# Data Analytics without Borders: Multi-Layered Insights for Syrian Refugee Crisis

Özgün Ozan Kılıç[1], Mehmet Ali Akyol[1], Oğuz Işık[1], Banu Günel Kılıç[1], Arsev Umur Aydınoğlu[1], Elif Surer[1], Hafize Şebnem Düzgün[2], Sibel Kalaycıoğlu[1], Tuğba Taşkaya Temizel[1]

[1]Middle East Technical University, 06800, Çankaya, Ankara
[2]Colorado School of Mines, Brown Hall 268, CO 80401, USA
ttemizel@metu.edu.tr

**Abstract.** This study aims to shed light on various aspects of refugees' lives in Turkey using mobile call data records of Türk Telekom, which is enriched with numerous local data sets. To achieve this, we made use of several techniques in addition to a novel methodology we developed for this particular domain. Our results showed that refugees are highly mobile as a survival strategy, a significant number of whom work as seasonal workers. Most prefer to live in relatively cheap neighborhoods, close to city transport links and fellow refugees. The ones living in low-status neighborhoods appear to be introvert, living in a closed neighborhood. However, the middle and upper class refugees appear to be the opposite. Fatih, İstanbul was found as an important hub for refugees. Finally, the officially registered refugee numbers do not reflect the real refugee population in Turkey. Due to their high mobility, refugees lag behind in keeping up-to-date information about their residential address, resulting in a significant discrepancy between the official numbers and the real numbers. We believe that policy makers can benefit from the proposed methods in this study to develop real-time solutions for the well-being of refugees.

**Keywords:** health · education · unemployment · social integration · safety and security.

## 1 Introduction

The civil war in Syria has caused one of the biggest forcibly displaced population in human history [1]. Turkey has become the main destination for Syrian refugees, with around 5 million. Although there are camps built for the refugees with better living conditions than urban areas [2], more than 90% of the Syrian population in Turkey live outside formal camps within host communities, the reasons for which are given as overcrowded camps, illegally entered individuals not being allowed to register to a camp, family ties, and financial independence [3]. The status of Syrian refugees under temporary protection is shaped within the framework of "Temporary Protection Regulation." It is stated that under this regulation, the problems such as education, health, work permit, and access to social services and assistance are solved. They are also treated the same as the Turkish citizens in accessing such rights given that they are registered with Ministry of Interior Directorate General of Migration Management (DGMM).

At the beginning, Syrian refugees were mainly located in the Southeast Anatolian region bordering Syria. However, over time and with the influx of arriving refugees, they expanded to other regions as well, covering the Mediterranean, Aegean, Central Anatolia, and Marmara regions—İstanbul having the highest number of refugees. So far, Turkey has provided exceptional support to Syrian refugees [4]; however, the problems are mounting. They can be summarized as income, unemployment, education, health, housing, and social tensions [3, 5, 6].

The Syrian refugees have impacted the economy [7]. For instance, “around 1.8 million of the Syrian refugees are of working age” [8]. Although some entrepreneurial efforts have been observed and some of the refugees are skilled, most of the refugees are employed as unqualified labor. Through supplying inexpensive informal labor for labor-intensive sectors, refugees displaced native workers, both formal and informal unemployment rates have increased, and furthermore, it was observed that in these sectors the prices had fallen around 4% [9]. At the beginning of the refugee crisis, the Turkish economy had been experiencing a transition from being a low-wage country to one based on skilled labor. With the arrivals of Syrian refugees, this transition has started to decelerate as they have offered cheap low-skilled labor to the job market, which took advantage of their vulnerabilities. In particular, several refugees found jobs as seasonal workers or in small industrial areas (“sanayi siteleri”) [10].

At the time of refugees’ arrival, Turkish cities were undergoing a profound transformation in terms of housing. Illegal settlements (“gecekondu”) have started to be demolished and TOKI (Governmental Mass Housing Administration) aimed to regulate the housing market [11] which provided partial solutions to the problem. Refugees arrived when Turkey was still struggling with its urbanization problems. Therefore, refugees found safer places in the fragmented cities easily. A good evidence of it is that they settled into still-untransformed poor and environmentally low-quality districts, which are very close to city centers. These districts provided life-saving pockets for refugees where they can survive easily.

Big data have recently started to be used to address big social and environmental challenges in developing countries [12]. With ethical and privacy issues on mind, humanitarian use of private data such as mobile call data records has a great potential in improving society [13]. Data for Refugees, which is a good example of “big data for good”, is a research challenge aiming to provide better living conditions to Syrian refugees in Turkey [14]. In this research challenge, we investigated the mobility patterns of refugees from different points of views in order to provide multi-layered insights for the Syrian refugee crisis. We found out that refugees are highly mobile in Turkey as a survival strategy. We also carried out detailed analyses based on three different districts and cities, which we chose according to our previous results. When enriched with our secondary data sets, we saw that those living in low-status neighborhoods are introvert unlike refugees living in middle and high-status neighborhoods.

As a result, the repeatability and the reproducibility of the proposed methods can be beneficial to policy makers to obtain real-time insights about seasonal workers and to arrange services such as mobile health and education services on time. We also put some of our interactive visualization tools online on http://d4r.metu.edu.tr. The project website also provides detailed information about each step we have carried out in our analyses with several examples.

## 2 Technical Description

D4R Challenge provided three main data sets (DS1, DS2, and DS3) along with some helper files, which were collected from 992457 customers of Türk Telekom, of which 184,949 are tagged as "refugees", and 807508 as Turkish citizens in 2017. As the paper provides a description of all features, sampling strategies and anonymization methods in depth [15], we will skip those in this paper and describe the other data sets (hereafter called "secondary data sets") we used.

*Primary data sets*

- Data Set 1 (DS1) comprises annual antenna traffic between each site.
- Data Set 2 (DS2) includes cell-tower identifiers of randomly chosen active users' hourly based two-week call detail records. A user's either incoming or outgoing call traffic is provided but not both.
- Data Set 3 (DS3) consists of randomly chosen refugee and non-refugee annual call traffic but with reduced spatial resolution (district level).
- BTS Locations (BL) comprises cell tower locations in latitude and longitude.
- District Mapping (DM) maps the district IDs used in DS3 to district names.
- City Mapping (CM) maps the city IDs used in DS3 to city names.

*Secondary data sets*

- Neighborhood-level, district-level, and city-level geospatial data sets indicating the administrative borders in Turkey (GSD) obtained from various official sources and government agencies and used for any information that we needed to filter by a geographic region.
- Neighborhood-level population data and various statistics from 2013 to 2017 for various cities (PD) obtained from Turkish Statistical Institute (TurkStat).
- Coordinates of houses and workplaces for rent and for sale in various neighborhoods in İstanbul (FRE), scraped from Hürriyet Emlak [16].
- Rental fees and other relevant information such as the area of use and building age for various neighborhoods in İstanbul districts (RF), scraped from Hürriyet Emlak [16].
- The results of 2017 Address-based Population Registry System where the education levels of residents are given at neighborhood level (APRS2017).

Some of the terms defined in this study are borrowed from Ahas et al. [17]:

*Mobile Operator User (MOU):* A subscriber of the mobile operator, which is present in the DS2 and DS3.

*Home-Time Anchor Point (HAP)*: An everyday anchor point, at which the probable home location of a person is identified based on the model.

*Work-Time Anchor Point (WAP):* An everyday anchor point, at which the probable work-time location of a person is identified based on the model. As the demographics of MOU are not known, it is not possible to determine whether that person is indeed working, studying or unemployed. Therefore, we called it work-time anchor point to refer to the most probable working time of day of that specific MOU.

Hereafter, we used the abbreviations in the parentheses provided for the data set descriptions in the forthcoming sections.

We mainly used R to manipulate, analyze, query, and visualize data. We also utilized GIS applications such as ArcMap and QGIS to match and enrich the data sets with our secondary data sets, some of which were collected through web scraping and/or Google Places API through Python/R. We benefited from MySQL and NoSQL approaches such as Hive that work inside distributed frameworks that focus on big data, such as Hadoop, to store and manipulate data particularly for organizing and querying data. The network analyses were made in Pajek.

## 3 Pre-processing

Data quality issues with respect to base transceiver stations (BTS) were handled. Each BTS was assigned to a neighborhood-level, district-level, and city-level administrative units in Turkey by the coordinates, using GSD. The ones that do not fall inside the administrative borders of Turkey or the ones that do not have coordinates in the first place were discarded. This way, a total number of 98854 BTS were matched with city-district pairs. In addition, it was noticed that the city and district columns in BL are not entirely correct. As it can be seen in Fig. 1, there are several BTS coordinates, labeled as located in İstanbul, were not within the administrative district. Those were corrected using GSD.

