

Article

Commuting Analysis of the Budapest Metropolitan Area Using Mobile Network Data

Gergő Pintér *  and Imre Felde 

John von Neumann Faculty of Informatics, Óbuda University, Bécsi út 96/B, 1034 Budapest, Hungary

* Correspondence: pinter.gergo@uni-obuda.hu

Abstract: The analysis of human movement patterns based on mobile network data makes it possible to examine a very large population cost-effectively and has led to several discoveries about human dynamics. However, the application of this data source is still not common practice. The goal of this study was to analyze the commuting tendencies of the Budapest Metropolitan Area using mobile network data as a case study and propose an automatized alternative approach to the current, questionnaire-based method, as commuting is predominantly analyzed by the census, which is performed only once in a decade in Hungary. To analyze commuting, the home and work locations of cell phone subscribers were determined based on their appearances during and outside working hours. The detected home locations of the subscribers were compared to census data at a settlement level. Then, the settlement and district level commuting tendencies were identified and compared to the findings of census-based sociological studies. It was found that the commuting analysis based on mobile network data strongly correlated with the census-based findings, even though home and work locations were estimated by statistical methods. All the examined aspects, including commuting from sectors of the agglomeration to the districts of Budapest and the age-group-based distribution of the commuters, showed that mobile network data could be an automatized, fast, cost-effective, and relatively accurate way of analyzing commuting, that could provide a powerful tool for sociologists interested in commuting.

Keywords: mobile network data; call detail records; data analysis; human mobility; urban mobility; social sensing; urban geography; urban sociology; commuting; sustainability



Citation: Pintér, G.; Felde, I.

Commuting Analysis of the Budapest Metropolitan Area Using Mobile Network Data. *ISPRS Int. J. Geo-Inf.* **2022**, *11*, 466. <https://doi.org/10.3390/ijgi11090466>

Academic Editors: Wolfgang Kainz and Hartwig H. Hochmair

Received: 20 May 2022

Accepted: 18 August 2022

Published: 29 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Hungary is a typical capital-oriented country. Budapest is the political, economic, logistical, and cultural center of the country, where almost 18% of the population live [1]. In 2017, the population of Budapest was 1,749,734, and the population of Pest county was 1,247,372. In the agglomeration of Budapest, 837,532 people lived according to the Hungarian Central Statistical Office (HCSO) [2]. The Danube river divides the city into the Buda and the Pest side.

Due to its central role, Budapest attracts a workforce from a relatively large area. This process establishes a contact between the capital and the surrounding settlements, called commuting. According to Kiss and Matyusz [3], commuting is the relation between two locations. The inhabitants of one source location travel to work/study to another location; this is called “out-commuting”. The target location receives a workforce that is called “in-commuting”. The loss of the source location follows from the fact that (i) the out-commuter does not use the local resources, (ii) does not create value, (iii) loosens their relation with the source location, as it is partly relocated to the target location, and (iv) although it brings back income, that is partly spent elsewhere. On the other hand, the target location (i) gains human resources, (ii) can create more value in place, (iii) the local relations and society become stronger, and (v) the local consumption increases.

Kiss and Matyusz state that, although commuting is an important and common phenomenon, its measurement by questionnaires (e.g., [4]) is occasional and inadequate to understand the detailed tendencies [3]. Commuting is analyzed by census data (e.g., [3,5–10]) in detail, but that is performed only once in a decade (in Hungary) and thus cannot follow sudden but permanent changes. For these reasons, commuting should be examined more continuously. This paper aims to provide a methodological overview for this task using mobile network data and at the same time, a case study about the commuting in the Budapest Metropolitan Area.

As questioning the population is a slow, tedious and expensive task, it would be obvious to automate the process with the available infocommunication technologies (ICT). In this study, the application of call detail record (CDR) processing is proposed to examine commuting, and the findings are validated by the results of studies that analyzed commuting using census data. In many cases, the findings are presented in a form as close to those results as possible to aid the comparison.

This paper is closely related to the authors' previous works. In [11], the home and work detection framework was introduced with some macrolevel commuting analysis as a validation. However, in this study, the commuting tendencies of the Budapest Metropolitan Area are analyzed in more detail. In [12], the temporal differences in the mobile network activity between the regions of Budapest are demonstrated, along with the typical times when a group of subscribers "wakes up" from a mobile network perspective. Some analyses regarding the length of the working hours are also presented.

The contributions of this paper are briefly summarized as follows:

1. Using anonymized mobile network data, the commuting tendencies of the Budapest Metropolitan Area are analyzed.
2. The settlement level commuting trends are compared to the former census-based commuting studies.
3. The scripts and queries used for the study have been made available.

The rest of this paper is organized as follows. First, we provide a brief literature review in Section 2. The utilized data and the methodology are described in Section 3; then, in Section 4, the results of this case study are discussed. Finally, in Section 5, the findings of the paper are summarized.

2. Literature Review

Identifying the home and work locations of a cell phone subscriber is a common and crucial part of the CDR processing, as a good portion of the people live their lives in an area that is determined by only their home and workplace [13,14]. Since these locations fundamentally determine people's mobility customs, the commuting trends can be analyzed between these locations. Commuting is studied using mobile network data within a city [14,15] or between cities [16–19] and also examined by social network data, such as Twitter [20–22].

Several approaches are known to determine the home location. Vanhoof et al. compared five different home detection algorithms (HDA), selecting the home cell by (i) the most activity, (ii) the greatest number of distinct days with phone activities, (iii) the most activities within a time interval (between 19:00 and 7:00), (iv) the most activities within a spatial perimeter, and (v) the combination of the temporal and spatial constraints [23]. The method shown in this paper uses the most activities within a time interval (Section 3.3).

While Vanhoof et al. compared different algorithms, Pappalardo et al. [24] compared different types of mobile network data (CDR, extended detail record (XDR) and control plane record (CPR)) to estimate the home locations. Furthermore, they validated the estimated home locations of sixty-five subscribers with the known geographical coordinates of their residence location. They found that XDRs should be preferred when performing home location detection.

Csáji et al. determined the subscribers' most common locations and, based on weekly calling patterns, identified the home and work locations, which showed a strong correlation

with population statistics [25]. They also found that the commuting distances could be well explained by a gravity model.

Besides the home and work locations—and the distance between them—other indicators are used to measure the mobility of subscribers. One of the most popular one, the radius of gyration, defines an area of a circle where an individual is usually located [26]. Pappalardo et al. used the radius of gyration to separate subscribers based on their mobility customs and defined two classes: returners and explorers [13]. While in the case of returners, the radius of gyration was dominated by their movement between a few preferred locations, explorers tended to travel between a larger number of different locations. To demonstrate this dichotomy, they defined the k -radius of gyration, which refers to the gyration radius of the k most frequent locations. The gyration radius of a two-returner is determined by the two most frequented locations, usually the home and work locations [13], so this method can also be used as a home detection algorithm.

Xu et al. determined the home locations and then applied a modified standard distance to measure the spread of each subscriber's activity space [27]. These activity spaces were broadly similar in concept to the radius of gyration, but the center was the home location, instead of the locations' center of mass.

As a part of the Data for Development (D4D) challenge, multiple mobile network datasets from the Ivory Coast (and later Senegal) were made available for research. "The goal of the challenge was to help address society development questions in novel ways by contributing to the socioeconomic development and well-being of the Ivory Coast population" [28]. The papers presented at the resulting conference (NetMob 2013) dealt with tracking mobility, economic and social activity, and epidemics [29]. Besides the scientific benefits, ethical and privacy concerns were also raised [29]. Using the same data sets, Dong et al. measured the efficiency of the transportation networks [30], and Šćepanović et al. analyzed the commuting patterns in the Ivory Coast and used them as a measure of poverty [31].

Jiang et al. identified daily activity patterns (motifs) that could extend the home-work location-based daily routine [14]; the home locations were validated with census and household travel survey results. Shi et al. used a kernel density map approach, from which they distinguished three types of community patterns: single-centered, dual-centered, and zonal patterns. Yin et al. separated the different types of activity (home, work, leisure, and school) with chains of activity [32], providing different approaches for a similar purpose. Diao et al. applied a regression model to travel survey data to predict the activity type (e.g., home, work, or social) of the mobile phone location data by considering the temporal distributions of different activities [33].

Mamei et al. computed origin–destination flows with road network mapping and also validated the home location estimation with census data [17]. Sakamane et al. inferred the street-level route choice of the subscribers for modeling transportation [34]. After the home and work locations were estimated, a set of potential route choices were obtained from the Google Maps Directions API, including choices of car and public transport. As opposed to [17,34], this study did not take into consideration the means of transport, only the origin–destination pairs.

Dannemann et al. [35] partitioned the city of Santiago (Chile) into several communities and identified the socioeconomic composition of these communities based on the home–work trajectories. Hadachi et al. reconstructed trajectories and extracted the origin–destination matrix to analyze commuting patterns [36]. They validated the detected homes, which showed a strong correlation between the home registration data, but not the commuting patterns.

Burgdorf et al. analyzed long-distance (over 50 km) passenger transport in Germany, including the means of transport [18]. They used mobile network data as an alternative to the traditional approach, using official statistical data, and household and roadside surveys. Burgdorf et al. found that the mobile network approach was broadly consistent

with the reference for the overall values, but there were significant deviations on the level of origin–destination pairs [18].

Hadachi et al. examined commuting to Tallinn, Estonia from the neighboring municipalities [36]. Šveda and Barlík also analyzed the commuting flows from the municipalities to Bratislava, Slovakia. However, these studies did not validate the commuting ratios with a census.

Pálóczi analyzed the country-wide commuting in Hungary, using the methods of complex network analysis, based on the data from the 2011 census [6]. He considered settlements as nodes, and commuters as a directed edge between the nodes, then applied the disparity (Y_2) parameter [37], which was developed to measure the heterogeneity of weighted relations. If Y_2 is close to one, it means that one destination dominates commuting. Pálóczi demonstrated that the out-commuting dependency was not the greatest around Budapest, but around Győr and Székesfehérvár [6], and it was also high at the centers of the employment regions (e.g., chief towns of the counties).

