 doi:10.20944/preprints202206.0347.v1

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; felde.imre@uni-obuda.hu
* Correspondence: pinter.gergo@uni-obuda.hu

Abstract: The analysis of the human movement patterns based on the mobile network data makes it possible to examine a very large population cost-effectively, and 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 and propose an automatized alternative to the current, questionnaire-based method. Commuting is predominantly analyzed by the census, but that is performed only once in a decade in Hungary. To analyze commuting, the home and the work locations of the subscribers are determined based on their appearances during and outside the working hours. The home locations 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 has been found that commuting analysis based on mobile network data strongly correlates with the census-based findings, even though home and work locations have been estimated by statistical methods. All the examined aspects, including commuting from sectors of the agglomeration to the districts of Budapest and demographic distribution of the commuters, show that mobile network data can be an automatized, fast, cost-effective, and relatively accurate way of commuting analysis, that could provide a powerful tool to the 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

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 lived [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 Hungarian Central Statistical Office (HCSO) [2]. The river Danube 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 makes 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 to another location; this is called "out-commuting". The target location receives a workforce, which is called "in-commuting". The loss of the source location is: (i) the out-commuter does not use the local resources, (ii) does not create value, (iii) loosens their relation, as it is partly relocated to the target location, and (iv) although 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 is occasional and inadequate [3]. Commuting is predominantly analyzed by the census data, but that is performed only once in a decade, thus cannot follow sudden, but permanent changes. They also stress, that commuting should be examined continuously, and its methodology should be established.

As questioning the population is a slow, tedious and expensive task, it would be obvious to automate the process with the available info-communication technologies (ICT). In this study, the application of CDR processing is proposed to examine commuting, mainly to Budapest, 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 some previous works. In [4], the home and work detection framework has been introduced with some macro-level commuting analysis as a validation. However, in this study, the commuting tendencies of the Budapest Metropolitan Area are analyzed in more detail. In [5], 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 detected population shows a strong correlation with population statistics.
3. The settlement level commuting trends do not just approximate the last census, but also fit the two-decade trends.

The rest of this paper is organized as follows. First, a brief literature review in Section 2. The utilized data is described in Section 3, the methodology is introduced in Section 4 then, in Section 5 the results of this case study are discussed. Finally, in Section 6, the findings of the paper are summarized.

2. Literature Review

Identifying the home and work locations of a 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 [6,7]. Since these locations fundamentally determine the people's mobility customs, the commuting trends can be analyzed between these locations. The commuting is studied, using mobile network data, within a city [7,8], or between cities [9–11], and also examined by social network data, such as Twitter [12,13].

Csáji et al. determined the subscribers' most common locations, and based on weekly calling patterns, identified the home and work locations [14]. Home locations showed a strong correlation with population statistics. 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 [15].

Xu et al. determined the home locations and then applied a modified standard distance to measure the spread of each subscriber's activity space [16]. Pappalardo et al. used the Radius of Gyration to separate the subscribers based on their mobility customs, and defined two classes: returners and explorers [6]. While in the case of returners, the radius of gyration is dominated by their movement between a few preferred locations, the explorers tend 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, that is usually the home and work locations [6], so this method can also be used as a home detection algorithm.

Pappalardo et al. [17] compared the estimated home locations of sixty-five subscribers with the known geographical coordinates of their residence location, using different types of mobile network data: CDR, eXtended Detail Record (XDR) and Control Plane Record (CPR). It has been found that XDRs should be preferred when performing home location detection.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1

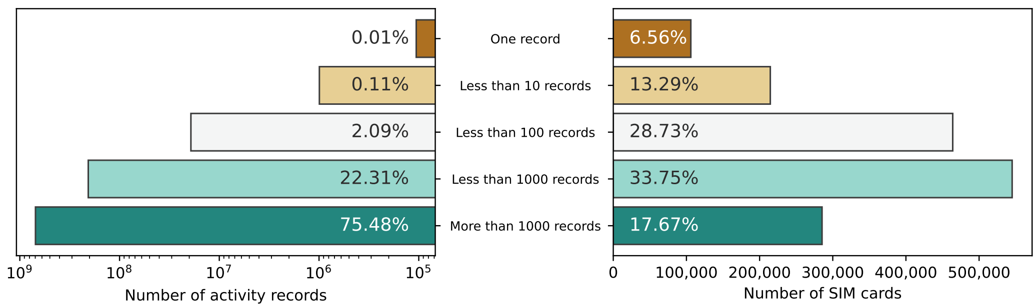


Figure 1. The SIM cards in the "April 2017" data set categorized by the number of activity records. The SIM cards over a thousand records (17.7%) provide the majority (75.48%) of the activity.

Vanhoof et al. compared five different home detection algorithms (HDA), selecting the home cell by (i) the most activity, (ii) the most 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 [18].

Jiang et al. identified daily activity patterns (motifs), that can extend the home-work location-based daily routine [7], the home locations were validated with census and household travel survey results. Yin et al. separated the different types of activity (home, work, leisure, school) with chains of activity [19], providing different approaches for a similar purpose.