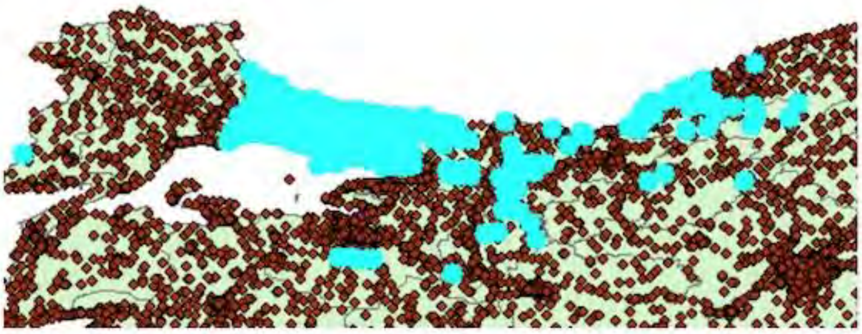

**Fig. 1.** BTS coordinates provided in BL marked as located in İstanbul but not within the administrative district of İstanbul are highlighted on the map.

DS1 included BTS data, which have zero number of calls or no entries in DS1 in different days or times of days. Fig. 2 shows one of the BTS' call numbers, some of which are not present in DS2. Apart from BTSs that have no data at all (more than half of the whole BTS population), a BTS has 82 missing days at best and 364 missing days at worst in total while the median number of the missing days is 92. Considering that some of the most missing ones belong to urban and dense urban areas (as also noted by the data set itself), it suggests that at least some of the missing days might be related to data quality issues rather than the lack of mobile traffic. This led us to work with average call numbers per BTS but not with cumulative sums, where the days with no data are excluded from calculations.


**Fig. 2.** The number of calls per day in July, 2017 in Çaykara, Trabzon measured in BTS number 5066930. The red line shows the total outgoing refugee call and SMS traffic and the blue line shows the total outgoing non-refugee call and SMS traffic. There are no instances in the data set for a number of days.

## 4 Methodology

The mobility of refugees is analyzed in five folds. In the first approach, we aim to understand how comparable the registered Syrian population in each city of Turkey is with that of call records in city level obtained from BTS call statistics. With the second approach, we analyze city networks connected by refugee mobility. With the third approach, we investigate the monthly refugee movement per district to be able to understand the influx of refugees over time. In particular, we address which districts are increasingly attracting refugees and whether there are any specific districts attracting refugees at a specific time of the year. In the fourth approach, we identify the most probable work and home-time anchor points from two-week data on selected rural and urban locations using mobility statistics and clustering algorithms. This output enabled us to identify the popular places for work and home of refugees. Here, we developed a new method to identify possible work and home time locations. Finally, we explore three different locations to understand the mobility patterns of refugees in depth using several statistics.

### 4.1 Background: Approaches for Determining Meaningful Places

Meaningful places or meaningful locations are defined as regularly visited places, which have a meaning for a person [18]. Mobile positioning data have been increasingly used for determining meaningful locations. In particular, finding work-time and home-time anchor points have been extensively studied in the literature [17, 19]. The common assumption in finding these places relies on a frequentist approach. First, a specific time interval is defined for work and home time periods. Second, the number of call days and the total number of calls are considered within these time intervals to identify significant anchor points. For example, if mobile calls are made from a specific BTS repeatedly during the home time, that BTS is marked as a potential home-time anchor point. In urban areas, several BTS can be located at the same site. Such BTS locations can be clustered using Hartigan's algorithm or k-medoid, and site information can be utilized when calculating mobile call metrics. Networks are also constructed including these frequent nodes (a node corresponds to a cluster comprising one or more BTS locations) to correctly identify home and work locations [19]. Finally, several metrics such as regularity, entropy, or radius of gyration (RoG) can be computed to understand the behavior of citizens [20].

Mobile traffic signatures, defined as the typical activity pattern of the mobile demand at one specific geographic zone, have been recently used to investigate the relationship between urban fabrics such as touristic and leisure places, and mobile network usage [21]. Different metrics were proposed based on voice and text traffic volume over weekends and weekdays some of which take into consideration the seasonality.

### 4.2 Comparison of Registered Syrian Refugee Population with the Number of Calls

Directorate General of Migration Management of Turkey published the registered Syrian population in each city of Turkey for 2017 [22]. According to the statistics, the highest numbers of refugees are located in İstanbul, Şanlıurfa, Hatay, and Gaziantep with 479555, 420532, 384024, and 329670, respectively. In this part of the study, we aggregated BTS call statistics at city level for refugees as follows: First, we calculated the total number of refugee calls and such for each day per BTS. Then, we calculated the monthly average from the daily totals per BTS in each month while discarding the days with no calls due to the aforementioned data quality problem. Then, we computed the sum of each BTS over a year to obtain the cumulative annual use of each BTS, which is denoted as $BR_i$. Each BTS is geospatially associated with a district and city using GSD. Finally, for each city, we summed up all BTS aggregated call data, which is denoted as $C_i$. The vertical percentage of refugee calls for each city is found using $VPCR_i = C_i/(\sum_{i=1}^{81} Ci)$, where 81 is the total number of cities in Turkey. Then, we made use of the registered Syrian population $PR_i$ and the total population of each city in Turkey $P$ to calculate the vertical percentage of the registered refugee population with $VPR_i = PR_i/P_i$. Finally, we divided $VPCR_i$ by $VPR_i$ to obtain the magnitude of the difference between the two statistics, which is denoted as *MD*. The map plotted in Fig. 3 shows the results. Although the registered Syrian refugees are officially reported low in Antalya, the total number of refugees' calls in Antalya is quite higher than expected. The second highest city is Kilis, which shares a border with Syria. There also seems to be an undocumented influx to East and West Black Sea regions, although it is not as significant as the previously mentioned ones, which will be examined closely in the following sections.


**Fig. 3.** The ratio of the total refugee calls per city in 2017 to the official residency records, higher numbers (represented with lighter colors) indicating a higher refugee influx compared to official figures (higher than expected)

The results show that there is a discrepancy between the number of registered users and the call data in certain cities. They indicate that although some refugees change their residency addresses, they do not inform the state agencies about this change on time. An expert working in Refugees and Migrants Solidarity Association we consulted also stated that they were suspecting of numerous undocumented refugees living in Antalya but they are not sure about how big the discrepancy is.

### 4.3 Analysis of City Networks Connected by Refugee Mobility

In order to understand how Syrian refugees use space in Turkey, incoming and outgoing calls data in DS3 were used to form 1-mode and 2-mode networks of refugees and the cities in which they have made phone calls. Initial basic statistics showed that refugees are highly mobile. Out of 37300 refugee MOUs, 53.8% visited only one, 19.9% two, 9.3% three, 5.2% four and 11.8% five or more cities.

In network analysis, initially multiple lines were summed in the 2-mode network of refugees and the cities to obtain the total number of calls a refugee has made in a city. In order to determine the cities where refugees reside in, rather than visit briefly during a trip, the ties that indicate less than 100 phone calls in a city in one year were removed. Then, all line values were replaced with the value 1, since the focus of attention is the presence of a refugee in a city, not how many phone calls they have made. This 2-mode network was then used to obtain the 1-mode network of cities. In this network, a pair of cities is connected by refugees who have been to both cities. In the 1-mode network, the value of the line that connects any two cities is the total number of refugees who have been in those cities (aka "network traffic" in Figure 4).

In order to quantify the most important cities for refugee mobility, weighted degree centralities were calculated. Top 10 cities that receive the highest refugee traffic in descending order were found as İstanbul, Gaziantep, Ankara, İzmir, Mersin, Adana, Hatay, Antalya, Şanlıurfa, and Kocaeli. Then, the lines with values lower than 50 were removed to simplify the graph and the largest connected component was determined, which yielded 33 cities. Fig. 4 shows the resulting ties between these cities. The sizes of the vertices show the registered Syrian refugee population [22] in the corresponding cities with a minimum of 155 (Giresun), a maximum of 479,555 (İstanbul), and a median of 8120. The widths of the lines indicate the total number of refugees linking the cities, with a minimum of 50 (between İstanbul and Kayseri), a maximum of 318 (between İstanbul and Kocaeli), and a median of 79.5. Some cities, such as Edirne, Tekirdağ, Çanakkale, Aydın, Muğla, Antalya, Kastamonu, Samsun, Trabzon, Ordu, Giresun, Tokat, and Sivas have a small number of registered refugees. Yet, the analysis shows that refugees visit and stay in these cities. The majority of these travels are to/from İstanbul only. The strongest single link a city has (compared to its registered refugee population) belongs to Antalya, tied to İstanbul.

