MAPPING THE ECOLOGICAL INTERVIEW:
GEOGRAPHIC VISUALIZATION FOR EXPLORING QUALITATIVE AND QUANTITATIVE ACTIVITY SPACE DATA

A Thesis
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
Of the Requirements for the Degree
MASTER OF ARTS

By
Yinghui Cao
May, 2010

Thesis Approval(s):

____
Dr. Jeremy Mennis, Thesis Advisor, Department of Geography and Urban Studies

____
Dr. Jerry Stahler, Committee Member, Department of Geography and Urban Studies

ABSTRACT

There is ever greater recognition in the social and health sciences that neighborhood context should be incorporated in studying human behavior. In terms of analyzing urban human behavior, activity space data (i.e. data associated with an individual's home and routine activity locations) can offer diverse information. However, it is challenging to explore and interpret activity space data, which often involve qualitative and quantitative variables describing both the environment and people's travel-activities. This paper will present a prototype software environment to visualize qualitative and quantitative urban activity space data. A system is developed to visualize the data generated by the Ecological Interview, a method for collecting activity space data, social network data, and perception of place data. This application integrates data on adolescents' travel behaviors, as well as their individual, social and community resources, by mapping subjects' activity spaces using multiple visual attributes (e.g. symbolization, color). Users can investigate the activity space data through an interactive interface developed using a commercial GIS software package. The case studies show how the visualization assists in the exploratory studies of activity space data. The visualization can also be used in post-modeling analysis through in-depth investigation of the multiple attributes for the best-fit and worst-fit cases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Pages

CHAPTER 1
INTRODUCTION

There is ever greater recognition in the social and health sciences that neighborhood context should be incorporated in studying human behavior. In terms of analyzing and modeling urban human behavior, activity space data (i.e. data associated with an individual's home and routine activity locations) can offer comprehensive information for learning how human activities occur over urban space. However, exploring and interpreting activity space data is challenging for a variety of reasons. First, activity space data usually involve spatial and temporal data in various forms. Furthermore, activity information may be both quantitative and qualitative. Finally, data on activity patterns concern not only the physical paths of individuals but also other variables, such as individual characteristics (e.g. gender, age, and so on.).

As a promising approach in Exploratory Data Analysis (EDA), geovisualization can facilitate identification and interpretation of patterns and relationships in complex spatial data (MacEachren and Kraak, 1997). Geovisualization can be primarily used to store and manifest spatial-referenced data, usually incorporating with other attributes by different means. Through the visual demonstration of spatial distributions along with other potential indicators, people will be able to detect and understand spatial-related patterns and relationships in a more straightforward way. In this paper, we present a prototype software environment to visualize qualitative and quantitative urban activity space data. A system was developed to visualize the data generated by the Ecological Interview, a method for collecting activity space data, place-based egocentric social network data, and perception of place data (Mason et al., 2004; Mennis and Mason, in press). The Ecological Interview data were collected as part of a U.S. National Institutes of Health (NIH)-funded study investigating social and spatial contextual influences on adolescent substance use (#1R21DA020146-01A2). By mapping adolescents' activity spaces and detailed information associated with their activities there, researchers will be able to first get a holistic picture of an individual's spatial movement in his/her daily life, and then understand how the adolescent is socially associated with those places, what kind of influence the neighborhoods may have on him/her, as well as his/her own perception and emotional reaction to the activity space.

The second section discusses the significance of involving activity space data in learning human spatial behaviors and summarizes the use of visualization in dealing with the complex spatially referenced data. Section 3 introduces the variables and measures in the Ecological Interview dataset. Section 4 and 5 are about the conceptualization, operationalization and implementation of our geovisualization system. In section 6 we illustrate its functionality in activity space data exploration and analysis by conducting a case study with Ecological Interview data. The final section is a conclusion about the potential usefulness of geovisualization in understanding human travel-activities both quantitatively and qualitatively.

CHAPTER 2
BACKGROUND

2.1 Activity Space Data

Researchers have been working to understand and model human spatial behaviors using the data that describe their activity spaces for decades. However, such studies were greatly constrained in the past due to the limited kinds of data and the lack of effective geocomputational environment (Kwan, 2000a). First, human travel-activity patterns in the real world are far more complex than what were assumed. In many of the past studies, these complexities, such as the travel distances between locations on the basis of transportation network, and variations of speed during the travelling, were simply ignored throughout the urban environments (Miller, 1994). Second, previous researchers seldom involved the human explanation and perception of their travel-activity into the spatial analytical framework. Due to the complexities of human activities, individuals' cognitive knowledge of the urban space usually sheds light on our understanding of their activity patterns. Third, the spatial data framework describing urban environments is often largely limited or not in a suitable format for analyzing human behaviors at an individual level. The commonly accessible socio-economic and environmental data in many previous studies are zone-based data (e.g. census tracts), which is not helpful in network-based or point-based travel-activity analyses (Kwan and Hong, 1998).

Fortunately, recent development in data capture and management provides great opportunities to involve more applicable environmental elements in learning human spatial behaviors as well as more detailed travel-activity data (Kwan, 2000a). First, comprehensive and detailed digital datasets of metropolitan areas, such as digital transportation network and street centerlines, have been published by national mapping agencies and census bureau (Smith and Rhind, 1998). These kinds of information are becoming widely used in learning the urban environments (Longley, 1998). In addition, individuals' explanations of their spatial activities and associated urban spaces have drawn a large amount of interest in human behavior research due to the increasing applications of integrating qualitative methods in Geographic Information Systems (GIS).

As all these kinds of information are recently integrated in activity space data to portray both urban environments and individual behaviors, an effective and appropriate computational and analytical environment is demanded for handling data of such large amount and variety. To address this need, social scientists are using GIS packages that provide spatial analytical tools and user-friendly customization. A primary utilization of GIS software is to accurately represent the objective environment (Miller, 1998). For example, detailed distribution of land parcels and the transportation network can be represented in 2-D maps or 3-D simulations. Actual travel distance based upon the street network can thus be calculated to replace the straight-line distance between two locations (i.e. Euclidean distance) (Entwisle et al., 1997; Kwan 1999). Secondly, more elements of activity space data are allowed to be incorporated by GIS and 3D techniques. GIS has been designed for dealing with numeric spatial data in quantitative geographic research. However, a large number of possibilities have been presented lately to incorporate all kinds of other information in various formats. One of the most prominent methods is geovisualization (e.g. Kwan, 2003; Al-Kodmany, 2001). These studies provide plentiful representation paradigms that are useful for dealing with multivariate qualitative and quantitative activity space data.

2.2 Quantitative and Qualitative approaches in GIS

In social science, there are two common research methods, quantitative and qualitative research, depending on diverse research questions and objectives. Quantitative research aims to develop mathematical models that involve quantitative measurements of properties to explain a phenomenon. Conventional GIS was generally designed for analyzing spatial relationships of quantitative data (Burrough and Frank, 1996). Due to the numeric nature of quantitative variables, GIS can easily be utilized in quantitative research for data representation and exploration.

Differently, qualitative data portray information by a set of vocabularies in a way that is similar to human cognition and communication of knowledge (Renz, 2002). A variety of conceptual frameworks can be adopted in qualitative research, including human feelings, emotions, perceptions, cognitions, and so on (Winchester, 2000). Qualitative variables describing these features are quite useful in representing the real environment and understanding human experience. Hence, qualitative approach is often utilized to analyze and understand human behaviors and activities in the physical and social environment.

Qualitative data can be derived from all kinds of empirical materials such as documents, pictures, videos, and interview transcripts, etc. In many fields, in-depth interviews are employed with the aim to better understand environmental context and human perception (Frechtling et al., 1997).

There is no such routine, as in quantitative research, to integrate qualitative variables in geographic studies. Researchers thus have presented a large number of possibilities to apply GIS in qualitative research or to incorporate qualitative data in quantitative GIS research. One promising method of incorporating qualitative data is geographic visualization. The motivation of visualizing spatial qualitative information is to observe spatial clustering or distribution of a phenomenon that are reflected by such data. Researchers in the field of information visualization have been visualizing qualitative data for decades (Tufte, 1983; Orford, 1999). Especially in cartography, people have successfully incorporated human cognition and knowledge in various types of research by visualizing qualitative data. Examples involve concept maps and cognitive maps (Heer, 2005), spatio-temporal visualization (Kwan, 2003) and other spatial data representation using GIS (Al-Kodmany, 2001).

Due to different kinds of information that quantitative and qualitative data can reveal, quantitative and qualitative approaches both have merits and shortcomings. Numeric quantitative data can easily be handled by traditional quantitative methods, but they are quite weak in understanding and explaining the uncertainty and complexity in spatial knowledge reasoning. Descriptive information, in the form of qualitative variables, is more helpful in interpreting human cognition since humans are more familiar with thinking and communicating qualitatively and logically. In terms of applying GIS in learning human spatial behaviors, qualitative data thus should be properly integrated with conventional quantitative spatial analysis to better understand the complex spatial context and to interpret human perception in the urban environment (Frechtling, et al., 1997).

2.3 Geovisualization in Exploratory Analysis

With the assistance of computers, visualization can be defined as the use of interactive representations of abstract data to reinforce cognition, hypothesis building and knowledge reasoning (Dykes et al., 2005). For the purpose of dealing with geographical referenced data, geovisualization integrates approaches from information visualization, exploratory cartography, Exploratory Data Analysis (EDA) and GISystems to provide theories, methods and tools for the visual exploration, analysis and presentation of geospatial data (MacEachren and Kraak, 2001). The utility of geovisualization can be categorized briefly as: 1) geospatial data storage and cartographic presentation and 2) exploratory data analysis, data mining and knowledge discovery (Dodge et al., 2008). From an academic perspective, interactive geovisualization is commonly adopted at both levels. The primary focus of recent research in geovisualization has been its role as a promising EDA tool (MacEachren and Kraak, 1997). Data is stored in systems and represented to help researchers set up initial knowledge and evaluate final results; through the interactive operations with the graphics and data, researchers will be able to test hypotheses, explore potential information and construct knowledge.

Many techniques can be employed to implement geovisualization environments, such as 2-D or 3-D scatterplots and parallel coordinates (e.g. Edsall et al., 2000) for detection of spatial patterns, trends and outliers, as well as iconographic techniques and map-based techniques (e.g. Dykes, 1997; Peterson, 1999) for investigating particular cases in the spatial context. For cartographic visualization, creating appropriate representation system is one of the most important issues in the field of visualization. Useful representations should convey information properly, efficiently and accurately, and allow people to compute desired conclusions (Carpendale, M.S.T., 2003).

