38. Wissenschaftlich-Technische Jahrestagung der DGPF und PFGK18 Tagung in München – Publikationen der DGPF, Band 27, 2018

Modeling and visualizing the spatial uncertainty of moving transport hubs in urban spaces - a case study in NYC with taxi and boro taxi trip data

ANDREAS KELER[1]

Abstract: Mobility in urban environments is complex due to numerous interacting components, with many of those that are difficult to specify. Examples include the presence of transport hubs, which connect different modes of transport, public and private. The properties of these locations include a temporally changing surface area of operational function that is heavily dependent on the complex and dynamically changing human mobility. Besides public transport services with specified stations, there are taxi services that can be associated with established transport hubs in the way of assigning spatiotemporal service hotspots. This work proposes a technique for relating taxi trip origins and destination hotspots for gaining knowledge on the spatial uncertainties of transport hubs, more precisely their movements within specific times. The case study in NYC focuses on the services of yellow taxi and boro taxi trip data on Saturdays in June 2015. The outcomes of applying the technique are matter of further investigation of spatial uncertainty perception, representation, and visualization. In the stages of the approach, the outcomes of transport hub movements are related with more general functional transport regions resulting from NYC public transport services.

1 Introduction

Urban human mobility relies on the presence of transport hubs, which connect different modes of transport, public and private. This paper proposes a technique for relating taxi trip origins and destination hotspots for gaining knowledge on the spatial uncertainties of transport hubs, more precisely their movements within specific times. Coming from the assumption that different transport modes are usable as a sequence within the day, the taxi services can be seen as one of the key modes, especially in complex urban environments. By assigning functional transport regions, often specific bus routes or metro lines are extracted together with respective stations. This functionality in public transport services is expressible via visualizing accessibilities, as for example by using detailed isochrones. Besides this, the fact of having dynamically changing accessibility surfaces resulting from dynamically changing traffic situations is essential for further analyses.

1.1 Describing the moving transport hub phenomena

Several appearances within the passenger behavior are connectable with the phenomena of the moving transport hub. By definition, movement of urban citizens can be perceived and especially represented differently. One general differentiation consists of grouping movement flows into movements of participants of one mode of transport or into individual movement that is often

[1] Technical University of Munich, Chair of Traffic Engineering and Control, Arcisstraße 21
D-80333 München, E-Mail: andreas.keler@mytum.de

intermodal in urban environments. Keeping this in mind, the movement representation varies with the spatial and temporal scale of analyzing movements in urban environments. On a very general level, one can define inter-city movements, where every city is aggregated by a single symbolization. From this type of representation, it is possible to define functional transport regions that may connect several cities, which comes from early research in transport geography, a discipline originating from human geography. By changing the scale to higher-resolution movement between city districts, it is possible to detect inter-city movements that can also indicate the quality value of selected districts in living and working. At the next higher resolution level, it is feasible to analyze movement between different stations of selected mobility services, especially from the established public transport. Modes using trains such as subways and tramways that usually base on installations of expensive infrastructures can have several mobility stations, such as tramway stations. Some of these stations have connections to nearby situated stations of other transport modes, which can be referred to as transport hubs. These transport hubs are special cases of mobility stations that connect stations of different transport modes within the same space, often with multiple elevation levels. In the possibly most common way in urban environments, the transport hub has numerous transport stations: different spatially fixed stations with respective connections usually situated in walking distance, and, stations that are dynamically changing in space and base on local knowledge or are inferable from vehicle movement observations. The latter can base on the vehicle movements of vehicle services such as taxi services. In this case, it is possible to estimate typical taxi trip origin and destination hotspots, which are connected with the taxi customer behavior of entering and leaving the taxi. These hotspots, especially when generated from taxi movement data acquisitions, are dynamically changing point clusters that occupy selected areas in space. Within these areas, it is possible to have numerous built transport infrastructural elements. Therefore, the appearance of a transport hub is in general more associable with a polygonal then with a pointwise representation in space. These ideas can be explained via an example from the operation of daily bus routes of the public transport in Shanghai.

