Visual Analysis of Place Connectedness by Public Transport

Natalia Andrienko, Gennady Andrienko, Fabian Patterson, and Hendrik Stange

Abstract—The concept of place connectedness (traditionally termed 'accessibility') refers to the ability of people to reach various services and to participate in activities. Connectedness by public transport is especially important for underprivileged and elderly people, while active use of public transport by the general population contributes to reducing traffic congestions and air pollution in cities. Place connectedness analyses are performed for a variety of purposes. In communication with transportation experts, we performed conceptual modeling of the domain of problems related to place connectedness, defined the system of analysis tasks, and matched the tasks to visual analytics techniques that are capable to support them. In this paper, we introduce the task typology and present the visual analytics techniques using several example scenarios of place connectedness analyses.

1 Introduction

In human geography and transportation, the concept of accessibility refers to the ability of people to reach various places for using services or performing activities [1]. Analysis of accessibility, in particular, by public transport means, is very important for transportation and land use planning as well as for understanding and improving the living conditions of underprivileged and vulnerable populations [2]. Furthermore, good accessibility by public transport can decrease the use of private vehicles and thereby reduce air pollution, noise, and traffic congestions in cities.

Accessibility analysis requires data representing the (public) transport network. The preparation of such data has always been laborious and time consuming, but this is changing since the appearance of General Transit Feed Specification (GTFS) [3] data provided over the Internet for an increasing number of cities and regions. Accessibility analyses are usually conducted with the use of geographic information systems (GIS) for performing computations and visualization of results on maps. One kind of result is an area from within which a certain target or a set of targets (e.g., schools, supermarkets, hospitals, etc.) can be reached within a given time budget. This information is often represented on a map by isochrones [4], which are lines enclosing areas of equal travel time to or from the given target or set of targets. Another popular representation is a choropleth map showing for each location how long it takes to get to the nearest target [5]. Choropleth maps are also used to represent the counts of opportunities of particular types (e.g., jobs, child care facilities, etc.) that can be available within a predefined travel time from each location [6]. Map-based analysis is typically complemented with statistical analysis involving the use of statistical graphics.

When using a GIS and statistical software, the analyst has limited opportunities for interacting with visual displays, viewing data from different perspectives, performing various comparisons, and considering other kinds of information apart from accessibility indicators computed for places. Thus, important information is how the transport network links are used, what are their roles in enabling the access to the targets, where much time is spent on waiting for the next connection, and how changes in the public transport circulation may affect the accessibility. Inclusion of these kinds of information extends the commonly adopted meaning of the term 'accessibility'; therefore, we use instead the term *connectedness*, which encompasses both place- and link-based information. Another reason is that the terms 'accessibility' and 'reachability' may be understood as referring only to reaching target locations from other locations but not involving the opposite travel direction, while the term 'connectedness' may have a broader meaning.

In our work, we have been pursuing two goals:

1) to study and describe in a systematic way the space of tasks pertinent to the concept of connectedness analysis, including also tasks that are of potential interest but not commonly performed due to limited support by available tools, and
2) to define a suite of visual analytics techniques that are capable to support the system of analysis tasks thus identified.

In developing both the task typology and the visual analytics toolkit, we actively communicated with transportation domain experts, who judged our ideas and examples of tool use from the perspective of domain relevance and suggested us further possible analysis scenarios and potentially useful applications of the tools. In this manner, we have fulfilled both goals, and the results have been validated by the domain experts. These results are described in our paper, which has the following structure.

- *N. Andrienko and G. Andrienko are with the Knowledge Discovery Department of Fraunhofer Institute IAIS, Sankt Augustin, 53757, Germany and with City University London, UK.*
  E-mail: see http://www.geoanalytics.net/and
- *F. Patterson and H. Stange are with the Knowledge Discovery Department of Fraunhofer Institute IAIS, Sankt Augustin, 53757, Germany*

Manuscript received April 19, 2015; revised August 26, 2015.

After discussing the related work in section 2, we describe the data used for connectedness analyses (section 3). Section 4 presents several examples of analysis scenarios and then introduces the overall problem statement, the system of tasks, and the requirements for supporting these tasks. Section 5 presents the proposed exemplary set of visual analytics tools, and section 6 concludes the paper.

2 Related Work

A recently published survey [7] outlined the directions of the visual analytics research related to transportation. The following topics discussed in the survey are related to our work: (1) exploration of dynamic characteristics of movement along a selected route; (2) analysis of movements between two selected locations; (3) aggregative study of collective movements over a territory.

For exploring movement along a selected route (1), public transport delays can be shown in a matrix with columns corresponding to ordered stops along the route, rows to trips or time intervals, and colors in the cells representing the arrival time in respect to the schedule: red for late, yellow for punctual, and green for early [8]. For providing the geographic context, this information could also be represented by 3D stacking of trajectories with segments colored according to movement attributes [9]. This technique was used to represent counts of passengers in metro trains [10]. Based on smart card use data, which include the start and end locations and times of each trip, the probable trajectories for all trips are reconstructed as the optimal routes from the trip origins to the destinations and aggregated into passenger flows between metro stations by time intervals. The flows for a single selected line are visualized in a 2D matrix and in a map-based 3D display. Pensa et al. [11] propose a web-based visualization tool which displays bus ridership for selected time period on a map.

For analyzing movements between two selected locations (2), Palomo et al. [12] create a visual overview of the daily schedule of a single train line. The temporal density of the trips, the average speed, or other characteristics are shown by a heat map built with the use of a kernel density estimation technique. Wunderlich et al. [13] compare several variants of visualization design for representing chances to arrive in time for changing to a connecting service, taking into account schedules and statistics of delays. For two selected way points, Zeng et al. [14] show aggregates (counts and durations) of the trips arriving to the first point from all possible origins and the trips departing from the second point to all possible final destinations. The trip segments between the selected points are aggregated into flows by the origin-destination pairs, and the dynamic characteristics of these flows are shown in the display.

Studies of collective movements over a territory (3) require visualization of flows between places, which is a challenging problem, since the flows can massively intersect. Guo et al. [15] used small multiple maps, where each map shows the distribution of the destinations for a selected origin. Similarly, OD maps [16] consist of multiple matrices whose cells correspond to places. A special layout algorithm puts the places in the matrix cells so that their relative positions are similar to the relative positions in the geographic space. FlowStrates [17] visualizes temporal dynamics of origin-destination flows by connecting two maps with the origins and destinations through a time panel which represents flow dynamics. To exclude line intersections, departures from each origin or arrivals to each destination can be aggregated by direction and distance ranges and represented on a map by diagrams [18]. The aggregates are also used for clustering of time intervals to reveal periodic patterns and trends in the overall mobility.

For studying durations of movement from/to a selected place, a technique called isochrone map is used since Galton proposed it in 1881 [19]. An isochrone is a line connecting places that are reachable within the same time. In a PhD thesis [20] defended in 1911, Riedel described in detail how to compute isochrone maps based on public transport timetables. Currently, GIS software is used for this purpose [21], and several APIs, including open (e.g. OpenRouteService [22]) and commercial (e.g. Google Maps API [23]) services, provide tools for calculating one-to-one and one-to-many optimal routes using selected transportation modes at given times.

Several visualization approaches enhance the expressiveness of isochrone maps. Inspired by the cartogram techniques developed in cartography, Hong et al. [24] transform an isochrone map into a Traffigram where distances from a travel origin are proportional to the travel times. Traffigram allows generating different map distortions for different times of a day thus supporting side-by-side visual comparison. Composite flow maps by Cornel et al. [25] use flow line textures and glyphs for showing the variety of available travel choices, which can support assessment of the sustainability of the transportation in situations when some of the planned services are suspended. The TravelTime platform [26] allows to build a series of isochrone layers corresponding to different conditions (e.g., different times or different transportation modes) and superimpose them on top of a common map background for visual comparison.

While isochrone-based approaches are good for providing an overview of travel times, only a few implementations support explicit comparison of multiple situations. USA-Graph by Kamw et al. [27] applies polygon intersection operations that enable comparison of regions accessible within a given time budget. Stewart [28] explores what-if scenarios of changes in a public transport network and shows their impact on selected accessibility metrics such as accessibility to jobs or pharmacies. These approaches do not support inspection of details, detection of bottlenecks, and understanding of factors that could explain the differences. Our approach aims to fill this gap.

The closest to our work thematically is the approach where a tree of paths by public transport from a selected place to all possible destinations is shown in a temporal display, so that the branch lengths are proportional to the travel durations [29]. For selected destinations, more detailed information can be shown, which includes the public transport routes (lines) used on different trip segments and the variation of the time of waiting at the intermediate stops over a day. The geographic context is provided by a separate map window displaying either a usual map or a time cartogram, where distances between locations are proportional to the travel times rather than geographic distances.