Commuting is traditionally analyzed by census data or directed questionnaires. Questioning the population requires fieldwork, which is a slow, tedious, and expensive task. Therefore, it is only performed occasionally, although more frequent data could help to optimize the public transportation system, or monitor the structure of the society, for example. Using infocommunication technologies such as the mobile network can provide a more cost-effective, frequent, and countrywide analysis of commuting.

Detecting home locations by mobile network data is a common practice in the literature. Different HDAs have been described and compared, also using different types of mobile network data. The detected home locations are often validated by official population data. With workplace detection, commuting has been analyzed on a smaller and larger scale as well. In this study, not only the home and workplace detection is validated via census data, but also the settlement-level commuting tendencies themselves. In Hungary, commuting is predominantly analyzed by census data, which is used as a reference to validate the CDR-based results.

3. Data and Methods

Vodafone Hungary, one of the three mobile phone operators providing services in Hungary with a 25.5% market share in 2017 Q2 [38], provided an anonymized mobile network data set for this study. The observation period was one month, April 2017. The observation area was Budapest, the capital of Hungary, and its agglomeration, which is illustrated in Figure 1a in relation to Pest county. The areas of the capital and the agglomeration are 525 km², and 2538 km², respectively. Budapest is divided into seven district groups by the HCSO referenced later in this study and shown in Figure 1b.

The mobile network, during its operation, constantly communicates with cell phones. This communication can be divided into two categories: (i) the passive, cell-switching communication that keeps the cellphones ready to use the mobile network at any time, and (ii) the active, billed usage of the mobile network including phone calls, text messages or mobile internet usage. Therefore, the call detail records (CDRs) are collected for billing purposes and contain information about the subscriber, the time of the activity, and the place (via the cell), where the activity occurred, although the records include neither the type of the activity (voice call, message, data transfer) nor the direction (incoming, outgoing).

The data set contained 955,035,169 CDRs collected by the operator for billing purposes, from 1,629,275 SIM cards. Figure 2 shows the mobile network activity distribution between the activity categories of the SIM cards. The left part of the figure displays the number of CDRs that a group of SIM cards generated. The right part of the figure shows the number of SIM cards in that group. Only 17.67% of the SIM cards generating more than 1000 CDRs provided the majority (75.48%) of the mobile phone activity during the observation period. Figure 3 has the same structure and shows the distribution of the SIM cards by the number of active days. Only about one-third (33.23%) of the SIM cards have activity on at least 21 different days. Despite the relatively large number of SIM cards present in the data,

most of them were not active enough to provide enough information about their users' mobility habits.

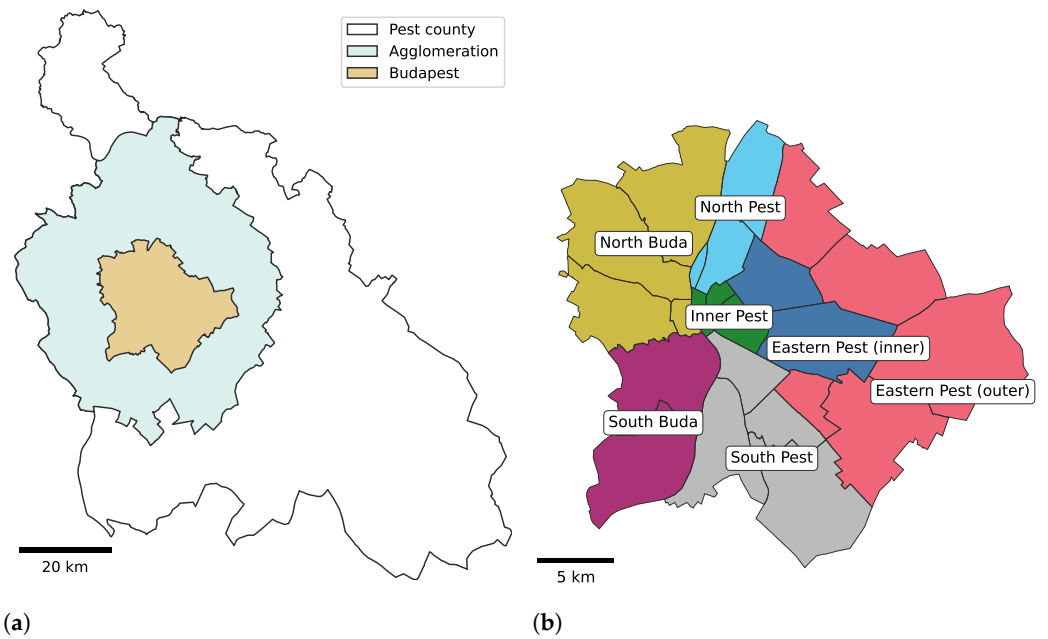


Figure 1. The observation area (a) including Budapest (brown), and its agglomeration (green) in relation to Pest county, and the district groups (b) of Budapest, defined by the Hungarian Central Statistical Office (HCSO).

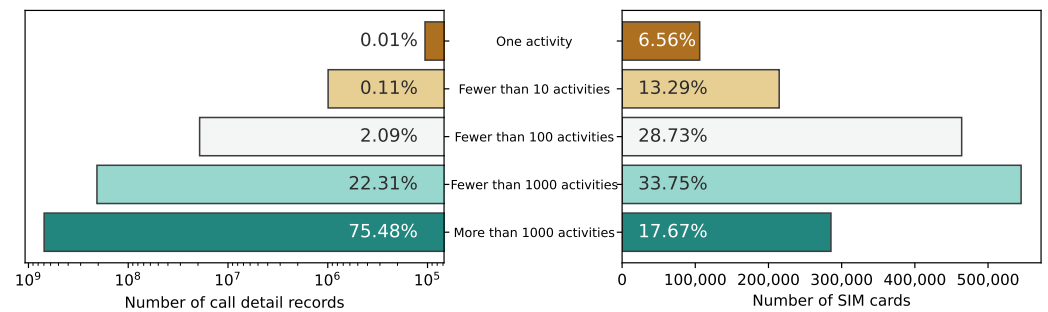


Figure 2. The SIM cards in the “April 2017” data set categorized by the number of CDRs. The left figure shows the number of CDRs and the right figure shows the number of SIM cards in each category. The SIM cards over a thousand records (17.7%) provide the majority (75.48%) of the activity.

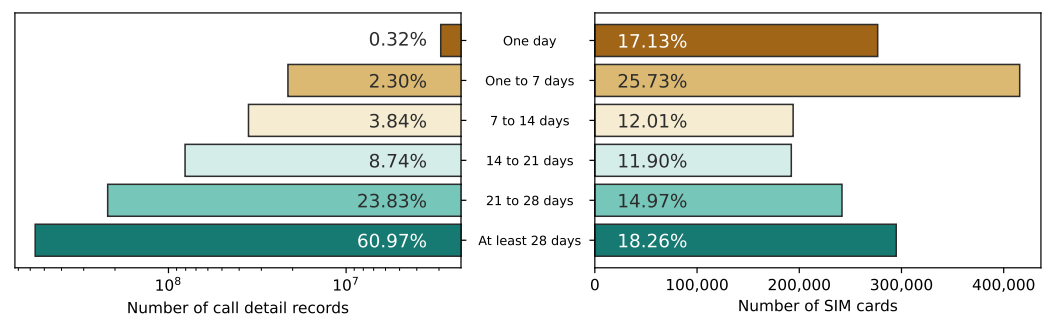
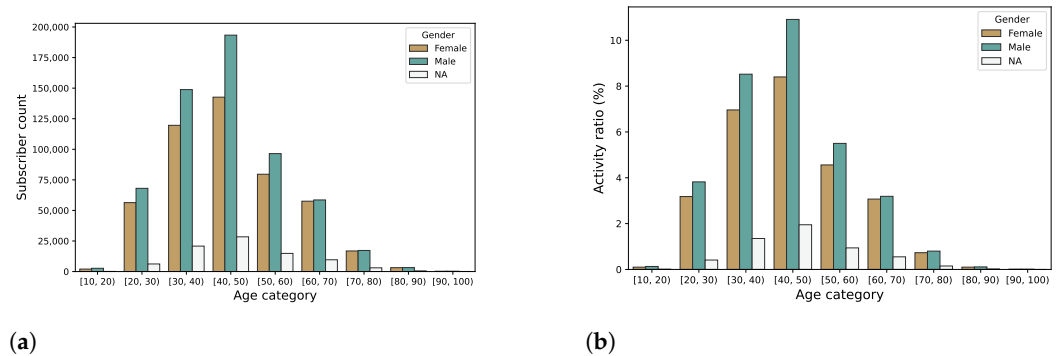


Figure 3. SIM card distribution of the “April 2017” data set, by the number of active days. The left figure shows the number of CDRs and the right figure shows the number of SIM cards in each category. Around 45.13% of the SIM cards have activity on at least 14 different days.

The provided mobile network data included information about the age and the gender of the cell phone subscribers in about 66.17% and 70.76% of the subscribers, respectively. Figure 4a shows the number of subscribers by age category and gender. Figure 4b shows the ratio of the generated CDRs of subscribers by age category and gender. Note that the subscribers without age information are omitted from the figures. Most of the Vodafone subscriptions—with information about the subscriber age—were owned by males in their forties. It is also notable in Figure 4b that most of the mobile network activity is generated by devices owned by subscribers in their thirties or forties.



(a) (b)
Figure 4. The number of cell phone subscribers (a), and the ratio of all mobile network activity (b) by age category, when the subscriber age is known.

3.1. Data Engineering

The high-level data processing workflow is illustrated in Figure 5. Our previous works [11,12,39,40] also relied on the first five stages. The cleaning stage simplified the provided information (Figure 5a). Originally, the device (or SIM) IDs were represented as long alphanumeric hashes, which were replaced by ordered numbers in order of appearance, partly to save disk space and partly because an integer comparison is faster.

