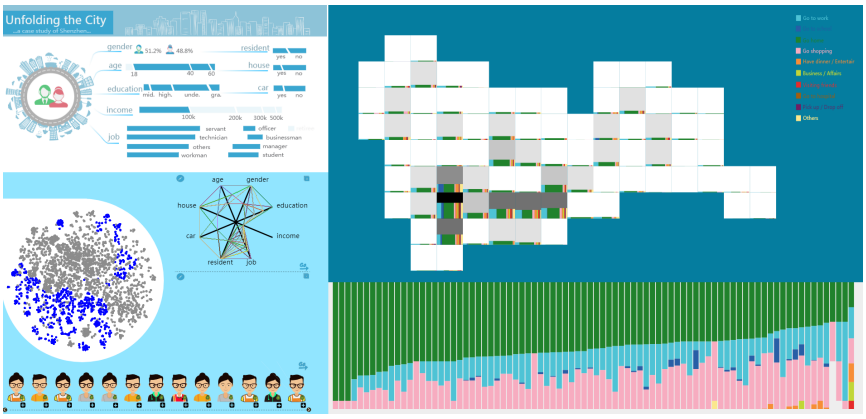

Fig. 6 [WQ: to be edit. System Interface: (a) an interactive infographics provides the basic statistic facets of individuals; (b) View1: a XXXX visualization gives an overview of individuals, where the similarity in high dimensional space is preserved in 2D space; (c) View2: data-driven profiles support users with organic individual perception; (d) View3: map view for movement of the chosen group; (e) View4: for ranking or comparison; (g) View5: top N kinds of citizens with social characteristics.]

Bar Glyph As shown in Figure 6(a), bar glyphs are used to encode quantitative relationships. For each characteristic, we compute the quantity and the percentage of each type, and encode it in the length of the bar. Interaction is also bound. When clicking a bar, filtering will be done, and the filtering result will be used in grouping phrase.

Octagon Glyph Glyph-based visualization **?** is the form of visual design to compose multivariable into a collection of unified visual symbols, known as a glyph. A glyph is intended for quick understanding and aligned comparison.

Cartoon Glyph Among glyph design, Chernoff Face (Chernoff, 1973) represents data variables by the different features of a cartoon face. Following the idea of Chernoff Face, we design a type of glyph, a graphical representation of people with specific individual characteristics. The idea behind using faces is that humans easily recognize faces and notice small changes without difficulty (C1). Those visual profiles are intended for intuitive visual understanding, from abstract to concrete and semantic understanding, to support users to target the interested individual groups effectively.

Figure 7 shows the legend for the user profile. The eight domains are encoded by visual symbols and organically organized in a human figure. There are basically three types of variables to drive the figure, i.e., the numeric, categorical and boolean ones. For numeric attributes such as income and education levels, visual symbol keeps consistent design but with changing visual variation, such as size, thickness. For the categorical attributes such as job, visual symbols are designed separately for better semantic meaning. For the boolean attribute like car, house, a symbol is designed to indicate its existence. With this consideration, the domains are designed as follows:

- **Gender** the gender is visually mapped to the hairstyle of the avatar.
- **Age** age is implied by the decoration on the hair. For the elder above 70, the hair is dyed to gray. For the youth beneath 18, a hair decoration is adapted for the different hairstyles of girls and boys.

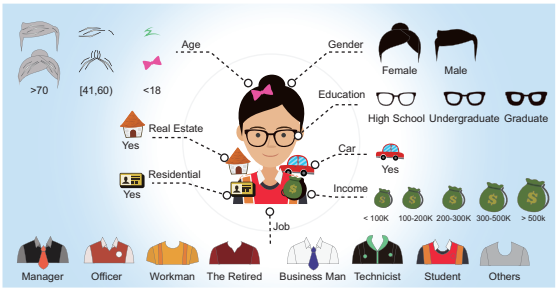

Fig. 7 **********

- **Education** The thickness of eyeglasses is used to indicate the different levels of education.
- **Job** The clothes is designed to imply the job of the individual. There are 9 types of clothes.
- **Belongings** for real estate, car, residential license are considered as the belongings to the individual, so we design each of them as an add-on decoration to imply whether the individual has it or not.
- **Income** a money symbol is used to represent the income, whose size encodes the income level. The more money an individual earns, the larger the symbol is.

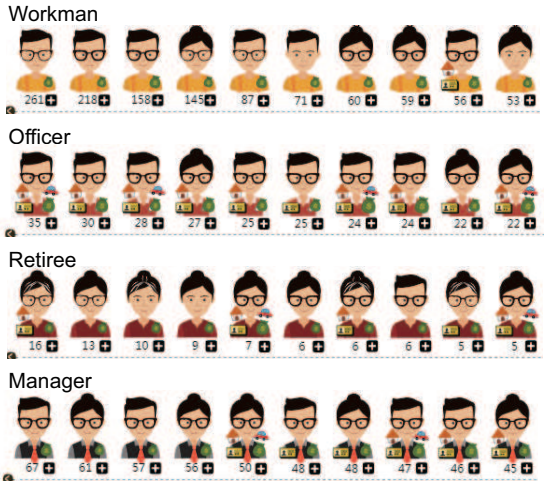

Fig. 8 Representative figures with the top 10 largest population in four jobs

With the visual mapping, the profile varies from individual to individual. By concretizing the attributes which otherwise is too abstract to percept, users can scan and search for interesting target effectively. Figure 8 lists the figures with the top 10 largest population in some job. The textual number beneath indicates the population. It is found that the majority of workmen earn a low salary and most of them have no residential license. On the contrary, for the officers, all of them have the residential license and most of them have a house.