Mamei et al. computed origin-destination flows with road network mapping, and also validated the home location estimation with census data [10]. Dannemann et al. [20] partitioned the city of Santiago (Chile) into several communities, and identified the socioeconomic composition of these communities, based on the home-work trajectories.

Pálóczi analyzed the country-wide commuting in Hungary, using the methods of complex network analysis, based on the data of census 2011 [21]. He considered settlements as nodes, and commuters as a directed edge between the nodes, then applied the disparity (Y_2) parameter [22], which was developed to measure the heterogeneity of weighted relations. If Y_2 is close to one, it means that one destination dominates the commuting.

Pálóczi demonstrated that the out-commuting dependency is not the greatest around Budapest, but around Győr and Székesfehérvár [21], but also high at the centers of the employment regions (e.g., chief towns of the counties).

3. Data

Vodafone Hungary, one of the three mobile phone operators providing services in Hungary with 25.5% market share in 2017 Q2 [23], provided an anonymized CDR data sets for this study. The observation area was Budapest, the capital of Hungary and its agglomeration. The observation period was one month, April 2017.

The CDR data set contains 955,035,169 activity records, from 1,629,275 SIM cards. Figure 1, shows the activity distribution between the activity categories of the SIM cards. Only 17.67% of all the SIM cards, that have more than 1000 activity records, provide the majority (75.48%) of the mobile phone activity during the observation period. Figure 2, 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 are not active enough to provide enough information about their mobility habits.

4. Methodology

The mobile network data processing framework has been introduced in a previous work [4], along with the home and work detection method. To analyze commuting, the home and the work locations of the subscribers have to be determined. As the CDRs are

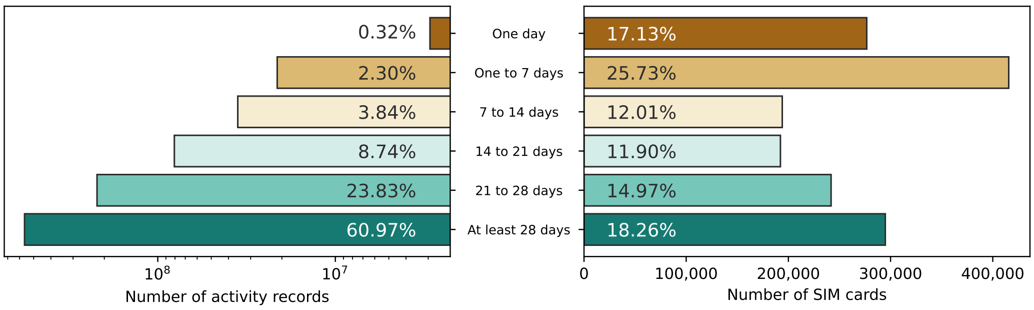


Figure 2. SIM card distribution of the "April 2017" data set, by the number of active days.

anonymized, do not contain information like a residential address. The workplace is even not mandatory for a subscription, the operators cannot have that information. After these locations have been determined, they should be validated. Although the applied approach is practically equivalent to what can be found in the literature, the validity of the results is hard to confirm. Pappalardo managed to validate the home locations in the case of sixty-five subscribers [17], but this is not possible in this case. So, the settlement and — in the case of Budapest — district-based population data [2] is applied from the HCSO.

4.1. Home and Work Locations

Most of the inhabitants in cities are spending significant time a day at two locations: their homes and workplaces. To find the relationship between these most important locations and Social Economic Status (SES), first, the positions of these locations have to be determined. There are a few approaches used to find home locations via mobile phone data analysis [16,24,25].

The work location is determined as the most frequent cell where a device is present during working hours, on workdays. Working hours are considered from 09:00 to 16:00. The home location is calculated as the most frequent cell where a device is present during the evening and the 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 is assumed that most of the activity is still generated from their home locations.

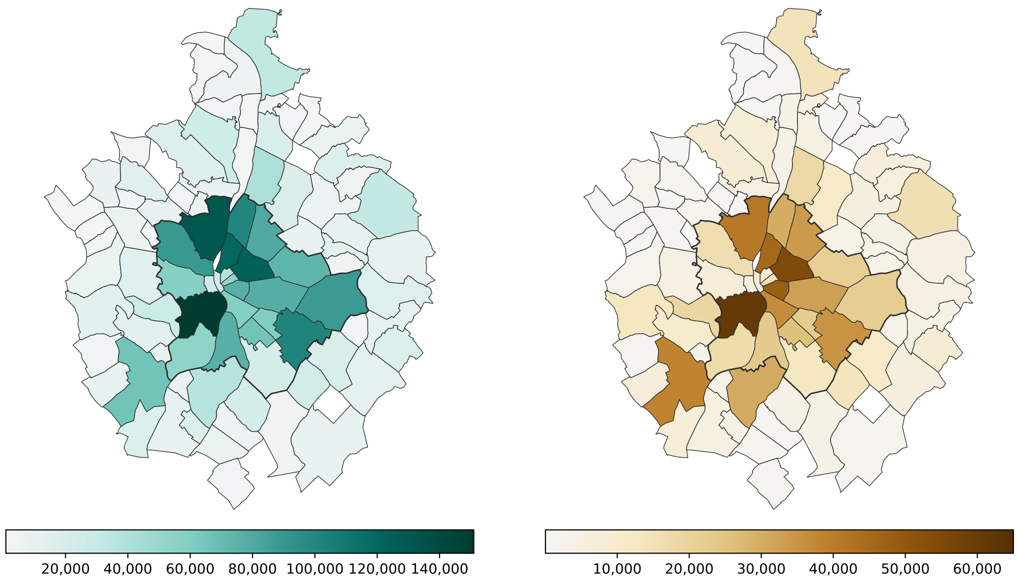
This method assumes that everyone works during the daytime and rests in the evening. Although in 2017, 6.2% of the employed persons worked regularly at night in Hungary [26], the current version of the algorithm does not try to deal with the night-workers. Some of them might be identified as regular workers but with mixed home and work locations.