Jacques Bertin (1983) developed a practical approach for his data graphics by defining a mark as something that is visual and can be used in cartography to show relationships visually. A mark can be a point, a line, an area, a surface or a volume. Marks can be differentiated by their relative placement on the map as well as by their visual characteristics such as size, shape, value, color and texture, and so on. According to Bertin's (1983) list of visual variable characteristics, a good representation should be selective, associative, quantitative, and aptly ordered for the purpose of visual interpretation. In addition, a visual variable should also be of proper length such that the changes in its value are recognizable with confidence. Bertin's (1983) theory of cartographic representation still plays a key role in the field of geovisualization today. Visual variables can vary from case to case depending on the characteristics of information that we want to represent.

Recent geovisualization research tends to incorporate more and more information in a single application. Temporal information has been integrated with spatially referenced data in so called a “space-time cube” (Hägerstrand, 1970) in many studies to learn the human behavior variations in both spatial and temporal dimensions (Kwan 2000b, 2003). Similarly, a large amount of other information can be added into the representation framework depending on researchers’ specific interests by different techniques such as multivariate mapping (Andrienko and Andrienko, 1999). As a special element, qualitative information, such as individual cognitive knowledge of travel-activities, can also be incorporated to represent the subjective environment (Ding and Kwan, 2004). However, many of these visualizations can only represent limited types of information at one time. For example, the 3D visualization with temporal information as the third dimension is highly constrained by the number of paths and the variety of properties to be visualized (Kwan, 1999; Shawn, 2008), since it is much harder for people to clearly recognize the spatial distribution within 3D spaces than on 2D maps. Therefore, it is necessary to develop a more practical, effective and efficient visualization framework for exploring the complex activity space data involving spatial, temporal and a wide range of other aspects in the forms of both qualitative and quantitative data.

CHAPTER 3
ECOLOGICAL INTERVIEW DATA

The sample was recruited from adolescent patients at a health care center managed and supervised by the Philadelphia Department of Public Health. The sample is composed of 301 adolescents ranging in age from 13 to 20. Sixty percent of the patients are female. The children come from different ethnic backgrounds, but are primarily African American (95%). Participants were approached in the waiting area. For a nominal monetary incentive, participants were asked to participate in the Ecological Interview. Parental consent was acquired for participating minors. All protocols were approved by the Philadelphia Department of Public Health and Villanova University Institutional Review Boards (IRB). The sampled individuals are all Philadelphia residents, with no major mental health disorder, and fluent in English. All participants were assessed regarding their substance use, mental health and spirituality, social network characteristics, and routine geographic locations that they regularly visit. After coding the qualitative interview data and geocoding the locations addresses in ArcGIS 9.3, the final dataset consists of 301 adolescents with 1027 valid location points and other individual and social network information available for data analysis.

Substance use involvement

Substance use involvement was measured using the Adolescent Alcohol and Drug Involvement Scale (AADIS) (Moberg and Hahn, 1991). AADIS is a highly accurate measure in separating adolescents who are free from any substance use disorders and those who are at least involved in one (Winters et al., 2001). The Ecological Interview generates AADIS scores ranging from 0 to 67. The 301 adolescents in the sample were then divided into two groups by a threshold of 36 for the AADIS value: 151 were labeled non-users (i.e. individuals with an AADIS value lower than the threshold) and the other 150 labeled users.

Activity spaces and transportation variables

The Ecological Interview captured adolescents' activity spaces to learn their travel-activity patterns. It generated a listing of the teen's weekly activity locations, as well as self-evaluations of these various geographical environments. The specific street addresses of the teenagers' home and routine locations were geocoded into a point feature class using ArcGIS 9.3. During the interview, researchers asked the teens which place is most risky (the place where you are most likely to engage in risky or dangerous activities, cause trouble, or do illegal activities), the safest (safest place from harm, danger, or the likelihood of engaging in risky or dangerous activities), the most important, the favorite and the most religious place to them. The adolescents were asked for explanations in their own words of what makes these locations risky, safe, and so on. This interview also captured the teens' traveling characteristics for their visiting routine locations, generalized as transportation variables, involving distance from home location to routine locations, how they travel to these locations, the time and day of a week of visiting, and how long they stay in these locations, in a typical week.

Demographic characteristics

Individual variables include the subjects' demographic characteristics, such as age, gender, race and parental education, and so on.

Social network

Egocentric social network characteristics were probed as a score addressing the risky and protective aspects of an adolescent's social network. Teens were asked to name up to five persons with whom they have contact at least once per month and with whom they have a "meaningful relationship," with corresponding locations associated, if in one of the five activity spaces for this child. The participants also had to provide information about their contacts' substance use, influence on their behavior, and types of activities, including positive activities such as receiving help with school or transportation, as well as negative activities such as engaging in illegal or dangerous behaviors. The social network quality score was then generated using the Adolescent Social Network Assessment (ASNA) (Mason et al., 2004). It is an aggregate measure of social influence towards people's deviant behavior, where high (positive) scores indicate the protective social influence and low (negative) scores indicate the negative influence on adolescents to use drugs or alcohol. Then the social network score at each activity location was calculated respectively, if associated with any of the five close contacts.

Geographic variables

Geographic data were utilized to describe the community characteristics of children's home and activity spaces. In this study, researchers adopted two types of geographic variables:

census data and other demographic and environmental data aggregated to neighborhood polygons and point data illustrating the distributions of various types of features, such as schools, libraries, police stations, crime incidents, alcohol sales, and other features we theorize influence substance abuse. Each variable is respectively associated with different types of activity locations for each sampled subject. The multiple variables to be visualized are listed in Table 1.

CHAPTER 4
CONCEPTUALIZATION AND OPERATIONALIZATION

The primary objective of this visualization system is to facilitate the management, retrieval and analysis of information for exploratory study of the activity space data and to shed light on further pattern recognition. A map-based visualization framework is adopted to implement this application system. The major challenges in visualizing the Ecological Interview dataset are, first, to visualize multiple attributes including both qualitative and quantitative data, and second, to provide interactivity to users. Multiple spatially referenced variables are represented in the map interface by a variety of visual attributes (e.g. color, symbol size, etc.). An interactive interface of the visualization application allows users to conveniently interact with different variables at different activity space locations.

4.1 Mapping Multiple Components

The Ecological Interview visualization system stores, organizes and manages both quantitative and qualitative Ecological Interview data. Quantitative codes are mostly categorical and continuous values. In maps, quantitative properties are commonly represented by distinct size, color and shape of the mapping symbols. Theoretically, more than one symbol and multiple visual attributes for a single symbol can be adopted in the same map simultaneously, such as using multiple proportional symbols in a polygon map with each symbol standing for one variable. But such maps should be well designed to avoid visual confusion and misunderstanding. Particularly for point maps, distinct visual attributes will be aggregated in a single point. These visual components (i.e. size, color and symbolization) should not be used at the same time. For instance, if we want to distinguish the size of a symbol, it is better to use identical shape to ensure that people are able to visually identify the differentiation of size. In addition to the obstacles in incorporating multiple quantitative data, qualitative information will be demonstrated in form of text at each point, leading to more difficulties in making clear and straightforward maps. Consequently, we will employ graphic symbols at different levels to explicitly visualize as much information as possible.

Table 2 shows a listing of the variables to be visualized in the Ecological Interview Visualization System and the corresponding visual components employed in the map interface. Two basic types of marks, circle points and lines, are adopted to represent the spatial placement of locations and trips. Different types of variables are associated with the mapping objects that they are logically associated with. For example, length of stay at each location will be displayed by different point symbol size. Technically, it can be represented by the size of line symbol as well. But the action of staying actually takes place at the location instead of during the trip. Although it is possible to exhibit the information correctly by both methods, users may have trouble making sense of such a representation intuitively, let alone interpret the combination of so many variables from a single map.

4.2 Symbology

The symbology adopted to visualize multiple variables in Ecological Interview data is shown in table 3. This representation system is designed following Bertin's theory of cartographic visual variables (1983). Color, size, position and shape of points and lines are used as basic visual variables. The symbol schemes are chosen with the rationale of making straightforward maps to facilitate the understanding of a wide range of information. We attempt to ease users' stress when facing an overwhelming number of. For example, the color for safe location is green and for risky location is red, which can be easily kept in mind according to our common sense. The attached symbols are also designed to help users understand the map intuitively, such as bed for weekend/holiday and hammer for weekday.

Figure 1 shows an example of a map of the activity spaces of a 16-year old female substance non-user. In the figures of this paper, the street map of Philadelphia will not be displayed as it should be in the Ecological Interview Visualization system in order to preserve the anonymity of the locations and thus to keep the confidentiality of the study subjects. The gray location is her home, red means risky location, green means safe, yellow means favorite and orange means important. The sun symbol at the upper left corner of her favorite location means she visits the locations regularly at night or at daytime. The hammer and bed symbols at the upper right corner indicate she visits her important and risky locations during weekdays or weekend. All the locations are represented by the same symbol of circle, while differentiated by its size. The distinct size shows that she usually stays longer at routine locations than at home. All triangle symbols attached at the locations stand for multiple responses to a certain attribute. The blue lines mean she visits these locations from home by motor vehicle, and green means by walking or by bicycle. The proportional symbol on the left of each circle indicates the relative value of the selected geographic variable, in this instance, density of drug sale offenses at this location compared to that at other spaces of this individual. A text bar will pop up when the mouse is pressed down at a location point, displaying the description of the location, adolescent's explanation for the location selection as well as her social network score at this place. This teenager said she felt unprotected at a club because "there is always someone getting shot around there," while unexpectedly, her social network quality is actually high there, indicating a protective social relationship she has there.

4.3 Objects and Features

For the purpose of implementation, the Ecological Interview Visualization System consists of several types of data objects to connect data with mapping objects:

1) The individual's activity locations: a location-based data layer contains all the subjects' home and routine activity locations, with geographic variables, selected transportation variables, social network scores and qualitative information at each type of activity space attached.
2) The individual's trips: a route-based data layer stores all the shortest paths from the subjects' home location to each routine location, with selected transportation variable attached.
3) The individual's attribute information: a subject-based table includes each subject's substance user's group involvement and demographic characteristics.

All three objects share the same unique identifiers: subject ID for each adolescent as well as the location ID for each activity space that he/she is involved. The location object contains all the physical locations derived from Ecological Interview and variables at location level. That includes most of the quantitative variables (i.e. geographic variables, transportation variables and social network scores) plus major qualitative variables (i.e. description and explanation of location selection). The route object stores all the shortest paths connecting each adolescent's home and activity spaces built upon the Philadelphia transportation network. The only variable associated with the paths is a numeric transportation variable, transportation mode. Individual records in both locations and trips objects can be identified by their location ID, since each location can only possibly have one trip from that individual's home. In contrast, the subject-based table provides information at the level of subjects (i.e. adolescents), hence should be identified by subject ID. Via these unique identifiers, we are able to retrieve data at both the location level and subject level.

4.4 System Capability

The capabilities of the Ecological Interview visualization system include:

1) Storing and organizing spatial locations and path information and associated spatially referenced variables.
2) Interactively querying and visualizing an individual's travel-activity attributes (including spatial and temporal properties and in-depth explanation), personal aspects, social network qualities and community resources at his/her activity spaces in

dynamic maps.