Some of the retired people are in old age. Most of the managers are at undergraduates, even graduates.

5.2 T-SNE Projection of Individuals

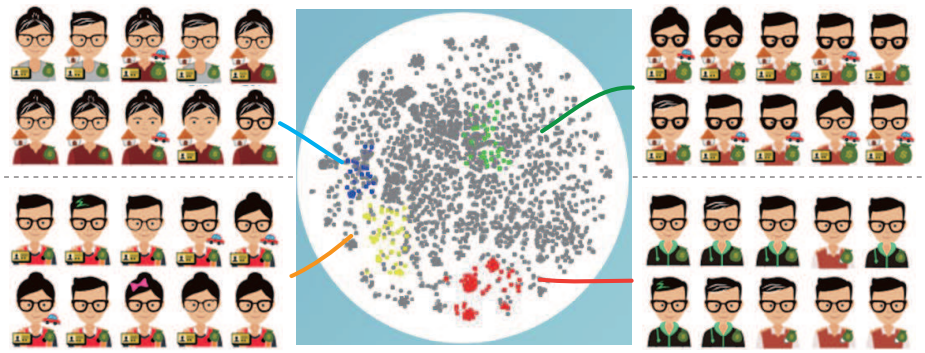

Fig. 9 t-SNE project with four groups of the interest: (a) t-SNE project

In our case, each individual can be denoted as a vector x_i with eight factors, and we get high dimensional data set $X = x_1, x_2, ..., x_n$. The dimensionality reduction techniques are to preserve the local structure of high dimensional data in low dimensional space, which are efficient approaches to provide a good overview of the multivarible individuals (*C2*). We adapt t-SNE project (**?**) to project X as two-dimensional dots $Y = y_1, y_2, ..., y_n$. One of steps in t-SNE is to compute the conditional probabilities to represent the similarity based on the distance between high dimenional datapoints. The conditional probability $p_{j|i}$ between x_j and x_i is given by:

$$p_{j|i} = \frac{exp(\left\|x_i - x_j\right\|^2 / 2\alpha_i^2)}{\sum_{k \neq i} exp(\|x_i - x_k\|^2 / 2\alpha_i^2)} \tag{1}$$

where α_i is the variance of the Gaussian that is centered on x_i. Specifically in our context, the high-dimensional Euclidean distances $\left\|x_i - x_j\right\|$ between x_i and x_j needs to be adpated the numerical and categorical characteristics. Characteristics such as age, income, are numerical and comparable, so the difference exactly explains when they are different. But the other characteristics, i.e. job, real estate, car, residential, are ordinal. There is not the numeric order. For example, the job distance from a manager to a businessman and a workman is not comparable, which is considered the same distance, i.e., set to 1.

As Figure 9shows, all 21435 volunteers are embedded in the 2D view, where the closer two dots are, the more similar they are in the eight characteristics. Figure 9 exemplifies four features groups of dots in a neighbor.

Multiple views of abstract view, t-SNE protection, and semantic data-driven profile visualization are coordinated in a Cross-filter machinesm (**?**). It allows end-users to interactive drill-down into individuals with interested characteristics from multiple perspectives(*C3*). Starting from the abstract criterion constraints, the scope of interest is narrowed down to individuals with(out) certain characteristics. And then further cross-filtering with semantically visual profiles can be performed to check the combination of 8 characteristical variables.

5.3 Location-based multidimensional grid visualization

We design a kind of glyph to express a high-dimensional travel indicators. It will be embed in the map view.

The elements can be represented in the glyph:

- value
- distribution
- proportion

6 Experts Feedback

[LM: Add a section to introduce the domain experts' feedback] We interview XX domain experts from the GIS field. XX of them are with ... The procedure went as following. We first introduce them the system an...

7 Case studies

In this section, we apply the method described above to study the case of Shenzhen. We introduce the several cases to demonstrate the usage and effectiveness of the system.

7.1 Case 1: Group by similar social characteristics

Use bar and tsne to find four groups. figure to show groups.

Their home locations and basis travel features.

7.2 Case 2: Spatial Attractiveness

Select two groups, and target purpose. figure to compare spatial attractiveness.

click a grid, show OD. figure

7.3 Case 3: Diversity of spatial attractiveness

Select group, figure to show diversity.

ranking, figure

two groups to compare, figure

8 Discussion and Conclusion