5. Results and Discussion

Figure 3, shows a comparison of the population between the HCSO and the CDR data. When comparing the two maps, there are a few differences: Some parts of Budapest are not dark enough on the CDR data 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.

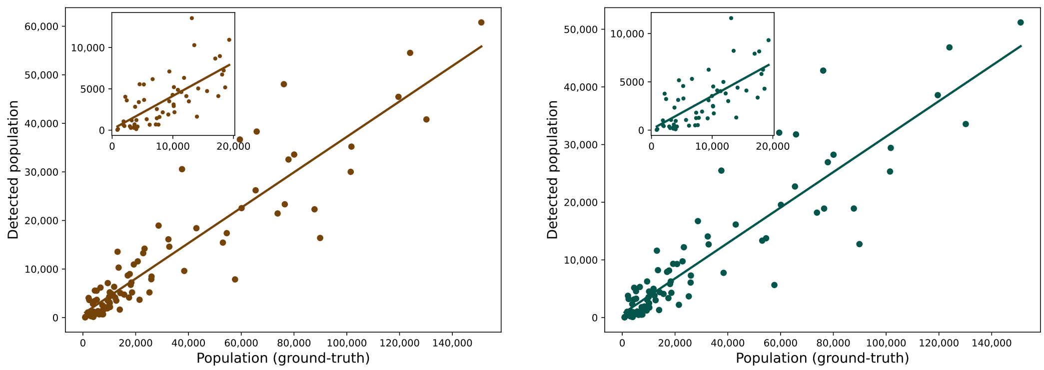
Apart from this, the findings based on CDRs correlate with the statistical data; the Pearson's R is 0.9213 (R^2 is 0.8488), counting every SIM cards. Figure 4a, shows this as a regression plot. As described in a previous work [27], the obtained Call Detail Records contain information (TAC) about the devices that use the mobile network. Based on this information more than 300,000 SIM cards have been identified, that do operate other types of devices like cellphones (e.g., 3G modem), indicating that they do not represent people. Figure 4b, illustrates the correlation without these SIM cards. Although the population values are decreased, the correlation does not seem significantly affected: Pearson's R is 0.9125.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1



(a) **(b)**

Figure 3. Comparing the ground truth [2] (**a**) and the estimated population based on mobile network data (**b**).



(a) **(b)**

Figure 4. Correlation between the population of the agglomeration and the districts of Budapest based on HCSO and mobile network data. In left figure (**a**), all the SIM card were used, in the right figure (**v**), SIM card that certainly operate in non-phone devices are excluded.

Numerically, the CDR data often shows significant mismatches, but they are 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 [23]. 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. As spatially more detailed market share is not available, it has to be supposed, 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.

5.1. 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 has changed. As part of the social distancing directive, to slow down the spread of the disease, working from home has come to the fore. Presumably, the prevalence of home-office will be higher than it was before the pandemic, as both the employers and the employees got used to this

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1

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, predates the pandemic, this question could only be answered in another work.

The workplaces have been determined, defined as the most frequent place where a subscriber appears during work hours. See Section 4.1 for the details. Querying the work locations of the inhabitants of an area, for example, a settlement can be the initial step of the commuting analysis. Figure 6, 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 the selected area itself, as many people work in the vicinity of their homes.

5.2. Connectivity

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 5, shows the connections between the districts of Budapest. The Buda districts are placed to the left, whereas Pest district are to the right and the colors of the nodes match with the district groups defined by HCSO [1]. 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 [21] could serve as a census-based reference.