It can be concluded that, on the one hand, there are visual analytics approaches that can support some tasks in connectivity analysis, on the other hand, the tasks and needs of this application domain have not been yet comprehensively studied and addressed. These limitations motivate our research, in which we are especially interested in defining and exploring possible approaches to supporting comparative analyses and "what if" modeling.

3 Data

For presentation of our concepts and approaches, we use data examples in the GTFS format, which is described below; however, the specific data format is not essential. It is only necessary that data contain information about each public transport trip, including the transportation mode, route identifier, carrier or trip identifier, and the sequence of stops with the times of arrival and departure.

3.1 Processing GTFS data

The General Transit Feeds Specification (GTFS) [3] is an open data format for public transportation schedules and associated geographic information. The GTFS data consist of several text files, which form an image of a relational database containing the public transport timetables for a certain transport network for a given time period. The files can be downloaded from TransitFeed.com [30]. Given a date, the following steps need to be done for generating trajectories from a collection of GTFS files:

1) Find the available services and their respective $service\_id$'s for the given date in the $calendar.txt$ and $calendar\_dates.txt$ files.
2) If there is a file $frequencies.txt$, which defines the frequencies of services for different time intervals, reconstruct the trips (i.e., stop sequences) and stop times of the services from these data.
3) A file $trips.txt$ may specify additional trips not covered by $frequencies.txt$, or the latter may be absent. Use the $service\_id$'s to extract all trips and their respective $trip\_id$'s from the $trips.txt$; then use the $trip\_id$'s to extract the stop times from the file $stop\_times.txt$.
4) Extract the coordinates of the stops from the file $stops.txt$.

This process largely relies on matching identifiers in different files; however, human-entered data often contain errors. For validating the data quality [31], it is necessary to check their spatial and temporal coverage using maps and time histograms. Further checks can be supported by redundant information in optional GTFS files that are often provided. Thus, files $routes.txt$ and $shapes.txt$, which list stops along each route, can be used for cross-checking the completeness of the generated trajectories.

An optional file $transfers.txt$ stores minimal feasible connection times between neighboring stops. If not available, the values can be estimated based on the distances and assumed walking speed.

3.2 Example data sets

The examples considered in this paper are based on data from two cities, Warsaw and Madrid, which have extensive multi-modal public transport networks with buses (including express and night services), trams, metro, and local trains. The city and suburban networks are integrated by shared stops and common fares, which enable economical and convenient connections. Both cities frequently provide actual timetables via TransitFeeds [32], [33]. The timetable for Warsaw covers 6,243 stops on 336 routes of bus, tram, metro, and local trains. The timetable for Madrid covers 6,030 stops on 228 routes of bus, tram and metro. Unfortunately, the timetables of the local railway lines are not available. Therefore, the examples based on the Madrid data do not reflect the complete picture, but they are used to illustrate the general approach.

4 Analysis Tasks and Requirements to Analysis Support

In this section, we define and discuss the tasks in connectedness analysis that can be performed solely based on transport network specifications and timetables of operation of transport services. We do not consider analyses requiring involvement of additional data, such as the distribution and demographic characteristics of the resident population, public transport use, travel demand, vehicle capacities, and transportation costs. Notes concerning the use of these data are made in Section 6.

4.1 Development of the task typology

No attempts of systematically defining the space of possible questions (tasks) related to place connectedness have been made so far. We did such an attempt in two international research and innovation projects. As we noted in section 1, the lack of convenient opportunities for interactive exploration limits the variety of tasks that can be performed in connectedness analysis. For this reason, the system of potentially meaningful task types cannot be fully defined based only on examples existing in the literature.

In our projects, the development of the task typology was done iteratively through repeated communication of visual analytics (VA) researchers with project partners professionally interested in connectedness studies. Several partners were from the city administrations of Warsaw and Dublin; they are further referred to as "domain experts". Other partners were from a company in Madrid providing analytical services in the domain of transportation; they will be called "transportation analysts". Additionally, several partners from Belgium, Israel, Italy, Greece and UK who specialize in big data analysis for transportation and mobility applications were involved in discussions of our work as it progressed.

The VA researchers started with building an initial prototype system supporting interactive exploration of place connectedness. They demonstrated the opportunities provided by the interactive operations to the domain experts. The expert's reaction was two-fold. On the one hand, they indicated the practical purposes for which the existing functions could be used. On the other hand, observing the existing opportunities provoked them to formulate further questions they deemed meaningful. The VA researchers documented the expert's feedback and transformed it to more general task formulations, as the initial statements were specific to the data examples used for the demonstration. After extending the functions of the software prototype to cover the new tasks thus elicited, the communication with the domain experts was repeated.

The current interest of the transportation analysts is study of the use of air transportation facilities. The partners provided examples of connectedness tasks relevant to their studies, which were also used in defining the task types and developing the software functionality. As a result of the process, the connectedness analysis software was gradually developed and the task typology was defined.

4.2 Examples of analysis scenarios

Here we describe several examples scenarios involving connectedness analysis that were discussed with the domain experts and transportation analysts and used in developing the general task typology.

Example 1. Choosing a place for living. A family planning to move to a big city is looking for a place to live. From several candidate places with suitable living conditions and prices, they want to choose a place that is best connected to areas of business and education activities in the morning, to recreation facilities in the evening, and to green spaces and water bodies on the weekend. For each candidate place, the family wants to see where they can get within a certain time budget, such as 30 minutes. They are also interested to know what transportation modes whey would mostly need to use. Tram and metro are preferred over bus, especially for traveling in busy hours, because their circulation is less affected by the traffic conditions on the streets.

Example 2. Planning a big public event. A city administration is organizing a big public event and wants the residents from the whole city to be able to conveniently get to the event place before the event begins and return home after the event ends. The organizers want to see which places will not be connected by the regular public transport or will require too much time to get. For serving these places, additional bus routes can be introduced. The organizers are also interested in detecting critical interchange nodes where passengers will have to wait long for the next connection, especially those nodes that will be used in many paths. To decrease the cumulative losses of time, the organizers may decide to increase the circulation frequency of some public transport routes.

Example 3. Exploring the reachability of medical facilities. The public health department of the city administration wants to know how well different areas of the city are connected to constantly operating medical facilities (hospitals) at times out of the usual working hours of the primary care physicians. The specifics of this scenario is that there are multiple target places, and connectedness of the other places to the nearest target place needs to be assessed. Apart from detecting places that are poorly connected to any hospital, the analysts want to see what area would be served by each hospital and to spot anomalies, such as areas connected to a spatially more distant hospital rather than to a closer one.

Example 4. Understanding choices between air and ground transportation. The airport administration is interested in increasing the use of the airport facilities; therefore, they wish that more people use air transportation. An analytical service company is asked to investigate whether the connectedness of people's home places to the airport and to the train station affects their choice of the transportation mode for medium-distance traveling. Data indicating the choices were retrieved from mobile phone use records of a set of travelers who used either air or rail transportation for getting to the same city [34]. From the data, the analysts could determine the home locations of the travelers and the transportation modes they used. The analysts want to identify the "catchment areas" of the airport and the train station (i.e., areas that are better connected to one of the two target places) and compare these to the spatial distributions of the homes of the flight and rail users. From the connectedness analysis perspective, the specifics of this scenario is comparison of connectedness to two target places.

Example 5. Examining the impacts of service disruption. The public transport services of a city are planning track reparation works on the metro line that goes to the airport. They need to identify the areas in the city whose connectedness to the airport will be severely impaired and thus understand where replacement transportation needs to be provided. They also want to see what alternative transportation channels would be used when the metro is not operating and to detect critical interchange nodes with long waiting time. The specifics of this scenario is comparison of connectedness under different conditions: with and without the metro.

4.3 Overall analysis problem

Generalizing from these example scenarios, we can state that the general problem in connectedness analysis is to study how well different places over a territory are connected to particular *target* places, such as airport, main station, hospitals, sport facilities, green spaces, etc. Connectedness of a place to a target can be characterized in terms of the following indicators:

- whether the place is connected to the target;
- travel duration from/to the place to/from the target;
- number of changes (characterize travel complexity);
- time spent on waiting for arrivals of transportation means;
- walk distance, when some parts of the trip are made by walking.

The first step in connectedness analysis is generation of optimal (e.g., fastest) paths through the given transport network between the target places and the other places involved in the analysis. Such optimal paths can be computed using any of the available web services (e.g. [22], [23]. We don't address this step in our work. From these paths, connectedness indicators for places and characteristics of the use of the nodes and links of the network are derived.