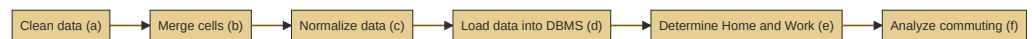


Figure 5. Workflow of the data processing.

The obtained data used property labels such as “MALE” and “FEMALE” to indicate gender, “CONSUMER” and “BUSINESS” for customer type, “PREPAID” and “POSTPAID” for subscription type, and the “UNKNOWN” string was used to denote unknown values. These strings were shortened, and the unknown values were represented with the proper “NULL” value to save disc space. The longitude and latitude values, provided in EPSG:4326 projection (also known as WGS 84) were rounded to 6 decimals because further decimals have no practical meaning in CDR positioning. The TAC is the first eight digits of the International Mobile Equipment Identity (IMEI) number that can identify the manufacturer and the model of the device wherein the SIM card is active. The TAC value was utilized in [12,40].

The cell density is proportional to the population density and inversely proportional to the cell sizes, which means that there are more and smaller cells downtown. There were both 2G and 3G cells active during the observation period. In 2017 Q2, the 4G cellular network technology was marginal in Hungary, and the 3G technology served 86.6% of the mobile phone traffic [38]. The overlapping technologies could also cause very close cell centroids (Figure 6a).

The close cells within 100 m were merged (Figure 5b) using the DBSCAN algorithm of the Scikit-learn [41] Python package. The number of records associated with a cell during the observation period was applied as weight. The less-active cells were merged into the more active ones similarly, as in [15]. After the merge, the Voronoi tessellation was applied to the merged cell centroids. Figure 6 shows the difference between the Voronoi polygons

before (Figure 6b) and after (Figure 6c) the cell merge. The merged cell polygons became larger on average and roughly covered a house block downtown.

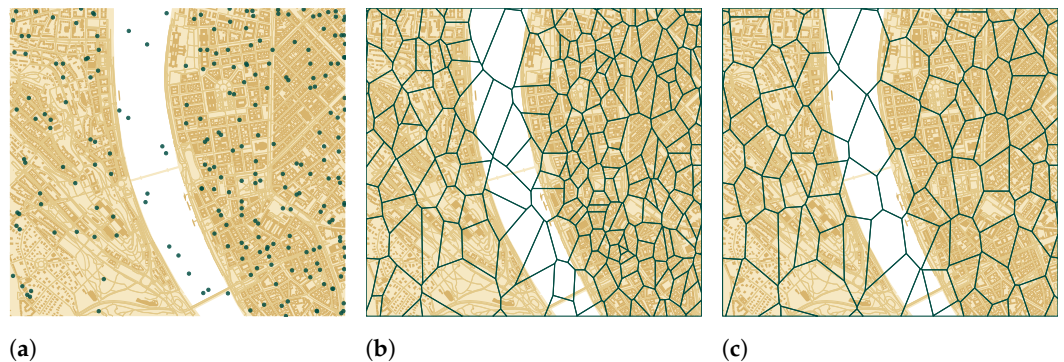


Figure 6. Mixed cell centroid (a), Voronoi polygons using the original cells (b), and Voronoi polygons after merging the close cells (c).

The obtained, “raw” data had a wide format (Figure 7a), which was normalized (Figure 5c) before importing into the database. The CDR table only contained the SIM ID, the timestamp, and the cell ID. A table was introduced to store the SIM-related properties (such as subscription type, customer type, age, gender) and another to store the cell properties (geographic coordinates). Figure 7b displays the normalized schema.

The rationale for the wide format may be that the subscriber data and the device can be changed within the observation period. This occurred about 3000 times during the observation period. The owner of the subscription can change its details and, of course, change their device if they bought a new mobile phone, for example. Subscriber and customer type were provided for every SIM, but age and gender were missing in many cases, presumably due to the privacy options requested by the subscriber. The age was available for 66.17%, and the gender was provided for 70.76% of the subscribers.

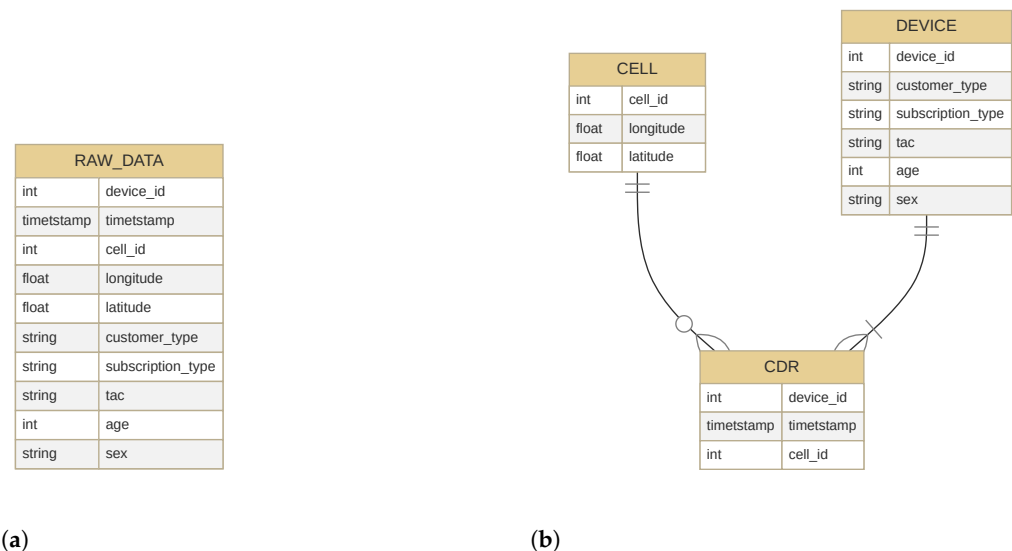


Figure 7. The schema of obtained, “raw” data (a), and the normalized data tables (b).

The normalized data were loaded into a PostgreSQL database (Figure 5d) using the schema of Figure 7b. Then, indexes were built for all columns of the database tables. The database management system is an important part of the data processing framework. For example, the first part of the home and work location estimation (Section 3.3) was implemented as an SQL query. The PostGIS extension was also used, which added support for geographic objects allowing location queries to be run in SQL.

3.2. Methodology

The mobile network data processing framework was introduced in the authors' previous work [11], along with the home and work detection method. The queries and scripts applied in this work have been published in a public repository [42], which contains the home/workplace detection queries and the plotting scripts. Although the mobile network data used in the study are not publicly available due to third-party restrictions, dummy data were generated by a (pseudo)random generator (also published) making the code executable. This also means that the figures in the repository are not valid, as only the dummy data were used as an input. The scripts have been published in a Jupyter Notebook format on GitHub (https://github.com/pintergreg/commuting_analysis, accessed on 11 August 2022).

In order to analyze commuting, the home and the work locations of the subscribers had to be determined (Section 3.3). As the CDRs were anonymized, they did not contain information such as a residential address. Even the workplace is not mandatory for a subscription, and the operators cannot have that information. After these locations were determined, they had to be validated. Although the applied approach was practically equivalent to what can be found in the literature, the validity of the results was hard to confirm. Pappalardo et al. validated the home locations in the case of sixty-five subscribers whose home addresses were known [24]. However, this was not possible in our case as no subscribers contributed their home addresses for this study. As a result, the settlement and—in the case of Budapest—district-based population data [2] were applied from the HCSO. The mobile-network-based results (e.g., detected population, detected settlement-level commuting tendencies) were compared to the ground truth (census-based values and studies based on the Hungarian census values) using standard statistical tools such as linear regression or Pearson's correlation coefficient.

Kernel density estimation (KDE) was used to analyze the most common target area where the inhabitants of a settlement commute. After the origin–destination matrix (home and workplaces) was determined, a settlement could be selected as an origin and the work locations (cell centroids) could be passed to the KDE algorithm. This analysis was performed in two ways: with and without the local workers, who worked in the same settlement where they lived, as local workers were not relevant for the commuting analysis.

The commuting between the districts of Budapest was also treated as a network, where the districts were nodes, and an edge represented a commuter that connected two districts. The network-based commuting or traffic flow analysis also has a long history (e.g., [43,44]). The NetworkX Python package [45] was used to build a graph, which was visualized by the chord diagram of the "mne_connectivity" package.

The commuting trends were also analyzed regarding the age groups. In [9], the commuting from the urban agglomeration to Budapest was presented using census data. Along with the obtained subscriber age information, the age-group-based analysis was performed on the mobile network data.

3.3. Home and Work Locations

Most of the inhabitants in cities spend a significant time each day at two locations: their home and workplace. To find the relationship between these most important locations and the social economic status (SES), first, the positions of these locations (cells) have to be determined. There are a few approaches used to find home locations via a mobile phone data analysis [27,46,47].

The work location was determined as the most frequent cell where a device was present during working hours, on workdays. Working hours were considered from 09:00 to 16:00. The home location was calculated as the most frequent cell where a device was present during the evening and at night on workdays (from 22:00 to 06:00) and all day on holidays. Although people do not always stay at home on the weekends, it was assumed that most of the activity was still generated from their home locations.

This method assumed that everyone worked during the daytime and rested in the evening. Although in 2017, 6.2% of the employed persons regularly worked at night in Hungary [48], the current version of the algorithm did not try to deal with night-workers. Some of them might be identified as regular workers but with mixed home and work locations.

Students, especially university students, might also be identified as workers with the university as a workplace. From a commuting perspective, this does not cause any problem, as from a transportation perspective, there is no difference between a worker and a student. Children usually do not own a subscription. Even if they have a cellphone, it might be registered to their parents. If these devices have enough mobile network activity, they might be identified as workers.

As only one workplace was selected (the most frequent cell during working hours), people with multiple workplaces or indefinite workplaces (e.g., couriers, taxi drivers, or letter carriers) might have an incorrect workplace detected. Note that this approach corresponds to common practice.

