Identification and Classification of Urban Centres

Using Public Transport Passenger Flows Data

Oded Cats, Qian Wang and Yu Zhao

Oded Cats (corresponding author) Department of Transport and Planning Delft University of Technology P.O. Box 5048, 2600 GA Delft, The Netherlands Phone number: +31 15 2781384 Fax number: +31 15 2787956 <u>o.cats@tudelft.nl</u>

Qian Wang Department of Transport, KTH Royal Institute of Technology Teknikringen 10, 100 44 Stockholm, Sweden <u>gianw@kth.se</u>

Yu Zhao Department of Transport, KTH Royal Institute of Technology, Sweden Teknikringen 10, 100 44 Stockholm, Sweden yuzh@kth.se

Keywords: Urban Centres; Polycentric; Passenger Flows; Clustering: Urban Planning.

Abstract

Urban regions worldwide tend to evolve from monocentric to a more complex distribution of activities. Measuring urban structures is thus essential for supporting an evidence-based spatial planning policy. This paper presents a methodology to identify and classify urban centres based on automatic transport flow data. Urban activity centres are identified based on the spatial distribution of travel flows. Urban centres are then classified and characterized based on the temporal distributions of incoming and outgoing flows, postulating that urban structure and the spatial distribution of activities are manifested in distinctive travel patterns throughout the day. The two-stage methodology was applied to Metropolitan Stockholm in Sweden using multi-modal public transport passenger flows. Stockholm was known for its long-term monocentric planning with a dominant central core and radial public transport system. Although the regional planning policy embraces a shift towards a polycentric planning policy, the results indicate that this has not been realized insofar.

1. Introduction

Urban regions worldwide tend to evolve from monocentric to a more complex distribution of activities. At the inter-urban scale, a polycentric policy was adopted by the European Union to support a balanced territorial development (Walsh 2012). At the intra-urban scale this pattern is driven by urban sprawl, suburbanization, the emergence of specialized employment clusters, shopping centres or other big visitor attractors located on the urban fringe or the combination of several interlining trends. In some cases, such developments may amount to the emergence of a polycentric or multi-centric urban structure (Davoudi 2003, Baum-Snow 2010). While these multi-centric concepts remain contested in the urban geography literature, they are essentially defined by the plurality of urban centres, often driven by the decentralization of people, jobs, and services from the core area to sub-centres. The emergence and the spatial distribution of urban centres are facilitated by the underlying transport network. This paper proposes a methodology to identify and classify urban centres based on transport flow data. The spatial and temporal mobility patterns facilitate the analysis of the centres and their relation to the urban activities and transport networks.

Anas et al. (1998) classified centres into two types - old towns incorporating an expanded urban area and newly spawned centres located at nodes of a transportation network. The growth of the second group is seen as the most popular pattern in changing the urban landscape. There is an extensive literature on the impact of urban forms and land-use distribution on travel patterns, such as the influence of urban agglomeration (Garcia-Palomares 2010) and polycentric urban structure (Schwanen et al. 2001, Casello 2007) on modal split and commuting patterns. Based on a spatial economics model, Louf and Barthelemy (2013) showed that the monocentric regime becomes unstable as the population grows and that the number of sub-centres grows sub-linearly with population size. External economics of scale underlie the emergence of urban clusters as the spatial concentration of specialized economic activities could foster innovation, spill-over effects and agglomeration benefits. In contrast to these studies, Shearmur and Coffey (2002) questioned the attempts to generalize spatial distribution trends and to imply that urban areas converge to a common spatial development trajectory.

Measuring urban structures is essential for supporting an evidence-based spatial planning policy. The identification of urban centres and their clustering and characterization will provide planners and policy makers with a better understanding of the existing metropolitan structure and enable them to assess how well it corresponds or diverts from planning policies. As pointed out by Meijers (2008) in the inter-urban context, there is an empirical deficit in the context of spatial planning development that should be addressed by applying a more analytical approach. Previous studies stressed the difficulty of obtaining flow data for analysing urban structure and using transport network attributes or topology indicators as proxies (e.g. EPSON 2004, Silva et al. 2014). However, the growing availability of 'big data' in the transport sector and in particular travel flow data facilitates the spatial and temporal analysis of urban activity.

This paper presents a two-stage methodology for identifying and classifying urban centres. Urban activity centres are identified by clustering transport nodes according to their spatial proximity and travel flows. A sensitivity analysis is performed to investigate the implications of different parameter values on the number of centres, their spatial distribution and the distribution of passenger flows. Centres are then classified based on their time-dependent flow profile including magnitude,

directness and the distribution of incoming and outgoing flows. It is postulated that urban structure and the spatial distribution of activities are manifested through time-dependent flow profile because urban centres with distinguished functions will yield distinctive travel patterns throughout the day. We use the term urban centre rather than sub-centre because the method could be applied in different spatial contexts and no particular structure – monocentric or polycentric –is assumed from the outset. An application to Stockholm County in Sweden demonstrates the potential to gain better understanding on the urban structure by applying the proposed two-step method with the use of pervasive data. This allows a direct method to investigate the emerging global order from numerous individual travel decisions rather than derive it by approximating generation and attraction rates for various land-uses.

The remainder of this paper is organized as follows. The following section reviews the literature on measuring urban structure based on urban geography and economics, spatial analysis and transport flow applications. We then propose a methodology for identifying and classifying urban centres based on time-dependent transport flows. Section 4 presents the context of our case study area and the data available for this analysis. The results are presented and discussed in the context of Stockholm regional and transport development in Section 5. The paper concludes with a discussion on the implications of our findings and study limitations.