4.4 The system of tasks

To describe the space of analysis tasks in a systematic way, we define tasks based on the data structure as in [35], [36]. The essence of the approach is consideration of the following facets:

- Analysis focus: the data component that is addressed by a task.
- Analysis level: synoptic or elementary (detailed). The synoptic level refers to sets of data items considered as wholes, while the detailed level refers to individual data items.
- Analysis mode: characterization or comparison. Characterization means determining characteristics of one thing (a set of items or an individual item) while comparison means determining similarities and differences between two or more things.

Analysis is usually a process in which multiple tasks are performed, and the focus, level, and mode repeatedly change. Thus, each of our example scenarios involves several analysis tasks.

4.4.1 Data structure and analysis foci

Data that are used in connectedness studies include two major components: spatial and temporal. The spatial component, in turn, involves two kinds of spatial entities: *places* and *links* between places, i.e., transportation channels [37]. We use the term "links" to refer to both immediate connections between neighboring places and more complex paths between more distant places. The paths may include interchange nodes where passengers change from one public transport route to another. Accordingly, the *spatial focus* in connectedness analysis may be on the places or on the links:

- *Place-centered tasks* are concerned with the connectedness indicators for places.
- *Link-centered tasks* are concerned with the use of transportation channels, modes, and routes and interchanges between these.

All our example scenarios involve place-centered tasks, and examples 1, 2, and 5 involve also link-centered tasks.

For each place, there are two kinds of links: out-links, which enable getting from this place to other places, and in-links, which enable getting to this place from other places. Hence, connectedness of a place has two aspects: out-connectedness and in-connectedness. As it is too hard for human comprehension to consider these two aspects simultaneously, the analysis usually focuses on one of them at a time. The decomposition into out-connectedness and in-connectedness applies to both place-centered and link-centered tasks.

The *temporal focus* means characterizing different time periods in terms of connectedness between places. This can be done using various summary indicators, such as the fraction of unconnected places and the statistics of the travel time and waiting time.

4.4.2 Analysis level

In describing analysis tasks dealing with spatio-temporal data, the distinction according to the analysis level is applied independently to the spatial and temporal components of the data. This means that a task may be synoptic with regard to one component and elementary with regard to the other component [36], [38].

With respect to time, elementary tasks refer to selected time moments of trip starts or ends, whereas synoptic tasks refer to time intervals and are concerned with the variation of the connectedness over the intervals. *With respect to space*, synoptic tasks refer to sets and subsets of places or links, whereas detailed tasks refer to individual places or links. Place-centered synoptic tasks consider the distribution of connectedness indicators over a territory. Link-centered synoptic tasks involve observation of the major (most actively used) transportation channels and modes and determination of their criticality (in particular, the possible consequences of these being not operational). Other tasks of this group are concerned with the interchanges between transportation routes and modes: how actively they are used and how much time is lost for waiting.

Synoptic tasks with respect to space can be subdivided into overall and intermediate level tasks. Overall level tasks refer to the whole set of entities (i.e., places or links), and intermediate level tasks refer to subsets. Intermediate level tasks may be performed with the aim to refine the understanding (mental model) of the connectedness over the territory. For this purpose, subsets of places may be selected based on the spatial locations (e.g., places in particular districts in a city) or connectedness characteristics (e.g., places with long travel time or multiple changes). Subsets of links may be selected based on the transportation modes, or involvement of particular public transport routes, or the number of paths using these links, or the characteristics of the paths using these links (e.g., travel duration, distance, number of changes, waiting time, etc.).

Elementary tasks with respect to space may be performed for detailed consideration of places with poor connectedness (unconnected or requiring too much travel time to get) and links with outlying characteristics, such as high criticality (being used in very many paths), long waiting time in interchange nodes, extreme lengths, or unexpected movement directions. Such places and links are detected while performing synoptic tasks. Detailed analysis may involve examination of relevant portions of the original data describing the public transport trajectories.

4.4.3 Analysis mode

Characterization tasks refer to a specific setup, which includes the following elements:

- the target place or set of places;
- the travel direction: to/from the target(s);
- the time of departure or arrival;
- optionally, conditions and constraints, such as exclusion of some public transport routes or transportation modes.

Comparison tasks refer to two or more distinct setups. While all combinations of setups are theoretically conceivable, practically meaningful are comparisons between setups differing in only one of the four elements: place, direction, time, or conditions.

Our example scenario 4 involves comparison of connectedness to different target places, and example 5 involves comparison of connectedness under different conditions. At the first glance, example 1 also involves comparison between several target places; however, it is done at a rather high level of abstraction. The family does not need to know exactly which place in the city is better connected to which target place, but they need to derive an overall judgment of the goodness of each target place and compare the resulting judgments rather than connectedness maps. Hence, in terms of our typology, the family performs characterization tasks.

4.5 Requirements to analysis support

The process of connectedness analysis consists of repeated fulfillment of the following steps: (1) specify the setup, i.e., the target place(s), travel direction, and departure or arrival time, as well as constraints, such as the maximal waiting time for the next connection and the maximal time or distance of walking that can be involved in a path; (2) compute the paths between the target place(s) and the other places and derive relevant data for places and links, including the connectedness indicators for the places (see section 4.3), the transportation means and intensities of the link use (i.e., the counts of the paths going through each link), and summary statistics of the waiting times at the interchange nodes; (3) analyze the data obtained in step 2. Steps 1 and 3 are performed by a human analyst, whereas step 2 is done automatically.

In step 1, the analyst needs interactive controls for making the necessary settings. The analyst should also be able to choose a subset of places for which the connectedness to the target place(s) will be studied, when there is no need in considering all available places.

In step 3, various analysis tasks are performed; hence, the whole variety of tasks defined in section 4 need to be supported by visual displays and interaction tools. Geographic map displays play the key role since the studied phenomenon is intrinsically geographic. Due to the complexity of the information that is analyzed, there is no way to visualize the data so that both synoptic and detailed tasks could be effectively performed. Synoptic tasks with respect to space need to be supported by visualizations enabling overall view of the spatial distribution of data over the territory or sub-territory under study. Simultaneous representation of multiple attributes or of place- and link-related data in the same display may be counter-productive for analysis since the display may be too overloaded and cluttered and therefore hard to comprehend. Hence, the analyst should be able to select what information to see next as well as to open several displays showing different kinds of information.

Besides the spatial distribution, the analyst should be able to see the statistical distribution of the derived attribute values for the places and links. This can be supported by frequency histograms or other statistical graphics.

Detailed tasks require visualization of selected individual paths, and they may also require representation of detailed data from the original transportation schedules used for path generation. To enable detailed tasks, there must be interactive facilities for selection of places or paths to be examined and for extraction of relevant portions of data from the schedules. Furthermore, within synoptic displays, places and links that may require detailed examination should attract the attention of the analyst, such as places with very long travel times or interchanges with very long wait times.

Intermediate level tasks need to be supported by tools for selection of data subsets, in particular, subsets of paths. In response to a selection, corresponding derived data must be presented; for that, some data need to be dynamically re-computed, e.g., the link use intensities and the waiting time statistics.

Synoptic tasks with respect to time require compact representation of the connectedness variation over a time interval. Statistical summaries characterizing the variation of attribute values for places and links (minimum, maximum, mean, variance, quartiles, mode, etc.) can be derived from the data referring to different time moments within the interval. For places, an important characteristic is the number of connectedness gaps, i.e., times when the place is not connected to the target place(s) under the given constraints. The statistical summaries from time intervals can be visualized on maps in ways similar to the visualization of attributes referring to time moments. The analyst should be able to see details on demand, i.e., the whole time series for selected places or links.

5 Tools for Connectedness Analysis

Here we present a possible set of computational, visual, and interactive techniques that can support the tasks and scenarios identified in the previous section. This should not be considered as a description of a concrete system. We attempt to present the techniques in a generic way, by describing their capabilities rather than specific implementation. The illustrations should thus be treated merely as examples of possible realization of the required capabilities.

5.1 Source data and path generation

The primary data for connectedness analysis consist of the following datasets: 1) geographic positions of public transport stops; 2) trajectories of public transport means, which include the identifiers of the visited stops and the times when the stops are visited; 3) a set of places whose connectedness needs to be studied. These data are used for generation of possible paths between places, and it is the set of the paths that is used in the further analysis. As noted earlier, paths can be generated using existing algorithms, software tools, or web-based services.

To specify the setup for path generation (section 4.5), the analyst uses interactive controls for selecting the target place(s) and setting the movement direction (i.e., whether the target places are trip destinations or origins), the time of departure or arrival, and the constraints, such as the maximal waiting and walking time. The selection of target places can be supported in two ways. First, the analyst can select places by direct interaction on a map display, e.g., by clicking or dragging. Second, places can also be selected using tools for interactive filtering, e.g., based on attribute values.