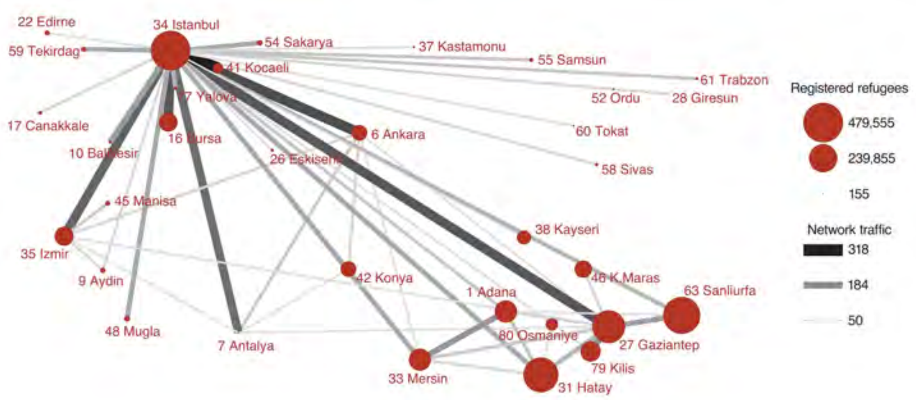


**Fig. 4.** Network map indicating the refugee links between cities (lines) along with their registered Syrian refugee populations (vertices)

### 4.4 Investigating monthly refugee movement per district

In this section, monthly refugee mobility is studied. Some refugees are known to be highly mobile and work as seasonal agricultural workers in Turkey. To be able to identify districts receiving the highest refugee influx, we calculated the sum of $BR_i$ for each district, which is denoted as $BRD_i$. As we will study the density and distribution of the outgoing refugee calls in DS1, we need to consider districts that have a significant number of call data. This is due to the fact that the mobility pattern of subscribers without heavy phone usage can be properly characterized, is indeed questionable [23]. The study reported that 17% and 38% of subscribers in their CDR data set had two or fewer records and fewer than seven CDR, respectively. In addition, some BTS records are not present in DS1. As it is not possible to understand the reason of omissions (whether it is a measurement problem or indicating zero call entries), we performed filtering, which resulted in discarding two-thirds of district data. As a rough cut off point, the districts with total outgoing refugee calls in 2017 lower than 3650 were filtered out to improve calculation time and reliability of the analysis results. This is due to the fact that to make an inference with a few calls can produce biased estimations. This threshold is determined using the mobile phone usage distribution statistics provided in [23].

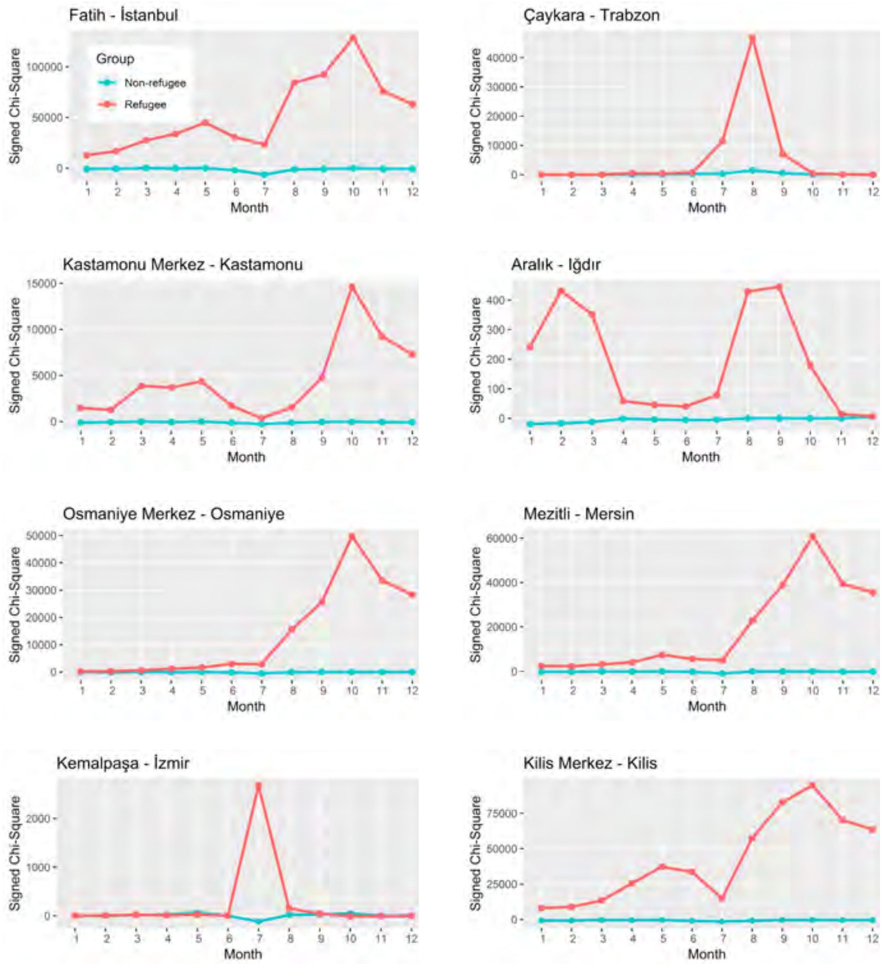


**Fig. 5.** Some of the districts with significantly higher outgoing refugee calls and/or monthly changes that can be explained

By looking at the monthly SCS values and their first order differences, we highlighted some of the districts that feature significant changes as seen in Fig. 5. Later, we consulted experts to understand the reason for the influx of patterns. Fatih-İstanbul was reported as an attraction center for refugees, the exact reasons for which are not clear. However, it is advocated that the low prices in accommodation, Fatih's easy access to other districts and the existing local community's religious background might have attracted them. This district is studied in depth in the forthcoming section. The agriculture expert we consulted informed us of many changes in the districts coincide with the harvest times. To be specific, the harvests of Çaykara-Trabzon in July and August are hazelnut and tea, Kastamonu in September and October are sugar beet, garlic, and paddy. In addition to that, the harvests of Iğdır in August/September are sugar beet and cotton while the harvest of Osmaniye in October is peanut. Wealthy refugees visit summer locations as frequently as non-refugees. We see an interesting peak in Kemalpaşa-İzmir in July and a smaller peak in August. In addition, for February/March in Iğdır, there is another one we observed. However, at the time of writing this report, we still could not reach an authority who could explain the exact reason for this refugee influx in July for Kemalpaşa and in February/March for Iğdır. On the other hand, for the smaller peak in Kemalpaşa in August, we were told by the expert that the reason of the increase in August could be attributed to the religious festival for Alevis, which is called Hamzababa Anma Törenleri (Hamzababa Commemoration) taking place between 28 and 29th of August each year. For the peak in July for Kemalpaşa, the experts from numerous local municipality services we consulted, speculated that it is a high season for harvesting and refugees might have stayed in this district for temporary accommodation. The agriculture expert also suggests the refugee influx in Mezitli is based on tourism and the fact that the peak happens in October does not contradict with it since Mersin is known for its scorching heat and humidity which shifts the tourism season towards fall. Kilis has a border with Syria, constantly receives refugees. We have provided the results of some districts in Fig. 3. Many more others can be viewed on our web site using an interactive tool.

### 4.5 A New Method to Understand Possible Work-Time and Home-Time Locations

In this section, we investigate refugees' meaningful places, specifically work-time and home-time mobility patterns, which are quantified using well-known measures used in the literature [24]. In particular, different aspects of irregularity are studied. We extended a well-known method in the literature, which was described in Section 4.1. The main contribution of our proposed method is rather than using a pre-determined time to identify work and home time locations, we identified important anchor points automatically using an algorithm. Many people such as white-collar workers have very structured daily lives. Their work hours are usually between 9:00 A.M. to 6:00 P.M. However, a garbage collector visits several districts to collect recycled materials such as glass, paper or plastics, while a seasonal agricultural worker might have changed his location within a week and moved to another district. Some people might have been working at different times of days, which we came across in the data sets considerably. Therefore, it may not be convenient to use a pre-determined time to find WAP and HAP locations. In addition, a static threshold for number of calls or number of call days is generally used to filter out the MOUs as their HAP and WAP information cannot be obtained due to limited number of call data. Instead of it, we used a clustering algorithm in this study. Later, we made use of DS2 to test our methodology.

The pseudocode of the algorithm is provided in Table 1. The first step of the algorithm involves finding the idle hours of a given user. As there might be idle hours during the working-time, we made use of a median filter, where each hour is replaced with the median value of a moving filter. Finding the first quartile enables us to identify the start and end points of a continuous time block. For the home-time period, we assume that MOU can be highly likely to be at home just before the idle time period so we chose a time interval starting three hours before the starting point of the idle time and ending at the end point of the idle time. Likewise, we considered four hours after the idle time as a starting time of a work-time period as we assumed that these four hours will highly likely to include home and commute locations. It is not meant the real work-time will start at that point rather we aim to discard noisy data. These numbers are obtained empirically based on DS2. Finally, the closer BTS locations are clustered. As mentioned by Isaacman et al. [25], in urban areas the BTS can be as dense as 200 meters and in suburban areas, they can be 3-5 kilometers apart. Therefore, we have tried different values for the radius, such as 200 meters, 500 meters, and 1 kilometer and the radius of 1 km gave more meaningful results.