2. Literature review

The two prominent approaches for identifying urban centres could be classified into morphological and functional methods. These two approaches essentially correspond to the analysis of densities or mobility patterns, respectively. Most previous studies undertook a morphological approach. Subcentres were identified by setting cut-off values for population density (Giuliano and Small 1991, Rodrigues da Silva et al. 2014), employment densities (Casello and Smith 2006, Louf and Barthelemy, 2013) or the ratio between the two (Shearmur and Coffey 2002). Gordon et al. (1986) and McDonald (1987) suggested to examine not only whether the density exceeds a certain absolute or relative value compared with the entire urban area but also whether it amounts to a local maximum compared with the surrounding area. Several studies defined sub-centres based on estimating the residuals of an employment density function and computing indexes of spatial correlation (McMillen 2003, Riguelle et al. 2007, Adolphson 2009). Rodrigues da Silva et al. (2014) also used data on the density of transport infrastructure as a proxy for travel patterns but concluded that this was a poor indicator of spatial urban evolution.

Functional urban areas are defined by both US and European authorities based on population distribution and inter-area movements (OMB 2000, OECD 2013). However, the morphological approach is restricted to a static representation of densities without considering the movement and exchange of people, goods or information between centres. Employment densities are considered by morphological studies as proxies for commuting trips, disregarding other trip purposes and hence may not capture adequately the intensity of urban activity across the region. An alternative approach considers the functional relations between centres based on the analysis of commuting patterns. The functional approach considers a more balanced flow distribution the blueprint of a polycentric structure, whereas the morphological approach regards a more balanced distribution of centres in terms of size and geographical distances as an indication of a polycentric form (ESPON, 2004). Although these perspectives may seem virtually equivalent, previous studies found urban areas that

their morphological and functional analysis diverge, presumably due to the underlying transport network structure (Hall and Pain 2006, Burger and Meijers 2012).

As pointed out by Shon (2005), while many studies investigated the commuting patterns that emerge in different urban structures, there were only few studies that attempted to study the urban structure based on the prevailing travel patterns. Interaction indices based on the magnitude and direction of commuting flows were analysed by Gordon et al. (1986), Burns et al. (2001), Roth et al. (2011) and Veneri (2013). This approach enables not only the identification of sub-centres but also the analysis of the relations between those centres. A common limitation that many of the previous studies share is their restriction to employment or commuting data for performing morphological or functional analysis, respectively, while work-related activities may constitute only a small portion of all urban interactions and movements (e.g. Schwanen et al. 2001). Furthermore, with the exception of Roth et al., data on commuting patterns was extracted from travel habit surveys which are costly and not as widely available and regularly updated as information on land-use distribution.

While the number and location of sub-centres can provide insights on the distribution of activities in the urban space, the function of each sub-centre within the urban system is not revealed. There is relatively limited research that aims to systematically classify urban centres. The study by Giuliano and Kenneth (1991) grouped centres by applying a hierarchal clustering based on the mixture of industries. Van der Laan et al. (1998) defined four categories of urban structure based on the combination of the degree to which commuting is oriented towards the centre vs. the sub-urban areas.

Transport systems worldwide are increasingly equipped with automated data collection methods. Data concerning travellers' flows is collected using various methods including GPS, plate recognition and ticket validation. This abundant data is collected, updated and directly available. This paved the way to a rapid increase in research concerned with unravelling travel patterns based on automatically collected mobility data. Liu et al. (2009) and Hasan et al. (2013) investigated the temporal and spatial variations of destination choices based on smart card data from London and Shenzhen, respectively. Few recent studies used public transport flows at the urban area level to shed light on the underlying urban structure. Roth et al. (2011) analysed smart card data from London underground system to identify sub-centres and revealed the polycentric structure of the city. The clustering process is based on the Euclidean distance and passengers' inflows. Based on data from the same system, Ceapa et al. (2012) observed distinctive temporal patterns for different stations located in residential, business and transport hub areas.

This paper contributes to this research domain by formulating and applying an integrated methodology for identifying and classifying urban centres based on incoming and outgoing public transport passenger flow counts. The proposed method is functional in the sense that it reflects movements of people and analyses the urban structure in terms of flows of people. It also enables a dynamic perspective of centres' intensity by conducting a time-dependent analysis. Moreover, it does not require extensive data processing or modelling assumptions such as those involved in path inference methods from entrance and exit records. However, the data does not contain information on the direction of travellers' movement and hence cannot identify the interaction between any pair of centres. The details of the two-stage procedure are presented in the following section.

3. Methodology

Public transport passenger flows are used in this study to identify activity centres. The individual stations in the public transport network are clustered according to their locations and loaded passenger flow. The identified clusters are thereafter classified into classes based on their time-dependent travel pattern. Before turning into the formulation of the two-stage identification and classification method, we define some notations that are used throughout this section.

Each station $s_i \in S$ is associated with alighting (incoming) and boarding (outgoing) passenger flows $f_{i,t}^a$ and $f_{i,t}^b$, respectively, where *S* denotes the set of stations and *t* is the respective time-window. The term 'station' refers in this paper to all boarding and alighting locations in the public transport system, regardless of the transport mode. A distance matrix *D* is constructed, where each entry d_{ij} , $s_i, s_j \in S$ is computed based on the geographical coordinates of stations. Given that the identification method resulted with a centre set *C*, the joint alighting and boarding flows for centre $c_m \in C$ and time-window *t* are denoted by $f_{m,t}^a$ and $f_{m,t}^b$, respectively. *C* is partitioned into a mutually exclusive and collectively exhaustive class set *U*, where each class $u_k \in U$ is defined based on the temporal flow distribution pattern. The numbers of stations, centres and classes are thus denoted by |S|, |C| and |U|, respectively.