5.2 Exploration of place attributes

Place attributes include the connectedness indicators and statistical summaries derived from time series of indicators. The attributes are numeric; the values can be represented on a map using techniques suitable for numeric attributes. One example is shown in Fig. 1, center: the travel time from a candidate place for living (scenario 1 in section

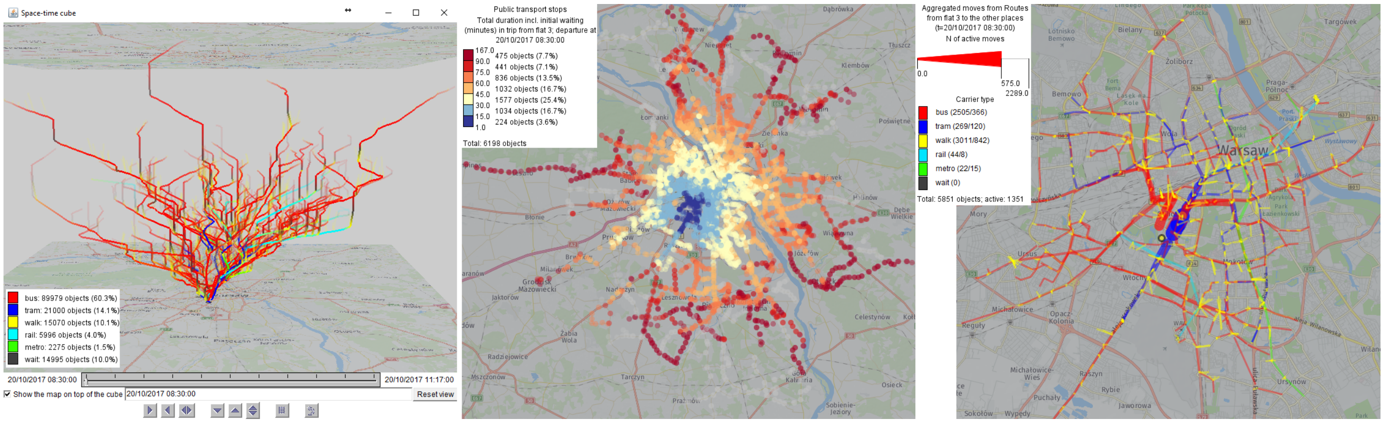

Fig. 1. Left: Connectedness analysis begins with computing optimal paths by public transport. A set of paths originating from a single place is shown in a space-time cube; the vertical axis directed upwards represents time. The colors of the path segments correspond to transportation modes; black vertical lines signify waiting at interchange nodes. Middle: The coloring of the dots on the map shows the travel duration to each place from the chosen origin. Right: A subset of paths with the travel duration under 30 minutes is shown in an aggregated way, with the line widths encoding the counts of the paths going between the places, and the line colors represent the transportation modes.

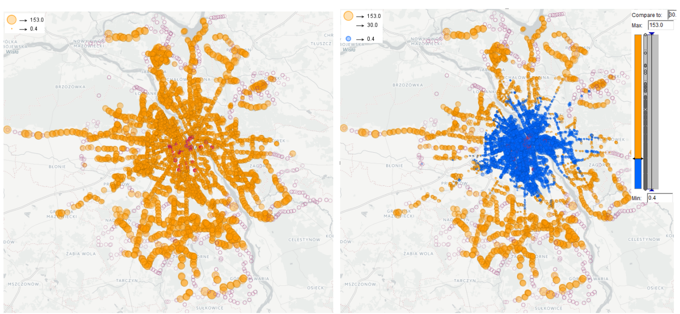

Fig. 2. Left: An example of visualizing a connectedness indicator (travel time to a hospital) by proportional sizes of circle symbols. Right: Application of interactive operation “visual comparison”: two color hues are used to represent values above and below an interactively selected reference value (30 minutes).

4.2) is represented by dot colors. Another possibility is demonstrated in Fig. 2, left: the travel time to the nearest hospital (scenario 3) is represented by proportional sizes of circle symbols. With this representation, the analyst can employ the interactive operation of visual comparison [39], in which the analyst selects a reference value, and two color hues are used to represent the values below and above this value. In Fig. 2, right, blue circles mark the places from which the nearest hospital can be reached within 30 minutes, and orange circles correspond to larger travel durations. The circle sizes are proportional to the absolute differences from the reference value. The reference can be set using a slider or a text edit field, or by clicking on a place in the map (the corresponding value becomes the reference). The dot plot on the right allows the analyst to spot extreme values and see how far they are from the others, The dot plot is linked to the map: when the mouse cursor is pointing on a dot, the corresponding place is highlighted in the map, and vice versa. Another possible representation with similar capabilities could be a frequency histogram.

Travel durations or other numeric information can also be represented, as in Fig. 10, by creating a raster of average values, which involves spatial aggregation and smoothing. A raster is a fine grid with cells containing numeric values, which are visually represented by color and/or shade variation. In computing a raster from values associated with points (like stops), each point contributes to the cell containing it and also to the neighboring cells within a chosen radius (e.g., 500m). A weighting function (kernel) is applied that decreases the contribution as the distance to the cell increases.

Representation of several connectedness indicators in the same map can obstruct perceiving the overall spatial distribution, and the display is more cluttered than when a single attribute is shown. Still, it may be useful to look at certain indicators in combination. For example, the analyst may wish to see the proportions of the waiting and walking time in the total travel durations. A suitable representation that can simultaneously show the total durations and the proportions is a map with pie charts. Thus, the map in Fig. 3 exhibits groups of places with long waiting or walking time. Many such groups are aligned along major roads or railways. In the city center, there are many places from which hospitals can be reached by walking only. Other kinds of map-located diagrams can also be used for looking at several attributes together.

Depending on the analysis scenario, places may also receive qualitative attributes. Thus, in scenario 3, paths from different places to the hospitals reachable in the shortest time are generated (Fig. 4, left). Each path connects some origin place to one of the available hospitals; the identifier of the destination hospital is attached to the origin place. This information can be shown on a map using a representation technique suitable for qualitative attribute values, for example, color coding, as shown in Fig. 4, right. Particular hospitals can be interactively selected for seeing the spatial distribution of the places connected to these hospitals.

5.3 Exploration of temporal variation of place connectedness

For studying how place connectedness varies over time (i.e., for synoptic tasks with respect to time), the path generation process is performed multiple times. For this purpose, the analyst sets the time interval from which the departure or arrival times will be taken and the length of the time step between the taken times. For example, in

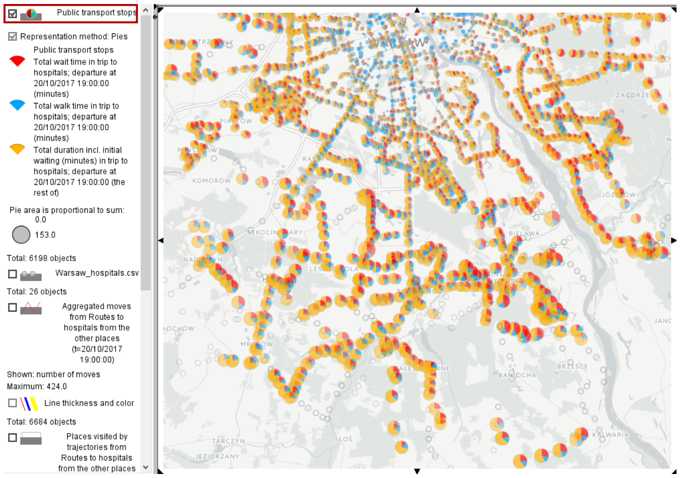

Fig. 3. An example of representing several connectedness indicators by diagrams. The overall sizes show the total travel time; the red and blue segments show the proportions of the waiting and walking time, respectively.

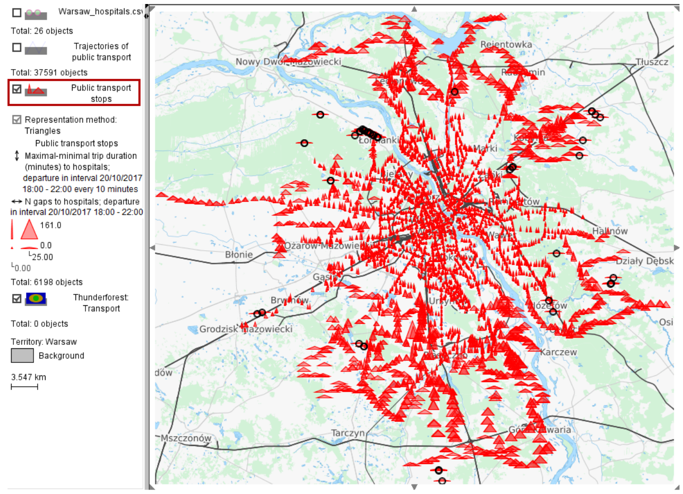

Fig. 5. The triangular symbols represent the differences between the maximal and minimal travel durations by the heights and the numbers of connectedness gaps by the widths. The black circles mark the places that are fully unconnected.