**Table 1.** Algorithm

| Algorithm to detect work-time and home-time patterns |
|---|
| input: caller_id *CID* in DS2<br>output: *WAP* and *HAP*<br>1. Retrieve *CDR* of *CID*<br>2. Calculate hourly call counts $hourly\_calls_i$ from CDR, where $i$>=0 and $i$<24<br>3. If there is no CDR in any $hourly\_calls_i$, assign it to zero.<br>4. Apply median filter with a window size three on $hourly\_calls_i$, to obtain filtered data, denoted as $filtered\_hourly\_calls_i$.<br>5. Sort the $filtered\_hourly\_calls_i$ in descending order.<br>6. Obtain the first quartile of $filtered\_hourly\_calls_i$ , which is denoted as *fq*.<br>7. Find the minimum and maximum hours in *fq*, denoted respectively as $fq_{min}$ and $fq_{max}$ respectively<br>8. Determine the start and end points of WAP and HAP as follows:<br>*Let wtp_start* be the work-time period start time, where $wtp\_start = fq_{max}+4$<br>Let *wtp_end* be the work-time period end time, where<br>$wtp\_end=wtp\_start + 6$<br>Let *h_start* be the home-time period start time, where $h\_start = fq_{min} -3$<br>Let *h_end* be the home-time period end time, where $h\_end = fq_{max}$<br>9. Find the most used BTS in terms of calls days between *wtp_end* and *wtp_start*, which is denoted as $BTSW_{max}$<br>10. Apply Hartigan's leader algorithm [26] to all BTS between *wtp_end* and *wtp_start*.<br>11. Select the cluster in which the most used BTS resides, which is denoted as *WAP*.<br>12. Find all BTS on the same calls days with *WAP* between *h_end* and *h_start*, denoted as $BTS_{HomeAll}$<br>13. Find the most used BTS in $BTS_{HomeAll}$, denoted as $BTSH_{max}$<br>14. Apply Hartigan's leader algorithm to all BTS between *h_end* and *h_start*<br>15. Select the cluster in which the most used BTS resides, which is denoted as *HAP* |

Then, we have extracted 31 number features from two-week CDR data of each MOU. Some of these features are borrowed from Soto et al. [27]. The complete list of features can be found on our website. The significant features selected by the process described below and our consequent statistical analyses are as follows:

- N_call_days/N_call_days_home_time/N_call_days_work_time: The number of unique days, unique days recorded within HAP, unique days recorded within WAP respectively
- N_calls: The number of calls
- N_city: The total number of cities the MOU is seen

- N_district: The total number of districts the MOU is seen
- RoG [27]
- Entropy_bts: Entropy of MOU based on the BTS footprint
- Entropy_district: Entropy based on the district footprint
- Entropy_district_home_time: Entropy calculated within HAP (based on district information)
- Entropy_district_work_time: Entropy calculated within WAP (based on district information)
- Entropy_cluster_home_time: Entropy calculated within HAP (based on the clusters formed after Hartigan's leader algorithm)
- Entropy_cluster_work_time: Entropy calculated within WAP (based on the clusters formed after Hartigan's leader algorithm)
- Ref_nonref_ratio: Ratio of calls made/received to/from refugees to non-refugees
- Total_movement: Total distance of travel made by MOU, calculated by summing the distance between each following BTS used by the MOU
- Work_home_dist: Haversine distance between the WAP and HAP

After deriving these features, we first applied sparse K-means (SK-means) algorithm using R's RSKC package [28] with different parameters to obtain the most relevant, uncorrelated and non-redundant features. As a result, we ended up with the following features (sorted in descending order according to their significance values): n_call_days, n_calls, n_call_days_work_time, n_call_days_home_time, entropy_cluster_home_time, entropy_cluster_work_time, entropy_district, entropy_district_home_time, rog_work_time, rog_home_time, n_city, and work_home_dist to cluster the MOUs. Finally, these features are used as inputs to Self Organizing Map (SOM) [29], which is a type of artificial neural network to visualize high dimensional data in a low-dimensional grid. SOM produced different mobility clusters and we selected the instances falling into the cluster nodes, where there is sufficient number of call days and number of calls, which are important to detect HAP and WAP more accurately. Finally, we have identified regions mostly preferred for work and living purposes by refugees. Note that two week data is quite limited to estimate WAP and HAP of an individual accurately. Hence, if there are not sufficient numbers of data points for a MOU, the algorithm cannot successfully identify these locations. The detailed steps of the algorithm are provided in the Appendix section including how to interpret clusters with examples as well.

Since we did not have a ground truth in hand for evaluation, we scraped seemingly the biggest online real estate website in Turkey, Hürriyet Emlak (2018), and obtained a total of 701 house/workplace ads for rent/for sale that correspond to the neighborhoods in Fatih, İstanbul (FRE). The sampling was neighborhood-wise stratified (to ensure each neighborhood is represented) and ad-wise systematic (to ensure that the price range and distribution are reflected) as every $n^{th}$ ad was recorded from the list of ads ordered by price while while $n = \lfloor ad\_size/sample\_size \rfloor$. The sampling of the ads for a specific neighborhood was only applied if the neighborhood had more than 10 ads for houses or workplaces, separately. After collecting 338 houses and 363 workplaces, we sampled the house and workplace locations that fall inside Fatih, İstanbul (by also using GSD) from the cluster of refugees that we are most certain. Again, we used stratified sampling to sample exactly 15 (house & workplaces in total) coordinates from each neighborhood and we obtained 338 houses and 517 workplace locations (855 points in total). The difference in numbers of points (701 vs. 855) is due to different sample sizes, Hürriyet Emlak not having all the neighborhoods available for filtering, and lack of ads for certain neighborhoods. Then, we compared the locations of the ads with refugees' predicted work/home locations. As can be seen in Fig. 6, some locations such as İskenderpaşa neighborhood (central region) are both residential and working places. The shops are mainly located on the sideways of the main street. However, like in Tahtakale neighborhood (the northeastern region of Fatih), some are mostly business related locations while some are mostly residential places as in Şehremini or Silivrikapı neighborhoods (southwestern region). Our results coincide with the ad types.

**Fig. 6.** HAP and WAP locations obtained from refugees' call data (indicated with light orange and light gray) along with the houses and workplaces obtained from Hürriyet Emlak (indicated with orange and dark gray)

## 5 Investigating Regions in Depth: Case Studies for Çaykara/Trabzon, Fatih/İstanbul, and Mezitli/Mersin

In this section, based on our previous findings, we explore three districts and cities in detail. We look into the characteristics of these locations and attempt to understand refugees' living conditions.

### 5.1 Findings for Çaykara/Trabzon

Our initial analysis showed that Çaykara, Trabzon has a significant peak in August. An agriculture expert we consulted suggested that it is most probably related to the harvest of tea (which also gives its name to the district) and hazelnut. We also confirmed it with "Development Workshop" reports [10]. The report states that the Black Sea region in Turkey receives seasonal Syrian agricultural workers between August and September, first starting to work in gardens in coastal regions then move innermost.

We analyzed DS2 and DS3 to find the refugees that were present around the specific BTS (#5066930) in August and understand where they come from. We found that people present in the area were also highly present in İstanbul and Mersin. The coordinates where the phone calls occurred also come right on top of the usual intercity bus route between İstanbul and the Black Sea region. The seasonal workers here might be later switching to Mezitli, Mersin region (which can explain its peaks in October) once the harvest is done. When we looked at DS3 to find the refugees that were in Çaykara in August, we found 306 callers (people with outgoing call data in DS3) and 275 callees (people with incoming call data in DS3). Interestingly, the most common district that these callers and callees present in 2017 was Fatih, İstanbul (68.4% and 60%, respectively). However, based on their most frequent call locations, it seems like these refugees do not live in Fatih, İstanbul. They either live in Ortahisar, Trabzon (a coastal district) or various districts in İstanbul, such as Kağıthane (a district known until recently for its low living standards but undergoing a rapid transformation in some parts).

### 5.2 Findings for Fatih/İstanbul

İstanbul appeared to be the top refugee location according to Directorate General of Migration Management of Turkey (DGMM 2017) and in DS1. The first thing to be noted as for the distribution of refugee calls in İstanbul where almost 40% of registered refugees live, is the high concentration of such calls on the European side (see Fig. 7). Given the fact that most of the city's population, central business activities, job opportunities, and urban amenities are on the European side, this concentration of refugee calls is in harmony with the basic characteristics of the human geography of the city.