It also has to be noted that home and work locations could not be detected for SIM cards with very few occurrences during the observation period—in other words with not enough mobile network activity. SIM cards without home and workplaces were omitted from the commuting analysis. 68.08% of the SIM cards had a detected home and workplace.

4. Results and Discussion

Figure 8 shows a comparison between the population registered by the HCSO and the detected population using CDR data. When comparing the two maps, there are a few differences: some parts of Budapest are not dark enough on the CDR data. For example, Districts 16 and 17 seem not as populated as in the HCSO data, but the divergent districts are on the Buda side: Districts 2 and 3. The difference may have ensued from the inaccuracy of the home detection in that area or simply from the different preferences in mobile operators of the inhabitants of the Buda Hills.

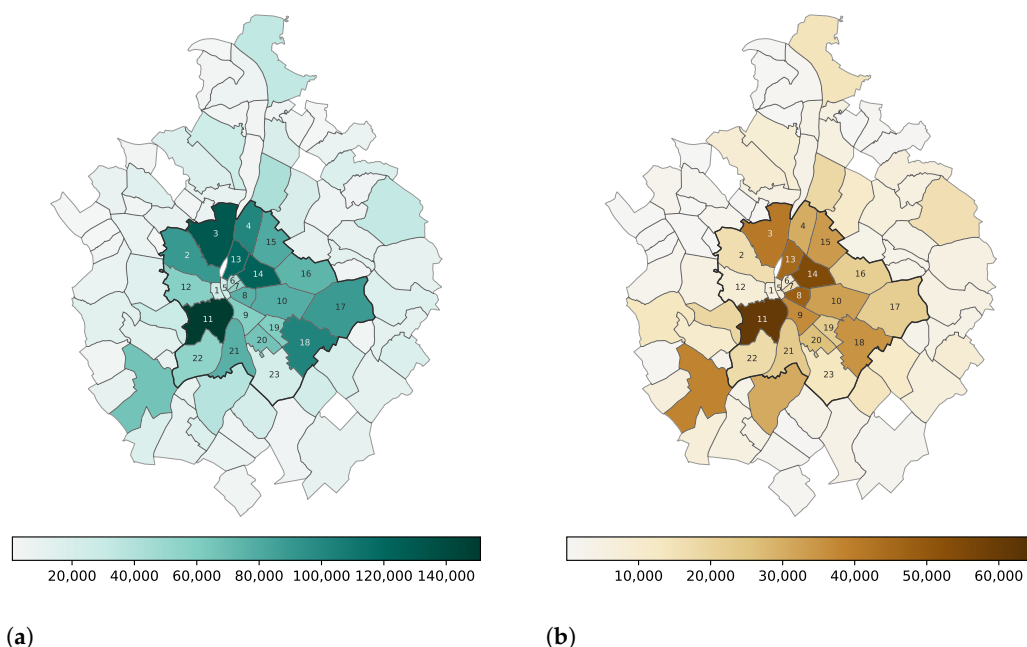


Figure 8. Comparing the ground truth (the population registered by the HCSO [2]) (a) and the detected population based on mobile network data (b). The district numbers of Budapest are also displayed.

Apart from this, the findings based on CDRs correlated with the statistical data; the Pearson's R was 0.9213 (R^2 was 0.8488), counting every SIM cards. Other studies

found similarly strong correlation values between population statistics and the detected homes [36,49]. Figure 9a shows this as a regression plot. As described in the authors' previous work [40], the obtained call detail records contained information (TAC) about the devices that used the mobile network. Based on this information, more than 300,000 SIM cards were identified that operated other types of devices than cellphones (e.g., 3G modem), indicating that they did not represent people. Figure 9b illustrates the correlation without these SIM cards. Although the population values were decreased, the correlation did not seem significantly affected: Pearson's R was 0.9125.

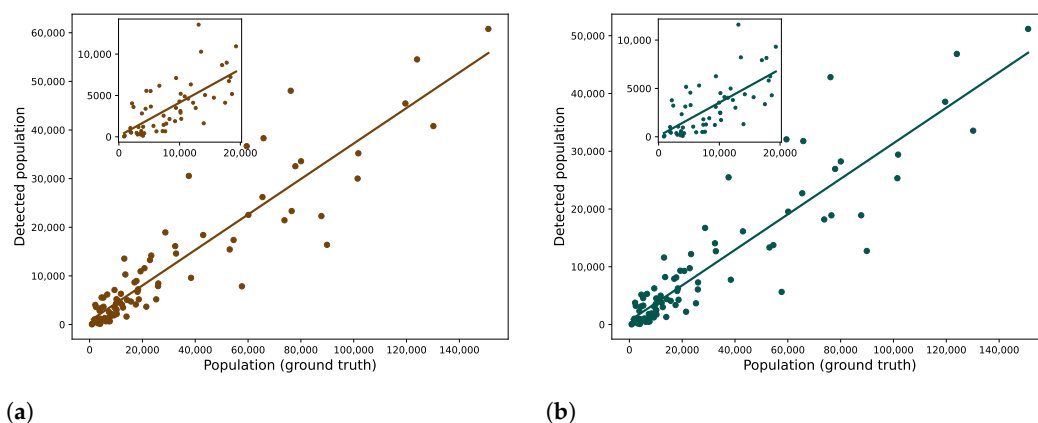


Figure 9. Correlation between the population of the agglomeration and the 23 districts of Budapest based on HCSO (ground truth) and mobile network data. The markers represent the settlements or districts (in the case of Budapest). In left figure (a), all the SIM cards were used, in the right figure (b), SIM cards that certainly operated in nonphone devices were excluded.

Numerically, the CDR data often showed significant mismatches, but they were not easy to objectively compare. The available mobile network data originated only from one operator, which had about 25% market share in the observation period [38]. This market share is about the subscriptions, not the number of unique people. Furthermore, it also has to be noted that this ratio represents a nationwide value. Since spatially, a more detailed market share was not available, it was assumed that Vodafone Hungary had the same market share in every subregion to make this comparison. Although this is unlikely, one-fourth of the population values can be used as a rule of thumb.

4.1. Work Locations

Along with the home locations, the workplaces are the most important element of the mobility and commuting analysis. During the COVID-19 pandemic, this changed. As part of the social distancing directive, to slow down the spread of the disease, working from home came to the fore. Presumably, the prevalence of a home office will be higher than it was before the pandemic, as both the employers and the employees became used to this situation, but many scopes of activities will still require a work location, so the importance of this topic will remain the same. However, as the data sources used in this work predated the pandemic, this question can only be answered in another work.

The workplaces were determined by considering the most frequent place where a subscriber appears during work hours. See Section 3.3 for the details. For example, querying the work locations of the inhabitants of an area, a settlement can be the initial step of the commuting analysis. Figure 10 shows the typical workplaces of three selected settlements and one district of Budapest, using Gaussian kernel density plots, in two different versions: with (left column) and without (right column) those subscribers who work in their home settlement. When the local workers are included, the darkest areas are within the selected area itself, as many people work in the vicinity of their homes. Excluding the local workers, Budapest becomes the main target for commuting.

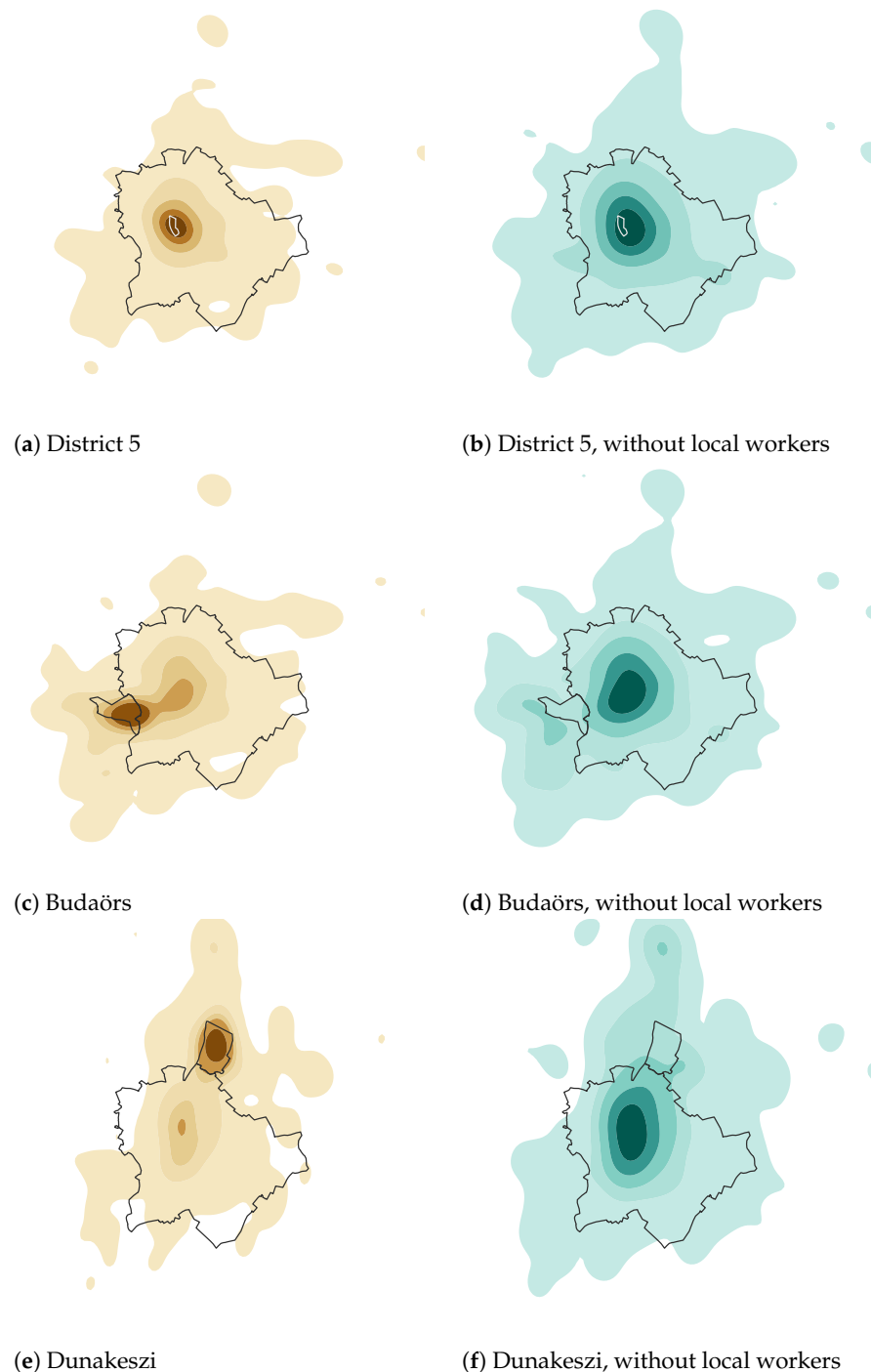


Figure 10. Using kernel density plots (with Gaussian kernel) to display the typical working locations for three selected settlements and a district of Budapest, with (a,c,e) and without (b,d,f) local workers. The administrative boundaries of Budapest and the selected settlements are also displayed.