3.1 Identification

The identification method consists of hierarchal clustering and descending composition of passenger flows and is inspired by Roth et al. (2011) and Ozus et al. (2012). A cluster could be conceptualized as consisting of a primary-centre station and a number of member stations. The purpose of identification is to group member stations to the closest cluster-centre station which dominates the cluster in terms of passenger flows, while not exceeding a maximum distance ρ between each member station and the cluster-centre. The clustering algorithm is not exhaustive so that insignificant or remote stations need not be clustered when identifying urban activity centres. A threshold δ for the accumulated share of clustered-passenger-flow (CPF) is used as a termination criterion for the identification algorithm.

Figure 1 presents the sequence of steps for identifying urban centres. In this study, the total public transport passenger flow per station, $f_i = \sum_t (f_{i,t}^a + f_{i,t}^b)$, is used in the clustering algorithm. The method involves the specification of two parameters ρ and δ . A sensitivity analysis is necessary for specifying the values of these parameters using criteria such as the number of clusters and the partition consistency. The distance-based algorithm adopted in this study clusters a station to the nearest cluster-centre. Alternatively, one could also use a flow-based algorithm which gives priority to the largest flow when clustering. However, this could result with violating centres' geographical integrity.

The algorithm runs through the flow-descending station set and it uses a list l to label stations as follows: $l_i = 0$ if station s_i has not been assigned yet to a cluster; $l_i = 1$ if station s_i is a member of a cluster; $l_i = 2$ if station s_i is a centre of a certain cluster. The algorithm searches for a station s_j which is the closest to s_i among those that have a higher ranking. In the case of the highest-ranked station, $s_j = s_i$ and s_j becomes the cluster-centre of the first cluster, c_1 . For all other stations, s_i becomes a member of the cluster which s_j is its cluster-centre if $d_{ij} \le \rho$. Otherwise, s_i forms the cluster-centre of a new cluster. In order to monitor the accumulated share of clustered flows, CPF, a

set of stations denoted *P* includes all the identified cluster-centres and stations within distance ρ from them (even if they have not been assigned to a specific cluster yet). The algorithm breaks thus when $CPF < \delta$.



Figure 1: Flowchart of the clustering algorithm for identifying urban centres

3.2 Classification

Given a centre set C obtained from the previous stage, the task now is to classify each of the centres in C into a smaller number of classes based on similar characteristics. The agglomerative hierarchical clustering approach is applied for classification. A detailed description of this approach and its properties can be found in Hartigan (1975). The classification stage consists of the following four steps:

i. Calculating cluster dissimilarity

An indicator $y_{m,t}(f_{m,t}^a f_{m,t}^b)$ is created for centre c_m so that the dissimilarity in temporal characteristics between a pair of centres $\{c_m, c_n\}$ is measured by $d_{c_m, c_n} = \sum_t |y_{m,t} - y_{n,t}|$, the pairwise distance.

ii. Determining a hierarchical tree on the basis of a linkage criterion

The construction of a hierarchical binary cluster tree structure requires a *linkage function*, which is a function of the pairwise distances (further detailed on linkage rules can be found in Milligan, 1980 and Tan et al, 2006). Different linkage criterions yield different hierarchical binary cluster trees. The tree could be illustrated graphically by a *Dendrogram*, in which the horizontal axis represents the indices of elements in the data and the vertical axis represents the height of U-shaped link between the elements. The height of the links corresponds to the *Cophenetic distance* as calculated by the linkage function.

The *Cophenetic correlation coefficient* could be used for validating how well the *Cophenetic distances* in the tree reflect the original distance data (Tan et al.2006). The closer the coefficient, γ , is to 1, the more accurately the tree represents the dissimilarities in the data. Comparing the coefficients by different *linkage function* allows us to determine which linkage criterion performs better.

iii. Constructing agglomerative class

Given the hierarchical cluster tree, there are two methods to partition data points into classes, one is natural division and the other is arbitrary clusters. The natural division method uses *inconsistency coefficient*, which compares the height of the current link and the average height of neighbouring links further down in the tree. By defining a threshold β as the cut-off value, the method partitions the links which have consistency coefficients greater than β .

The arbitrary cluster partitions the data into a pre-defined number of classes. This method could be illustrated by placing a horizontal line crossing the U-links in a *dendrogram*. Unlike the natural division method, this method allows us to specify any number of classes.

iv. Assessing the classification results

The centres in set *C* are assigned into a class set *U* and a total intra-distance is computed as the sum of all the pairwise distances, $d_{intra} = \sum_{u_k \in U} \sum_{c_m, c_n \in u_k} d_{c_m, c_n}$. The total intra-distance depends on the clustering method and the number of classes yielded. The marginal effect on the total intra-distance of adding one or more classes has to be investigated. The total intra-distance is therefore used to assess the classification method and the appropriate number of classes.

The classification result also yields the inter-distance between each pair of classes. If the value of indicator y for class u_k is notated as $y_{k,t}(\dots, y_{m,t}, \dots), \forall c_m \in u_k$, the pairwise distance is $d_{u_k,u_l} = \sum_t |y_{k,t} - y_{l,t}|$. The total inter-distance is the sum of the pairwise distances $d_{inter} = \sum_{u_k,u_l \in U} d_{u_k,u_l}$. The better the classification algorithm is the intra-distances become smaller compared to the inter-distances.