Fig. 1 shows an example from BURGFELD & LIEBHARDT (2015) of such an appearance, mainly consisting of the comparison between two officially assigned bus routes (a) and not-preprocessed GNSS positioning data from respective buses from February 28, 2007 (b). Both bus routes in Shanghai, 20 and 36, are highly influenced by daily traffic congestion events. This makes it understandable that at certain times of working days, bus movements differ from official routes due to traffic congestion avoidance strategies. More precisely, these traffic congestion avoidance patterns appear often as slight variations of the route, as taking the byroad that is less congested. This is pictured in Fig. 1 by the large variations of spatial distribution of tracked bus movements, especially for bus route 20 (red).

In another way, we can speak of moving transport hubs, when bus stations are shifting in space within specific time windows. The example in Fig. 1b shows that buses of route 20 (in red) are moving around a larger area in the northwestern part of the map view, which is not assigned as an official part of the bus route. Serving specific needs of commuters, together with local knowledge can result in specific movement patterns. These patterns can appear around only one specific station or include two or more stations of different mobility services and modes of transport.

In another way, moving transport hubs are definable as slight variances of the mobility hotspots within the transport hub area.
These connections are in some cases visible, as in the specific example of bicycle parking stations above metro stations underground. The concentrations of bicycles, namely the bicycle densities within the transport hub vary spatially and, in some cases, spatiotemporally, always depending on an enormous selections of different factors influencing the urban mobility in general.

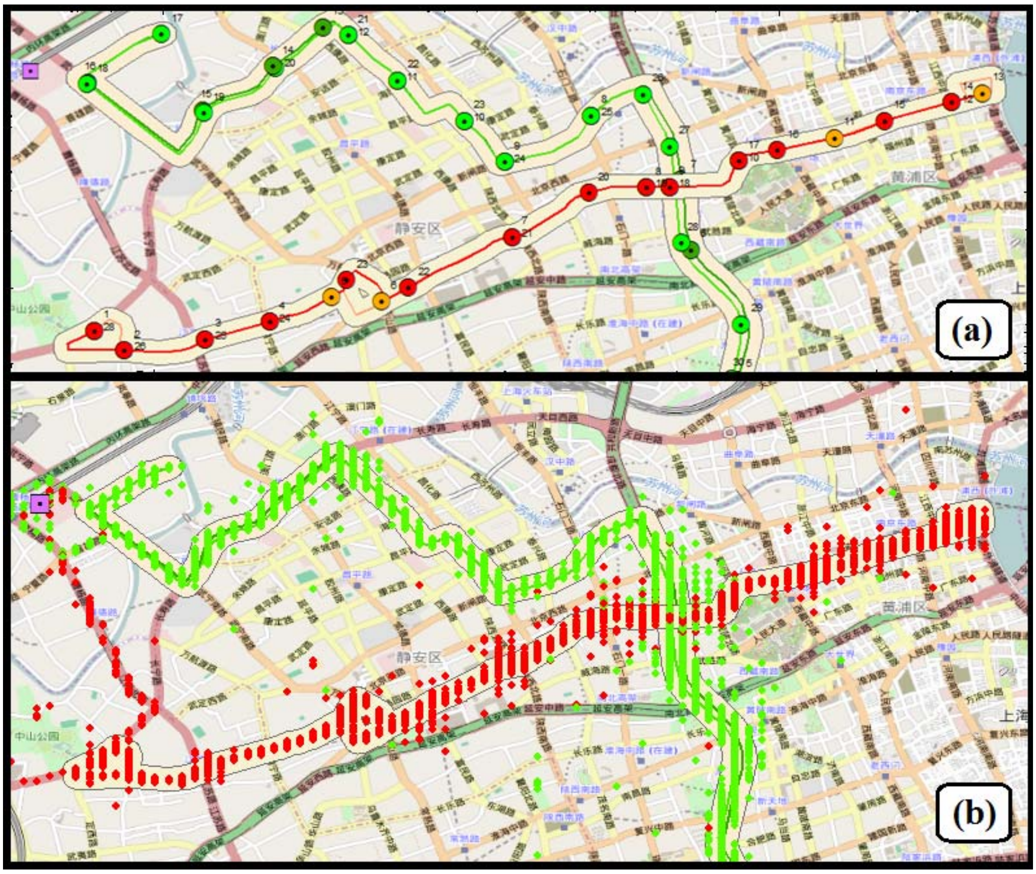

Fig. 1: Example for the presence of moving transport hubs with a comparison of (a) the official bus routes 20 (red) and 36 (green), and, (b) plotted raw positioning data[2] from respective bus movements from February 28, 2007, in Shanghai (BURGFELD & LIEBHARDT 2015)