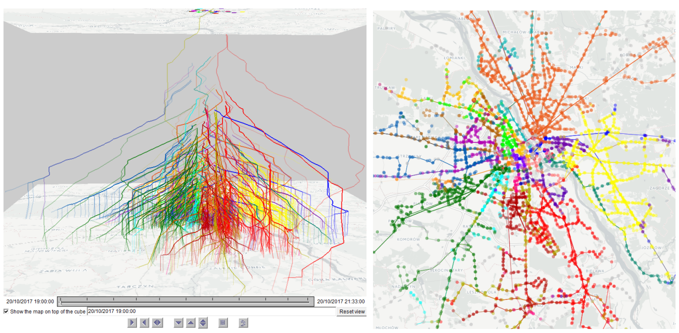

Fig. 4. Left: The space-time cube shows paths to hospitals. Right: The coloring of dot symbols on the map represents the destination hospitals for the places.

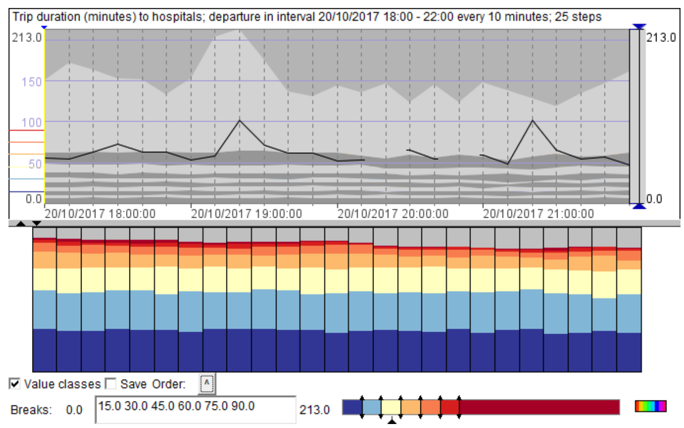

Fig. 6. The overall variation of the travel durations over a time interval is represented by a decile graph (top) and a temporal histogram (bottom). One selected time series is shown in the upper graph by a black line.

studying place connectedness to hospitals (scenario 3), the analyst may wish to see how the connectedness varies for different departure times in the interval 18:00-22:00 with a step of 10 minutes. From the results of all runs, time series of connectedness indicators are generated and summary statistics are computed, including the mean, minimal, and maximal travel durations from or to each place and the number of connectedness gaps, i.e., times when there were no paths from or to a place. The entire time series cannot be represented in a map in a comprehensible way, but the analyst may look at the spatial distribution of the summary indicators. Apart from the visualization techniques discussed in section 5.2, the visualization demonstrated in Fig. 5 may be useful. The triangular symbols represent the differences between the maximal and minimal travel durations to hospitals by the heights and the numbers of connectedness gaps by the widths. Tall symbols indicate the potential for travelers to save time by choosing an appropriate moment for departure. Availability of routing services can significantly help people who need to travel from places with large variation among the travel durations. The same applies to places that have connectedness gaps but aren't fully unconnected.

The places that are fully unconnected or have very limited connectedness can be easily located on the map using interactive selection (brushing), which can be done in a statistical display like a histogram of gap counts. Thus, in Fig. 5, black circles mark the places that are not connected to hospitals, i.e., where the number of connectedness gaps equals the length of the time series.

In time-focused tasks, the analyst disregards the spatial distribution of the connectedness indicators and looks either at the overall characteristics of the temporal variation for all places (synoptic level with respect to space) or at the variation for selected places (elementary level with respect to space). Such tasks are supported by temporal displays. An example is presented in Fig. 6. The display consists of two sections, a decile graph (top) and a temporal histogram (bottom), with a shared horizontal axis representing time. The vertical axis in the decile graph corresponds to the value range of the attribute being visualized (travel duration in this example); the vertical dimension in the histogram represents the number of places. In the upper graph, the lighter and darker stripes represent the value intervals between the deciles (each decile includes 10% of the available values). The graph in Fig. 6 shows that from 80% of the places hospitals can be reached in less than 50 minutes during the whole interval. The values in the 9th decile range from 42 to 64 minutes, whereas the range of the upper decile is very wide, reaching the maximal trip duration 213 minutes. The analyst may select particular places for viewing their individual time series, which can be represented by lines on top of the decile graph.

In the temporal histogram (Fig. 6, bottom), the bars corresponding to the steps of the time series are divided into colored segments based on a division of the attribute value range into intervals. The colors from dark blue to dark red correspond to the trip durations below 15 minutes, 15-30, 30-45, 45-60, 60-75, 75-90 minutes, and over 90 minutes. The heights of the segments are proportional to the counts of the trips with the respective durations. The gray segments represent the counts of the connectedness gaps, i.e., the places from which no hospital can be reached. We see that the number of gaps increases over time, especially significantly in the middle of the time interval (at time steps 20:10 and 20:20), whereas there is no temporal trend for any of the trip duration intervals.

5.4 Exploration of link attributes

5.4.1 *Dynamic updating of path-derived data*

To perform intermediate level tasks, the analyst may need to select a subset of paths and see the place- and link-related data corresponding to these paths. For example, these may be paths going to/from specific places, or paths with specific characteristics in terms of duration, length, waiting time, or walking distance, or paths going through specific transportation channels, or paths using a specific public transport route. When a subset of paths is selected, the analyst must be able to see only the relevant places and links, which is achieved by automatic or user-controlled filtering of the sets of places and links based on the path selection. Moreover, the attribute values of the places and links must be consistent with the selection. While the connectedness indicators of the places remain the same unless another set of paths to/from these places is computed, the link-related attributes, which characterize the paths going through the links, need to be updated according to the path selection, i.e., new values of the attributes must be derived from the subset of the paths. For convenience of the analyst, the updates are done automatically in response to each operation of path subset selection by means of interactive query tools, which enable place-based, link-based, and attribute-based selection.

5.4.2 *Exploring attributes of transportation channels*

Data characterizing transportation channels can be shown on a map using the flow map technique [40], in which aggregate movements are represented by flow symbols (curved or straight lines, with or without arrows at the ends) whose widths are proportional to the movement volumes. In Fig. 1, right, a flow map represents aggregated paths from one candidate home place (scenario 1). Each flow symbol corresponds to one *immediate link*, i.e., a link connecting two places, where the second place is directly (i.e., without intermediate stops) reached from the first one. The width of a flow symbol is proportional to the number of paths that use the corresponding immediate link. Colors of flow symbols can represent the transportation modes that are used on the links. In scenario 1, the family choosing a place for living evaluates the candidate places, among other criteria, in terms of the diversity of the transportation modes available. In the given example, they see that the candidate place is connected to others by two transportation modes, tram (blue) and bus (red), which is good. They also see that metro (green) and rail (cyan) services are reachable in relatively short time. The map in Fig. 1, right, represents the link-based aggregates derived from the subset of the paths with the durations below 30 minutes; this is a result of dynamic update of link attributes.

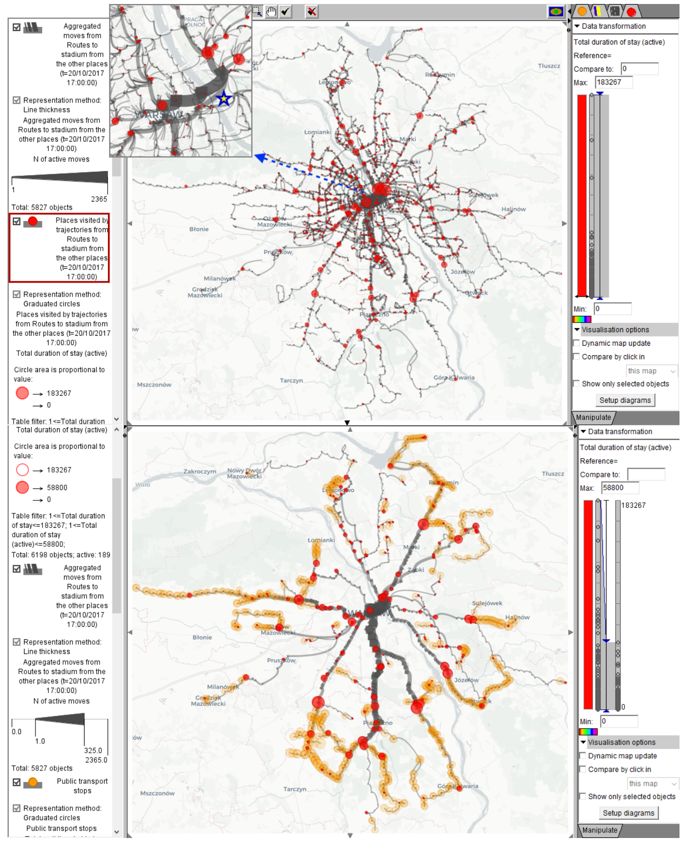

Fig. 7. The total waiting time spent at the interchanges is encoded by proportional sizes of red circle symbols. Top: The map shows data derived from the entire set of paths to a public event site. Bottom: The map shows data derived from the subset of paths where the waiting time exceeds 20 minutes. The orange circles mark the origin places of the selected paths; the circle sizes are proportional to the waiting time in the respective paths.