4. Case study

4.1 Regional planning in Stockholm

The proposed method is applied to the case study of Metropolitan Stockholm (in Swedish: Storstockholm) defined as Stockholm County. Stockholm is positioned between lake Mälaren and the Baltic Sea and is built up on a big archipelago. Large green areas (30% of the area), lakes and waterways (additional 30%) form geographical barriers that divide the built-up area. With 2.16 million inhabitants and the fastest growth rate, it is the largest metropolitan area in Sweden. The county covers 6,500 square kilometres (approximately 105 km north to south and 60km east to west) and includes 26 municipalities surrounding the Swedish capital. In the late 19th century and early 20th century, the city grew gradually from the old town to encompass neighbouring islands. Stockholm is positioned between built u

Stockholm is famous for its long-term monocentric planning with a dominant central core and the planning of relatively dense satellite rail-bound towns (within and beyond the municipality boundaries) throughout the 20th century (Cervero, 1995). The inseparable urban and transport planning in Stockholm are a prime example of a radial public transport system which is primarily oriented towards regional accessibility rather than providing local coverage (Derrible and Kennedy, 2010). Börjesson et al. (2013) concluded based on an ex-post analysis that the land-use development along the long-stretched rapid public transport corridors led to a more dispersed region in the case of Stockholm. Satellite towns were developed along the rapid public transport corridors and thereof creating a public transport metropolitan area which promotes suburb to centre commute.

Unlike most European metropolitan areas (Riguelle et al. 2007, Veneri 2013), Stockholm primarily expended by developing satellite towns along its expending rapid public transport system rather than the evolution of a pre-existing hierarchical urban system and the absorption of nearby towns. Nevertheless, the trend observed by Veneri (2013) across Europe towards integration rather than expansion is also observed in Stockholm. The satellite towns led to the decentralization of population but only in the last two decades were also followed by the development of employment and shopping centres. This is equivalent to the congregation of urban structures in proximity to ring roads in Los-Angeles (Giuliano et al. 2012), in both cases planning and urban economics favour locations with superior regional accessibility.

Since the turn of the century there has been a noticeable shift towards developing sub-centres, promoting a more balanced distribution of activities. Johnson (2012) and Schmitt et al. (2013) provided recent reviews of the urban planning context in Stockholm and how the polycentrality strategy was embedded into the planning process since it was firstly introduced in the regional development plan in 2001. Subsequent plans identified 7-8 regional sub-centres that will promote a more multi-centre urban structure around the primary core. Adolphson (2009) investigated the urban structure changes in the Stockholm region between 1991 and 2004 based on the residuals of

an employment density function and found an increase from 5 to 7 urban centres as well as an increase in the spatial dispersion of the primary centre.

The strategy of promoting the development of urban centres within the greater Stockholm area is motivated by the need to relief the city centre from the negative externalities associated with oversaturation. It was argued that sub-centres will make the urban system more robust and diverse. While radial commuting patterns still dominate passenger flows, a more polycentric structure is promoted by regional planning guidelines as well as direct attempts to support the development of strategic nodes (Stockholm City 2011). This policy is supported by the development of a corresponding transport infrastructure - a cross-radial light rail train, several extensions of the metro system and increasing the capacity of the commuter train system - designed to support a stronger network of strategic nodes in Metropolitan Stockholm and reduce travel barriers.

4.2 Public transport data

The closely inter-linked development of Metropolitan Stockholm and its public transport system motivates the analysis of the urban structure based on public transport passenger flows. The average number of public transport trips per person per day in Stockholm County is 0.63 of which the lion share is performed by either metro (44%) or bus (41%). Moreover, 70% of the inhabitants of the central part of Stockholm County (including all the metro stations) use public transport at least several times a week and 54% of all trips are carried out by public transport, with this percentage increasing to 80% for trips with a destination in the regional core (SL 2013).

Passenger flows were available for each of the 12,757 stations located within the Stockholm County boundaries for all modes. The public transport system in Stockholm County consists of commuter train, metro, bus, tram, light rail and local trains. Regional and national trains are not included in this analysis. The number of boarding and alighting passengers for 5 time periods – night (0-6), AM peak (6-9), midday (9-15), PM peak (15-18), evening (18-24) - on an average weekday were computed for each station based on automatic passenger counts (APC), fare validation and gate counters (estimated number of metro entrees and exits). The matrix of geographical distance were calculated based on station coordinates.

Figure 2 presents the cumulative distribution function of urban public transport passenger flows. It is evident that flows are very unevenly distributed over stations, with 80% of the passengers' activity taking place in only 5% of the stations. A power law relation is observed between passenger flow and number of stations implying a linear function for the log-log plot, in line with previous results for the patterns reported by Lee et al. (2008) and Lin and Ban (2014) concerning passenger flows in metro and air traffic networks, respectively.



Figure 2: Cumulative density function of passenger flows as function of share of stations

5 Results

5.1 Identification

The identification algorithm is applied for the case of Stockholm. A sensitivity analysis was carried out in order to assess the impact of the two parameters – the maximum distance between the primary centre of each urban centre and the farthest station, ρ , and; the minimum share of flows that are assigned to centres, δ - specification and select their values. Similarly to the morphological cut-off methods discussed in Section 2 (Riguelle et al.2007, Anas et al. 1998), the number and boundaries of centres may depend on the threshold used.

The range of maximum distance was set to 1 to 3km with 0.5km intervals since the largest district in Stockholm municipality is 2460 hectors, roughly equal to a circle with a radius of 3 km. The share of passenger flow was set tested for 50,60,70 or 80 %. It is expected that smaller values of ρ and higher values of δ will result in more centres. In addition, the results of the distance-based identification algorithm were compared with a flow-based algorithm (see Section 3.1). For this case study, the two algorithms always yield the same number of centres but the distance-based algorithm. Hence, the distance-based algorithm resulted in more evenly distributed centres.