1.2 Motivation

Urban environments are characterized by complex and dynamically changing human mobility. Vehicle movement trajectories of urban vehicle fleets, as taxi fleets, can help extracting this knowledge. Taxi trip data is usable in many different applications of different research domains.
Due to often-immense sizes of the data extracts, many applications make only use of taxi trip origin and destination points. In an aggregated view, these points may be clustered into origin

[2] 'SUVnet-Trace Data' was obtained from the Wireless and Sensor networks Lab (WnSN) at Shanghai Jiao Tong University, http://wirelesslab.sjtu.edu.cn/taxi_trace_data.html

and destination hotspots, which are connectable, especially, when there are different taxi fleets, with specific operational restrictions. This promotes the automated deduction of local knowledge on smaller scales of investigation. In a simple way, this deduction may consist of spatial queries for the presence of mobility service stations within agglomerations of origins and destinations of different taxis of different taxi fleets.

The question here is if it is possible to define the moving transport hub phenomena as a general collection of different appearances of urban mobility services, as pictured in Fig. 2 by the tree cases. Case 1 of Fig. 2 shows a size variation and movement of the defined taxi service hotspot polygons t1 and t2 around only one railway station. Dependent on the inspected time windows that correspond with t1 and t2, there are different conditions and traffic situations that influence the shapes of the two taxi service hotspot polygons. These polygons are then connected with on specific railway station that is situated within both. For case 1, this is the only condition for defining one transport hub polygon, which is definable via the polygon intersection, where the railway station is situated. Case 2 in Fig. 2 shows the appearance of having two intersecting taxi service hotspot polygons that have an intersection area without the presence of mobility stations. Nevertheless, mobility stations are respectively included in every polygon. One definition of a moving transport hub for case 2 would consist of a spatially larger area of influence (combination of t1 and t2) and one core intersection area. One important aspect of case 2 is the condition of having connections between the mobility stations in walking distance. Additionally, it is possible to reason on the connectivity on the operational level, as for example the specific official public bus route and respective driving directions of operating buses. Case 3 is defined, when these mentioned conditions do not apply. For case 3 in Fig. 2 the mobility stations are within the polygons t1 and t2 but are for example situated on different road segments, which indicates greater walking distances. Case 3 appears often at complex signalized intersections and indicates additional waiting times between specific mobility stations within the moving transport hub.

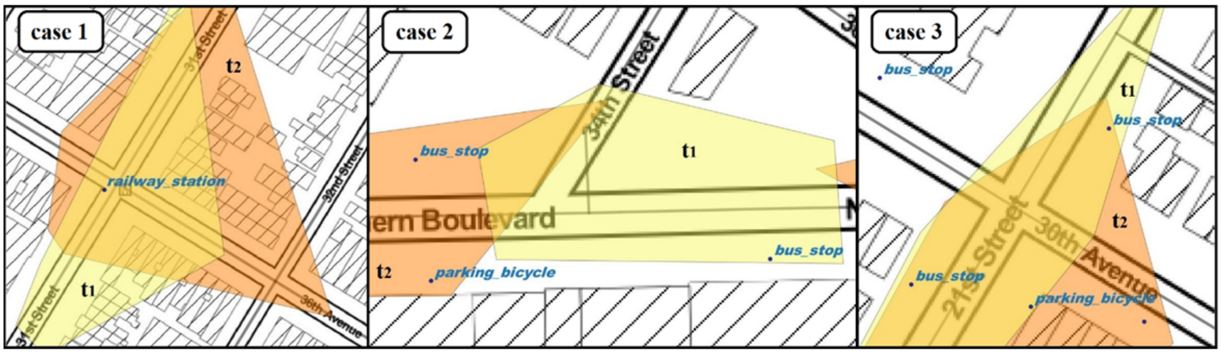

Fig. 2: Proposed definition of the urban transport hub phenomena as a collection of different appearances

All three cases in Fig. 2 represent the general ideas on modelling the moving transport hub phenomena. Understanding time-dependencies and dynamics of moving transport hubs can be beneficial for planning, decision-making, and various ongoing research in connection with smart cities and intelligent transportation systems (ITS). Besides gaining possible benefits for microscopic traffic simulations, there are options for improving activity models of agent-based traffic simulations, which is partially also the aim of this work.