“Dynamic” means that maps can be changed depending on users’ different definitions instead of over time. Users are free to visualize different individuals, different types of locations and associated trips, and a wide range of geographic variables. According to different selections, the information displayed in the map will be correspondingly adjusted and the extent of view will be automatically zoomed to the selected activity spaces.

As shown in Figure 2, the interactive interface activated by the “individual visualization” button provides several options for users’ free selection. Analysts are allowed to choose a subject ID from the check list, trip(s) to one or more activity locations as well as one geographic variable from the list for visualization. By clicking the button of “Show Qualitative Data,” users will easily view the qualitative explanation and adolescents’ subjective feelings at each location through the text bar popping up when the location is clicked. The “Show legend” tool in the tool bar allows users to view the whole legend of the visual variables. In the table of content, users can easily review the multiple properties (categorical or numeric values) represented by the visual attributes for each location and path.

CHAPTER 5
IMPLEMENTATION

The Ecological Interview Visualization system is implemented by customizing ArcGIS 9.3. The two mapping objects, location and trip shapefiles, are organized and stored in the application as two layers. This section addresses the techniques that are used in processing the mapping objects and implementing the visualization application.

5.1 Generation of Location and Path Objects

The point shapefile with all the activity locations are derived from the geocoding results of street addresses acquired from Ecological Interview. The line shapefile containing all the travelling paths for each individual's home to his/her activity locations are generated using a Python script. With the location shapefile as input, the script calls the network analysis tool embedded in ArcGIS 9.3 repeatedly to calculate travelling routes between each pair of locations based upon the Philadelphia transportation network. This is known as Geoprocessing provided by ArcGIS desktop software. Geoprocessing allows people to automate and document multiple-step GIS procedures known as workflows via model or script. ArcGIS 9.3 supports many of today's most popular scripting environments, such as Python, VBScript, JScript, and Perl (ESRI). We choose to use Python for the present research because it is an easy-to-learn programming language similar to C. Python uses simple readable code to document complicated data structures and make them straightforward to work with. Python can also be seamlessly integrated with more sophisticated programming developer environment such as Java. In addition, Python is free and has a widespread community.

5.2 Implementation of Visual Components

The Ecological Interview Visualization system is developed based on the platform of ArcGIS 9.3 (ESRI, Inc.) by its embedded customization environment, Visual Basic for Applications (VBA). ArcGIS software provides a variety of resources that enable users to customize by their specific preference. The functionality of ArcGIS Desktop can be extended by using one of the developer toolkits and a development language. People can use the embedded VBA environment to develop simple scripts or as an interactive prototyping environment for more complicated development with a COM-based language like Visual Basic 6, Visual Basic .NET, or Visual C++.

This application is constructed using a variety of ArcObjects. ArcGIS is built and extended by the ArcObject software components. The ArcGIS family relies on ArcObjects to provide data management, map presentation functionality, and more. ArcObjects include a wide range of programmable components, such as data frames, layers, features, tables, cartographic symbols and the pieces that make up these things including points, lines, polygons, records, fields, colors and so on (Robert Burke, 2003). All ArcObjects classes have interfaces and all ArcObjects variables are pointer variables that point to the interfaces. Many ArcObject interfaces are used for the implementation of the Ecological Interview Visualization system. First, ArcObjects are used in managing and operating the maps, layers, features and tables during the interactive data retrieval procedure. These interfaces include *IMap*, *IGeoFeatureLayer*, *IFeatureClass* and *IFeatureCursor*, and so on. Second, the visual attributes of lines and points, as well as the text labels for displaying qualitative information, are controlled by a set of ArcObjects interfaces, such as *ISimpleLineSymbol*, *ICharacterMarkersymbol*, *ITextElement*, and *IElementCollection*. Third, *IUniqueValueRenderer* interface is adopted to visualize individual's activity spaces by controlling a renderer where the symbols are assigned to the activity locations of this specific individual.

The user accessible tools and functions in ArcGIS desktop are mostly constructed from the corresponding ArcObjects. By calling ArcObjects in the programming, people can customize the parameters more flexibly. For example, in order to use the multiple symbols at a single point to represent multiple variables, people should manually change the visual attributes of the point symbol in ArcMap. However, the customized application would easily implement such symbology by setting parameters in the programming according to specific values of the multiple variables at each location. The use of ArcObjects in programming also makes it possible for interactive visualization. The visualization is refreshed according to every change made by users on the mapping interface, such as the change of subject ID, the type of location, or the geographic variable. Users' selections of visualization parameters will be passed from the visualization interface to the background program, and translated into parameters for each ArcObject. Therefore, a new map will be produced for the selected subject, type(s) of location, and geographic variable, and the map will also be automatically zoomed in to the selected spaces.

CHAPTER 6
CASE STUDY WITH ECOLOGICAL INTERVIEW VISUALIZATION SYSTEM

The main task of the Ecological Interview Visualization System is to store the activity space data and allow people to easily retrieve the information both numerically and qualitatively. By visualizing the spatial distribution of adolescents' activity spaces and detailed information associated with their activities, we will be able to first get a holistic picture of an individual's spatial movement in their daily life, and then understand how the adolescent is socially associated with those places, what kind of influence the neighborhoods may have on him/her, and how he/she feels and behaves at the activity spaces. The in-depth investigation can shed light on potential underlying relationships within the dataset, which will contribute to hypothesis building as well as post-modeling analysis. This section will demonstrate several case studies using this visualization system for better illustration of its usage in the research.

6.1 Exploring an individual's activity space data

Since we are dealing with spatially referenced data, the spatial dispersion of a person's activity can usually reveal meaningful information about his/her daily life. Figure 3 shows a 17-year-old female who travels every week faraway to her risky, important and favorite locations in northern Philadelphia. Her explanation of visiting those places tells us that she probably has most of her family living in that area. So she might spend a lot of time there, which is a relatively safe area compared to her own house in west Philadelphia according to the density of drug sale offenses. Another example given by Figure 4 is a 17-year-old female user. She travels a long way to her risky place frequently during the weekday, and she has a positive network score there. Simply looking at these numbers, it is difficult to know what is going on at this place. But her articulation explains that she treats her school as the most risky place and she "only talks to two people there", somewhat indicating a special psychological feeling which may be related to her mental health and misbehavior. The next example in Figure 5 is a 19 year old male who identifies three locations as his father's, mother's and girlfriend's houses. Although his father lives in a relatively safer neighborhood than where he lives, he considers there the most risky because it is full of "drama and violence" and his father is probably an alcohol or drug user himself, given the negative network score. Also from his explanation we can tell that his mother does not exert positive influence on him either and he is a teenage father. Hence, we can somewhat understand his lifestyle and the possible reason why he is involved in substance abuse.

Sometimes we can obtain substantial information relevant to the adolescents' delinquency. Through Figure 6, the 18-year-old boy tells us he is an ex-offender and is currently on parole. He has some negative social relationships at a bar close to his home. He is quite resentful with the problematic environment around his house so he spends lots of time with his girlfriend who probably uses drug or alcohol. Figure 7 show the activity space for an 18 year old girl who seems to be raised in a quite protective family, at a relatively safe neighborhood according to the low density of drug sale offenses and highly positive social network effect. However, she tells us that she “smoke weed” at her friend’s house, where she is positively related with someone. Regardless of all the protective social influence, her explanation is straightforward enough to announce that she is a substance user.

By integrating all kinds of activity space data, we will be able to extract plenty of information simultaneously. It is helpful in detecting implicit or profound relationships between variables, and in constructing more reasonable hypotheses in the further step. For example, in Figure 8, the 16-year-old girl shown has non-negative network scores in all of her activity locations, thus indicating a high influence of non-use. Unfortunately, this girl is a substance user. Therefore, we would expect other factors that influence her to use drugs or alcohol. We can see that she regularly visits her brother’s house during the weekd where she has a “fear of being raped “and is exposed “to drugs”, despite of her indication of no social relationships there. So visiting time at the risky place might be a useful predictor and should be verified by further analysis. The other suspected factor is the geographic variable, density of drug sale offenses, due to its high value at her home while she stated a protective social network there.

6.2 Post-Modeling Analysis

The Ecological Interview Visualization System can as well been used in post-modeling data exploration and analysis. Post-modeling analysis refers to the interpretation and explanation of the performance of an existing statistical model usually by thorough investigation of special cases or unexpected patterns. Using the visualization system we will be able to verify outliers efficiently while incorporating not only the information included in the model but also the one that is excluded. It can explain to us why these outliers do not fit well in the model and whether the predictor is powerful or not in general cases. Furthermore, we can usually reveal unexpected relationships between excluded variables.

As an example, we built a simple model by regressing a single variable, social network score at risky place, on the substance abuse variable. We adopted the continuous AADIST variable instead of the substance user group for the purpose of better observation of outliers. The modeling results show that risky social network score is negatively and significantly related to adolescents' involvement in substance abuse. It means that the lower the risky social network score is, the higher the AADISTOT value is. In other words, if a teenager has negative social network score at the risky activity location, he/she is more likely to be a substance user. Figures 9-12 show four worst-fit cases. The first two are predicted to be nonusers while they are actually users, and the last two cases are vice-versa.

The person shown in Figure 9 identifies her risky, important, and favorite places all at her grandmother's house. She considers these places risky because she "gets into a lot of trouble there" and is "locked up there because of drug activity". It is actually not the sort of risky influence that we define in this study. Therefore, although she has a positive social network score, she is not necessarily a non-user. As a matter of fact, the prediction is quite inaccurate due to her misspecification. Her resistance to her grandmother's protection of her from using drugs should be a valuable piece of information and might be related to her psychological health, indirectly indicating her misbehavior. In Figure 10, the boy has a negative social network at his risky place, but its influence is not strong due to its low absolute value. However, he is actually seriously involved in using drugs or alcohol with a very high value of AADIST (AADIST = 67) which is far greater than the predicted value derived from the risky social network score. Hence, there might be other influential factors revealing his severe substance abuse, such as he regularly visits and stays for more than two hours at the risky place where drug activity and violence usually happen, or he possibly has some sort of mental or emotional disorder due to his own statement that “no one knows where” he lives.

Figures 11 and 12 display the activity space data for two non-users whose social network scores at risky places are both 0. The null value of social network score means that the adolescents do not identify any close social contacts at those locations, while not necessarily meaning that they are not influenced by the social environment at all. Therefore there is a possibility of unbiased modeling since the actual missing data are treated as a value of 0 for a continuous variable. Given our data collection procedure, there is a handful of missing data that should not be neglected in risky social network score. Meanwhile, we discover that the two non-users both name their risky places to be their schools, which are in the category of somewhat protective environment for many other youth.