The number of centres and the standard deviation of passenger flows among centres for different parameter-combinations are presented in Figure 3. It is evident, that the larger the radius - fewer centres and a higher standard deviation are obtained. In addition, the higher the threshold δ is - the larger number of centres and the lower the standard deviation are. Though it is expected that covering a greater share of the flows or reducing the geographical coverage of a centre implies more centres, the decrease in standard deviation is a property of the geographical dispersion of flows. The more smaller clusters are generated, the more heterogeneous the share of flows become due to the greater geographical dispersion. Note however that this may not necessarily be the case if a more even multi-polar spatial distribution of flows exists.



Figure 3: Number of centres and standard deviation of centres' flows for combinations of maximum distance and minimum share of assigned flows

In addition to the number of centres and the variation of passenger flows among the centres, the marginal contribution of each centre should be significant. The cumulative distribution function of the share of flows associated with each centre for each partition was studied and revealed that many (12 out of 20) of the parameter specifications resulted with the smallest centre contributing less than 1% of the total flows. This implies the generation of minor centres which are not of regional importance. In contrast, (3 out of the 20) partitions with lower thresholds and/or higher radiuses resulted with the smallest centre being relatively large and similar in size to the second smallest centre, thus suggesting that additional centres could potentially be included. This approach does not require local knowledge but rather rely on relative size. Note that the distribution of percentage flow contribution is on its own an indicator of the urban activity concentration.

The partition yielded by $\rho = 1.5$ km, $\delta = 0.6$ was selected from the remaining parameter specifications because it obtained the lowest flow variability. Figure 4 presents the number of stations and the accumulated share of passenger flows (out of those included in the partition) where centres are shown in a descending order of flows. It is evident that centres containing a higher number of stations are in most cases associated with larger passenger flows, there are exceptions to this rule (e.g. Solna Centrum vs. Fruängen). The 3 largest centres (T-Centralen, Fridhemsplan and Stockholms Södra) account for more than 50% of the clustered flows indicating that passenger flows and urban activities are concentrated in a relatively limited area in line with the power law characterizing station flows.



Figure 4: Number of stations and share of passenger flows included in each centre

The 17 centres identified in Stockholm by applying the distance-based clustering identification algorithm with $\rho = 1.5$ km, $\delta = 0.6$ are displayed in Figure 5 using ArcMaps. The stations that constitute the cluster-centre are highlighted and we refer to centres by the names of the respective cluster-centre stations. With the exception of one centre (Märsta, a commuter train station), all centres are located within the core of the Metropolitan Stockholm and their centre-stations are metro stations. The centre of gravity is T-Centralen (the only station where all metro lines intersect) with seemingly orbiting rings of centres in the inner-city (Tekniska Högskolan, Karlaplan, Stockholms Södra, Fridhemsplan), inner/closer-suburbs (Universitetet, Ropsten, Gullmarsplan, Liljeholmen, Alvik, Solna Centrum) and outer/further-away-suburbs (Mörby Centrum, Fruängen, Skärholmen, Brommaplan, Kista). The centres are confined to a relatively small part of Metropolitan Stockholm the far-most centres are Märsta and Fruängen which are 36 km bird-distance apart - in agreement with the findings of Vasanen (2012) from three large urban areas in Finland that most of the trend towards polycentricity was limited to a relatively small part of the respective regions. Note that most of the centres lie west of T-Centralen and the inner-city due to the existence of greater physical barriers in the eastern part of the Stockholm archipelago. In the following section, these centres will be classified by analysing the temporal patterns of passenger flows.



Figure 5: GIS visualization of distance-based algorithm with $\rho = 1.5$ km, $\delta = 0.6$

5.2 Classification

The agglomerative hierarchical clustering method was applied to the 17 centres identified in Stockholm. A preliminary investigation of passenger flow profiles showed that night flows are insignificant and thereof only passenger flows between 6:00-24:00 are considered. Three alternative temporal-profile indicators, $y_{m,t}(f_{m,t}^a f_{m,t}^b)$, are examined in this study:

- A+B : the total passenger flow during a certain time interval, $f_{m,t}^a + f_{m,t}^b$; an absolute measure of activity at an urban centre
- *A-B*: the difference between incoming and outgoing flows, $f_{m,t}^a f_{m,t}^b$; an absolute measure of urban centre attraction
- (A-B)/(A+B): the rate of net incoming flows of the total flows, $(f_{m,t}^a f_{m,t}^b)/(f_{m,t}^a + f_{m,t}^b)$; a relative measure of urban centre attraction that can range from -1 (solely a source, origin) to +1 (solely a sink, destination)

While A+B corresponds to the overall 'size' of a centre, A-B and (A-B)/(A+B) indicate the role of the centre as a 'source' or a 'sink' in the network, in absolute and relative terms, respectively. Computing these indicators and constructing the time profile for each centre enables us to examine the daily distribution and apply the classification method in order to find common patterns which characterize a set of centres.

For each indicator, a hierarchal clustering tree visualized as a *dendrogram* is created by minimizing the sum of pair-wise intra-distances and maximizing the sum of pair-wise inter-distances as explained

in Section 3.2. The *dendrograms* presented in Figure 6 illustrate how centres are grouped based on their degree of similarity with respect to the temporal profile of each indicator. The distance between any pair of stations is depicted by the value corresponding to the lowest intersection of the respective branches. It is apparent that the various indicators yield different hierarchical cluster trees. Unsurprisingly, A+B results in the most pronounced hierarchy, where the central business district and commercial centre around T-Centralen stands out as it has the greatest total flows throughout the day. This singular centre is followed by the centres of Fridhemsplan and Stockholms Södra and then Tekniska Högskolan, Gullmarsplan and Liljeholmen. Compared with A+B and A-B, a profoundly different pattern emerges for (A-B)/(A+B) which results with two distinctive groups of classes apart from the centre of Fruängen which differs significantly from all other centres.