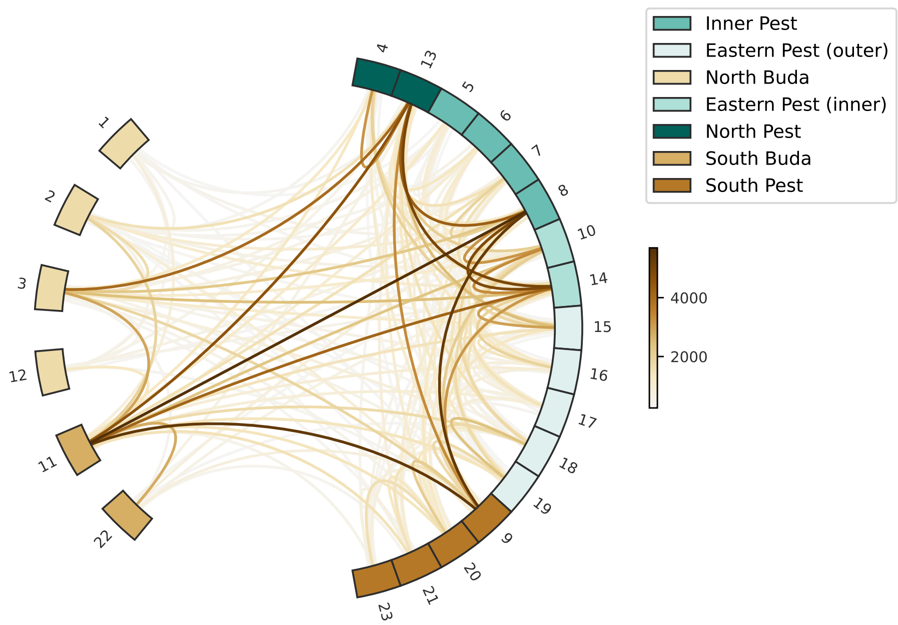


Figure 5. Connectivity between Budapest districts, using the home and work locations. A link originates from the home district to the work district, excluding the ones who where lives. The weak links are omitted to improve visibility. The district nodes are colored by the district groups.

5.3. 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 is performed on the mobile network data and the information processed from the census. In Hungary, a census is obtained every

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1

10 years and a micro-census with a 10% corpus at halftime. The last census was performed in 2011, while the last micro-census was in 2016. Note that, the next census should have been performed in 2021, but it was postponed due to the COVID-19 pandemic. Based on these surveys, commuting to Budapest (and generally in Hungary) is analyzed in studies like [3,21,28,29]. These studies are used as the reference for comparing the results.

Figure 7, shows the comparison between the CDR and the census-based [29, Figure 1] 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.

Good agreement mostly within Budapest has been 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, covers mainly the area of Budapest, and its close vicinity. It also contains phone activities from the surrounding county, but by moving away from Budapest, the available data decreases.

The fraction of workers who have their homes in the same district is very close to 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 show the best match to the census data (where the CDR should have the best quality), while the agglomeration is somewhat overestimated in many districts.

5.4. The Agglomeration

In [28], there is a more detailed analysis in regard to the commuting from the agglomeration, that is divided into six sectors and the commuting is examined by origin (home sector, occasionally by towns) and destination (district group of Budapest).

Figure 8, shows the commuters' distribution in the districts of Budapest, from the six sectors of the agglomeration, based on the CDR evaluation. In representation, Figures 8a–8f are analogous to [28, Figures 2–9], 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, between censuses: 1990, 2001 and 2011 [28]. The same analysis has been made using CDR processing, and six settlements of the 13 thoroughly analyzed of [28] are presented in this study. The results are summarized in Figure 9, compared with the censuses. It contains a settlement from every sector of the agglomeration, so it also serves as a more focused analysis of Figure 8. 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 9a), albeit North Pest, outer Eastern Pest, and South Pest are not significant commuting destinations, census data show an increasing tendency, which is confirmed by the mobile network data. The CDR based results of South Buda and Inner Pest also fit the trend, but in an opposite tendency. The most considerable discrepancy lies 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 census 2011 and the mobile network data is 0.8976.

Dunakeszi (Figure 9b) is east of the River Danube 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 have an increasing tendency, the inner Eastern Pest and North Buda do not show so clear tendencies. The correlation coefficient (Pearson's R) between the census (2011) and the CDR based results is 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 show increasing trends (Figure 9c). The correlation coefficient, in the case of Vecsés, is 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. Besides that, Inner Pest received considerable in-commuters, but its importance seems to be decreasing. The rest of the district groups had roughly the same trends (Figure 9d). Dunaharaszti has the strongest correlation out of the examined settlements: Pearson's R is 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 fit into the trends of the last three censuses, although Eastern and South Pest seem overestimated, and North Buda underestimated by the mobile network data based approach (Figure 9e). The correlation with the ground-truth is 0.8488 (Pearson's R).

In the case of Szentendre (Figure 9f), the mobile network based results might show the most significant discrepancy. Still, the correlation coefficient (Pearson's R) is 0.9127. As located in the Northwestern sector, west of the Danube, the most obvious destination for commuting is North Buda. According to census data, it had the most in-commuters, even with a slightly increasing tendency However, the CDR based results underestimate it, whereas Eastern and South Pest seem overestimated. The result of Inner Pest lags behind the census the latest census data, but that fits 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.

5.5. Demography

As the available mobile network data contains information about the age and the gender of the subscribers — in the case of the 66.17% and 70.76% of the subscriptions, respectively — the commuting trends can be studied by age-groups.

Koltai and Varró provide reference data for this analysis [29, Table 1]. Figure 10a, shows the distribution of the commuters by age categories and the sector as the home location. Only those commuters are examined who work in Budapest.

It is not clear from the paper, what is 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 10b), the 60+ means over 60 and less than 100. However, there are not many subscribers over 70, only 2.48% of SIM card owned by people older than 70 years.