5.4.3 *Exploring attributes of interchange nodes*

The attributes of interchange nodes include the counts of the paths that go through the nodes and statistics of the waiting time, such as the minimum, maximum, mean, and total (i.e., the sum of the waiting time in all paths). A possible way of visualizing the waiting time is by proportionally sized symbols, as in Fig. 7, where they are drawn on top of a flow map showing the link use intensity . This example corresponds to scenario 2 (section 4.2). The public event organizers explore the paths to the event site by which people could arrive at the site before the planned time of event opening. The upper map corresponds to the whole set of paths from all places. The inset in the upper left corner shows an enlarged map fragment, with the star indicating the event site.

The lower part of Fig. 7 demonstrates an effect of dynamic updating of link data in response to a selection of a subset of paths. The organizers want to see where the time is lost in the paths involving long waiting. They have selected the paths where the waiting time exceeds 20 minutes. The map has been automatically updated. It shows only the links involved in the selected paths. The widths of the flow symbols represent the intensity of the link use in the selected paths, and the circle sizes represent the total waiting time in the selected paths. Please note that the maximal widths of the flow lines and the maximal sizes of the circles have been adjusted to the decreased ranges of the attribute values using interactive controls. The images in Fig. 7 show (on the right of the map) the controls associated with the circles, and similar controls for the flow symbols are available in another tab.

Additionally to the link data, the lower map in Fig. 7 shows also the places the selected paths originate from, with the corresponding waiting time represented by proportional sizes of orange circles.

5.5 Comparative analysis

In comparison tasks (section 4.4.3), the analyst examines differences regarding the following aspects: path existence, place connectedness indicators, and link use characteristics. Visual comparison of the first two aspects is illustrated in Fig. 8. The example refers to scenario 4 (section 4.2). The analyst compares the connectedness of places in Madrid to the airport, which is located on the northeast of the city, and to the main train station, which is located in the center. The comparison is done based on a set of paths arriving at the airport by 8:00 and a set of paths arriving at the train station by 8:30 (assuming that the pre-boarding procedures take less time in the train station than in the airport). The left image demonstrates the comparison of path existence. The places are represented by dots with colors showing whether they are connected to both the airport and the train station (green), to one of them (purple and blue), or to none of them (red). On the right, the differences in the travel durations to the train station and to the airport are encoded by color hues (blue for negative and orange for positive) and circle sizes, which are proportional to the absolute values of the differences. According to Gleicher et al. [41], this method of supporting visual comparison is classified as explicit encoding.

Comparison of link characteristics is not easy to do within a single map, where the flow symbols would overlap and intersect. The main supporting technique is juxtaposition [41] of two or more maps, as in Fig. 9 demonstrating comparison of connections to the airport (left) and train station (right). We see that there are two major channels used for getting to the airport, a metro line (red) and an express bus (blue; the long straight line means the absence of intermediate stops). The main channel for getting to the train station is the metro, where many paths converge.

5.6 "What if" analysis of the impact of traffic disruption

When exploring link-related data, the analyst can not only see the transportation modes used on the links but also access (by map-based direct interaction) detailed data, such as the public transport routes (lines) serving the links. In the

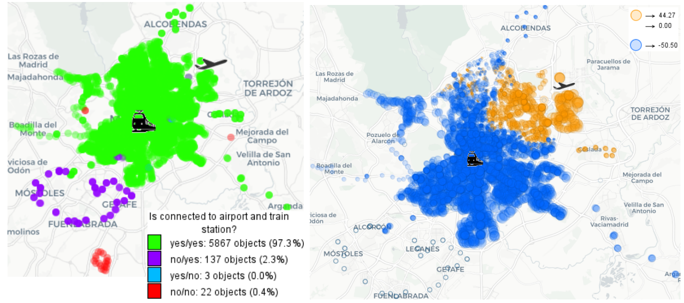

Fig. 8. Comparison of place connectedness to the airport and train station in Madrid in the morning. Left: The dot colors indicate to which targets the places are connected. Right: The colors and sizes of the circles show the differences in the travel time.

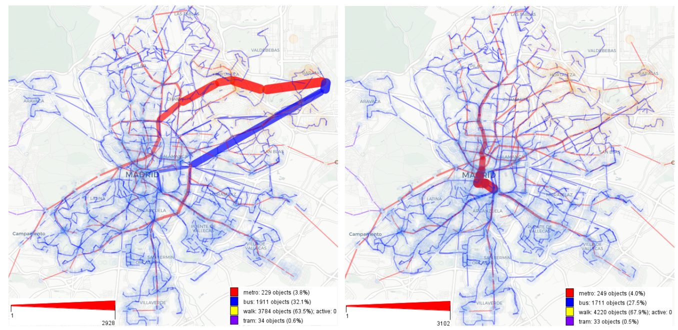

Fig. 9. Comparison of the link use and transportation modes in traveling to the airport (left) and train station (right).

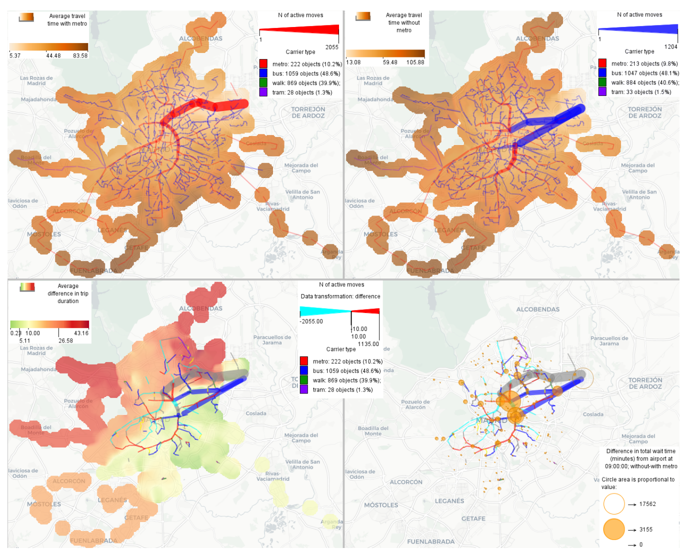

Fig. 10. Top: Link use in the trips departing from the airport at 09:15 in cases 1 (left) and 2 (right), i.e., with and without the metro line 8. The background shading represents the average travel time. Bottom left: The background shading shows increased average travel durations in case 2 compared to case 1; green shades represent increases by less than 10 minutes. Bottom left and right: The line widths represent increases and decreases of link use. The lines representing increases are colored according to the transportation modes used in case 2. Decreases are shown with the opposite colors: cyan, yellow, and green instead of red (metro), blue (bus), and magenta (tram). Links that disappear in case 2 are painted in gray. Bottom right: The orange circles show the increases of the total wait time at the interchange nodes.

connectedness to the airport (Fig. 9), a great role is played by the metro line 8, which takes more than a half of the paths from all places to the airport. The chain of thick red flow symbols corresponds to this metro line. Similar and sometimes even more prominent patterns exist also for other times of arrival to the airport, as well as for traveling from the airport. Thus, in Fig. 10, top left, which corresponds to the departure from the airport at 9:15, the prevalence of the use of metro (red) over the use of bus (blue) is even higher than in Fig. 9.

As maintenance works need to be done on the metro line 8, the analyst wants to see how the exclusion of this line will affect the connectedness to the airport throughout the city (scenario 5 in section 4.2). The trajectories of the metro line 8 are excluded by interactive filtering. The paths to and from the airport are built using the restricted set of the public transport trajectories. The analyst compares the aggregated results for the cases 1 and 2, with and without the use of the metro line 8, respectively. The comparison is illustrated in Fig. 10. The maps on top show the link uses in cases 1 (left) and 2 (right); the walked links are hidden for reducing the display clutter. The background shading represents the average travel time. Please note that the color scale is applied to different value ranges: 5.37-83.58 minutes on the left and 13.08-105.88 on the right.