Figure 6: Dendrogram based on average linkage criterion of each indicator [TC:T-Centralen; SS: Stockholms södra; RS: Ropsten; FP: Fridhemsplan; LH: Liljeholmen; TH: Tekniska högskolan; GP: Gullmarsplan; MC: Mörby centrum; KP: Karlaplan; AV: Alvik; FG: Fruägen; MT: Märsta; KT: Kista; SH: Skärholmen; BP: Brommaplan; UT: Universitetet; SC: Solna centrum]

Based on the inconsistency coefficient of the hierarchical trees, the set of 17 urban centres was partitioned into 6-8 classes for each indicator. The number of clusters was selected separately for each indicator based on analysing the marginal contribution of an additional class to the total intradistance (d_{intra}) to minimize variations within classes. The natural division method proved to provide superior results than the arbitrary cluster partition for all three indicators and was therefore chosen for classifying the centres. Figure 7 presents the temporal profile of each centre for each indicator groups into the identified classes (left) and their corresponding spatial distribution in Metropolitan Stockholm (right).

The centres are grouped into 6 classes based on their total flow profiles (*A+B*). It is evident that the classes are grouped into very cohesive classes as the magnitude of total flows follows almost the exact same ranking throughout the day and with visible qualitative gaps between classes. The total flow during the PM peak is higher than during the AM peak for centres that belong to the four highest-flow classes, although to a different extent. In contrast, the AM and PM peaks are of similar magnitude and an overall more balanced distribution is observed on the two lower-flow classes which include 10 out of the 17 centres.

A less obvious classification prevails for the net incoming flow (*A-B*) which results with 8 classes based on the directness and daily distribution of incoming and outgoing flows. T-Centralen has again a unique pattern as it gradually turns from a major sink to a substantial source. All of the other innercity centres (Fridhemsplan, Stockholms Södra, Karlaplan and Tekniska Högskolan) have a PM peak that mirrors their AM peak, the blueprint of a commuting destination with balanced flows during the midday and a small net outgoing flows during the evening. A similar pattern with a smaller magnitude is observed for the more peripheral employment centres of Kista and Universitetet. The opposite pattern is observed for residential areas such as Gullmarsplan, Brommaplan and Fruängen which form a class. More balanced patterns characterize areas with diverse activities such as Liljeholmen and Skärholmen, while the former is an example of mixed-use development (mixed offices and dense housing, shopping and entertainment, industry), the latter encompass large areas with distinctive activities (the largest shopping area in Sweden and a suburban multi-storey neighbourhood).

The classification based on the temporal profile of relative attraction, (*A-B*)/(*A+B*), results in a partitioning that is significantly different from the two absolute indicators, splitting the 17 urban centres into 7 classes. This indicator enables to consider the relative attraction deficit and hence reveal similarities in directness and function that are not apparent when analysing absolute flows. Unlike its exclusivity in previous indicators, T-Centralen and Tekniska-Högskolan follow the same trend, suggesting that the north-eastern districts of the inner-city (Östermalm and Vasastan) offer similar urban properties to those of the commercial city centre, although with a lower level of intensity. A similar profile is observed for the urban centre of Stockholms Södra, Fridhemsplan and Kista. The former two are additional inner-city centres, while the latter is the only centre outside of the inner-city that prevails as a cluster with inner-city qualities although with a currently secondary intensity. Centres with either substantial employment and limited leisure and night-life activities (Universitetet and Karlaplan) or almost exclusively residential areas (Fruängen) are characterised by a very polarized profile.







A-B





Figure 7: Temporal profiles (left) and maps (right) based on A+B (top), A-B (middle) and (A-B)/(A+B) (bottom)

The distinctive daily distribution of travel patterns reflects the distinguished functions of the urban centres. The superposition of the classification yielded by the three indicators provides an integrated perspective on the function of each urban centre and its similarity or dissimilarity to other centres. Centres which are similar with respect to one indicator may vary greatly with respect to another indicator. For example, the centres Liljeholmen and Tekniska Högskolan, the centres with the 5th and 6th highest flows, belong to the same branch of A+B as they exercise a very similar daily distribution of total flows. However, these two centres are dissimilar on A-B and (A-B)/(A+B) with Tekniska Högskolan acting as a sink in the AM peak, balanced incoming and outgoing flows in the midday and being a source in the PM peak and evenings, while Liljeholmen has very balanced flows throughout the day. This suggests that although these centres are of major regional centres and have a similar daily distribution of intensity in terms of people circulation, they have different functional roles. In other words, the morphological and functional classifications may differ. Tekniska Högskolan has a temporal flow distribution characteristic of a commuting destination while Liljeholmen encompasses more diverse activities. The former includes the north-eastern districts of Stockholm inner-city which include major public establishments while the latter lies south-west of Stockholm inner-city and contains a mixed-use shopping, office and residential areas. Both centres include public transport interchange hubs to local commuter trains and light rail train, respectively, as well as bus terminals.

6 Conclusions

This paper proposes a two-stage methodology for identifying and classifying urban centres and applied it to Metropolitan Stockholm based on multi-modal public transport passenger flows. Based on a sensitivity analysis, the distance-based clustering algorithm resulted with 17 urban centres when specifying a radius of 1.5 km and assigning 60% of all passenger flows in the case study area. In contrast, Stockholm is expected to have only 2-3 sub-centres in addition to its primary centre based on the equilibrium number of sub-centres as a function of population and commuting costs derived by McMillen and Smith (2003) based on an empirical analysis of 62 large American cities. However, the geographical and planning circumstances in Stockholm promoted the generation of a relatively large number of sub-centres which are physically separated. Conducting a morphological analysis of Stockholm, Adolphson (2009) identified 7 urban centres – the largest one covering 7 of the centres identified in this paper and 2 additional one-to-one correspondence. However, the four smallest morphological centres were not detected in our analysis, whilst 7 of the centres found in our analysis were not identified as centres by Adolphson. This may indicate that similarly to Greater London and in contrast to most regions in the Netherlands (Burger and Meijers 2012), Stockholm is more functionally polycentric than morphologically polycentric.