4.2. Commuting between Districts

As the origin and the destination of the commuting are determined, it is possible to build a network, for example, considering the districts of Budapest nodes that are connected by the commuters. Figure 11 shows the connections between the districts of Budapest. The Buda districts are placed to the left, whereas the Pest districts are to the right, and the colors of the nodes represent the district groups defined by HCSO [1] and match the colors of Figure 1b. The edges represent commuters between districts, removing self-links, and the weight of the edges denotes the number of commuters. The weight is expressed by colors,

using darker colors for the stronger edges. The weakest links ($w < 250$) are omitted to improve visibility.

Extending this topic to the level of the agglomeration, or the country, could be another research direction: for example, to analyze the in-commuting and out-commuting. Pálóczi's work [6] could serve as a census-based reference.

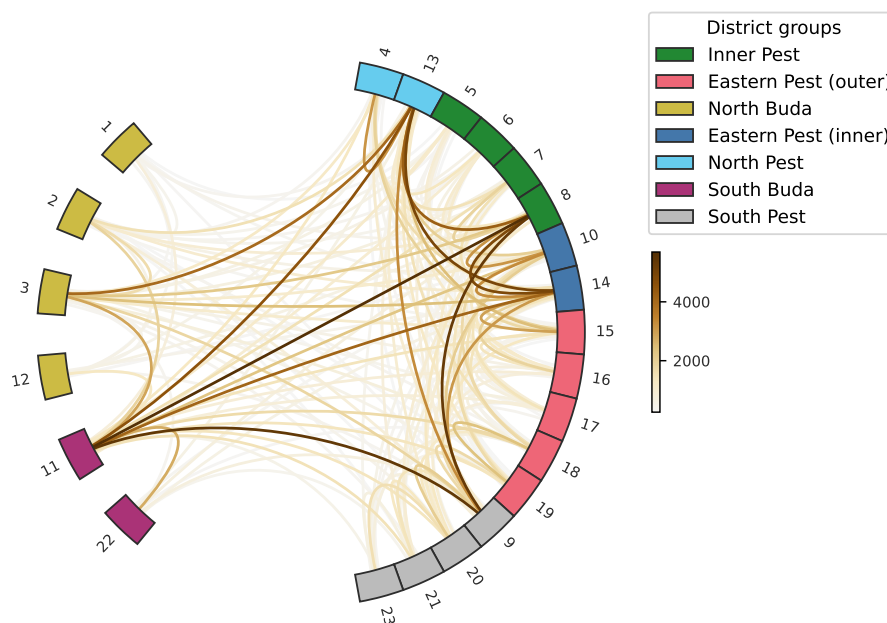


Figure 11. Connection by commuters between Budapest districts (numbered nodes) based on the home and work locations. A link represents how strong the commuting between two districts is. The weak links are omitted to improve visibility. The district nodes are colored by the district groups.

4.3. Validation by Census

In order to verify the reliability and accuracy of the method proposed for the home and work location estimation, a comparative study was performed on the mobile network data and the information processed from the census. In Hungary, a census is obtained every ten years and a microcensus with a 10% corpus at halftime. The last census was performed in 2011, while the last microcensus was in 2016. The next census should have been performed in 2021, but it was postponed until October 2022 due to the COVID-19 pandemic. Based on these surveys, commuting to Budapest (and generally in Hungary) is analyzed in studies such as [3,6,7,9]. These studies were used as a reference for comparing the results.

Figure 12 shows the comparison between the CDR and the census-based (Figure 1 in [9]) traveling ratios of the commuters by the districts of Budapest and the home location category. People who work in Budapest are represented, and the home location can be (i) the same district where one works, (ii) another district of Budapest, (iii) the agglomeration, and (iv) other settlements outside the agglomeration.

A good agreement mostly within Budapest was found on the proportions of the commuters. The most significant difference can be seen with the “outside agglomeration” category. This deviation, however, originated from the content of the data source, as the mobile network data used in this study covered mainly the area of Budapest and its close vicinity. It also contained phone activities from the surrounding county, but by moving away from Budapest, the available data decreased.

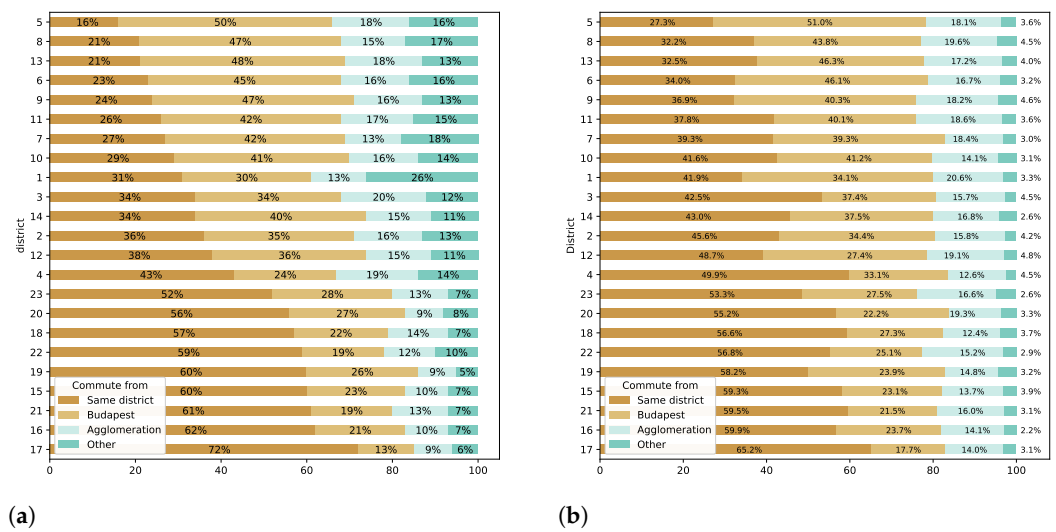


Figure 12. Comparison between the census-based (a) (Figure 1 in [9]) and the CDR (b) commuting ratios for the districts of Budapest, from the same district, other parts of Budapest, the agglomeration or out of the agglomeration.

The fraction of workers who have their homes in the same district was very close to that of the census data in the outer districts (15–23) but generally overestimated in the core districts (1, 5–9) and the inner districts (2–4, 10–14). The workers from other district groups showed the best match to census data (where the CDR should have the best quality), while the agglomeration was somewhat overestimated in many districts.

4.4. The Urban Agglomeration

The validation presented in this study used the results of [7], where a detailed analysis was presented in regard to commuting from the urban agglomeration to Budapest. The urban agglomeration was divided into six sectors, and the commuting was examined by origin (home sector, occasionally by towns) and destination (district group of Budapest).

Figure 13 shows the commuters’ distribution in the districts of Budapest, from the six sectors of the agglomeration, based on the CDR evaluation. In representation, Figure 13a–f are analogous to Figures 2–9 in [7], and show to which districts the inhabitants commute from the given sector of the agglomeration.

Lakatos and Kapitány analyzed the commuting tendencies of some settlements to the districts of Budapest among censuses from 1990, 2001, and 2011 [7]. The same analysis was conducted using CDR processing, and 6 settlements of the 13 thoroughly analyzed in [7] are presented in this study. The results are summarized in Figure 14, compared with the last three censuses. It contains a settlement from every sector of the agglomeration, so it also serves as a more focused analysis of Figure 13. The location of the settlements in relation to Budapest is also displayed on small maps to give context to the findings. From towns west to the capital, the most common commuting targets are the Buda-side and the inner districts, for example. Moreover, in many cases, the mobile-network-based findings, which are six years older than the last census, indicate a clear continuation of the previous tendency.

In the case of Budaörs (Figure 14a), albeit North Pest, outer Eastern Pest, and South Pest were not significant commuting destinations, census data showed an increasing tendency, which was confirmed by the mobile network data. The CDR-based results of South Buda and Inner Pest also fitted the trend, but in an opposite tendency. The most considerable discrepancy lay in the cases of North Buda and the inner Eastern Pest district groups. The Pearson correlation coefficient, regarding all the six district groups, between the 2011 Census and the mobile network data was 0.8976.