In the juxtaposed maps in Fig. 10, top, the differences in the use of the metro line 8 and bus routes are very prominent while other differences can be hardly seen. Also, the differences in the travel time are not very clear. In such cases, comparison can be supported by explicit encoding of the differences, as in Fig. 10, bottom. On the left and right, the link use values for case 1 have been subtracted from the values for case 2. The line widths show how much the link use will increase or decrease. The increase is shown by the color corresponding to the transportation mode in case 2, and the decrease is shown by the opposite color; cyan, yellow, and green are opposite to the red (metro), blue (bus), and magenta (tram), respectively. Links that disappear in case 2 are shown in black. This difference map reveals an increased use of the metro lines connected to the airport bus routes 200 and 203.

On the bottom left, the technique of explicit encoding is also used for showing the differences in the travel time. Here, the travel time for case 1 are subtracted from the travel time in case 2. Areas where the travel time will highly increase have become prominent on the map. Explicit encoding can also reveal the interchange nodes where the waiting time will significantly increase. In Fig. 10, bottom right, the increases of the total waiting time are represented by proportional sizes of orange-colored circles. The most significant increase of the waiting time will happen at the airport, where all passengers will have to wait for the buses 200 and 203. The prominence of the corresponding symbol has been reduced by interactive selection of a smaller value range. This operation reveals several interchange nodes with long waiting time for the next connections after using the bus routes 200 and 203.

Based on this comparison, the analyst may understand that the increases of the travel and waiting time can be much reduced first of all by providing additional travel opportunities in the northern part of the city, which can not only reduce the travel time from this area but also the cumulative waiting time at the interchange nodes, since people traveling from the north will not have to go there for taking one of the regular airport buses.

6 Discussion and conclusion

The primary goals of our research have been to develop a conceptual framework for place connectedness studies and to define a general system of relevant analysis tasks. A following goal was to find, investigate, and generalize the ways in which visual analytics can support these tasks taking advantage of synergistic combinations of visual, interactive, and computational techniques. Consideration of several example scenarios of problem solving helped us to distil the key concepts and to see the variety of possible tasks, which provided a basis for our following work on task generalization and systematization. However, our work was not oriented to a specific application and did not aim at developing software tools for solving particular problems. The result of our work is a general framework for using visual analytics techniques and workflows in place connectedness studies.

The framework development and evaluation were conducted in continuous close contact with transportation domain experts and other partners whose practical activities or research interests were related to transportation and human mobility. The motivating scenarios, the system of tasks, the supporting techniques and workflows of their use were iteratively discussed with the partners. The feedback received helped us to elaborate the scenarios, refine the system of tasks, and enhance the visual analytics workflows and their components. Importantly, the development of the task typology motivated the partners to think about further analytical questions and potential analysis scenarios. The demonstration of the analytical techniques and workflows convinced the partners of the power and utility of visual analytics approaches and motivated them to think of various possible ways of utilizing this power in practical applications.

As it was discussed with the partners, there exist many categories of potential users with different information needs, skills, and time constraints, and different use situations characterized by varying hardware (CPU, memory, screen size and resolution), interaction modalities, and bandwidth capabilities. Respectively, development of a single one-size-fits-all application is not feasible. The partners imagined a range of innovative applications that could be based on the approaches we had developed. Thus, ticket terminals at stops may display actual accessibility maps for fixed-time tickets. Web-based routing tools may suggest optimal times for starting planned trips or suggest an optimal destination of a given type, e.g., an easy-to-access health care practitioner with a certain specialization. Business may optimize their working schedules for making travelling more comfortable for their employees or customers.

Given the goals and character of our research, it would be irrelevant to compare the proposed techniques and workflows against any specific solution. However, we can make a general conclusion that visual analytics approaches can be used to enhance the tools utilized currently by supporting iterative and incremental analysis processes, enabling not only observation of overall connectedness patterns but also active exploration of these patterns, and orienting further steps of problem solving. More specifically, we can highlight the following advantages of visual analytics approaches with regard to different categories of tasks.

The studies of place connectedness (accessibility) reported in the literature involve mainly place-centered tasks (section 4.4.1), which are traditionally fulfilled with the use of maps representing place connectedness indicators. We do not propose any radically new means for supporting place-centered tasks. Substantial differences between the use of a GIS and a visual analytics system can lie in the ease of, first, obtaining the necessary maps and graphics and, second, accessing detailed information and performing comparisons by interacting with the maps and graphics.

Link-centered tasks (section 4.4.1) are not common in the literature on accessibility analysis. We defined the classes of link-centered tasks in discussions with transportation domain experts. To support the analysis of link use, we propose flow maps, which can show quantitative and qualitative information, such as the path counts and the transportation modes for the network links. Whilst the flow map technique is a well known way of representing aggregated movements, it has not been widely used in connectedness studies. Characteristics of interchange nodes can be shown on top of a flow map by symbols or diagrams.

We emphasize the importance of intermediate level tasks (sections 4.4.2, 4.5), in which the analyst focuses on subsets of places, links, or paths. Such tasks are supported by interactive filtering and dynamic re-aggregation functions. Comparison tasks are supported by dynamic computation and explicit visual encoding of differences on maps and other displays. Such facilities are common for visual analytics systems but are not available in the software currently used for connectedness analyses.

However, the use of visual analytics approaches does not only bring benefits but also involves certain costs. These approaches can be perceived as complex by general users and, indeed, they require training, but the same applies to any methods for non-trivial analysis that require human reasoning rather than allow purely automatic application. The complexity of the visual analytics approaches is a natural consequence of the possibility to perform deeper and more comprehensive analyses involving multiple types of data and multiple representations supporting different tasks and perspectives. It is typical for visual analytics that displays are information-rich and may include several information layers. To avoid visual clutter and information overload, the displays need to be carefully designed. Particularly, the display content and level of detail need to be easily adaptable to different analysis subtasks and operations performed along the analysis process. Hence, to be practically utilizable, visual analytics techniques need to be well-designed for specific applications and appropriately taught to target users.

Our research contribution consists of (1) comprehensive and systematic analysis of a problem domain where visual analytics approaches can provide new opportunities to researchers and practitioners, (2) definition of required tool capabilities for performing analyses in this domain, and (3) description of a possible set of visual analytics techniques fulfilling these requirements.

The limitation of our work is that it does not consider analysis tasks requiring involvement of additional data apart from the public transport network and timetables; however, some general notes can be made. There are several categories of relevant additional data: resident population characteristics, travel demands (which are usually estimated using surveys), public transport characteristics, such as travel costs and capacities of vehicles, and ridership. The latter two categories can be incorporated in the analysis in a straightforward way by computing and visualizing additional attributes for places (travel costs) and links (aggregate capacities and ridership). The population information can be used in intermediate level tasks, in which the analyst may select places with particular population characteristics, e.g., high number of poor families. Travel demand data often have the form of origin-destination matrices specifying the estimated number of travelers between areas. These data complement the path-derived aggregates by telling how many times each path is expected to be used. On this basis, additional place- and link-related attributes can be computed and analyzed.

A possible extension of this research would be consideration of the variety of possible what-if scenarios in connectedness analysis. A typology of such scenarios could inform and orient further works on visual analytics support to connectedness analysis.

In conclusion, we can state that the problem domain of connectedness (accessibility) studies can greatly benefit from using visual analytics techniques and tools. In communication with domain experts, we have elicited the requirements to support of the tasks and outlined a prototypical set of tools addressing these requirements. We have also identified potentially interesting directions for further research.

ACKNOWLEDGMENTS

This work was partially funded by Fraunhofer Cluster of Excellence on "Cognitive Internet Technologies", by EU in projects Track&Know (grant agreement 780754) and SoBigData Research Infrastructure (grant agreement 654024), and by German Science Foundation DFG in priority research program SPP 1894 on Volunteered Geographic Information.

REFERENCES

[1] T. Litman, *Evaluating Accessibility for Transportation Planning: Measuring People's Ability to Reach Desired Goods and Activities*. Victoria Transport Policy Institute, 2012.

[2] K. Fransen, T. Neutens, S. Farber, P. De Maeyer, G. Deruyter, and F. Witlox, "Identifying public transport gaps using time-dependent accessibility levels," *Journal of Transport Geography*, vol. 48, pp. 176–187, 2015.

[3] "GTFS specification," https://developers.google.com/transit/gtfs/reference/, accessed: 23.02.2018.

[4] L. Wang, Y. Liu, Y. Liu, C. Sun, and Q. Huang, "Use of isochrone maps to assess the impact of high-speed rail network development on journey times: a case study of Nanjing city, Jiangsu province, China," *Journal of Maps*, vol. 12, no. sup1, pp. 514–519, 2016.