This paper goes beyond measuring the intensity of urban centres in an attempt to characterize urban centres on the basis of temporal public transport passenger flow distribution. The 17 centres were classified based on the temporal profile of 3 flow indicators. The results indicate that the central business district and the commercial centre still dominates and acts as a magnet of intensive urban activities. Surrounding inner-city districts offer similar qualities, primarily the north and north-eastern districts, although with a lower magnitude. There is only one urban centre beyond the inner-city boundaries, Kista (12 km from the city centre) that emerges as an important attractor of regional importance although of a secondary order. Since its early development in the 1980s, Kista has become the largest corporate and information technologies area in Sweden. In addition, centres directly south of the inner-city, Gullmarsplan and Liljeholmen, where a mixed-use and large

entrainment facilities were constructed in recent years, grow in importance and present a balance between incoming and outgoing flows throughout the day. Other urban centres that were identified by the planning authorities have not insofar emerged as important activity centres (e.g. Ropsten, Märsta), remain 'bedroom communities' (e.g. Fruängen, Brommaplan) or do not even constitute a centre based on our analysis (e.g. Fleminsberg, Barkarby-Jakobsberg). These results indicate that Stockholm has not yet been realized a polycentric or even multi-centric urban structure. Hence, the aim to release pressure from central part by polycentric structure has not been achieved yet. There are however indications that two kinds of secondary centres exist: (a) specialized sub-centres emerging in the periphery and; (2) mixed-used centres where the inner-city expends beyond the waterways. These processes are facilitated and expected to further ascend with an increased crossradial accessibility, not dissimilar from the role that highway ring roads played in the decentralization of cities in the United States (Baum-Snow 2010, Giuliano et al. 2012). This however may have negative consequences on the modal shift as empirical evidence indicates that the decentralization of urban areas and the emergence of sub-centres results with a modal shift towards the private car (Schwanen et al. 2001).

The methodology presented in this paper could be applied to a range of spatial units and systems (national, international, airports, maritime traffic). The proposed identification and classification methods are expected to produce reasonable results even in the lack of familiarity with the study area and could be automated. These advantages could facilitate their applicability to analysis geographical areas worldwide (McMillen and Smith 2003).

The analysis of urban form should ideally encompass all relevant travel flows, regardless of the transport mode. This study postulates that since urban and public transport planning are inseparable in the case Stockholm and public transport is the primary travel mode, public transport travel patterns facilitate the analysis of the urban structure. Notwithstanding, the analysis of single-mode flows can potentially introduce a bias due to the relations between destination, departure time and mode choice. For example, peak periods and centres that are more accessible by public transport may be overrepresented in passenger flows. Future studies may conduct an analysis based on the analysis of the integration of public and private transport modes. Furthermore, the analysis of directional data (e.g. GPS, smart cards) will enable the analysis of functional relations between urban centres (e.g. the formation of clusters of centres). A directional data will also allow to separate the intra-centre flow in order to find how much external flow each centre attracts, similarly to the concept of 'surplus of centre importance' used in Burger and Meijers (2012) analysis. Finally, investigating the evolution of urban structure over long periods will allow examining the trend and whether it is in line with the planning policies. Such an analysis could be compared to findings on the evolution in morphological and functional terms by Rodrigues da Silva et al. (2014) and Shon (2005), respectively.

Acknowledgements

The authors are grateful to Stockholm County Traffic Administration (SLL) for providing the data that enabled this study and in particular to Tengblad Kée and his colleagues.

References

Adolphson M. (2009). Estimating a polycentric urban structure. Case study: Urban changes in the Stockholm region 1991-2004. *Journal of Urban Planning and Development*, 135, 19-30.

Anas A., Arnott R. and Small K. (1998). Urban Spatial Structure. *Journal of Economic Literature*, 36 (3), 1426-1464.

Baum-Snow N. (2010). Changes in transportation infrastructure and commuting patterns in US metropolitan areas, 1960-2000. *The American Economic Review*, 100 (2), 378-382.

Burger M. and Meijers E. (2012). Form follows function? Linking morphological and functional polycentircity. *Urban Studies*, 49 (5), 1127-1149.

Burns M.C., Boix M. and Roca J. (2001). Contrasting indications of polycentricism within Spain's Metropolitan Urban Regions. *Eight European Estate Society Conference*, Spain.

Börjesson M., Jonsson D. and Lundberg M. (2013). An ex-post CBA for the Stockholm metro. *CTS* working paper 2013:34. Available at: http://econpapers.repec.org/paper/hhsctswps/2013_5f034.htm

Casello J.M. (2007). Transit competitiveness in polycentric metropolitan regions. *Transportation Research Part A*, 41 (1), 19-40.

Casello J.M. and Smith T.E. (2006). Transportation activity centers for urban transportation analysis. *Journal of Urban Planning and Development*, 132, 247-257.

Ceapa I., Smith C. and Capra L. (2012). Avoiding the crowds: Understanding tube station congestion patterns from trip data. *UrbComp – International Workshop on Urban Computing*, 134-141, Beijing.

Cervero R. (1995). Sustainable new towns: Stockholm's rail-served satellites. Cities, 12 (1), 41-51.

Davoudi S. (2003). Polycentricity in European Spatial Planning: From an Analytical Tool to a Normative Agenda. *European Planning Studies*, 979-999.