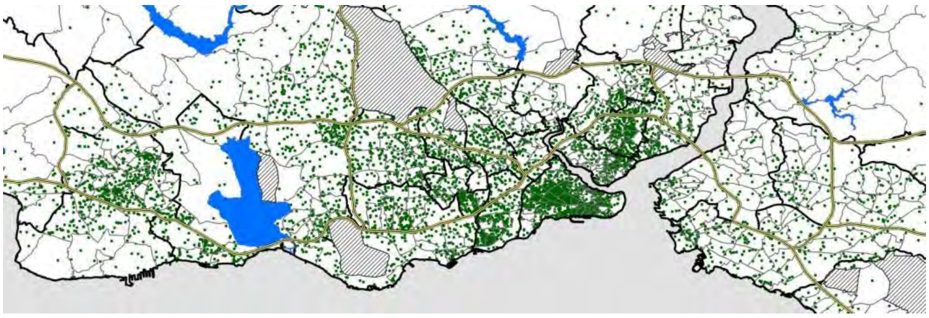

**Fig. 7.** Dot density map of refugee calls (each dot represents 1000 calls) obtained using DS1

On the European side, refugees seem to prefer such districts as (from west to east) Esenyurt, parts of Güngören, Bağcılar, Zeytinburnu, Fatih, Beyoğlu, Şişli, and Beşiktaş districts. Of these districts, Şişli and Beşiktaş are known to be the parts of the modern city center. The neighborhoods preferred by the refugees are close to the main transport routes and provide easy access to some important urban amenities such as city center, business districts, and entertainment facilities. These locations are also mostly low-income areas of the city. In an attempt to prove this claim, we use APRS2017 data set. The map in Fig. 8 shows the distribution of university graduates by neighborhoods as a percentage of total population above 6 years of age together with the dot density map of refugee calls on the European part of the city. We know from previous studies that education level is almost perfectly correlated with income level, meaning that the higher the education level in a given neighborhood, the higher the income of its residents [30]. A comparison of the maps in Fig. 6 makes it clear that refugees tend to conglomerate in areas where the residents are low-income groups.

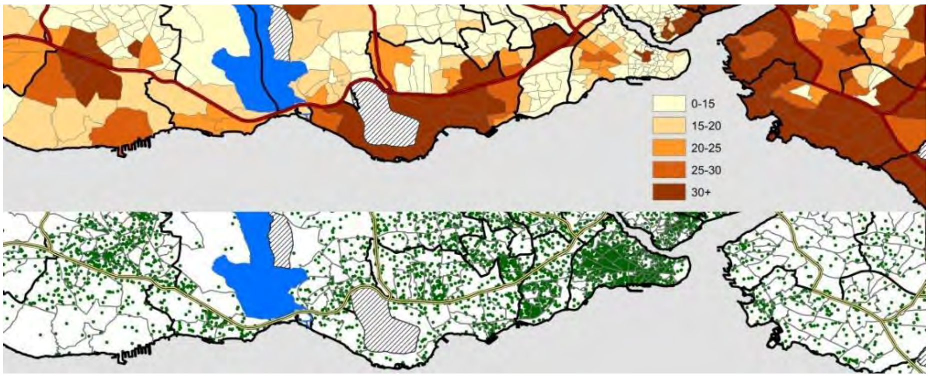


**Fig. 8.** The percentage of university graduates compared to the dot density map

**Table 2.** Monthly rents for each square meter of housing units in selected neighborhoods

| District | Neighborhood | Monthly rent per $m^2$ (TL) | District | Neighborhood | Monthly rent per $m^2$ (TL) |
|---|---|---|---|---|---|
| Selected neighborhoods with high concentration of refugees | | | | | |
| Bağcılar | 100. Yıl | 11.9 | Güngören | Abdurrahman Nafiz Gürman | 17.7 |
| Beyoğlu | Hüseyinağa | 19.2 | Güngören | Mehmet Nesih Özmen | 17.2 |
| Beyoğlu | Katip Çelebi | 32.1 | Güngören | Tozkoparan | 15.4 |
| Beyoğlu | Kocatepe | 24.1 | Güngören | Güven | 14.8 |
| Beyoğlu | Kuloğlu | 30.5 | Zeytinburnu | Beş Telsiz | 27.6 |
| Beyoğlu | Şehit Muhtar | 28.4 | Zeytinburnu | Çırpıcı | 27.3 |
| Esenyurt | Barbaros Hayrettin | 15.0 | Zeytinburnu | Merkezefendi | 26.7 |
| Esenyurt | Cumhuriyet | 14.4 | Zeytinburnu | Seyitnizam | 25.4 |
| Esenyurt | Fatih | 8.1 | Zeytinburnu | Sümer | 16.2 |
| Esenyurt | Mehterçeşme | 9.8 | Zeytinburnu | Nuri Paşa | 18.5 |
| Esenyurt | Mevlana | 12.2 | Fatih | Mevlanakapı | 19.0 |
| Esenyurt | Talatpaşa | 9.4 | Fatih | Şehremini | 20.1 |
| Esenyurt | Yenikent | 10.9 | Başakşehir | Başakşehir | 17.7 |
| Selected neighborhoods with low concentration of refugees | | | | | |
| Beşiktaş | Akatlar | 77.9 | Beşiktaş | Levent | 92.1 |
| Beşiktaş | Arnavutköy | 126.2 | Beşiktaş | Ulus | 87.8 |
| Beşiktaş | Gayrettepe | 58.9 | Şişli | Şevketpaşa | 66.7 |

The zones in İstanbul where refugees are concentrated offer relatively cheap housing opportunities for its residents, where the housing rents are generally low. In order to reach a better understanding of refugee concentration areas in İstanbul, we compiled the monthly rents for 3186 housing units advertised on 6 September 2018 in Hürriyet Emlak (RF), for selected neighborhoods of İstanbul (see Table 2). The results show that the areas preferred by refugees for housing purposes are low-housing cost areas, with rents in some cases 4 to 5 times lower than those in middle and upper class areas characterized by the absence of refugees. We can also add from the existing literature on Syrian refugees in Turkey [31] that even this is a partial and misleading picture of the housing conditions of refugees both in İstanbul and Turkey as a whole, as the figures published in real-estate agents' websites are only for those housing units on the "formal" housing market and refugees do have to live in sub-standard housing units not preferred by the local population.

To complement this study, we also investigated Fatih in street level using the proposed algorithm to understand in which parts of Fatih they usually live and work. The results show that HAP of non-refugees is clustered in the more expensive areas of Fatih district. However, HAP of refugees is clustered relatively in poorer areas.

As for the distribution of refugees in İstanbul, the following conclusions can be drawn:

- Refugees tend to live close to fellow refugees (evidenced by the exceptionally high concentrations of refugee calls in some parts of İstanbul);
- Refugees prefer those areas of the city where the poor and low-income groups live (evidenced by the comparatively low-education levels of refugee concentration zones in İstanbul);
- Refugees tend to concentrate in low property value areas of the city (evidenced by housing rent values of refugee concentration zones in İstanbul);
- Refugees live in inner city areas, close proximity to main transport lines for easy access to job opportunities and urban amenities (evidenced by the general distribution of refugees in İstanbul).

Fatih, known for its rather traditional and religious residents since the Early Republican Period, is the most refugee-saturated district not only in İstanbul but in the whole country. It is believed that many Syrian refugees started their new life in here, thanks to relatively cheap rental prices of basement floor flats, with hopes to hold on to İstanbul where there are more jobs and the city life is dominant.

### 5.3 Findings for Mezitli/Mersin

Mezitli, a district of Mersin, appeared to have received a significant refugee influx according to our chi-square analysis results. Therefore, in this section, we aim to study Mersin in depth. Using our algorithm, we obtained HAP and WAP locations for each MOU who has been seen at least two times according to their calls records in Mersin. We have selected the MOUs after we have clustered them using SOM. The results of the SOM clustering can be seen in Fig. 9.

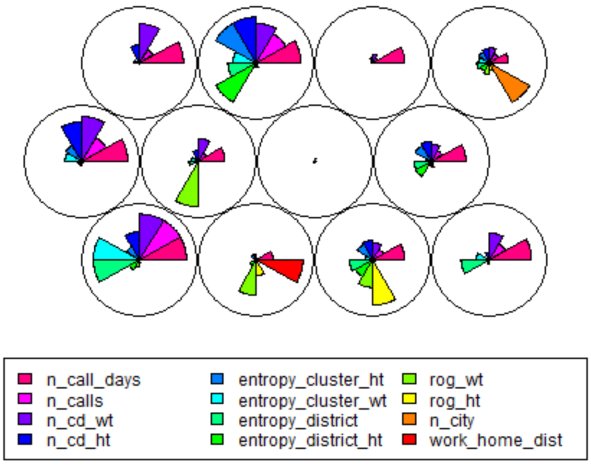


**Fig. 9.** Codes plot of the SOM clustering of refugees in Mersin

While selecting the nodes, we have considered the ones having sufficient number of call days and number of calls, as we have discussed before. As a result, we have selected MOUs contained by the 1st, 5th, and 10th nodes (enumeration starts from the bottom left and ends at the top right).

Using the GSD data set, we mapped each HAP and WAP locations to neighborhoods in Mersin. Furthermore, using the PD data set, we labeled each quarter as low-status (LSN), middle-status (MSN), and high-status neighborhoods (HSN) in seaside districts of Mersin, which are Mezitli, Akdeniz, Yenişehir, and Toroslar, according to their education level and youth population ratio. Among these districts, Mezitli is one of the high-status neighborhoods in Mersin. The locations of the HAP and WAP of the selected MOUs can be seen in Fig. 10.