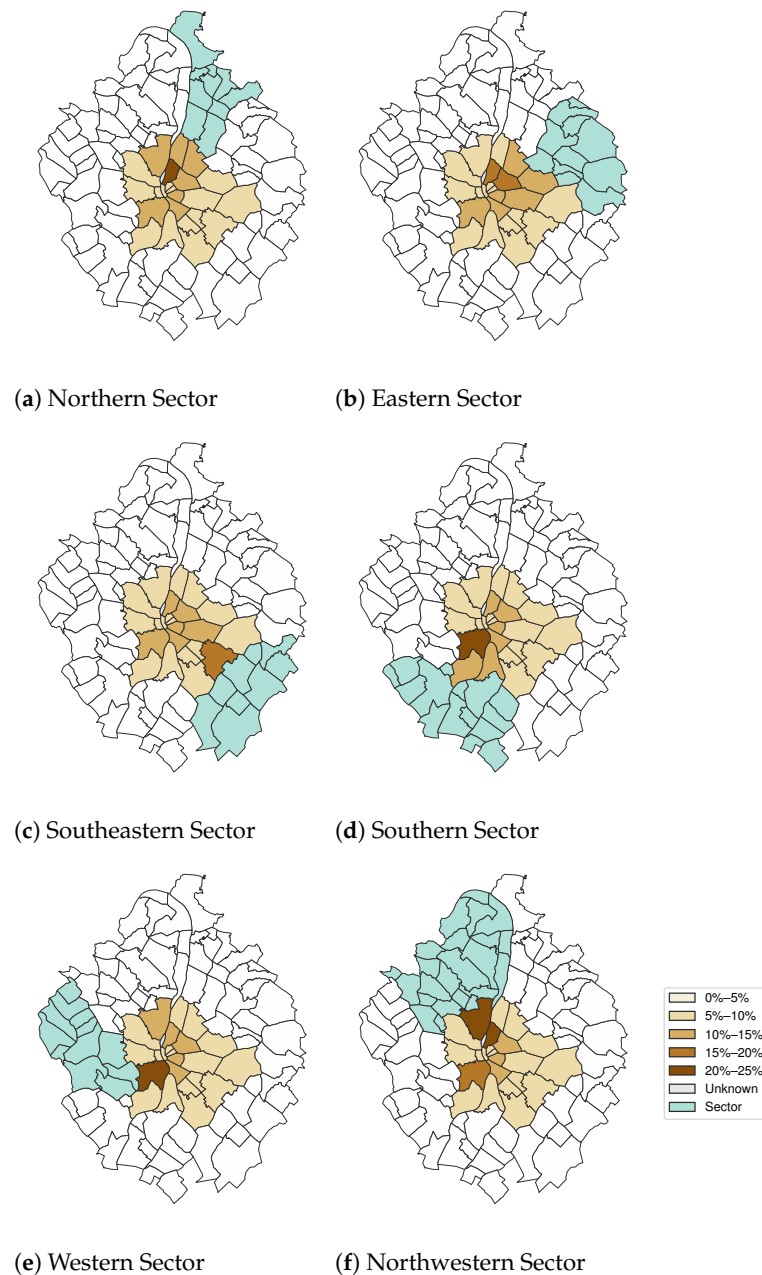


Figure 13. Commuting from the six sectors of the urban agglomeration, based on CDR evaluation.

Dunakeszi (Figure 14b) is east of the Danube river and north of Budapest, which implies the dominance of North Pest as the commuting destination, although its importance has been decreasing over the last few decades, as well as Inner Pest. While South Buda, South Pest, and the outer Eastern Pest had an increasing tendency, the inner Eastern Pest and North Buda did not show such clear tendencies. The correlation coefficient (Pearson's R) between the 2011 Census and the CDR based results was 0.9416.

Vecsés is in the southeastern sector of the agglomeration, from where the majority of the commuters work in the inner and outer Eastern Pest, Inner Pest, and South Pest regions. North Pest and Buda was not a notable destination for the commuters, but the results showed increasing trends (Figure 14c). The correlation coefficient, in the case of Vecsés, was 0.924.

Dunaharaszti is in the Southern sector of the agglomeration and east of the Danube. Consequently, the main destination of the commuters was South Pest. Moreover, Inner Pest received a considerable number of in-commuters, but its importance seemed to be

decreasing. The rest of the district groups had roughly the same trends (Figure 14d). Dunaharaszti had the strongest correlation out of the examined settlements: Pearson’s R was 0.971.

Érd has the largest population (65,857 in 2017 [2]) in the agglomeration and also in the Southern sector. The detected commuting ratios fitted into the trends of the last three censuses, although Eastern and South Pest seemed overestimated, and North Buda underestimated by the mobile-network-data-based approach (Figure 14e). The correlation with the ground truth was 0.8488 (Pearson’s R).

In the case of Szentendre (Figure 14f), the mobile-network-based results might show the most significant discrepancy. Still, the correlation coefficient (Pearson’s R) was 0.9127. Located in the Northwestern sector, west of the Danube, the most obvious destination for commuting is North Buda. According to the census data, it had the most in-commuters, even with a slightly increasing tendency. However, the CDR-based results underestimated it, whereas Eastern and South Pest seemed overestimated. The result of Inner Pest lagged behind that of the latest census data, but that fitted into the trend.

These detailed results demonstrate the applicability of the CDR processing for commuting analysis. It would be interesting to compare these results with the next census. That would reveal how precisely these findings fit into the trend of the changing commuting customs of the population of the agglomeration.

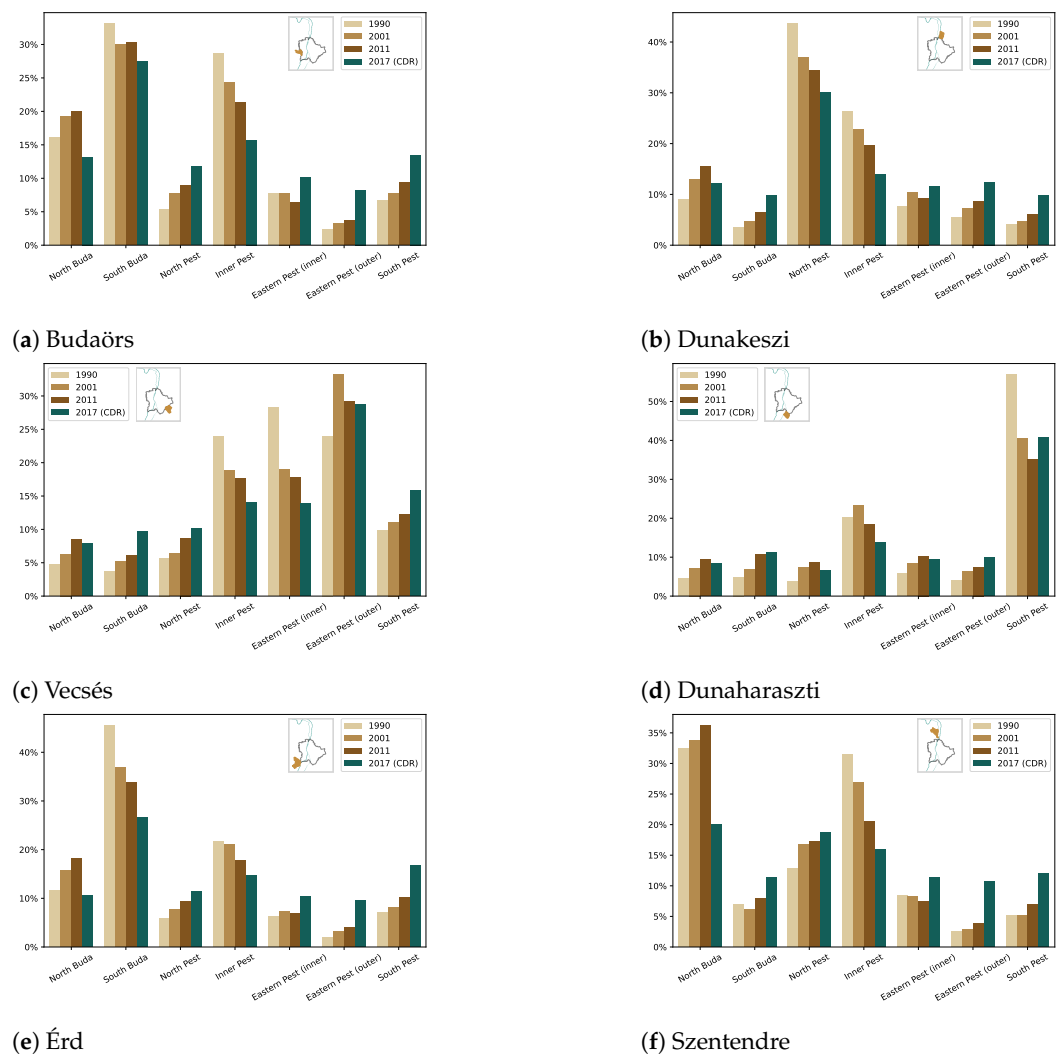


Figure 14. Commuting to the seven district groups of Budapest from selected settlements of the agglomeration, comparing census (1990, 2001, and 2011) and mobile network data. Next to the legends, the location of the settlements in question is displayed on a map.

4.5. Commuting by Age Groups

As the available mobile network data contained information about the age and the gender of the subscribers—for the 66.17% and 70.76% of the subscriptions, respectively—the commuting trends could be studied by age groups.

Koltai and Varró provided reference data for this analysis (Table 1 in [9]). Figure 15a shows the distribution of the commuters by age categories and the sector as the home location. Only those commuters who work in Budapest were examined.

It was not clear from the paper what was the upper limit of the “60+” age category. The people who usually go to work are assumed to be younger than 65 years old (the current retirement age in Hungary), although people can work over 65. In the CDR-based figure (Figure 15b), the 60+ means over 60 and less than 100. However, there were not many subscribers over 70, only 2.48% of the SIM cards were owned by people older than 70 years.

Comparing data obtained by the microcensus and the cellular information, good agreements (Pearson’s R is 0.8977) were found on the trends and measures of the distribution of the commuters by age categories. The most significant difference between the census and the CDR-based data were within the “60+” and the “50–59” categories. The number of people in their fifties seemed underrepresented by the CDR data, while the “60+” category was overrepresented, which might have been caused by the different interpretation of the upper limit. On the other hand, the values were very similar in the other categories. Based on the similarity of the results (Figure 15), it was confirmed that mobile network data can be a reliable method for commuting analysis even regarding demographic features.