Derrible S. and Kennedy C. (2010). Characterizing metro networks: state, form and structure. *Transportation*, 37, 275-297.

EPSON (2004). Potentials for polycentric development in Europe. Project report.

Garcia-Palomares J.C. (2010). Urban sprawl and travel to work: the case of the metropolitan area of Madrid. *Journal of Transport geography*, 18, 197-213.

Giuliano G., Redfearn C. Agarwal A. and He S. (2012). Network accessibility and employment centres. *Urban Studies*, 49, 77-95.

Giuliano G. and Small K.A. (1991). Subcenters in the Los-Angeles Region. *Regional Science and Urban Economics*, 21, 163-182.

Gordon P., Richardson H.W. and Wong H.L. (1986). The distribution of population and employment in a polycentric city: The case of Los Angeles. *Environment and Planning A*, 18, 161-173.

Hall P. and Pain L. (2006). *The polycentric metropolis: Learning from mega-city regions in Europe*. London: Earthscan.

Hartigan J. A. (1975). *Clustering Algorithms*, Wiley, New York.

Hasan S., Schneider C.M., Ukkusuri S.V. and Gonzalez M.C. (2013). Spatiotemporal patterns of urban human mobility. *Journal of Statistical Physics*, 151 (1-2), 304-318.

Johnson G. (2012). Stockholm 2030. In: *The Future of Cities and Regions*, Eds. L. Bazzanella et al. Springer Geography, pp. 65-98.

Lee K., Jung W-S., Park J.S. and Choi M.Y. (2008). Statistical analysis of the Metropolitan Seoul subway system: Network structure and passenger flows. *Physica A*, 387 (24), 6231-6234.

Lin J. and Ban Y. (2014). The evolving network structure of US airline system during 1990-2010. *Physica A*, 410, 302-312.

Liu L., Hou A., Biderman A., Ratti C. and Chen J. (2009). Understanding individual and collective mobility patterns from smart card records. *Proceedings of the* 12th International IEEE Conference on Intelligent Transport Systems, St. Louis, USA.

Louf R. and Barthelemy M. (2013). Modeling the polycentric transition of cities. *Physical Review Letters*, 111.

McDonald J.F. (1987). The identification of urban employment subcenters. *Journal of Urban Economics*, 21, 242-258.

McMillen D.P. (2003). Identifying sub-centres using contiguity matrices. Urban Studies, 40 (1), 57-69.

McMillen D.P. and Smith S.C. (2003). The number of subcenters in large urban areas. *Journal of Urban Economics*, 53, 321-338.

Meijers E. (2008). Measuring polycentricity and its promises. *European Planning Studies*, 16 (9), 1313-1323.

Milligan G.W. (1980). An examination of the effect of six types of error perturbation on fifteen clustering algorithms. *Psychometrika*, 45(3), 325-342.

OECD (2013). Definition of functional urban areas for the OECD metropolitan database.

OMB – Office of Management and Budget (2000). Standards for defining metropolitan and micropolitan statistical areas. *Federal Register*, 65, 249, 82228-82238.

Ozus E., Akin D. and Ciftci M. (2012). Hierarchical cluster analysis of multicenter development and travel patterns in Istanbul. *Journal of Urban Planning and development*, 138(4), 303-318.

Riguelle F., Thomas I. and Verhetsel A. (2007). Measuring urban polycentrism: A European case study and its implications. *Journal of Economic Geography*, 1-23.

Rodrigues da Silva A.N, Manzato G.G. and Pereira H.T.S. (2014). Defining functional urban regions in Bahia, Brazil, using roadway coverage and population density variables. *Journal of Transport Geography*, 36, 79-88.

Roth C., Kang S.M., Batty M. and Barthelemy M. (2011). Structure of urban movements: Polycentric activity and entangled hierarchical flows. *PLoS ONE*, 6(1), e15923.

Schmitt P., Harbo L.G., Dis A.T. and Henriksson A. (2013). Urban Resilience and Polycenricity: The case of Stockholm Urban Agglomeration. In: *Resilience Thinking in Urban Planning*, A. Eraydin and T. Tasan-Kok (Eds.), 106, 197-209.

Schwanen T., Dieleman F.M. and Dijst M. (2001). Travel behaviour in Dutch monocentric and policentric urban systems. *Journal of Transport Geography*, 9, 173-186.

Shearmur R. and Coffey W.J. (2002). A tale of four cities: intrametrpolitan employment distribution in Toronto, Montreal, Vancouver and Ottawa-Hull, 1981-1996. *Environment and Planning A*, 34, 575-598.

Shon J. (2005). Are commuting patterns a good indicator of urban spatial structure? *Journal of Transport Geography*, 13, 306-317.

Stockholm City, the City Planning Administration (2011). The walkable city, Stockholm city plan.

SL – AB Storstockholms Lokaltrafik (2013). *Fakta om SL och länet 2012* [In Swedish – Facts on SL and the county 2012].

Tan P., Steinbach M. and Kuman V. (2006). Introduction to data mining. Addison-Wesley, Boston, MA.

Van der Laan L. (1998). Changing urban systems: An empirical analysis at two spatial levels. *Regional Studies*, 32 (3), 235-247.

Vasanen A. (2012). Functional polycentricity: Examining metropolitan spatial structure through the connectivity of urban sub-centres. *Urban Studies*, 49 (16), 3627-3644.

Veneri P. (2013). The identification of sub-centers in two Italian metropolitan areas: A functional approach. *Cities*, 177-185.

Walsh C. (2012). Territorial Agenda of the European Union 2020: Towards an Inclusive, Smart and Sustainable Europe of Diverse Regions. *Planning Theory and Practice*, 13 (3), 493-496.