[5] S. Farber, M. Z. Morang, and M. J. Widener, "Temporal variability in transit-based accessibility to supermarkets," *Applied Geography*, vol. 53, no. Supplement C, pp. 149 – 159, 2014.

[6] S. Farber and L. Fu, "Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis)investments over time," *Computers, environment and urban systems*, vol. 62, pp. 30–40, 2017. [Online]. Available: http://dx.doi.org/10.1016/j.compenvurbsys.2016.10.005
[7] G. Andrienko, N. Andrienko, W. Chen, R. Maciejewski, and Y. Zhao, "Visual analytics of mobility and transportation: State of the art and further research directions," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 8, pp. 2232–2249, Aug 2017.
[8] J. D. Mazimpaka and S. Timpf, "A visual and computational analysis approach for exploring significant locations and time periods along a bus route," in *Proceedings of the 9th ACM SIGSPATIAL International Workshop on Computational Transportation Science*, ser. IWCTS '16. New York, NY, USA: ACM, 2016, pp. 43–48.
[9] C. Tominski, H. Schumann, G. Andrienko, and N. Andrienko, "Stacking-based visualization of trajectory attribute data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 12, pp. 2565–2574, Dec 2012.
[10] M. Itoh, D. Yokoyama, M. Toyoda, Y. Tomita, S. Kawamura, and M. Kitsuregawa, "Visual exploration of changes in passenger flows and tweets on mega-city metro network," *IEEE Transactions on Big Data*, vol. 2, no. 1, pp. 85–99, March 2016.
[11] S. Pensa, E. Masala, M. Arnone, and A. Rosa, "Planning local public transport: A visual support to decision-making," *Procedia - Social and Behavioral Sciences*, vol. 111, no. Supplement C, pp. 596 – 603, 2014, transportation: Can we do more with less resources? 16th Meeting of the Euro Working Group on Transportation. Porto 2013.
[12] C. Palomo, Z. Guo, C. T. Silva, and J. Freire, "Visually exploring transportation schedules," *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 170–179, Jan 2016.
[13] M. Wunderlich, K. Ballweg, G. Fuchs, and T. von Landesberger, "Visualization of delay uncertainty and its impact on train trip planning: A design study," *Computer Graphics Forum*, vol. 36, no. 3, pp. 317–328, 2017.
[14] W. Zeng, C.-W. Fu, S. Mueller Arisona, A. Erath, and H. Qu, "Visualizing waypoints-constrained origin-destination patterns for massive transportation data," *Computer Graphics Forum*, vol. 35, no. 8, pp. 95–107, 2016.
[15] D. Guo, J. Chen, A. M. MacEachren, and K. Liao, "A visualization system for space-time and multivariate patterns (vis-stamp)," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1461–1474, Nov 2006.
[16] J. Wood, J. Dykes, and A. Slingsby, "Visualisation of origins, destinations and flows with OD maps," *The Cartographic Journal*, vol. 47, no. 2, pp. 117–129, 2010.
[17] I. Boyandin, E. Bertini, P. Bak, and D. Lalanne, "Flowstrates: An approach for visual exploration of temporal origin-destination data," *Computer Graphics Forum*, vol. 30, no. 3, pp. 971–980, 2011.
[18] G. Andrienko, N. Andrienko, G. Fuchs, and J. Wood, "Revealing patterns and trends of mass mobility through spatial and temporal abstraction of origin-destination movement data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 23, no. 9, pp. 2120–2136, Sept 2017.
[19] F. Galton, "On the construction of isochronic passage-charts," *Proceedings of the Royal Geographical Society and Monthly Record of Geography*, vol. 3, no. 11, pp. 657–658, 1881. [Online]. Available: http://www.jstor.org/stable/1800138
[20] J. Riedel, *Anregungen für die Konstruktion und die Verwendung von Isochronenkarten*. Universität Leipzig, 1911.
[21] D. O'Sullivan, A. Morrison, and J. Shearer, "Using desktop GIS for the investigation of accessibility by public transport: an isochrone approach," *International Journal of Geographical Information Science*, vol. 14, no. 1, pp. 85–104, 2000.
[22] "OpenRouteService," https://openrouteservice.org/, accessed: 10.04.2019.
[23] "Directions API," https://developers.google.com/maps/documentation/directions/start, accessed: 10.04.2019.
[24] S. R. Hong, Y.-S. Kim, J.-C. Yoon, and C. R. Aragon, "Traffigram: Distortion for clarification via isochronal cartography," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '14. New York, NY, USA: ACM, 2014, pp. 907–916.
[25] D. Cornel, A. Konev, B. Sadransky, Z. Horváth, A. Brambilla, I. Viola, and J. Waser, "Composite flow maps," *Computer Graphics Forum*, vol. 35, no. 3, pp. 461–470, Jun. 2016.
[26] "TravelTimePlatform," https://www.traveltimeplatform.com/, accessed: 10.04.2019.
[27] F. Kamw, S. AL-Dohuki, Y. Zhao, T. Eynon, D. Sheets, J. Yang, X. Ye, and W. Chen, "Urban structure accessibility modeling and visualization for joint spatiotemporal constraints," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–13, 2019.
[28] A. Stewart, *Visualizing urban accessibility metrics for incremental bus rapid transit projects*. Massachusetts Institute of Technology, 2014.
[29] W. Zeng, C. W. Fu, S. M. Arisona, A. Erath, and H. Qu, "Visualizing mobility of public transportation system," *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1833–1842, Dec 2014.
[30] "TranseetFeeds," https://transitfeeds.com/, accessed: 23.02.2018.
[31] G. Andrienko, N. Andrienko, and G. Fuchs, "Understanding movement data quality," *Journal of Location Based Services*, vol. 10, no. 1, pp. 31–46, 2016.
[32] "TranseetFeeds ZTM Warsaw," https://transitfeeds.com/p/ztm-warszawa/720/latest, accessed: 23.02.2018.
[33] "TranseetFeeds EMT Madrid," https://transitfeeds.com/p/emt-madrid/212/latest, accessed: 23.02.2018.
[34] P. García, R. Herranz, J. Javier, G. Andrienko, N. Adler, and C. Ciruelos, "Big data analytics for a passenger-centric air traffic management system. a case study of door-to-door intermodal passenger journey inferred from mobile phone data," in *6th SESAR Innovation Days (SIDs)*, 2016.
[35] J. Bertin, *Semiology of Graphics*. University of Wisconsin Press, 1983.
[36] N. Andrienko and G. Andrienko, *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2005.
[37] G. Andrienko, N. Andrienko, P. Bak, D. Keim, and S. Wrobel, *Visual Analytics of Movement*. Secaucus, NJ, USA: Springer-Verlag New York, Inc., 2013.
[38] A. Koussoulakou and M. Kraak, "Spatia-temporal maps and cartographic communication," *The Cartographic Journal*, vol. 29, no. 2, pp. 101–108, 1992.
[39] G. Andrienko and N. Andrienko, "Interactive maps for visual data exploration," *International Journal of Geographical Information Science*, vol. 13, no. 4, pp. 355–374, 1999.
[40] M.-J. Kraak and F. Ormeling, *Cartography: visualization of spatial data*, 2011.
[41] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, "Visual comparison for information visualization," *Information Visualization*, vol. 10, no. 4, pp. 289–309, Oct. 2011.

Gennady Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems in Germany and part-time professor at City University London. He was a paper chair of *IEEE VAST* conference (2015–2016) and an associate editor of *IEEE Transactions on Visualization and Computer Graphics* (2012–2016). Currently, he is an associate editor of two journals, *Information Visualization* and *International Journal of Cartography*, and editorial board member of *Cartography and Geographic Information Science*.

Natalia Andrienko is a lead scientist responsible for visual analytics research at Fraunhofer Institute for Intelligent Analysis and Information Systems in Germany and part-time professor at City University London. Results of her research on spatio-temporal data analysis have been published in two monographs "*Exploratory Analysis of Spatial and Temporal Data: a Systematic Approach*" (Springer 2006) and "*Visual Analytics of Movement*" (Springer 2013). Natalia Andrienko is an associate editor of *IEEE Transactions on Visualization and Computer Graphics* since 2016 and editorial board member of *International Journal of Geographical Information Science*, *Journal of Geographical Systems* and *International Journal of Cartography*.

Fabian Patterson is a data scientist at Fraunhofer Institute IAIS. He is working on software tools for the analysis and visualisation of data streams associated with moving entities in the context of smart cities and maritime security.

Hendrik Stange is a senior data scientist at the Fraunhofer Institute IAIS. He is heading big data analytics projects in automotive, health care, logistics, security, and customer goods. As a practicing project manager, he is responsible for business innovation and method development supporting market leaders and SME in becoming data-driven enterprises.