Comparing data obtained by the micro-census and the cellular information, good agreements (Pearson's R is 0.8977) have been 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 are within the "60+" and the "50–59" categories. The number of people in their fifties seems underrepresented by the CDR data, while the "60+" category is overrepresented, that might be caused by the different interpretation of the upper limit. On the other hand, the values are very similar in the other categories. Based on the similarity of the results (Figure 10), it is confirmed that mobile network data can be a reliable method for commuting analysis even regarding the demographic features.

5.6. Limitations

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

6. Conclusion

In this study, the evaluation of the subscribers' home and work locations are presented, and the results were compared to ground truth. Though the detected population numerically differs from the actual population, the distribution across the settlements shows a strong correlation. This can 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 can be an effective solution for commuting analysis. The findings are presented in a form as close to the results of other studies, that examined the commuting in the agglomeration of Budapest, as possible to aid the comparison.

It was examined to which district people commute from the sectors of the agglomeration. In the case of some selected settlements, the destination districts of the commuters are also presented in contrast to the last three censuses. It was found that mobile network based results fit into the three-decade tendencies. The commuters were also analyzed by age groups, which also show good agreement with the census-based studies.

These results confirm that mobile network data is capable of commuting analysis. Using activity records from all the operators of a country, a more precise and representative analysis could be performed. Given the fact 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.

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

Funding: This research was supported in part the by 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 author would like to thank Vodafone Hungary for providing the Call Detail Records used in this study.

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

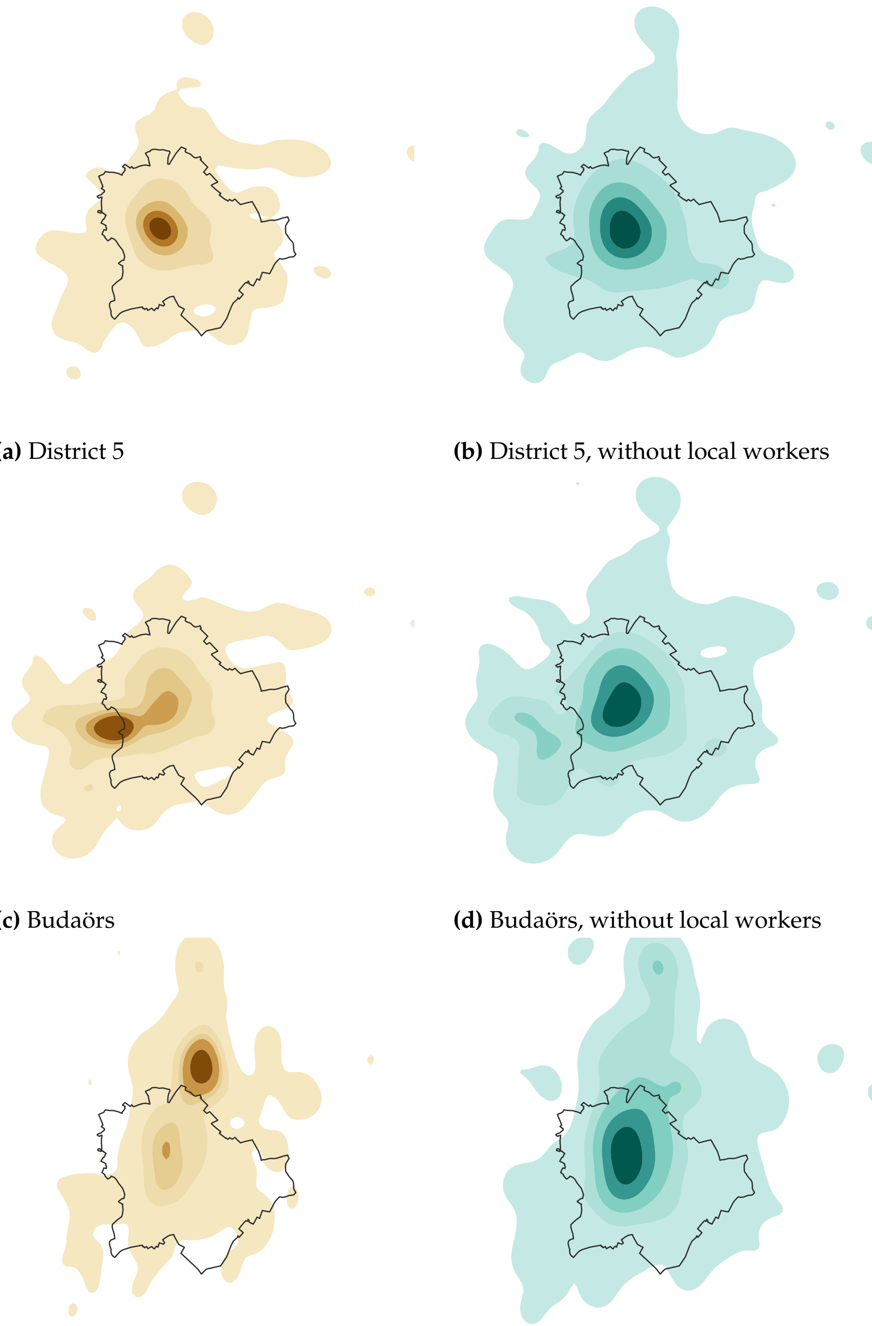
The following abbreviations are used in this manuscript:

CDR	Call Detail Record
CPR	Control Plane Record
HCSO	Hungarian Central Statistical Office
HDA	Home Detection Algorithm
ICT	Information and Communication Technology
SES	Social Economic Status
SIM	Subscriber Identity Module
TAC	Type Allocation Code
XDR	eXtended Detail Record

References

1. Központi Statisztikai Hivatal. *Budapest – Gazdaság és társadalom*; Központi Statisztikai Hivatal: Keleti Károly utca 5–7., 1024 Budapest, Hungary, 2018. ISBN: 978-963-235-541-2.
2. Központi Statisztikai Hivatal. Calculated population data by settlement - Resident population in Hungary (2017 - 2020). (Online; Accessed on 22 June 2021).
3. Kiss, A.; Matyusz, Z. Az ingázás, mint forgalomkeltő tényező. *Munkaügyi szemle* **2015**, *59*, 20–34.
4. Pintér, G.; Felde, I. Evaluating the Effect of the Financial Status to the Mobility Customs. *ISPRS International Journal of Geo-Information* **2021**, *10*, 328. https://doi.org/10.3390/ijgi10050328.
5. Pintér, G.; Felde, I. Awakening City: Traces of the Circadian Rhythm within the Mobile Phone Network Data. *Information* **2022**, *13*, 114. https://doi.org/10.3390/info13030114.
6. Pappalardo, L.; Simini, F.; Rinzivillo, S.; Pedreschi, D.; Giannotti, F.; Barabási, A.L. Returners and explorers dichotomy in human mobility. *Nature communications* **2015**, *6*, 8166.
7. Jiang, S.; Ferreira, J.; Gonzalez, M.C. Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. *IEEE Transactions on Big Data* **2017**, *3*, 208–219.
8. 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 Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks, 2017, pp. 43–48.
9. 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. *Development Engineering* **2018**, *3*, 133–165.
10. Mamei, M.; Bicocchi, N.; Lippi, M.; Mariani, S.; Zambonelli, F. Evaluating origin–destination matrices obtained from CDR data. *Sensors* **2019**, *19*, 4470.
11. 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. *Scientific reports* **2021**, *11*, 1–13.
12. Wang, Q.; Phillips, N.E.; Small, M.L.; Sampson, R.J. Urban mobility and neighborhood isolation in America's 50 largest cities. *Proceedings of the National Academy of Sciences* **2018**, *115*, 7735–7740.
13. Bokányi, E.; Juhász, S.; Karsai, M.; Lengyel, B. Universal patterns of long-distance commuting and social assortativity in cities. *Scientific reports* **2021**, *11*, 1–10.
14. 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. *Physica A: statistical mechanics and its applications* **2013**, *392*, 1459–1473.
15. Diao, M.; Zhu, Y.; Ferreira Jr, J.; Ratti, C. Inferring individual daily activities from mobile phone traces: A Boston example. *Environment and Planning B: Planning and Design* **2016**, *43*, 920–940.
16. 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.
17. 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 science* **2021**, *10*, 29.
18. Vanhoof, M.; Reis, F.; Ploetz, T.; Smoreda, Z. Assessing the Quality of Home Detection from Mobile Phone Data for Official Statistics. *Journal of official statistics* **2018**, *34*, 935–960.
19. Yin, L.; Lin, N.; Zhao, Z. Mining daily activity chains from large-scale mobile phone location data. *Cities* **2021**, *109*, 103013.
20. Dannemann, T.; Sotomayor-Gómez, B.; Samaniego, H. The time geography of segregation during working hours. *Royal Society open science* **2018**, *5*, 180749.
21. Pálóczi, G. Researching commuting to work using the methods of complex network analysis. *Regional Statistics* **2016**, *6*, 3–22.
22. Barthelemy, M.; Barrat, A.; Pastor-Satorras, R.; Vespignani, A. Characterization and modeling of weighted networks. *Physica a: Statistical mechanics and its applications* **2005**, *346*, 34–43.
23. 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, 23-25. Ostrom u., Budapest 1015, Hungary, 2019.
24. Ahas, R.; Silm, S.; Järv, O.; Saluveer, E.; Tiru, M. Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of urban technology* **2010**, *17*, 3–27.
25. 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. Springer, 2015, pp. 194–208.
26. Eurostat. Employed persons working at nights as a percentage of the total employment, by sex, age and professional status, 2020. (Online; accessed on 31 March 2020).
27. 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. https://doi.org/10.3390/info12110468.
28. 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 Statisztika* **2016**, *56*, 209–239. https://doi.org/10.15196/TS560206.
29. Koltai, L.; Varró, A. Ingázás a budapesti agglomerációban. *Új munkaügyi szemle* **2020**, *1*, 26–37.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1