CHAPTER 7
CONCLUSION AND DISCUSSION

The Ecological Interview Visualization System is developed as a prototype for the geovisualizing of activity space data. Given the special characteristics of variables in the Ecological Interview dataset, a 2-D map is adopted as the fundamental framework in this visualization. Unlike the space-time paths portrayed as trajectories in the 3-D space-time aquarium (Hägerstrand, 1970), the spatial and temporal attributes in the Ecological Interview dataset are used to quantitatively represent adolescents' travel-activity behaviors. Furthermore, 2-D representation can possibly manifest more information than visualization in 3-D space, according to human visual capability. The applications of 3-D space-time cube rarely represent more than one variable other than the spatial and temporal distribution (Kwan, 1999, 2000b; Ding and Kwan, 2004). In contrast, 2-D maps are more flexible in integrating diverse information for interactive data exploration (Andirenko and Andirenko, 1999).

The particular visual representation presented in this research is designed to effectively visualize the multiple variables of the Ecological Interview dataset, involving spatial and temporal, as well as qualitative and quantitative information. Using circle points representing activity locations and lines representing trips from home to activity spaces, quantitative transportation variables were integrated by the visual characteristics of the two marks. The symbology is designed to be selective, associative, quantitative, and aptly ordered, according to Bertin's theory of visual variables (Bertin, 1983). The visual characteristics are also selected to help people understand adolescents' travel-activities intuitively, efficiently and accurately. In addition, qualitative variables are associated with each individual and all of his/her activity spaces in the visualization application using dynamic text labels. Through the systematic representation, six quantitative variables are integrated to describe adolescents' travel-activities and the neighborhood environments; individuals' characteristics, their social network score as well as their self-perception and explanation of location selection are qualitatively represented to manifest the multi-level aspects of adolescents' activity spaces.

The Ecological Interview Visualization System can be used primarily to store, manipulate, and represent activity space data. Via the map interface, researchers are allowed to explore the complex dataset with both quantitative and qualitative information interactively and efficiently. With a preliminary focus of the spatial dispersion, this application aggregates information about adolescents' activity spaces and travel-activity patterns as much as possible. This integration is very helpful for exploratory data analysis which emphasizes the detection and discovery of unexpected knowledge. Through several case studies, we also demonstrate how this visualization system might be useful in exploring interesting cases, uncovering potential factors, understanding outliers and interpreting existing models.

However, the Ecological Interview Visualization system is still quite limited from several perspectives. First, the difficulties in collecting the elaborate and comprehensive activity space data may greatly limit the application of such visualization in studying human behaviors. Second, it is hard to observe or examine overall or partial trends or patterns using individual cases. More statistical visual tools may be incorporated to facilitate data exploration of general patterns that occur over groups of subjects. For example, density surfaces can be similarly used as in Kwan's visualization of travel-activity patterns (1999, 2000b, 2003), while the two axes of time and space can be replaced by many other attributes, such as social network scores and neighborhood environmental variables. Such tendency can be visualized while controlling for certain variables, such as age, sex and types of locations. There are many other statistical visualization techniques, such as scatterplots and parallel coordinates, which can also assist the observation and investigation of patterns. One of the most important issues of using visualization as an exploratory tool is to allow users freely choose the variables of their interest for visual testing.

Third, the Ecological Interview visualization is still limited in human-machine interaction. One possible development is to provide more options for researchers' interactive selection, such as allowing the mapping of multiple individuals' activity spaces by selected groups (e.g. age, gender and race, etc.), or integrating selected type(s) of activity spaces (e.g. risky or safe) for multiple adolescents. More modes of interaction should thus be created and the visual variables should be adjusted correspondingly for clearer representation of the activity spaces of a group of adolescents, which will significantly differ from mapping a single person. Finally, qualitative information shows critical power in researchers' understanding of data outliers or patterns according to the case studies in chapter 6; however, it is hard to draw explicit conclusion from the qualitative data since it varies from case to case. Consequently, it might be helpful to appropriately code the qualitative information to quantitative values, and incorporate both variables.

REFERENCES CITED

Al-Kodmany, K. 2001. Bridging the gap between technical and local knowledge: Tools for promoting community-based planning and design. *Journal of Architectural and Planning Research*, 18 (2):110-130.

Andrienko, G.L. and Andrienko, N.V., 1999. Interactive maps for visual data exploration. *International Journal of Geographic Information Science*, 13(4): 355-374.

Bertin J., 1967. *Semiology of Graphics: Diagrams, Networks, Maps*. University of Wisconsin Press.

Burrough, P.A., and Frank, A.U., 1996, Geographic Objects with Indeterminate Boundaries: Bristol, Taylor & Francis Inc.

Burke, R., 2003. *Getting to Know ArcObjects: Programming ArcGIS with VBA*. ESRI Press.

Carpendale, M.S.T., 2003. Considering Visual Variables as a Basis for Information Visualisation. *Computer Science TR #2001-693-16*. Dept. Computer Science, University of Calgary.

Ding, G.X. and Kwan, M.P., 2004. 3D-VQGIS: 3D Visualization and Qualitative Analysis of Geospatial Data. *Technical papers: 3rd International Conference on Geographic Information Science*, Maryland, USA.

Dykes, J.A., 1997. Exploring spatial data representation with dynamic graphics. *Computers & Geosciences*, 23(4): 345-370.

Dykes, J.A., MacEarchren, A.M. and Kraak, M.J., (Eds.), 2005. *Exploring Geovisualization*. Elsevier, USA.

Dodge, M., McDerby, M., Turner, M., 2008. *Geographic Visualization: Concepts, Tools and Applications*. Wiley, London.

Edsall, R.M., Harrower, M. and Mennis, J., 2000. Visualizing Properties of Temporal and Spatial Periodicity in Geographic Data. *Computers & Geosciences*, 26(1): 109-118.

Entwisle B, Rindfuss RR, Walsh SJ, Evans TP, Curran SR (1997) Geographic information systems, spatial network analysis, and contraceptive choice. *Demography*, 34(2):171-187

Frechtling, J., Sharp, L., and Westat, 1997, User-Friendly Handbook for Mixed Method Evaluations, v. 2004, no. 6-24.

Gahegan, M., 2001. Visual Exploration in Geography, in Miller, H.J. and Han, J. (Eds.), *Geographic Data Mining and knowledge Discovery*. Taylor & Francis, London and New York, 260-287.

Hägerstrand, T., 1970. What about people in the regional science? *Papers in Regional Science Association*, 24(1): 7-21.

Heer, J. and Boyd, D., 2005. Vizster: Visualizing Online Social Networks. *Proc. IEEE Symposium on Information Visualization*. IEEE Press, Piscataway, NJ, 33-40.

Longley, P.A., 1998. GIS and the development of digital urban infrastructure. *Environment and Planning B,* Anniversary Issue: 53-56

Kwan, M.P. and Hong, X.D., 1998. Network-based constraints-oriented choice set formation using GIS. *Geographical Systems*, 5:139-162

Kwan, M.P., 1999. Gender, the Home-Work Link, and Space-Time Patterns of Nonemployment Activities. *Economic Geography*, 75(4): 370-394.

Kwan, M.P., 2000a. Analysis of human spatial behavior in a GIS environment: Recent developments and future prospects. *Journal of Geographical Systems,* 2(1): 85-90.

Kwan, M.P., 2000b. Interactive Geovisualization of activity travel patterns using three-dimensional geographical information systems: a methodological exploration with a large data set. *Transportation Research C-EMER*, 8 (1–6): 185– 203.

Kwan, M.P. and Lee, J., 2003. Geovisualization of Human Activity Patterns Using 3D GIS: A Time-Geographic Approach, in Goodchild, M.F. and Janelle, D.J., Eds, *Spatially Integrated Social Science: Examples in Best Practice.* Oxford University Press, Oxford, 48-66.

MacEachren, A.M., Kraak, M.J., 1997. Exploratory Cartographic Visualization: Advancing the Agenda. *Computers & Geosciences*, 23 (4): 335-343.

MacEachren, A.M., Kraak, M.J., 2001. Research challenges in Geovisualization. *Cartography and Geographic Information Systems*, 28 (1): 3 –12.

Mason, M., Cheung, I., and Walker, L., 2004. Substance Use, Social Networks, and the Geography of Urban Adolescents. *Substance Use & Misuse*,39 (10–12): 1751–1777.

Miller, H.J., 1994. Market area delimitation within networks using geographic information systems. *Geographical Systems*, 1:157-173.

Miller, H.J., 1998. Beyond the isotropic plane: elements and tools for a geospatial analysis. Paper presented at the Varenius specialist meeting "Status and trends in Spatial Analysis". *The National Center for Geographic Information and Analysis,* December 10-12, Santa Barbara, CA.

Mennis, J. and Mason, M.J., in press. The influence of people and places on adolescent substance use: Integrating activity space and social network data for analyzing health behavior. *Annals of the Association of American Geographers*.

Moberg, D. P. and Hahn, L., 1991. The adolescent drug involvement scale. *Journal of Adolescent Chemical Dependency*, 2(1): 75-88.

Orford, S., Harris, R. and Dorling, D., 1999. Geography: Information Visualizationi n the Social Sciences. *Social Science Computer Review*, 17(3): 289-304.

Peterson, M.P., 1999. Active legends for interactive cartographic animation. *International Journal of Geographic Information Science*, 13(4): 375-384.

Renz, J., 2002, Qualitative Spatial Reasoning with Topological Information: Tokyo, Springer, v. 2293.

Shawn, S.L., Yu, H.B. and Bombom, L.S., 2008. A Space-Time GIS Approach to Exploring Large Individual-based Spatiotemporal Datasets. *Transactions in GIS,* 12(4): 425-441.

Smith, N. and Rhind, D.W., 1998. Characteristics and sources of framework data. In: Longley PA, Goodchild MF, Maguire DJ, Rhind DW (eds) *Geographical Information Systems: Principles, Techniques, Management and Applications*, vol. 2. John Wiley: New York, 655-666.

Tufte, E. R. 1983. *The visual display of quantitative information*. Cheshire, Conn. (Box 430, Cheshire 06410): Graphics Press.

Winchester, H.P.M., 2000, Qualitative Research and its Place in Human Geography, *in* Hay, I., ed., Qualitative Research Methods in Human Geography: Victoria, Australia, Oxford University Press, p. 1-21.

Winters, K., Botzet, A., Anderson, N., Bellehumeur, T., and Egan, B., 2001. *Screening and Assessment Study, Wisconsin Division of Juvenile Corrections, Alcohol and Other Drug Abuse*. University of Minnesota, Center for Adolescent Substance Abuse Research, Minneapolis.

Table 1. Multiple variables in Ecological Interview data.