1.3 Clustering vehicle movement trajectories and representation as moving polygons

One simple way to gain more knowledge on massive vehicle movement data set is using point clustering techniques, with and without respecting the temporal component.
Besides selecting suitable clustering techniques, it is important to specify the data representation of the vehicle movement. By selecting whole trajectories of vehicle trips, it should first be clear at which accuracy the trajectories were acquired and with which sensor. Additionally, the state of quality estimation of the vehicle movement data set is important. Raw and not-preprocessed trajectories might evolve as data with a considerable proportion of outliers that can go up to 50% (KELER 2017). Correcting these outliers might be complex, especially when applying map matching techniques. The large number of variations in implementing the latter can cause complex adaptations onto specific data sets. When raw vehicle movement data from urban environments includes systematic signal losses due to the presence of higher buildings, selected parts of the urban road network are fully or partially missing. One possibility for extending these missing partitions is the usage of different interpolation techniques that may help reconstruct missing trajectory elements.
Depending on the application and the way of analysis, extracts of trajectories can serve as data input. Extracting points or positions from a movement trajectory can facilitate selected applications as for example defining spatiotemporal trip destination hotspots (KELER & KRISP 2016). In case of taxis, the usual origin and destination correspond with taxi customer pick-ups and drop-offs.
After defining the trajectory point extracts, it is possible to segment the data sets into data extracts based on selected time window as for example hour-wise for the time of the day or day-wise for day of the week. By comparing the clustering results of these partitions, usually with the same clustering parameters, it is possible to formulate first insights after visualization and visual inspection of the outcomes. In a more general way, these clusters can be aggregated into convex hulls, which improves in many cases the visual inspection, reduces data volumes, and, allows the inspection of matching polygons on a more general level. Matching between different types of areal polygons relies on the Dimensionally Extended nine-Intersection Model (DE-9IM) as introduced by CLEMENTINI et al. (1994).

2 Methodological approach for modelling and visualizing spatially uncertain moving transport hubs

The methods of the proposed technique include density-based point clustering, convex hull estimations, polygon matching, and polygon intersection. The components are part of an extendible workflow, which is pictured in Fig. 3.
Fig. 3a shows the workflow components that make use of spatial positions of taxi trip origins and destinations as data input for computing density-based point clusters for a selected data partitions based on the OPTICS algorithm (ANKERST et al. 1999). The subsequent step bases on JARVIS

(1973) and consists of computing the convex hull for every density-based point cluster. The aim of the mentioned polygonization of point clusters is providing visibility at certain map scales by creating larger polygons for spatially extensive taxi trip origin or destination distributions at for example multiple connected road segments. These taxi trip origins and destinations are respectively clustered based on their spatial distribution within a selected time window as for a whole day.

The subsequent procedure consists of polygon matching in Fig. 3b. Polygon intersection in Fig. 3c is one part of defining moving transport hubs in the sense of analyzing the intersection patterns of consecutive polygons within a time series. Matching between different types of areal polygons relies on the Dimensionally Extended nine-Intersection Model (DE-9IM) as introduced by CLEMENTINI et al. (1994). The definition of moving transport hubs comes along with the step of inferring the indicators for the certain movement from different mobility services. In general, this step is not specified in detail and there are different possible options. One option is to extract information from OpenStreetMap that can indicate selected functionalities related to transport.