(a) District 5

(b) District 5, without local workers

(c) Budaörs

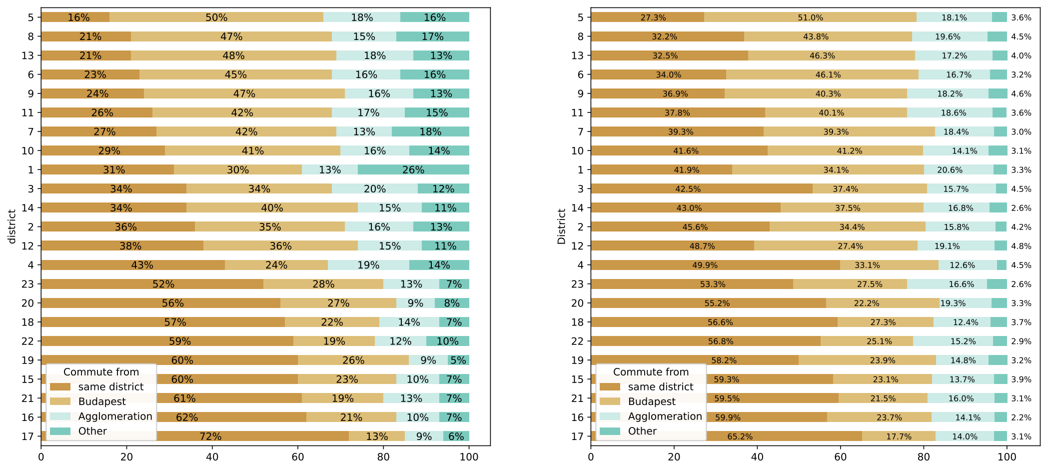
(d) Budaörs, without local workers

(e) Dunakeszi

(f) Dunakeszi, without local workers

Figure 6. 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**, **g**) and without (**b**, **d**, **f**, **h**) local workers.

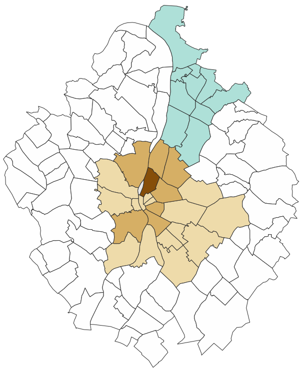
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1



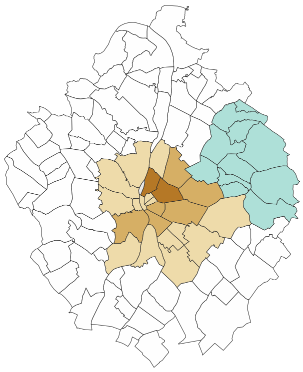
(a) **(b)**

Figure 7. Comparison between the census based (**a**) [29, Figure 1] and the CDR (**b**) commuting ratios to the districts of Budapest, from the same district, other parts of Budapest, the agglomeration or out of the agglomeration.

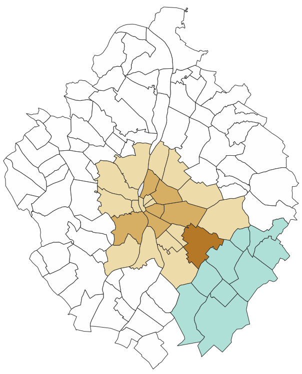
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1