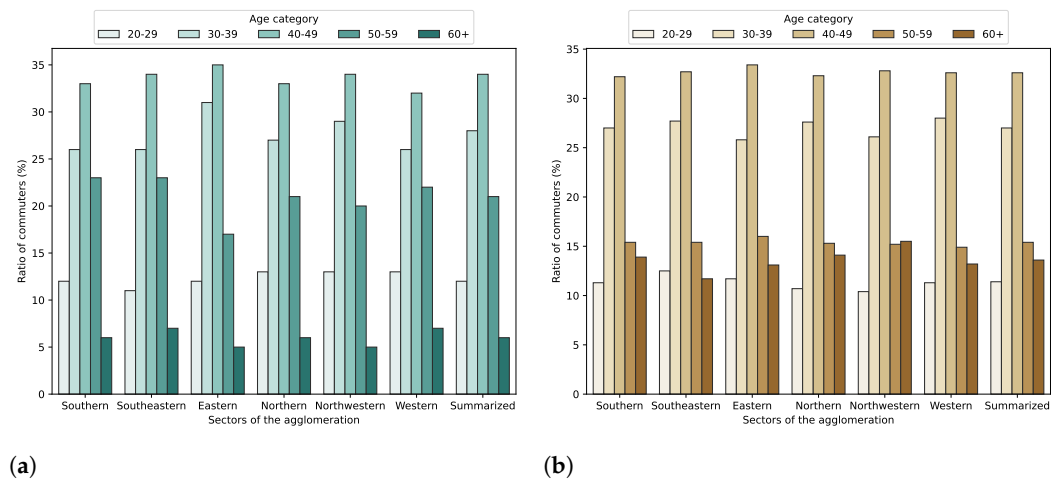


Figure 15. Distribution of the commuters by age categories and the sectors of the agglomeration (%). Comparison between microcensus data (Table 1 in [9]) (a) and mobile network data (b).

4.6. Limitations

This study used mobile network data from only one of three operators, so the results did not represent a full population. As children and elders do not (or are less likely to) own subscriptions, the data mainly represented the “working age” population. The OECD defines working age between 15 and 64 [50]. In 2017, 6,255,448 people lived in Hungary according to the HCSO, which is 63.85% of the full population [51]. Another limiting factor was that the provided data included age information for about 66.17% of the subscribers.

Moreover, the service provider preference was unknown by settlement level and by age group. Only the countrywide market share was known (25.5% in 2017 Q2 [38]), which, however, referred to subscriptions.

The evaluation and the validation were performed based on the results of other studies that analyzed commuting based on census data. With a direct access to the statistical data from HCSO and other sources, more and finer aspects of the validation could be performed.

Besides information availability, some socioeconomic groups of the society may have been excluded, as CDRs reflect structural inequalities [52].

4.7. Privacy Concerns

The authors could not identify the subscribers by the provided device ID. The timestamps were truncated to 10 s by the data provider to mitigate the identifiability of an activity (e.g., a voice call) in the data. As the call detail records were bound to a cell, the spatial resolution was determined by the area that a cell covered, which was roughly a house block downtown. There were some indoor cells, for example in malls, with a much finer spatial resolution (a building).

One of the main issues working with CDRs was that the activity was sparse. The cell-level location of the subscribers was known only when they actively used the mobile network (making calls, sending short messages, or transferring data). The privacy concerns would be much more serious if the cell-switching information had been included, in which case the location (cell) would be known every time a phone connects to another antenna. The cell-switching information would provide an almost continuous movement trajectory. Taylor formulated the same concern [29].

When a home or work location was detected, a cell was selected. A cell was represented by the centroid of the cell geometry (a circular sector from the top view). If a cell covered a housing complex, the subscriber could be any of its residents. This resolution could seem to be vague enough even with the age and gender information of the subscriber in an urban environment.

The authors are aware that prior knowledge about a specific movement in the observation area might help to decipher the anonymity if, for example, a subscriber were deliberately generating activity in well-recognizable places, or had a peculiar mobility pattern. This is in agreement with Sharad and Danezis, stating that it is possible to identify an individual in Orange's Data for Development (D4D) challenge Dataset 2 by using their movement records and a pre-existing location profile [53]. This could happen even in an urban environment, not only in rural areas. For example, the identity of a famous individual (e.g., politician, celebrity) whose workplace or home is publicly known can be attacked.

Nonetheless, the mobility and commuting analysis using mobile network data has great potential in social geography. The analysis presented in this paper could be executed at the operators—in a controlled environment—and only the (aggregated) results would be published, as the statistical offices publish results from the censuses.

5. Conclusions

In this study, the evaluation of cell phone subscribers' home and work locations were presented, and the results were compared to the population statistics. Though the detected population numerically differed from the actual population, the distribution across the settlements showed a strong correlation. This could be explained by the fact that the CDRs were obtained from only one mobile network operator.

Based on the home and workplace detection, it was demonstrated that mobile network data could be an effective solution for a commuting analysis. Besides the home locations, the settlement-level and the age-group-specific commuting trends were also validated from census data. The findings were presented in a form as close to the results of other studies that examined commuting in the agglomeration of Budapest as possible to aid the comparison. This work also served as a case study about commuting in the Budapest Metropolitan Area using mobile network data, with the queries and (plotting) scripts provided [42].

The district people commute to from the sectors of the agglomeration was examined. In the case of some selected settlements, the destination districts of the commuters were also presented in contrast to the last three censuses. It was found that the settlement level commuting trends based on the mobile network did not just approximate the last census

but also fitted the two-decade trends. The commuters were also analyzed by age groups, which also showed good agreement with the census-based studies.

These results confirmed that mobile network data can be used for a commuting analysis. Using mobile network data from all the operators of a country, a more precise and representative analysis could be performed. Given the fact that mobile networks are available in the most populated areas, mobile network data should be standardized for statistical and sociological usage while respecting privacy and personal data.

Understanding the commuting patterns in detail can help urban planning and public transport optimization. Providing viable public transport options to the workplaces can reduce the commuting time and decrease the number of cars in the cities. Once-in-a-decade questionnaires are not eligible to track the changes in commuting. Infocommunication technologies such as the mobile network can provide a more cost-effective and frequent analysis of commuting. Naturally, this application would require legal regulation, as a cooperation with the mobile operators and the Central Statistical Offices.

Author Contributions: Conceptualization, Gergő Pintér; methodology, Gergő Pintér; software, Gergő Pintér; validation, Gergő Pintér; formal analysis, Gergő Pintér; investigation, Gergő Pintér; resources, Gergő Pintér and Imre Felder; data curation, Gergő Pintér; writing—original draft preparation, Gergő Pintér; writing—review and editing, Gergő Pintér; visualization, Gergő Pintér; supervision, Imre Felder; project administration, Imre Felder; funding acquisition, Imre Felder. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Eötvös Loránd Research Network Secretariat under agreement ELKH KÖ-40/2020, and in part by the National Development, Research and Innovation Fund of Hungary under grant 2019-2.1.11-TÉT-2020-00204.

Data Availability Statement: The mobile network data used in this study are not publicly available due to third-party restrictions. For plotting the maps, OpenStreetMap data were used; these data are copyrighted by the OpenStreetMap contributors and licensed under the Open Data Commons Open Database License (ODbL).

Acknowledgments: The authors would like to thank Vodafone Hungary for providing the call detail records used in this study. We would also like to thank the three anonymous reviewers for their valuable comments on this article.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CDR	Call detail record
CPR	Control plane record
D4D	Data for Development
HCSO	Hungarian Central Statistical Office
HDA	Home detection algorithm
ICT	Information and communication technology
IMEI	International Mobile Equipment Identity
KDE	Kernel density estimation
SES	Social economic status
SIM	Subscriber identity module
TAC	Type allocation code
WGS 84	World Geodetic System, also known as EPSG:4326
XDR	Extended detail record