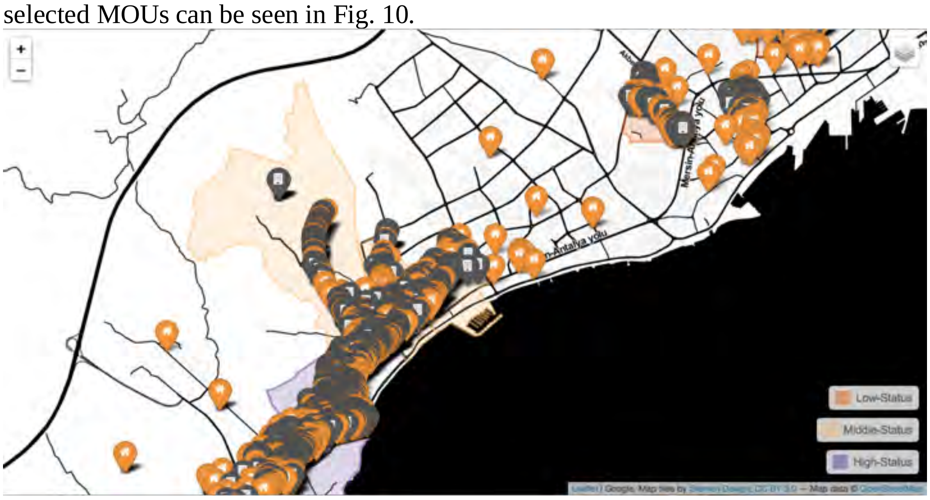


**Fig. 10.** HAP (marked with orange color) and WAP (marked with gray color) locations of selected MOUs in LSN (marked with light orange), MSN (marked with peach color), and HSN (marked with purple)

Considering all 31 features, we investigated whether there is a statistically significant difference between the neighborhoods using a pairwise Mann-Whitney U test, as it is a non-parametric test that can be used for non-gaussian distributions. We have presented the number of MOUs in each neighborhood (N), followed by the mean and median values of each feature per neighborhood in Table 3.

The results presented in Table 3 only shows the features that have been found as significant in comparisons, which are presented in Table 4, with p-value $< 0.05$. Additionally, when Bonferroni adjustment is applied, significance level drops to 1.66e-2 since we have made 3 different comparisons for each feature. Table 4 shows the comparison results, which were found significant. We make 2-sided comparisons with Mann-Whitney U Test and p-values implies the 2-tailed exact significance.

As a result, the test indicated that the distance between HAP and WAP locations is the highest for HSN indicating that MOUs living in HSN probably work in farther neighborhoods. Additionally, the entropy calculated based on different district visits (Entropy_district) for LSN is the lowest, which is also confirmed by the entropy based on the BTS visits (Entropy_bts), and for MSN it is the highest among others. Also, MOUs living in HSN appear to be more regular and MOUs living in the LSN appear to be more irregular than the others in their home-time periods based on their entropy based on BTS visits (Entropy_bts_home_time). One feature that can shed light on the introversion of the MOUs is the refugee to non-refugee call ratio (Ref_nonref_ratio) and in our tests, we have seen that MOUs in LSN have the lowest ratio whereas MOUs in HSN have the highest ratio, showing that refugees in LSN have the least interaction with the non-refugees and refugees in HSN have the most interaction with the non-refugees. Lastly, we have seen that the total movement (Total_movement) and RoG is the lowest for LSN indicating that the total movement and RoG increase as the level of education and youth population level increases.

**Table 3.** Descriptive statistics of LSN, MSN, and HSN for each feature presenting the number of MOUs (n), mean (μ), standard deviation ($\sigma$), min, max, and percentiles of the distributions.

| | | **n** | **μ** | **$\sigma$** | **Min** | **Max** | **Percentiles** | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **25th** | **50th (Median)** | **75th** |
| Work_home_dist | LSN | 176 | 0.57 | 1.6 | 0 | 14.20 | 0.09 | 0.15 | 0.35 |
| | MSN | 342 | 1.46 | 4.0 | 0 | 34.77 | 0.07 | 0.18 | 0.96 |
| | HSN | 619 | 1.53 | 3.4 | 0 | 26.68 | 0.14 | 0.36 | 1.40 |
| Entropy_district | LSN | 176 | 10.1 | 9.2 | 0 | 56.9 | 3.5 | 7.9 | 13.9 |
| | MSN | 342 | 17.7 | 15.0 | 0 | 71.3 | 5.9 | 13.0 | 26.2 |
| | HSN | 619 | 13.8 | 13.3 | 0 | 100.4 | 4.5 | 9.5 | 19.0 |
| Entropy_bts_home | LSN | 176 | 5.90 | 3.0 | 0.4 | 19.2 | 3.9 | 5.4 | 7.2 |
| | MSN | 342 | 5.20 | 3.2 | 0 | 22.4 | 3.0 | 4.7 | 6.5 |
| | HSN | 619 | 4.60 | 3.0 | 0 | 23.5 | 2.8 | 4.1 | 5.6 |
| Ref_nonref_ratio | LSN | 176 | 0.14 | 0.2 | 0 | 1.54 | 0.00 | 0.04 | 0.16 |
| | MSN | 342 | 0.33 | 0.8 | 0 | 6.75 | 0.02 | 0.09 | 0.28 |
| | HSN | 619 | 0.33 | 0.4 | 0 | 2.80 | 0.05 | 0.18 | 0.42 |
| Rog | LSN | 176 | 5.70 | 5.1 | 0.84 | 42.74 | 2.72 | 4.32 | 6.77 |
| | MSN | 342 | 6.30 | 4.6 | 0.41 | 36.52 | 3.54 | 5.04 | 7.80 |
| | HSN | 619 | 6.40 | 3.8 | 0.80 | 47.62 | 4.17 | 5.39 | 7.95 |
| Total_movement | LSN | 176 | 278 | 256 | 26 | 2161 | 118 | 211 | 373 |
| | MSN | 342 | 355 | 310 | 20 | 1604 | 134 | 277 | 462 |
| | HSN | 619 | 344 | 266 | 15 | 1970 | 172 | 271 | 421 |
| Entropy_bts | LSN | 176 | 35.3 | 21.8 | 4.7 | 166.4 | 20.2 | 30.6 | 44.7 |
| | MSN | 342 | 41.7 | 24.7 | 6.0 | 135.9 | 23.4 | 36.6 | 55.3 |
| | HSN | 619 | 43.0 | 25.3 | 3.3 | 158.5 | 25.2 | 37.5 | 53.8 |

**Table 4.** Mann-Whitney U Test results for the comparisons of LSN, MSN, and HSN for each feature (Comparisons read as follows; for all comparisons like "LSN vs HSN", the one with the lower median value written on the left hand size, that is, the median of LSN is less than the median of HSN for this example).

| | | Mann-Whitney U Test | |
|---|---|---|---|
| | | **U** | **p** |
| Work_home_dist | LSN vs HSN | 71765 | 6.27e-11 |
| | MSN vs HSN | 126860 | 3.42e-07 |
| Entropy_district | LSN vs MSN | 21408 | 7.28e-08 |
| | LSN vs HSN | 46842 | 2.27e-03 |
| | HSN vs MSN | 38784 | 7.28e-08 |
| Entropy_bts_home | MSN vs LSN | 35412 | 9.85e-04 |
| | LSN vs HSN | 71254 | 4.30e-10 |
| | MSN vs HSN | 92768 | 1.50e-03 |
| Ref_nonref_ratio | LSN vs MSN | 24361 | 3.56e-04 |
| | LSN vs HSN | 32263 | 4.40e-16 |
| | MSN vs HSN | 128090 | 6.58e-08 |
| Rog | LSN vs MSN | 25286 | 2.88e-03 |
| | LSN vs HSN | 41825 | 2.55e-06 |
| Total_movement | LSN vs MSN | 25756 | 7.20e-03 |
| | LSN vs HSN | 44189 | 1.31e-04 |
| Entropy_bts | LSN vs MSN | 25317 | 3.06e-03 |
| | LSN vs HSN | 43739 | 6.54e-05 |

# 6 Summary of Findings (SF)

**SF1. A place in the city—Place as a means of survival:** Finding an adequate place to live in is a key to success in the urban jungle. A place where they can interact with their peers, have easy access to urban facilities such as work and leisure is vital for their survival in the city. Using the established networks of solidarity among the refugee community, newcomers maximize their access to flows of information, which may from time to time play a life-saving role. Almost as a rule, the newcomers to a city tend to concentrate in particular parts of urban areas [32].

We have discovered in the second layer of our findings that Syrian refugees in Turkish cities tend to live in areas where [a] they can have maximum interaction with

Syrian community, which is crucial in access to information flows; [b] the rents are lower; [c] they can have easy access to urban facilities, namely in inner city areas.

**SF2. Networks—Sine qua non for survival:** All of the above—i.e. joining the complex web of relations characterizing seasonal agricultural work and finding the best place to live in a city—could not have been achieved without networks. Some of the hotspots we have discovered seem to function like hubs in a network. For example, Fatih district definitely plays a high-level hub role, as we have detected a non-negligible number of refugees (whose speculated homes are outside Fatih) coming to Fatih and also heading to Çaykara where they work for tea harvest.