Variable names	Variable descriptions	Value	Value description
Output			
Substance user group	Substance user or non-user	1	User
		0	Non-user
Demographic variables			
Gender	Individual's gender	0	Female
		1	Male
Age	Age at interview	13 to 20	continuous
Transportation variables			
Type of locations	Including home, safe, risky, important, favorite and religious	0	No
		1	Yes
distance	Shortest paths distance from home to routine locations calculated by python script based on the Philadelphia street network	continuous	meters
Time	Time in a typical day to visit the location	0	Nighttime
		1	Daytime
Day	Day in a typical week to visit the location	0	Weekend & holiday
		1	Weekday
stay	Length of stay in the location	1	< 30 min
		2	31min - 1hr
		3	1 - 2hrs
		4	> 2hrs
Transportation mode	Vehicles utilized to visit the location	0	Motor vehicles
		1	Walk/bike
Qualitative description	Qualitative explanation of the adolescents' selection of locations and feeling at the spaces	text	
Social network scores	an index score addressing the risky or protective aspects of an adolescent's social network at each type of activity spaces	-70 to 70	Positive score means protective influence from the social network, while negative score means risky influence
Geographic Variables	33 geographic variables reflecting community resources at each type of activity spaces	continuous	

Table 2. Multidimensional variables and corresponding visual attributes.

Mapping objects	Variables	Visual Attributes
Path	Transportation mode	Line color
Place	Type of routine locations	Symbol color
	Time of day	Attached symbol
	Day of week	Attached symbol
	Length of stay	Symbol size
	geographic variables	Proportional symbol
Text symbols	Substance user group	Text labels
	Demographic characteristics	Text labels
	Social network scores	Pop-up Text bars
	Qualitative text	Pop-up Text bars

Table 3. Listing of Symbology

Features	Visual Attributes	Values	Visual symbols
Transportation mode	Line color	Walk/bike	
		Motor vehicles	
		unknown	
Type of routine locations	Symbol color	Home	
		Safe	
		Risky	
		Favorite	
		Religious	
		Important	
Time of day	Attached symbol	Daytime	
		Nighttime	
		Multiple Response	
Day of week	Attached symbol	Weekday	
		Weekend/holiday	
		Multiple Response	
Length of stay	Symbol size	< 30 min	15
		31min - 1hr	25
		1 - 2hrs	35
		> 2hrs	45
geographic variables	Proportional symbol	scale	

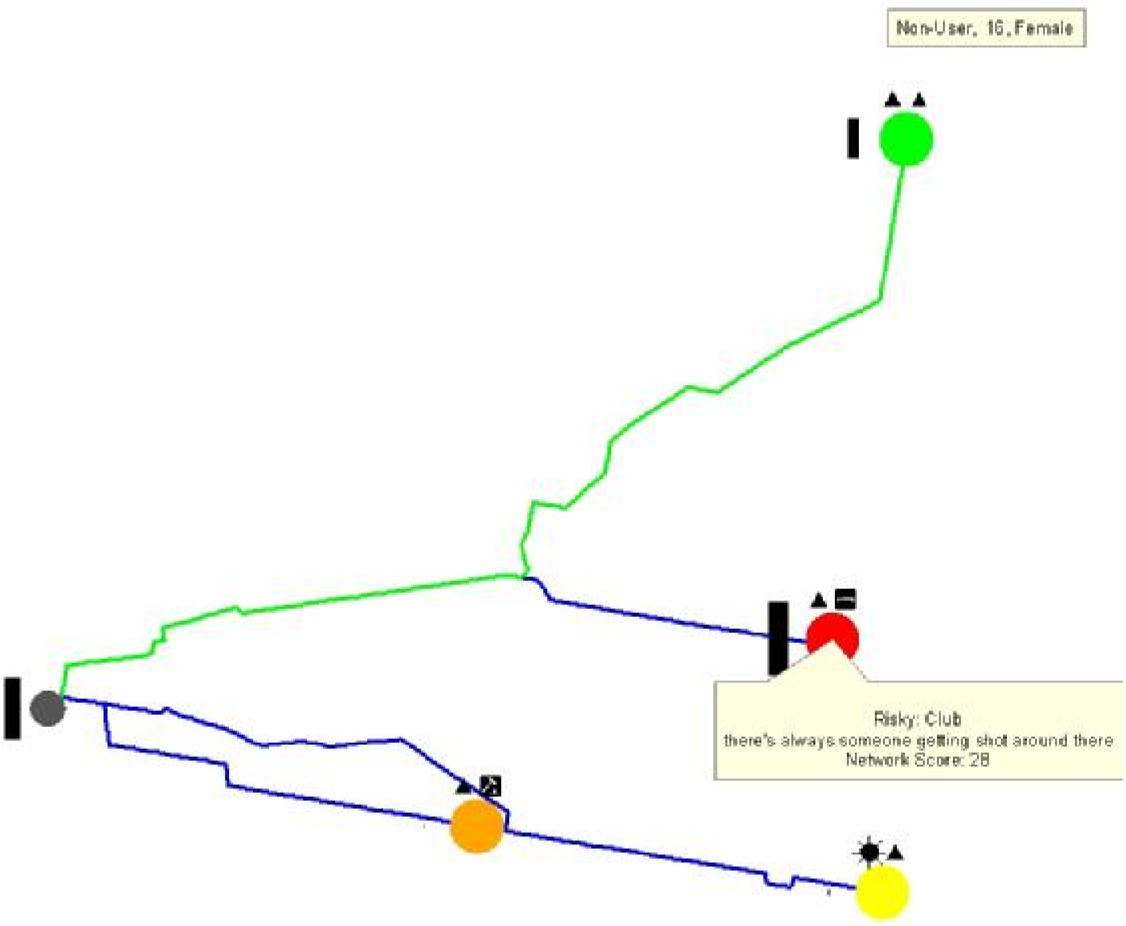

Figure 1. An example of all the travel-activities for a 16-year-old female non-user. No street map is shown as it should be in the Ecological Interview Visualization System for the purpose of preserving the privacy of the study subjects.