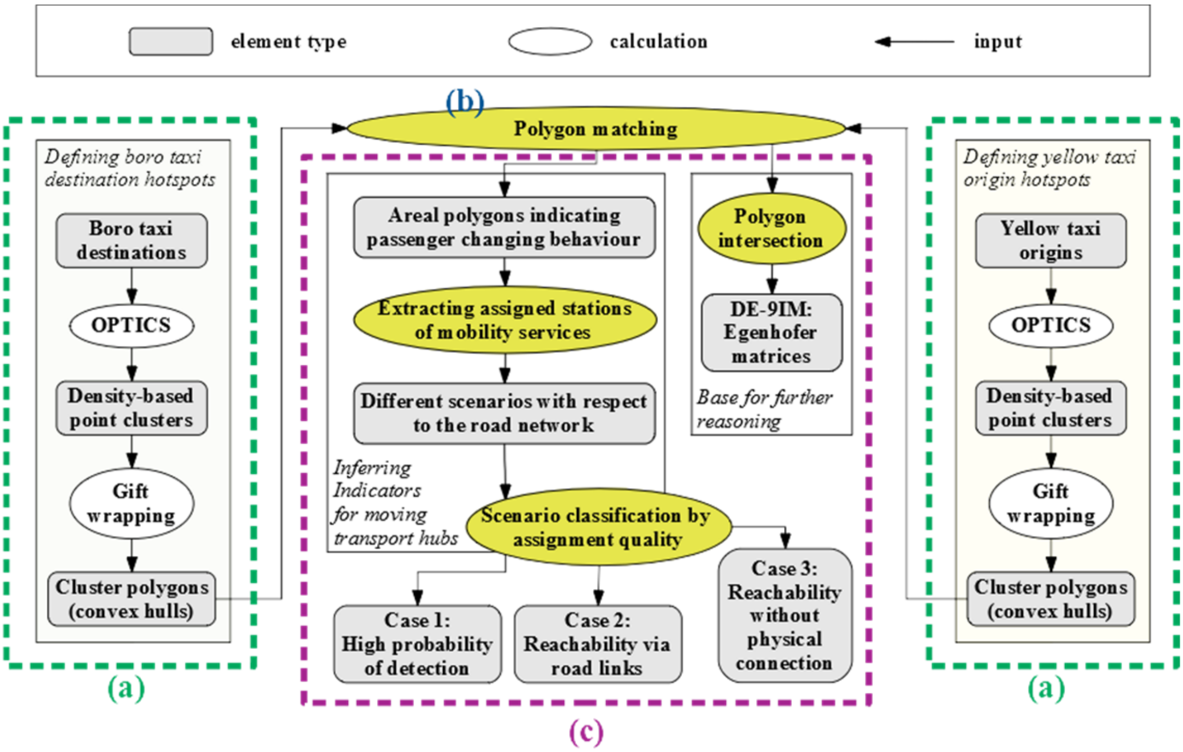

Fig. 3: Workflow of (a) generating origin and destination hotspot polygons, (b) its subsequent matching as consecutive sequences and same time windows, and (c) the definition of the moving transport hubs

By matching temporally subsequent taxi trip origin or destination hotspot polygons, it is possible gain more knowledge on the spatiotemporal distribution of service areas with similar functionalities. Fig. 4 shows four layers of boro taxi trip destination hotspots with introduced opacities. The darker the areas appear, the more common is the assignment as typical boro taxi trip destination hotspots. Besides the resulting boro taxi trip destination hotspots in Fig. 4, there are also extractable yellow taxi trip origin hotspots and functional transport regions resulting from public transport stations of subway and bus lines. This has a semantic connection to the restricted yellow zone and specific commuting patterns between Manhattan and the outer boroughs of NYC. The further steps include matching these spatiotemporal hotspots with characteristic transport hubs, and testing a specific visualization scheme for representing spatial and spatiotemporal uncertainties of the mentioned transport hubs. The outcomes of applying the techniques are matter of further inspections and visual analysis.

Fig. 4: Visualization of the polygon matching results, showing the spatial uncertainty of boro taxi trip destination hotspots for four successive time windows