References

1. Központi Statisztikai Hivatal. *Budapest–Gazdaság és Társadalom*; Központi Statisztikai Hivatal: Budapest, Hungary, 2018. ISBN: 978-963-235-541-2.
2. Központi Statisztikai Hivatal. Calculated Population Data by Settlement—Resident Population in Hungary (2017–2020). Available online: <http://statinfo.ksh.hu/Stainfo/QueryServlet?ha=NT6B01&lang=en> (accessed on 22 June 2021).
3. Kiss, A.; Matyusz, Z. Az ingázás, mint forgalomkeltő tényező. *Munkaügyi Szle.* **2015**, *59*, 20–34.
4. Giménez-Nadal, J.I.; Molina, J.A.; Velilla, J. Trends in commuting time of European workers: A cross-country analysis. *Transp. Policy* **2022**, *116*, 327–342. [[CrossRef](#)]
5. Salas-Olmedo, M.H.; Nogués, S. Analysis of commuting needs using graph theory and census data: A comparison between two medium-sized cities in the UK. *Appl. Geogr.* **2012**, *35*, 132–141. [[CrossRef](#)]
6. Pálóczi, G. Researching commuting to work using the methods of complex network analysis. *Reg. Stat.* **2016**, *6*, 3–22. [[CrossRef](#)]
7. Lakatos, M.; Kapitány, G. Daily Mobility of Labour Force (Commuting) and Travel in Budapest and in the Metropolitan Agglomeration Based on Data of the Population Census. Part II. *Területi Stat.* **2016**, *56*, 209–239. [[CrossRef](#)]
8. Maris, M.; Kovacik, M.; Fazikova, M. Commuting trends and patterns behind the regional imbalances in Slovakia. *Eur. J. Geogr.* **2019**, *10*, 23–36.
9. Koltai, L.; Varró, A. Ingázás a budapesti agglomerációban. *Új Munkaügyi Szle.* **2020**, *1*, 26–37.
10. Goel, R.; Mohan, D. Investigating the association between population density and travel patterns in Indian cities—An analysis of 2011 census data. *Cities* **2020**, *100*, 102656. [[CrossRef](#)]
11. Pintér, G.; Felde, I. Evaluating the Effect of the Financial Status to the Mobility Customs. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 328. [[CrossRef](#)]
12. Pintér, G.; Felde, I. Awakening City: Traces of the Circadian Rhythm within the Mobile Phone Network Data. *Information* **2022**, *13*, 114. [[CrossRef](#)]
13. Pappalardo, L.; Simini, F.; Rinzivillo, S.; Pedreschi, D.; Giannotti, F.; Barabási, A.L. Returners and explorers dichotomy in human mobility. *Nat. Commun.* **2015**, *6*, 8166. [[CrossRef](#)] [[PubMed](#)]
14. Jiang, S.; Ferreira, J.; Gonzalez, M.C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Trans. Big Data* **2017**, *3*, 208–219. [[CrossRef](#)]
15. Fiadino, P.; Ponce-Lopez, V.; Antonio, J.; Torrent-Moreno, M.; D’Alconzo, A. Call Detail Records for Human Mobility Studies: Taking Stock of the Situation in the “Always Connected Era”. In Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks, Los Angeles, CA, USA, 21 August 2017; pp. 43–48.
16. Zagatti, G.A.; Gonzalez, M.; Avner, P.; Lozano-Gracia, N.; Brooks, C.J.; Albert, M.; Gray, J.; Antos, S.E.; Burci, P.; zu Erbach-Schoenberg, E.; et al. A trip to work: Estimation of origin and destination of commuting patterns in the main metropolitan regions of Haiti using CDR. *Dev. Eng.* **2018**, *3*, 133–165. [[CrossRef](#)]
17. Mamei, M.; Biccocchi, N.; Lippi, M.; Mariani, S.; Zambonelli, F. Evaluating origin–destination matrices obtained from CDR data. *Sensors* **2019**, *19*, 4470. [[CrossRef](#)] [[PubMed](#)]
18. Burgdorf, C.; Mönch, A.; Beige, S. Mode choice and spatial distribution in long-distance passenger transport—Does mobile network data deliver similar results to other transportation models? *Transp. Res. Interdiscip. Perspect.* **2020**, *8*, 100254. [[CrossRef](#)]
19. Barbosa, H.; Hazarie, S.; Dickinson, B.; Bassolas, A.; Frank, A.; Kautz, H.; Sadilek, A.; Ramasco, J.J.; Ghoshal, G. Uncovering the socioeconomic facets of human mobility. *Sci. Rep.* **2021**, *11*, 1–13. [[CrossRef](#)] [[PubMed](#)]
20. Wang, Q.; Phillips, N.E.; Small, M.L.; Sampson, R.J. Urban mobility and neighborhood isolation in America’s 50 largest cities. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 7735–7740. [[CrossRef](#)] [[PubMed](#)]
21. Bokányi, E.; Juhász, S.; Karsai, M.; Lengyel, B. Universal patterns of long-distance commuting and social assortativity in cities. *Sci. Rep.* **2021**, *11*, 20829. [[CrossRef](#)]
22. Petutschnig, A.; Albrecht, J.; Resch, B.; Ramasubramanian, L.; Wright, A. Commuter Mobility Patterns in Social Media: Correlating Twitter and LODES Data. *ISPRS Int. J. Geo-Inf.* **2021**, *11*, 15. [[CrossRef](#)]
23. Vanhoof, M.; Reis, F.; Ploetz, T.; Smoreda, Z. Assessing the Quality of Home Detection from Mobile Phone Data for Official Statistics. *J. Off. Stat.* **2018**, *34*, 935–960. [[CrossRef](#)]
24. Pappalardo, L.; Ferres, L.; Sacasa, M.; Cattuto, C.; Bravo, L. Evaluation of home detection algorithms on mobile phone data using individual-level ground truth. *EPJ Data Sci.* **2021**, *10*, 29. [[CrossRef](#)] [[PubMed](#)]
25. Csáji, B.C.; Browet, A.; Traag, V.A.; Delvenne, J.C.; Huens, E.; Van Dooren, P.; Smoreda, Z.; Blondel, V.D. Exploring the mobility of mobile phone users. *Phys. A Stat. Mech. Its Appl.* **2013**, *392*, 1459–1473. [[CrossRef](#)]
26. Gonzalez, M.C.; Hidalgo, C.A.; Barabasi, A.L. Understanding individual human mobility patterns. *Nature* **2008**, *453*, 779. [[CrossRef](#)]
27. Xu, Y.; Shaw, S.L.; Zhao, Z.; Yin, L.; Fang, Z.; Li, Q. Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation* **2015**, *42*, 625–646. [[CrossRef](#)]
28. Blondel, V.D.; Esch, M.; Chan, C.; Clerot, F.; Deville, P.; Huens, E.; Morlot, F.; Smoreda, Z.; Ziemlicki, C. Data for Development: The D4D Challenge on Mobile Phone Data. *arXiv* **2012**, arXiv:1210.0137.
29. Taylor, L. No place to hide? The ethics and analytics of tracking mobility using mobile phone data. *Environ. Plan. D Soc. Space* **2016**, *34*, 319–336. [[CrossRef](#)]
30. Dong, L.; Li, R.; Zhang, J.; Di, Z. Population-weighted efficiency in transportation networks. *Sci. Rep.* **2016**, *6*, 26377. [[CrossRef](#)]

31. Šćepanović, S.; Mishkovski, I.; Hui, P.; Nurminen, J.K.; Ylä-Jääski, A. Mobile phone call data as a regional socio-economic proxy indicator. *PLoS ONE* **2015**, *10*, e0124160. [[CrossRef](#)]
32. Yin, L.; Lin, N.; Zhao, Z. Mining daily activity chains from large-scale mobile phone location data. *Cities* **2021**, *109*, 103013. [[CrossRef](#)]
33. Diao, M.; Zhu, Y.; Ferreira, J., Jr.; Ratti, C. Inferring individual daily activities from mobile phone traces: A Boston example. *Environ. Plan. B Plan. Des.* **2016**, *43*, 920–940. [[CrossRef](#)]
34. Sakamane, P.; Phithakkitnukoon, S.; Smoreda, Z.; Ratti, C. Methods for inferring route choice of commuting trip from mobile phone network data. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 306. [[CrossRef](#)]
35. Dannemann, T.; Sotomayor-Gómez, B.; Samaniego, H. The time geography of segregation during working hours. *R. Soc. Open Sci.* **2018**, *5*, 180749. [[CrossRef](#)] [[PubMed](#)]
36. Hadachi, A.; Pourmoradnasseri, M.; Khoshkhal, K. Unveiling large-scale commuting patterns based on mobile phone cellular network data. *J. Transp. Geogr.* **2020**, *89*, 102871.
37. Barthelemy, M.; Barrat, A.; Pastor-Satorras, R.; Vespignani, A. Characterization and modeling of weighted networks. *Phys. A Stat. Mech. Its Appl.* **2005**, *346*, 34–43. [[CrossRef](#)]
38. National Media and Infocommunications Authority, Hungary. *A Nemzeti Média-és Hírközlési Hatóság Mobilpiaci Jelentése 2015. IV.–2019. II. Negyedév*; Technical Report; National Media and Infocommunications Authority: Budapest, Hungary, 2019.
39. Pintér, G.; Mosavi, A.; Felde, I. Artificial intelligence for modeling real estate price using call detail records and hybrid machine learning approach. *Entropy* **2020**, *22*, 1421. [[CrossRef](#)]
40. Pintér, G.; Felde, I. Analyzing the Behavior and Financial Status of Soccer Fans from a Mobile Phone Network Perspective: Euro 2016, a Case Study. *Information* **2021**, *12*, 468. [[CrossRef](#)]
41. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
42. Pintér, G. Commuting Analysis of the Budapest Metropolitan Area using Mobile Network Data. *Preprints* **2022**. [[CrossRef](#)]
43. De Montis, A.; Barthélemy, M.; Chessa, A.; Vespignani, A. The structure of interurban traffic: A weighted network analysis. *Environ. Plan. B Plan. Des.* **2007**, *34*, 905–924.
44. Zhang, W.; Fang, C.; Zhou, L.; Zhu, J. Measuring megaregional structure in the Pearl River Delta by mobile phone signaling data: A complex network approach. *Cities* **2020**, *104*, 102809. [[CrossRef](#)]
45. Hagberg, A.A.; Schult, D.A.; Swart, P.J. Exploring Network Structure, Dynamics, and Function using NetworkX. In Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, 19–24 August 2008; Varoquaux, G., Vaught, T., Millman, J., Eds.; 2008; pp. 11–15.
46. Ahas, R.; Silm, S.; Järvi, O.; Saluveer, E.; Tiru, M. Using mobile positioning data to model locations meaningful to users of mobile phones. *J. Urban Technol.* **2010**, *17*, 3–27. [[CrossRef](#)]
47. Bojic, I.; Massaro, E.; Belyi, A.; Sobolevsky, S.; Ratti, C. Choosing the right home location definition method for the given dataset. In Proceedings of the International Conference on Social Informatics, Beijing, China, 9–12 December 2015; Springer: Berlin/Heidelberg, Germany, 2015; pp. 194–208.
48. Eurostat. Employed Persons Working at Nights as a Percentage of the Total Employment, by Sex, Age and Professional Status. 2020. Available online: https://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=lfsa_ewpnig&lang=en (accessed on 31 March 2020).
49. Šveda, M.; Barlík, P. Daily commuting in the Bratislava metropolitan area: Case study with mobile positioning data. *Pap. Appl. Geogr.* **2018**, *4*, 409–423. [[CrossRef](#)]
50. OECD. *Working Age Population (Indicator)*; OECD: Paris, France, 2022. [[CrossRef](#)]
51. Központi Statisztikai Hivatal. 22.1.1.3. Népeség Korév és Nem Szerint, Január 1. Available online: https://www.ksh.hu/stadat_files/nep/hu/nep0003.html (accessed on 20 May 2021).
52. Letouzé, E.; Vinck, P.; Kammourieh, L. The law, politics and ethics of cell phone data analytics. *Preprints* **2015**. 10.20944/preprints202206.0347.v1. [[CrossRef](#)]
53. Sharad, K.; Danezis, G. De-anonymizing d4d datasets. In *Workshop on Hot Topics in Privacy Enhancing Technologies*; PETS: Bloomington, Indiana, 2013; p. 10.