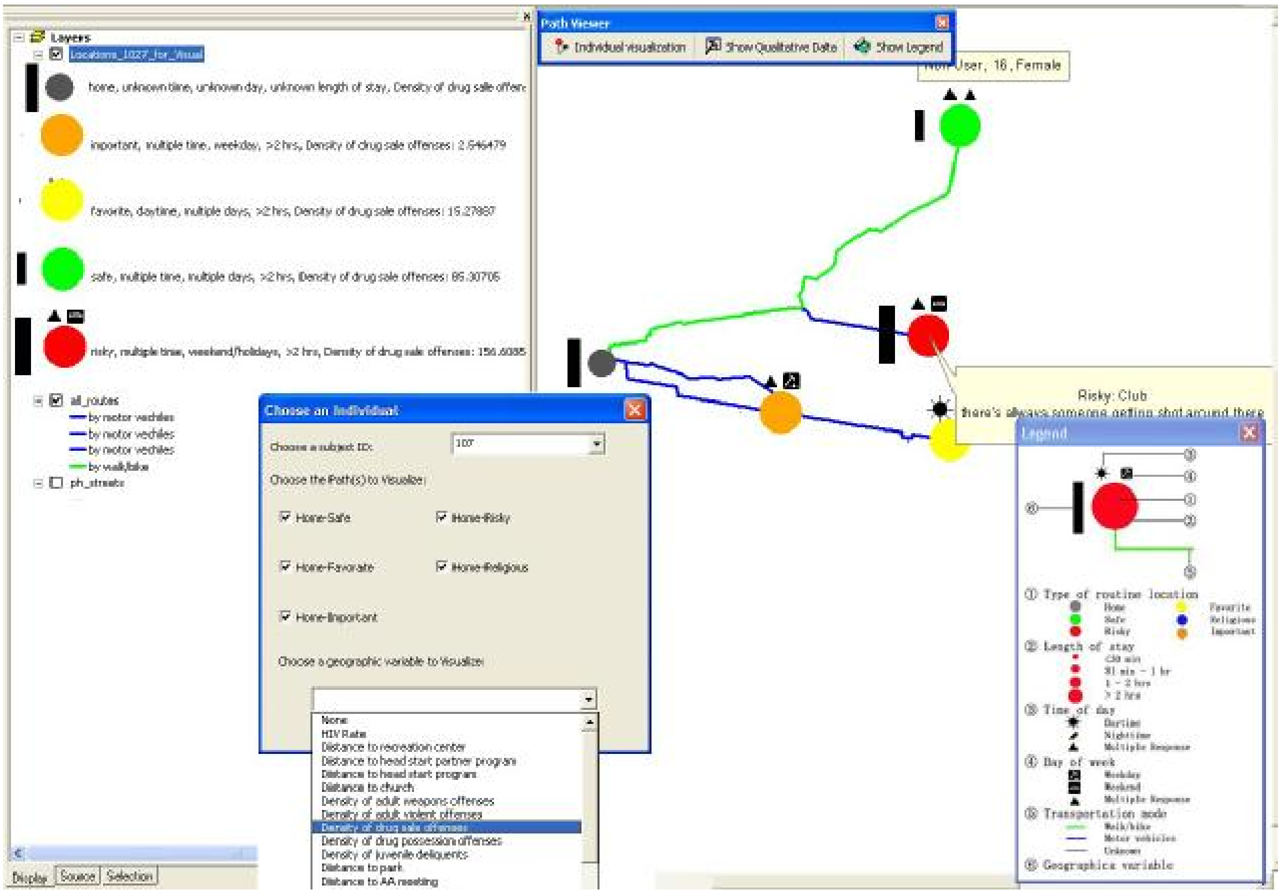

Figure 2. Interfaces of Ecological Interview visualization system

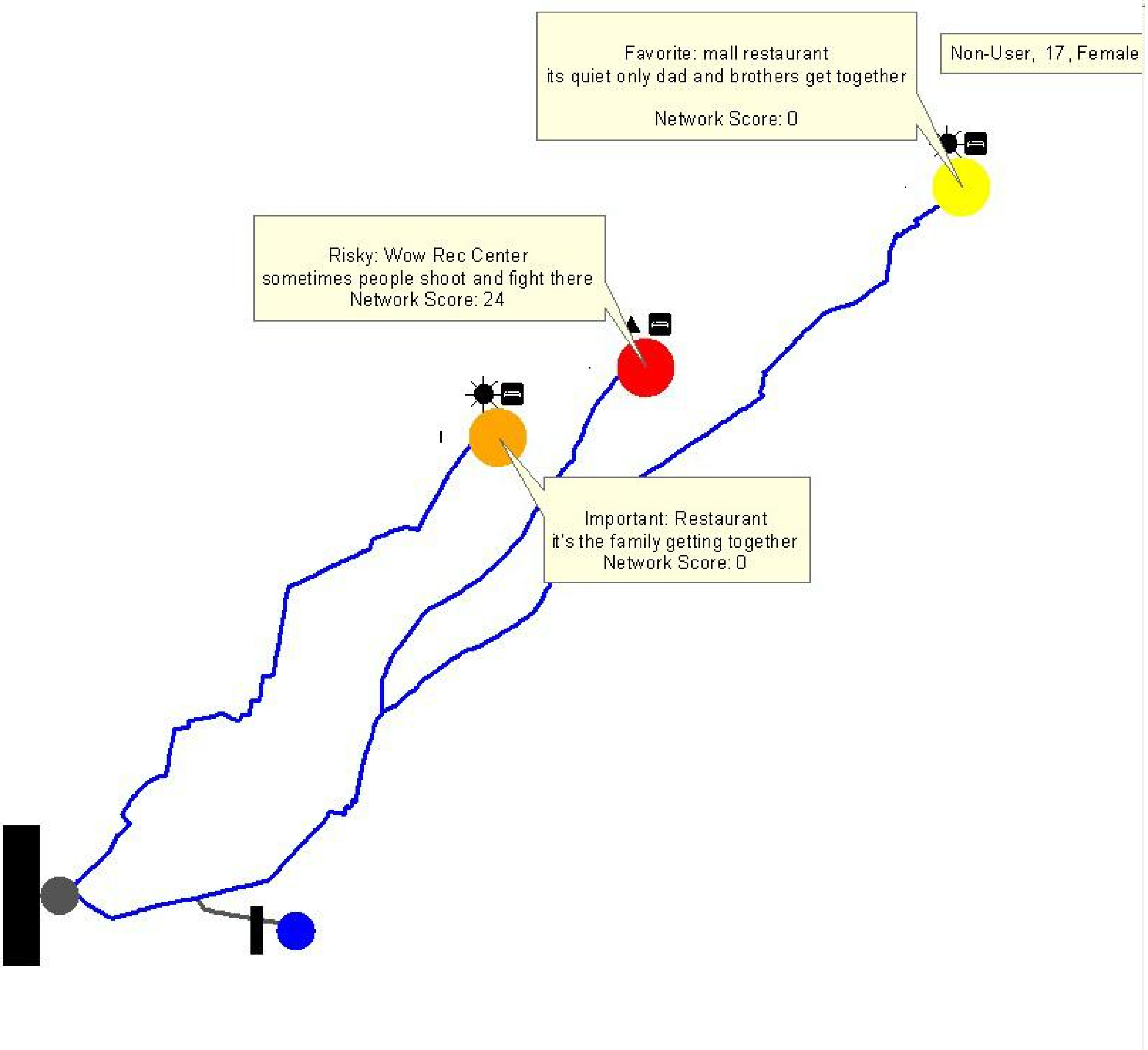

Figure 3. Activity space data of a 17-year-old female non-user (subject id = 288), with the geographic variable density of drug sale offenses

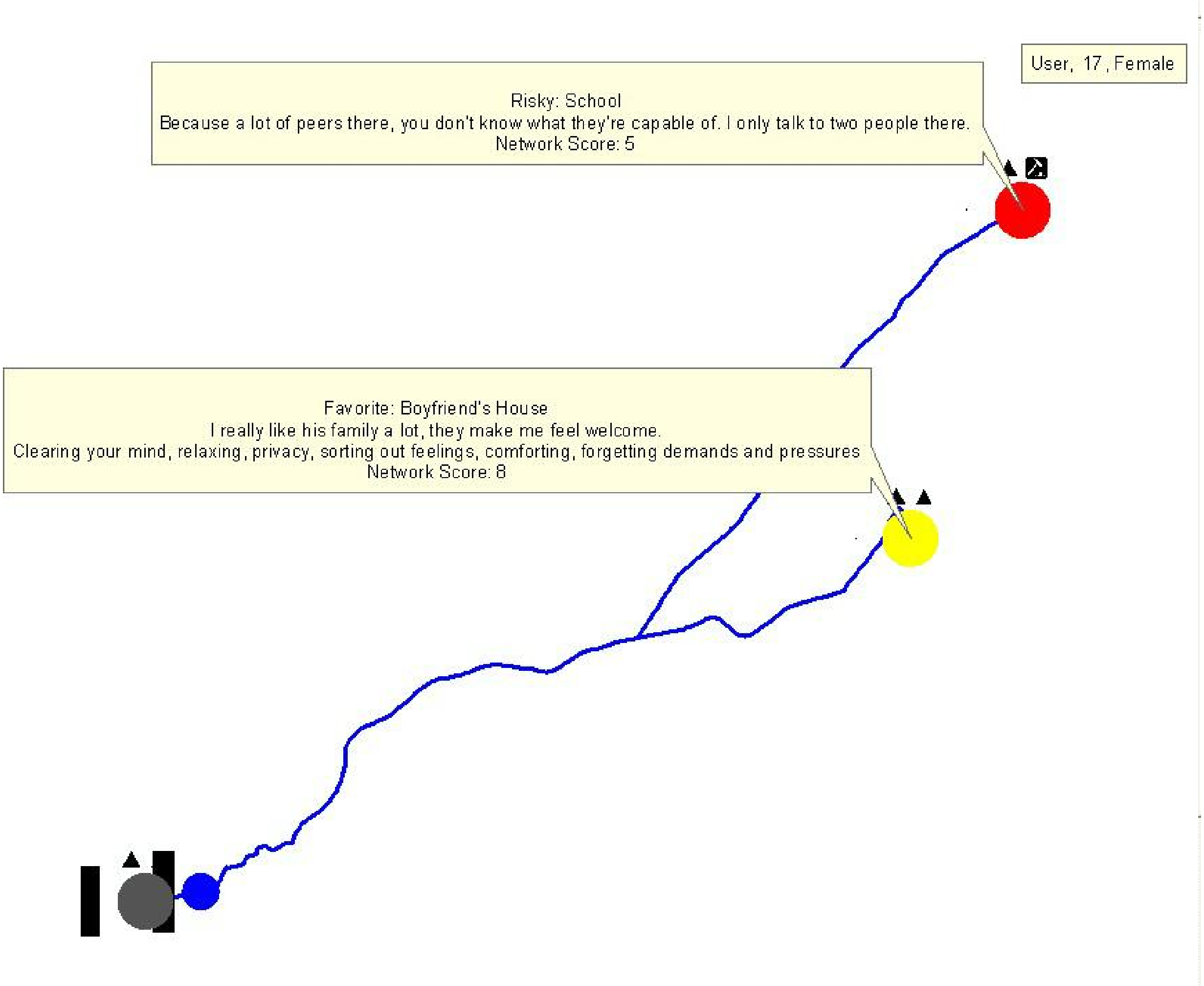

Figure 4. Activity space data of a 17-year-old female user (subject id = 70), with the geographic variable density of drug sale offenses

Figure 5. Activity space data of a 19-year-old male user (subject id = 111), with the geographic variable density of drug sale offenses

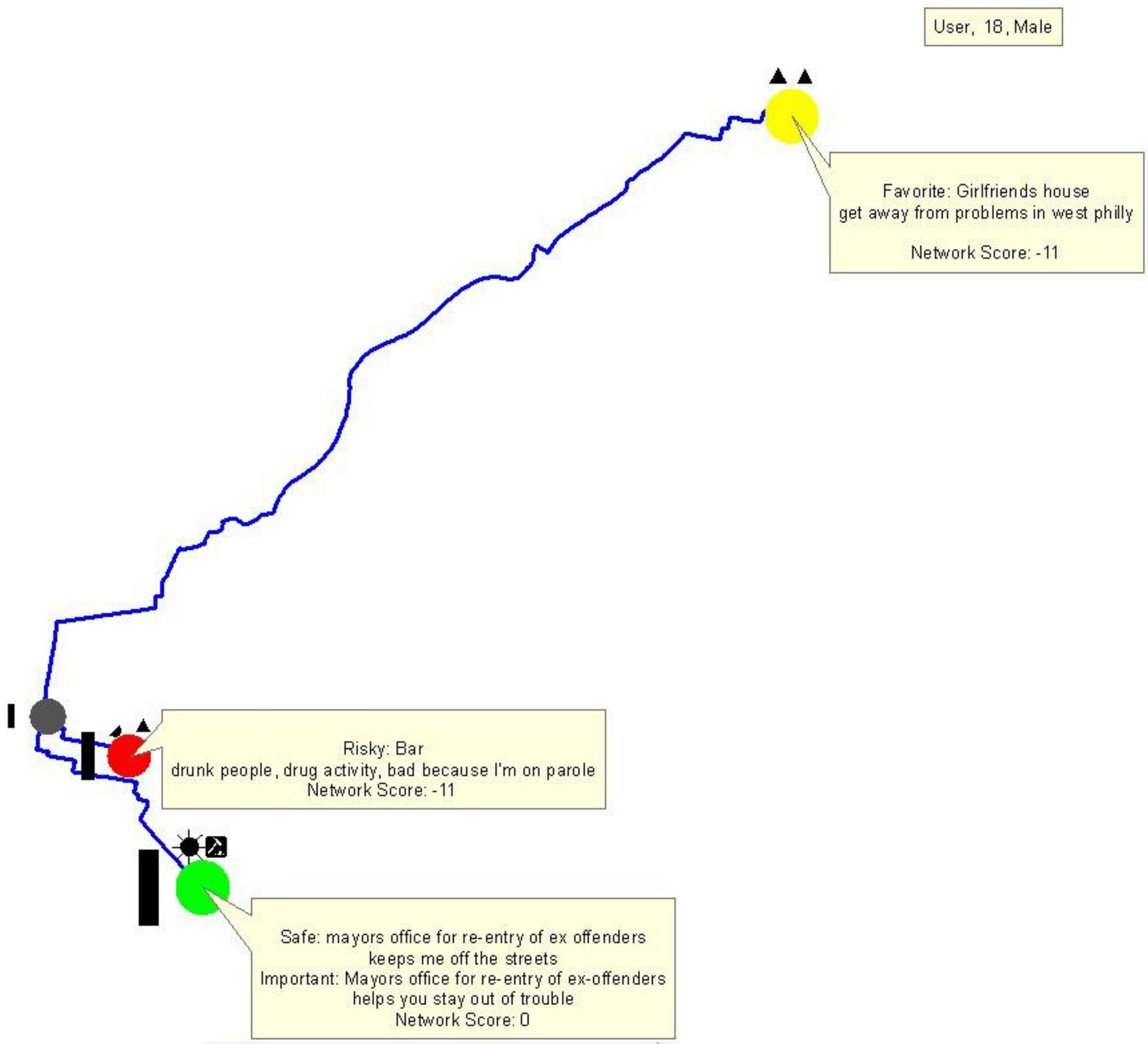

Figure 6. Activity space data of a 18-year-old male user (subject id = 193), with the geographic variable density of drug sale offenses