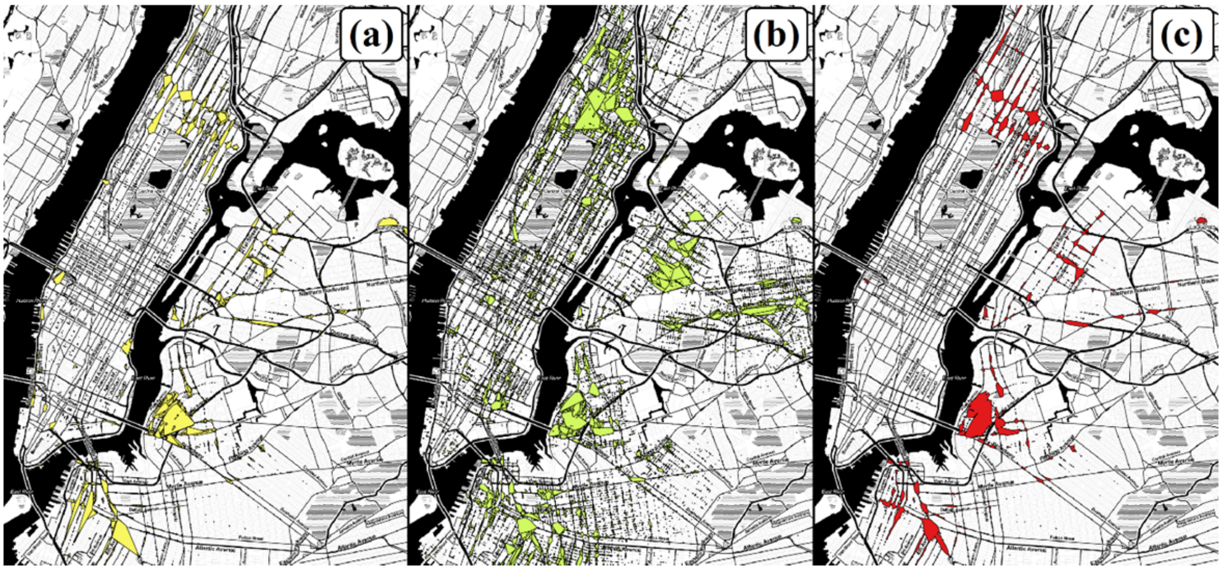

Fig. 5: Taxi hotspots polygons with (a) yellow taxi trip origins, and, (b) boro taxi trip destinations, and, (c) the intersection areas between yellow and boro taxi hotspots

Fig. 5 is interpretable as the merged product of all polygon layers within a time series. In case of Fig. 5, all four successive time windows of June 2015 are matched. Fig. 5a and 5b picture yellow taxi origin hotspots and green boro taxi destination hotspots. Fig. 5c shows the intersection product of both types of hotspots via red polygons. By further visual inspection, it is feasible to detect certain clusters in the spatial distribution of the intersection areas. One typical property of every intersection product between yellow taxi origin and boro taxi destination hotspots is the inclusion of the usually larger polygons around the border of the yellow zone, which has a restriction in customer pick-ups for boro taxi drivers. From this intersection product in Fig. 5c, it is possible to provide a preselection of traffic-related points of interest (POI's) or mobility service stations. Extracts from OSM serve as inputs for assigning those taxi hotspot intersection areas as moving transport hub, simply based on the presence of its existence in space. From the whole area of the State of New York, three selected types of points are extracted from OSM, which is pictured in Tab. 1.

Tab. 1: Selected information extracted from OpenStreetMap connected with functional transport regions and stations of mobility services in NYC

Type of information	POI's	traffic	transport
Feature classes or values	bicycle rental; car rental; car sharing	parking; parking bicycle; parking multistorey; parking underground	bus station; bus stop; ferry terminal; railway halt; railway station; taxi; tram stop
Number of extracted records	551	7 053	4 733

From a total number of 12 337 OSM point extracts 3 511 (= 28.5%) are matching with the intersection product in Fig. 5c. These extracted OSM points serve for the matching with every taxi hotspot intersection layer of every Saturday in June 2015.

3 Results and discussion

The presented methodology is tested with the mentioned selection of boro taxi and yellow taxi data from Saturdays in June 2015. Tab. 2 shows an overview on the used data partitions, the inferred taxi service hotspots, and, the generated moving transport hubs. By observing the low number of defined moving transport hubs in Tab. 2, it is possible to state that many yellow taxi origin and boro taxi destination hotspots are not located nearby stations of other mobility services in NYC.

Tab. 2 shows the unexpected outcomes of applying the workflow in Fig. 3. Despite the fact that only 33 496 yellow taxi trips appeared on June 6, all other yellow taxi trip on Saturdays are around thirteen times higher. Due to the fact that Tab. 2 only shows yellow taxi trip origins, we have higher origin densities at selected taxi stands, especially at the southern part of Manhattan, which is difficult to observe in Fig. 5a due to the smaller spatial extents of the hotspot polygons.