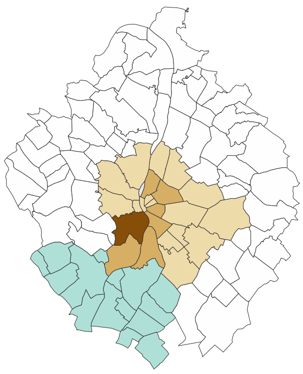
(a) Northern Sector



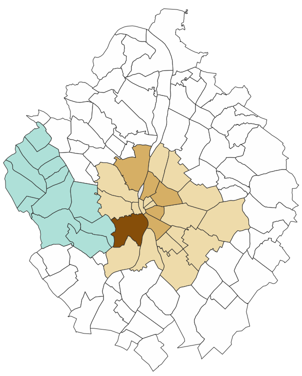
(b) Eastern Sector



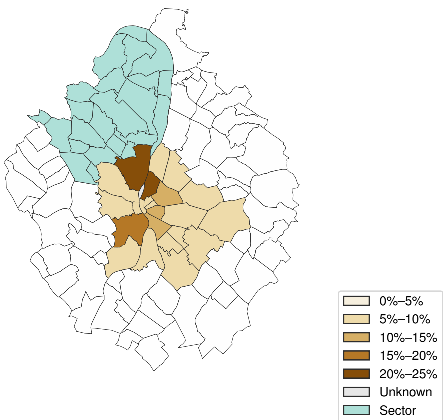
(c) Southeastern Sector



(d) Southern Sector



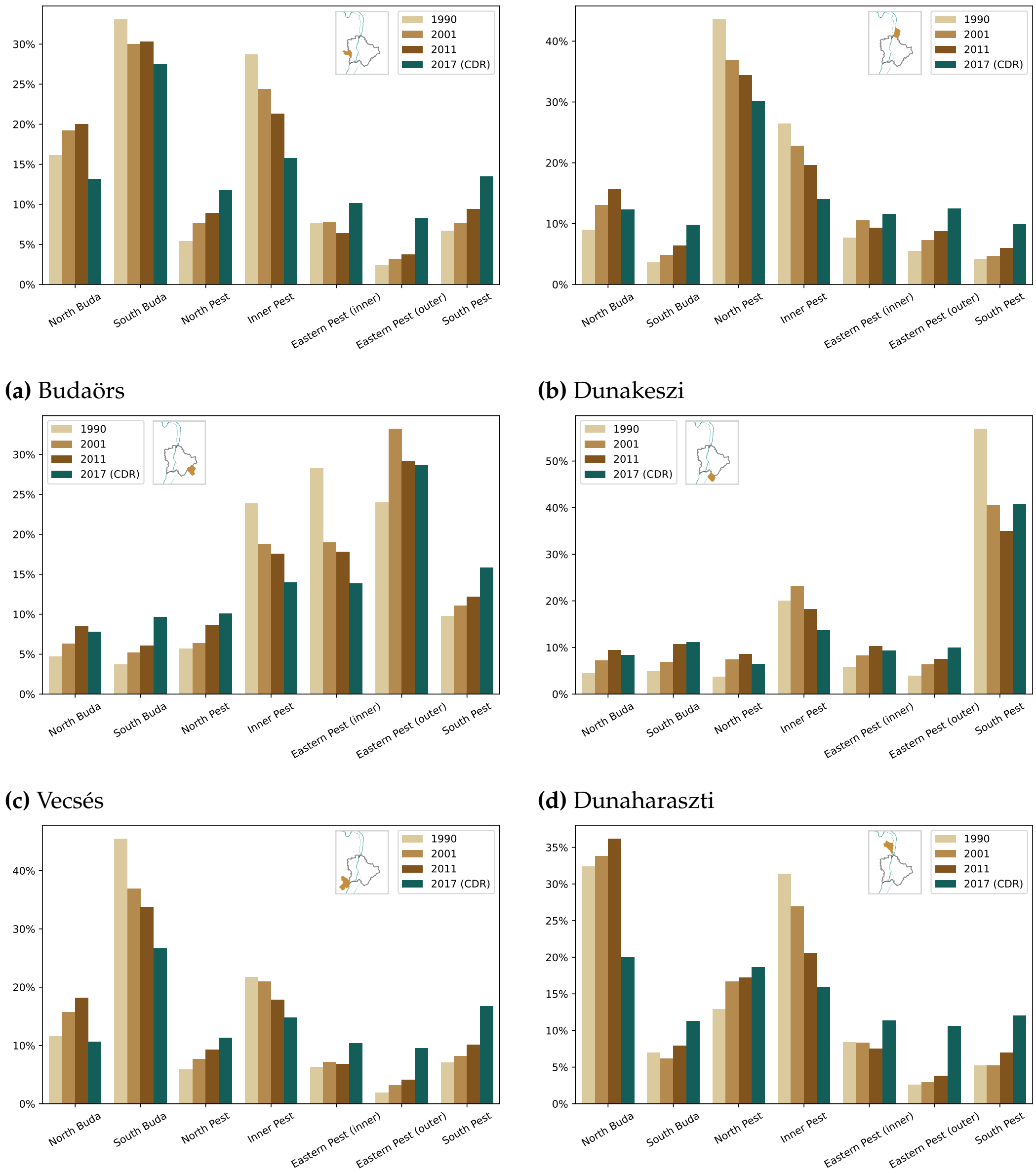
(e) Western Sector



(f) Northwestern Sector

Figure 8. Commuting from the six sectors of the agglomeration, based on CDR evaluation.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1



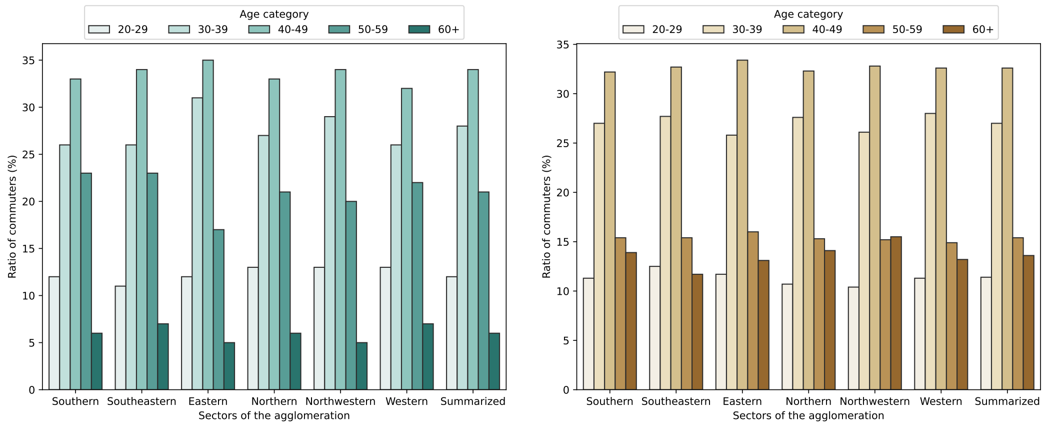
(a) Budaörs **(b)** Dunakeszi

(c) Vecsés **(d)** Dunaharaszti

(e) Érd **(f)** Szentendre

Figure 9. Commuting to the seven districts 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 in a map.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 27 June 2022 doi:10.20944/preprints202206.0347.v1



(a) **(b)**

Figure 10. Distribution of the commuters by age categories and the sectors of the agglomeration (%). Comparison between micro-census data [29, Table 1] (**a**), and mobile network data (**b**).