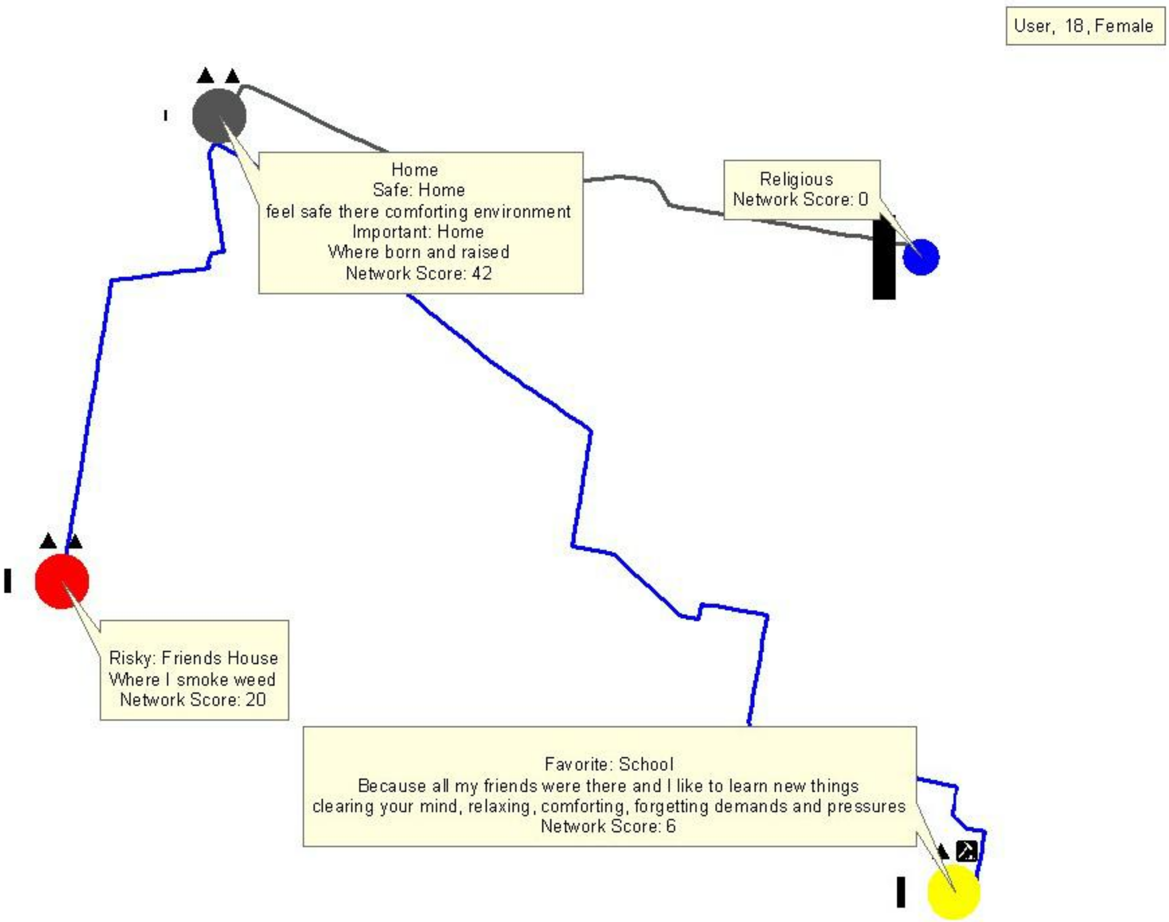

Figure 7. Activity space data of a 18-year-old female user (subject id = 73), with the geographic variable density of drug sale offenses

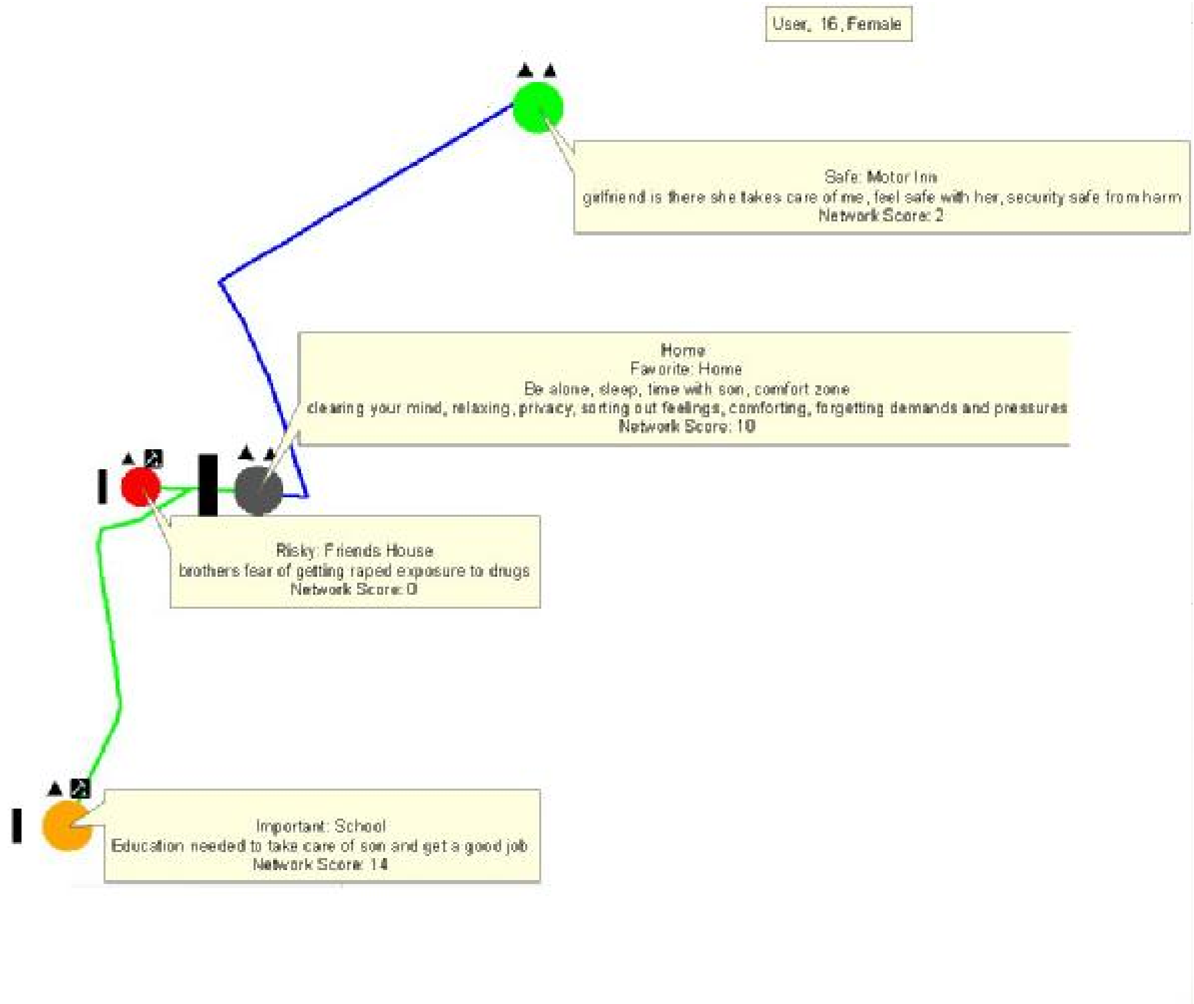

Figure 8. Activity space data of a 16-year-old female user (subject id = 20), with the geographic variable density of drug sale offenses

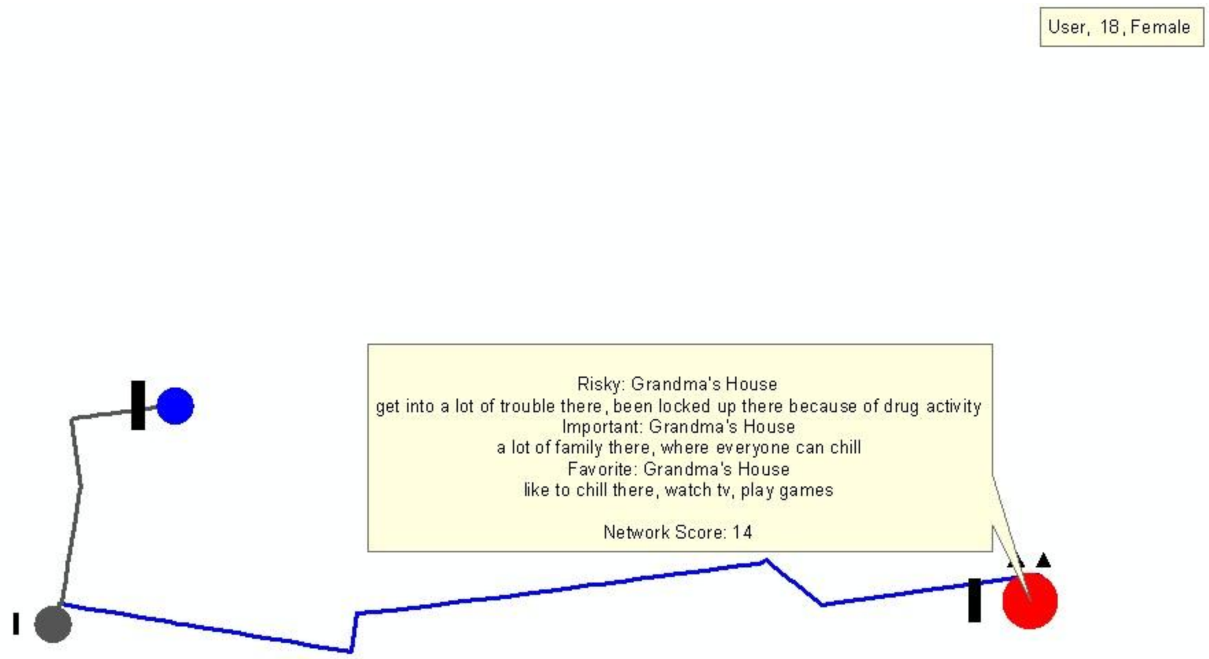

Figure 9. Activity space data of a 18-year-old female user (subject id = 300), with the geographic variable density of drug sale offenses