Tab. 2: Overview of inspected results for the Saturdays in June 2015 in NYC

2015	June 6	June 13	June 20	June 27
Boro taxi destination points (= total number of trips)	50 367	47 675	44 859	51 704
Boro taxi destination hotspots	5 006	4 919	4 769	5 233
Yellow taxi origin points (= total number of trips)	33 496	427 994	411 855	445 852
Yellow taxi origin hotspots	1 169	1 248	1 234	1 137
Number of defined (moving) transport hubs	172	178	167	179

The defined transport hubs are represented visually in Fig. 6 with an example from NYC showing an area at the Broadway. The example itself describes the dynamic size variations and spatial configurations of extracted polygons. Every defined moving transport can also be further intersected with each other for gaining typical moving transport hub patterns. In general, the example in Fig. 6 shows clearly the partially separation and aggregation of two nearby situated clusters of mobility stations: in the northwestern part and in the southeastern part of the map views. The transport hub areas on June 6 and June 27, 2015, have more the appearance of two separated polygons, whereas the transport hub area on June 13 and June 20, 2015 have respectively one more aggregated polygon.

Fig. 6: Visualization of the moving transport hub phenomena based on an example in NYC with four successive Saturday in June 2015 as time windows

Since the polygons base on the intersections of boro taxi destinations and yellow taxi origins, it is possible to imagine this example as a typical area for changing the mode of transport.

4 Outlook

First results of the presented approach show promising insights into the operational behavioral aspects of specific taxi drivers in urban environments. Since there are multiple examples that include the presence of mobility service stations and indicators, it is possible to assign moving transport hubs with different manifestations. Future work can focus on including information from traffic light signaling, together with respective average green times for every intersection, which may provide further insights into the analysis of transport hub movement. The following step would be including functional transport regions into the analysis on the level of specific bus or metro lines with the connectivity of specific stations. This would facilitate the classification into the introduced cases 1 to 3 (see Fig. 2) with knowing the accessible road segments. Furthermore, a segmentation into vehicle, bicycle and pedestrian lanes can improve the validation of extracted moving transport hubs.

One further future step consists of inspecting the correlation between the activity of appearing events that have a disproportionally large mentioning in social data extracts (e.g. from Twitter, FourSquare and Wikipedia). Knowing the typical or recurrent movement patterns of moving transport hubs can benefit the understanding of perceived attractiveness for selected demographical groups in urban environments. KELER et al. (2017) have for example defined measures of urban attractiveness for selected time windows in Augsburg, resulting in a spatially-high-resolution surface (1 m). The extension of this would be to include the dynamical aspect by generating a series of rasters as a time series. Building on this, it would be possible to not only estimate urban attractiveness, but as well to provide a connection to the usage of daily transport services in relation to the extracted moving transport hub phenomena. This might be one step in validating the proposed urban appearance, together with real world observations as for example via usage of multiple video devices.

5 References

ANKERST, M., BREUNIG, M.M., KRIEGEL, H.-P. & SANDER, J., 1999: OPTICS: Ordering Points To Identify the Clustering Structure. ACM SIGMOD international conference on Management of data, ACM press, 49-60.

CLEMENTINI, E., SHARMA, J. & EGENHOFER, M. J., 1994: Modelling topological spatial relations: Strategies for query processing. Computers & Graphics, **18**(6), 815-822.

BURGFELD, O. & LIEBHARDT, L., 2015: GI-Anwendungsprojekt - FCD-Analyse. Project report, Seminar GI-Projekt WS 2014/2015, January 27, 2015. Geoinformatics Group Augsburg, University of Augsburg.

JARVIS, R.A., 1973: On the identification of the convex hull of a finite set of points in the plane. Information Processing Letters, **2**, 18-21.

KELER, A., 2017: Ableitung von Verkehrsengpässen anhand Staudetektion mit Fahrzeugtrajektorien. Geoinformationssysteme 2017 – Beiträge zur 4. Münchner GI-Runde, Kolbe, T.H., Bill, R. & Donaubauer, A. (eds), Wichmann, Berlin/Offenbach, 172-184.

KELER, A. & KRISP, J.M., 2016: Is there a relationship between complicated crossings and frequently visited locations? – A case study with boro taxis and OSM in NYC. Proceedings of the 13th International Symposium on Location-Based Services, Gartner, G. & Huang, H. (eds), 186-190.

KELER, A., LI, X., MAZIMPAKA, J.D., WANG, L. & KRISP, J.M., 2017: Designing measures for urban attractiveness using VGI and Open Data-a case study in Augsburg. Workshop on semantic information in geosocial data for the mobile age (SemGeoSoc'17), September 21-22, 2017, Zürich, Switzerland.