**SF3. Better off Syrians—Internal divisions:** We know from other studies that an important portion of Syrians came to Turkey without a chance to turn their savings in their homeland into cash [31]. There are also some cases, though not many, where some Syrians arrive in Turkey with some accumulated wealth. This we believe creates a division within the Syrian refugee population, with better off Syrians living in comparatively higher status neighborhoods and having different mobility patterns, compared to the rest of the refugees. Mersin appeared to be a very interesting city where refugees from different socioeconomic levels are living. Our analyses showed that middle and high-status refugees use a very large space, travel long distances, have regularity in their mobility (regular work home patterns), whereas low-status refugees appear to be trapped in a small neighborhood meaning that traveling not very distant districts, having high irregularity in their mobility.

**SF4. Seasonal work—Geographical mobility as a survival strategy:** One of the most important findings of our study is the one that shows the movement of Syrian refugees among various districts of Turkey. The evidence to this fact is the unusually high number of calls made by refugees in certain parts of the country in certain months of the year. What seems to be an anomaly of the data set is, in fact, a practice that is very common in Turkish agriculture: the prevalence of seasonal work in agriculture. We know from a recent study that Syrian refugees have replaced seasonal Kurdish workers in the last few years especially in cotton and hazelnut harvests, two most common products utilizing seasonal labor [33]. They have also replaced Georgian migrants in tea harvest which, compared to other products, requires highly skilled labor.

The unusual peaks in the number of outgoing refugee calls in some regions in certain months attest to the fact that a significant portion of Syrian refugees are on the move in search of an adequate job and are in harmony with the harvest seasons of agricultural products. Their enhanced geographical mobility as a survival strategy shows the capability of Syrian refugees to adapt to new conditions. Furthermore, the fact that they have replaced (or are in the process of replacing) Georgian migrants in tea harvest is a testimony to their ability to alter and penetrate the existing networks.

## 7 Policy Recommendations

In this section, after a thorough analysis of D4R data, available literature reviews and investigation of available social protection mechanisms, we list and group the main vulnerabilities of Syrian refugees. In addition, we propose short-term and long-term solutions and policy recommendations. We investigate the problems followed by proposed recommendations under the subheadings of 1) Education & Employment 2) Safety & Security 3) Work, 4) Healthcare, and 5) Integration.

**Education & Employment:** As our findings demonstrate, Syrian refugees use geographical mobility as a survival strategy (SF4) which means spending short amounts of time in certain regions. Policies are required for both children and agricultural laborers. For children, the main problem is the access and participation to education. Harvesting times often coincide with the school period for some districts. Although we are not sure whether they are moving with their families (demographics of MOUs and their call networks are not provided in detail in the data sets), it is highly probable that there may be Syrian refugee children who are unable to benefit from education services. Furthermore, having one parent moving all year long is not a healthy environment to raise children. Studies also show that language barriers, insufficient salary, invisible costs of education (i.e. costs for course materials, transportation, and lunch costs), and distance from urban centers may result in education problems. As a remedy, introductory programs can be expanded in order to take language and cultural differences into account. Social assistance policies should specifically focus on the involvement of children in education. In line with our finding (SF4), mobile school programs such as TÜBİTAK 4007's science buses and science fairs that bring science to rural areas, can visit farm areas to help sustain continuity in children's education. As for adults, who are mostly seasonal workers in agriculture, problems are many: long working hours, low wages, health and hygiene problems, makeshift tents they have to live in, the lack of basic amenities, and so on. From a wider perspective, seasonal labor cannot be a long-term solution in the sense that it guarantees only (or even falls short of) a minimum level of subsistence. In addition, seasonal work is a strategy that may lead to the failure of coming generations, since it almost requires the movement of young members of the family who cannot attend their school during seasonal work, a fact that may lead to persistent long-term problems. It was also reported that in numerous stages of agriculture, different skills are required but refugees do not have the necessary background skills [10], which can be solved with specialized training programs. Generations may stick in seasonal work due to a lock-in network. As networks are critical (SF2), conditions of the agricultural work network should be improved and new occupational networks should be introduced. Agricultural occupations can be made more structured with better pay. For the latter, new job opportunities aiming to integrate Syrians into the labor force should be created. This may be achieved by offering vocational and entrepreneurship trainings.

**Safety & Security:** The results of our analyses showed that Syrian refugees tend to live close to other fellow refugees (SF2) and they live in poor housing conditions in underdeveloped areas of the cities (SF1). This creates ghetto-like communities and becomes a handicap on overall integration in Turkey. To address these vulnerabilities, "urban transformation" policies should be redesigned taking Syrian refugees into account and integration to the society should be aimed. Housing is a problem for seasonal workers. Recently, a project called METİP (Project for Improvement of the Working and Living Conditions of Seasonal Migratory Agricultural Workers) [34] has been put into effect by the Ministry of Labor and Social Security with the goal of improvement of the current living, shelter, transportation, education, health, security, social relations and social security status of the seasonal migratory agricultural workers who migrate to other cities with their families to be employed as agricultural workers. It is important that the continuity of the METİP project will be ensured in such districts.

**Healthcare:** As mentioned above, most of the Syrians are employed in seasonal work (SF4). Seasonal workers and women are the most vulnerable group in terms of accessing healthcare due to remoteness and language barriers. Seasonal changes in welfare, difficulties in access to childcare services, obligation to work during pregnancy and concern about nutrition quality are the main problems of women. Hygiene is a problem for everyone. Mobile health units that travel farms can be established as a remedy to provide the necessary healthcare services.

**Integration:** All of our findings are in fact intertwined and part of the bigger challenge that is going to take a long time for everyone involved to deal with – **integration**. For instance, just with being employed in seasonal work will not be enough for Syrians to integrate with the Turkish society. It also brings out the competition on limited resources with Turkish seasonal workers. Another example may be that if Syrian children do not receive the necessary education, their destiny will be low-paid jobs and no integration. Ghettos would be another obstacle for integration. Moreover, there are even fractions within the Syrian communities itself (SF3). Thus, even with the limited data provided to researchers in D4R, it is obvious that addressing the integration is the real challenge ahead of Turkey in the long run. In our opinion, in addition to NGOs and academics, government bodies and international organizations should work together to establish a big coalition to come up with an integration strategy and obtain a buy-in from the public to tackle the integration challenge, because the last challenge will require not only huge amounts of resources and wisdom but also empathy, compassion, and tolerance.

## 8 Acknowledgement

We would like to thank Türk Telekomünikasyon A.Ş. for the one-year anonymized mobile communication data they provided within the D4R Challenge.

## References

1. United Nations High Commissioner for Refugees (UNHCR): Global trends 2017: Forced displacement in 2017. http://www.unhcr.org/5943e8a34.pdf (2017). Accessed 31 August 2018
2. ORSAM (Ortadoğu Stratejik Araştırmalar Merkezi): Suriye'ye komşu ülkelerde Suriyeli mültecilerin durumu: Bulgular, sonuçlar ve öneriler. http://www.madde14.org/images/e/e5/OrsamSuriyeKomsu2014.pdf (2014). Accessed 31 August 2018
3. İçduygu, A.: Syrian refugees in Turkey: The long road ahead. Washington DC: Migration Policy Institute (2015)
4. Kaya, A.: Istanbul as a space of cultural affinity for Syrian refugees: "Istanbul is safe despite everything!" Southeastern Europe 41(3):333–58 (2017)
5. World Bank: Turkey's response to the Syrian refugee crisis and the road ahead. Washington DC: World Bank (2015)