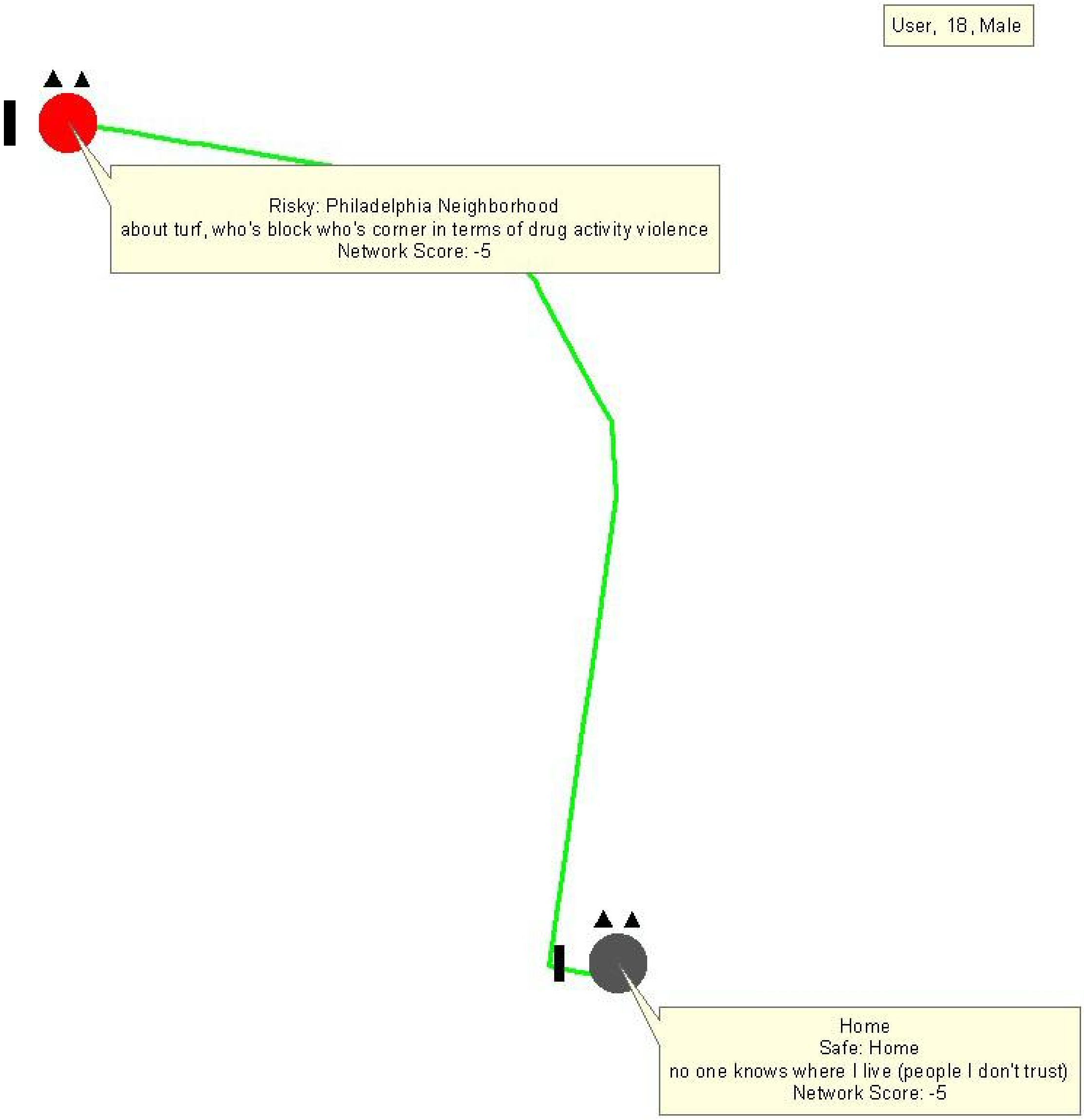

Figure 10. Activity space data of a 18-year-old male user (subject id = 326), with the geographic variable density of drug sale offenses

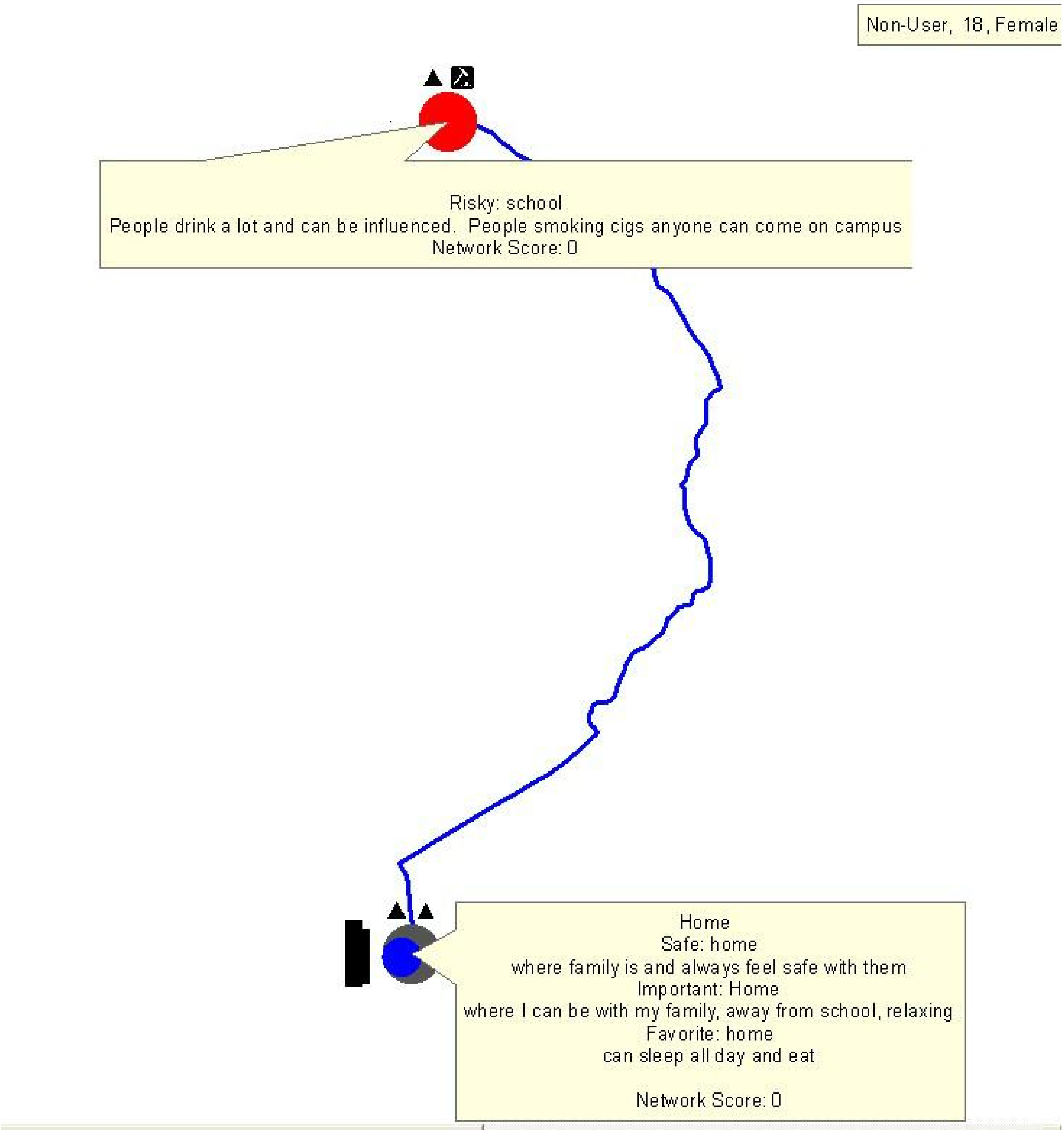

Figure 11. Activity space data of a 18-year-old female non-user (subject id = 305), with the geographic variable density of drug sale offenses

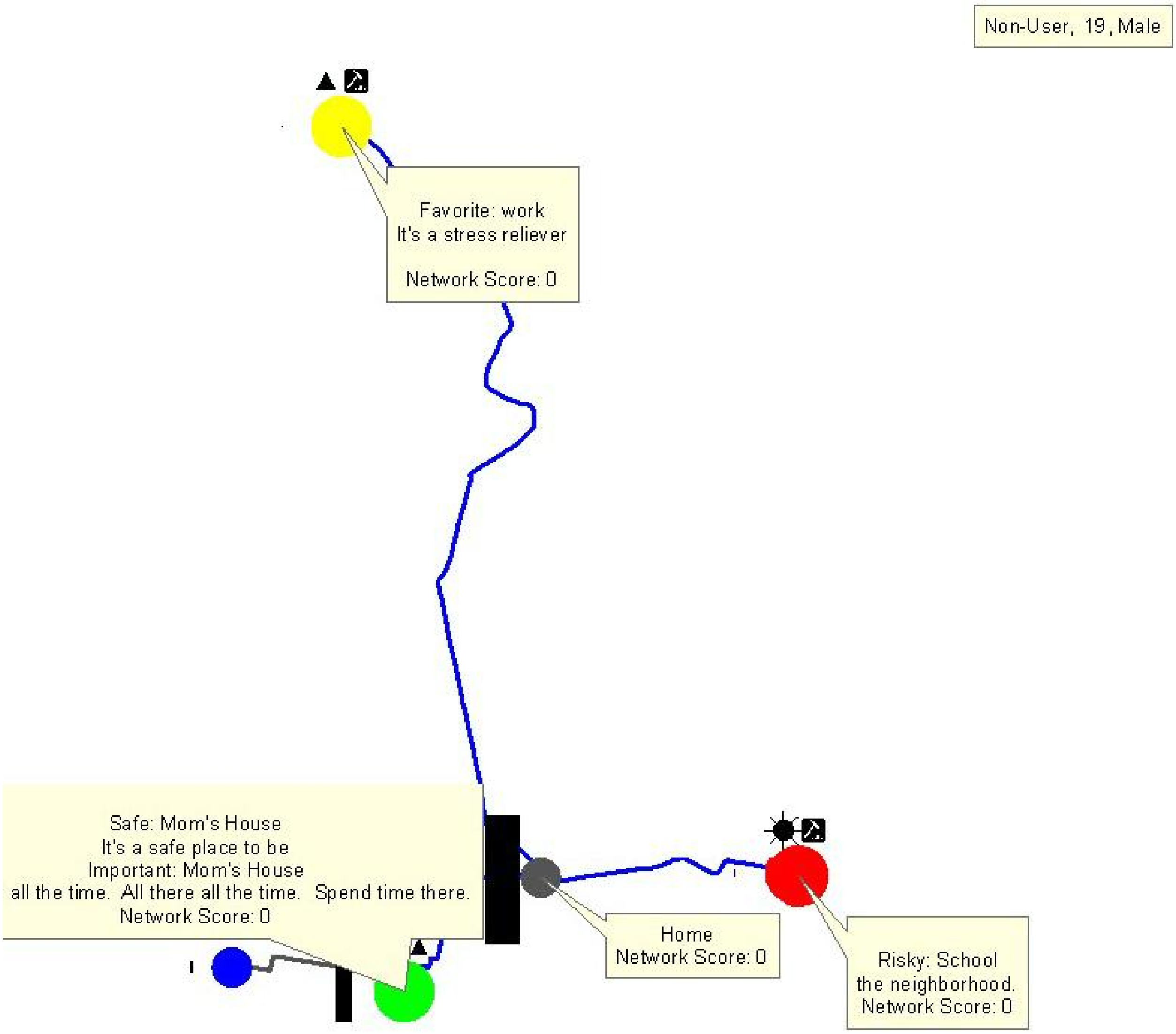

Figure 12. Activity space data of a 19-year-old male non-user (subject id = 291), with the geographic variable density of drug sale offenses