6. Stock, I., Aslan, M., Paul, J., Volmer, V., Faist, T.: Beyond humanitarianism: Addressing the urban, self-settled refugees in Turkey. Bielefeld: COMCAD (2016)
7. Cagaptay, S., Menekse, B.: The impact of Syria's refugees on Southern Turkey. Policy Focus 130. Washington, DC: The Washington Institute For Near East Policy. http://www.washingtoninstitute.org/uploads/Documents/pubs/PolicyFocus130_Cagaptay_Revised3s.pdf (2014). Accessed 31 August 2018
8. Asik, G.A.: Turkey badly needs a long-term plan for Syrian refugees. Harvard Business Review. https://hbr.org/2017/04/turkey-badly-needs-a-long-term-plan-for-syrian-refugees (2017). Accessed 14 September 2018
9. Balkan, B., Tumen, S.: Immigration and prices: Quasi-experimental evidence from Syrian refugees in Turkey. Journal of Population Economics 29:3:657–686 (2016)
10. Dedeoğlu, S.: Yoksulluk nöbetinden yoksulların rekabetine: Türkiye'de mevsimlik tarımsal üretimde yabancı göçmen işçiler mevcut durum raporu. http://www.ka.org.tr/TumYayinlar (2016). Accessed 12 September 2018
11. Isikkaya, A.D.: Housing policies in Turkey: Evolution of TOKI (Governmental Mass Housing Administration) as an urban design tool. Journal of Civil Engineering and Architecture 10:316–326 (2016)
12. Chandy, R., Hassan, M., Mukherji, P.: Big data for good: Insights from emerging markets. Journal of Product Innovation Management 34(5):703–713 (2017)
13. Alemanno, A.: Big data for good: Unlocking privately-held data to the benefit of the many. European Journal of Risk Regulation 9(2):183–191 (2018)
14. Türk Telekom: Data for refugees. http://d4r.turktelekom.com.tr/ (2018). Accessed: 31 August 2018
15. Salah, A.A., Pentland, A., Lepri, B., Letouzé, E., Vinck, P., de Montjoye, Y.A., Dong, X., Dağdelen, Ö.: Data for Refugees: The D4R Challenge on mobility of Syrian refugees in Turkey. arXiv preprint arXiv:1807.00523 (2018)
16. Hürriyet Emlak: Online real estate ads. https://www.hurriyetemlak.com (2018). Accessed 8 August 2018
17. Ahas, R., Silm, S., Järv, O., Saluveer, E., Tirui M.: Using mobile positioning data to model locations meaningful to users of mobile phones. Journal of Urban Technology 17(1):3–27 (2010)
18. Nurmi, P., Koolwaaij, J.: Identifying meaningful locations. In: Mobile and Ubiquitous Systems: Networking & Services, 2006 Third Annual International Conference, pp. 1–8. IEEE, San Jose (2006)
19. Jiang, S., Ferreira, J., González, M.C.: Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore. IEEE Transactions on Big Data 3(2):208–219 (2017)
20. Graells-Garrido, E., Peredo, O., García, J.: Sensing urban patterns with antenna mappings: the case of Santiago, Chile. Sensors 16(7):1098 (2016)
21. Furno, A., Fiore, M., Stanica, R., Ziemlicki, C., Smoreda, Z.: A tale of ten cities: Characterizing signatures of mobile traffic in urban areas. IEEE Transactions on Mobile Computing 16(10):2682–2696 (2017)
22. Directorate General of Migration Management of Turkey (DGMM): Hangi ilde ne kadar Suriyeli var? İşte il il liste. http://www.bik.gov.tr/hangi-ilde-ne-kadar-suriyeli-var-iste-il-il-liste/ (2017). Accessed 31 August 2018
23. Zhao, Z., Shaw, S.L., Xu, Y., Lu, F., Chen, J., Yin, L.: Understanding the bias of call detail records in human mobility research. International Journal of Geographical Information Science 30(9):1738–1762 (2016)
24. Yang, P., Zhu, T., Wan, X., Wang, X.: Identifying significant places using multi-day call detail records. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 360-366. IEEE, Limassol (2014)

25. Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., Varshavsky, A.: Identifying important places in people's lives from cellular network data. In: International Conference on Pervasive Computing, pp. 133–151. Springer, Berlin, Heidelberg (2011)
26. Hartigan, J.A.: Clustering algorithms. New York, NY, USA (1975)
27. Soto, V., Frias-Martinez, V., Virseda, J., Frias-Martinez, E.: Prediction of socioeconomic levels using cell phone records. In: International Conference on User Modeling, Adaptation, and Personalization, pp. 377–388. Springer, Berlin, Heidelberg (2011).
28. Kondo, Y., Salibian-Barrera, M., Zamar, R.: RSKC: An R package for a robust and sparse k-means clustering algorithm. Journal of Statistical Software 72(5):1–26 (2016)
29. Kohonen, T.: The self-organizing map. Neurocomputing 21(1–3):1–6 (1998)
30. Işık, O., Ataç, E.: Yoksulluğa dair: Bildiklerimiz, az bildiklerimiz, bilmediklerimiz. Birikim 269(268):66–86 (2011)
31. Eraydın, G.: Migration, settlement and daily life patterns of Syrian urban refugees through time geography: A case of Önder neighborhood. Unpublished PhD Thesis, Middle East Technical University (2017)
32. National Academies of Sciences, Engineering, and Medicine (NASEM): The integration of immigrants into American society. Washington, DC: The National Academies Press (2015). doi: https://doi.org/10.17226/21746
33. Kalkınma Atölyesi: Mevsimlik gezici tarım işçiliği izleme: Mevcut durum haritası (2012–2013) http://www.ka.org.tr/TumYayinlar (2013). Accessed 12 September 2018
34. Prime Minister's Office: Memorandum Circular n. 2017/6. Official Gazette of Turkey, 30043 (2017)

# 9 Appendix: Finding HAP and WAP

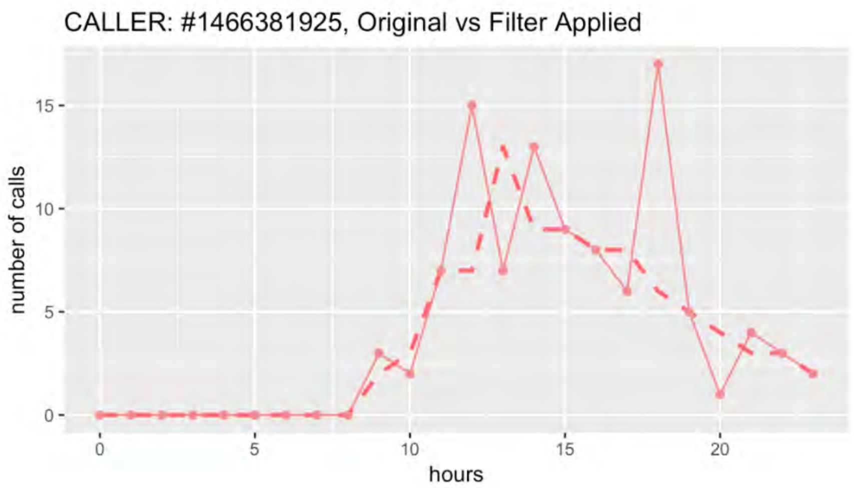


**Fig. a1.** The number of calls per hour for a MOU, original is shown with straight line and the median filter applied is shown with dashed line.

In order to extract the important places of individuals, we study their fine grained mobility using DS2. More specifically, we try to find WAP and HAP for each MOU. To be able to find those points, first we aggregate the hourly calls counts for each MOU and we apply a median filter with the bandwidth of three to hourly calls signal to smooth unexpected low and high values of call counts out. As depicted in Fig. a1. after the median filter is applied, the hourly call counts signal is smoothed.

Then, we sort the hours according to the number of calls made during that hour in ascending order. We select the hours in the first quartile and three hours before that as the Home-Time Period (HTP) assuming that this period is spent during the most probable HAP. After that, we find the Work-Time Period (WTP) by first adding four hours, which is allocated for preparation and commute, to the end of HTP and selecting the next six hours, which is assumed to be a safe period for work activities for most of the people, as the WTP. In Fig. a2, HTP and WTP have been marked on the filtered data.

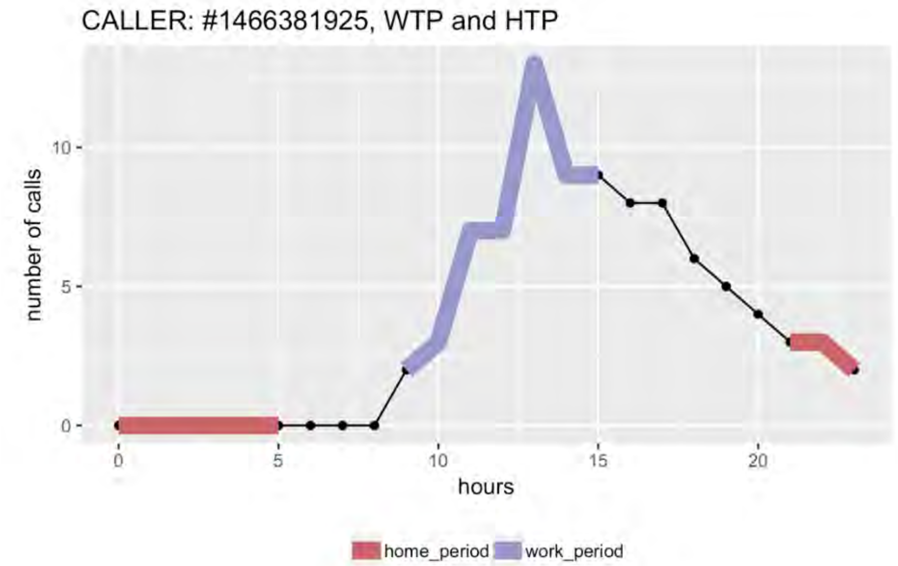


**Fig. a2.** The graph showing the Work-Time Period (WTP) and Home-Time Period (HTP) of a MOU. HTP included the hours in the first quartile (between 00:00 and 5:00) and three hours before that (between 21:00 and 23:59). WTP is found by first adding four hours to the end of HTP (9:00) and selecting the next six hours (end point is 15:00).

After finding the HTP and WTP, we derive the BTS used in those periods to find HAP and WAP. First, we find the HAP by sorting BTS by the number of unique days they used and we select the one with the highest number of unique day usage. In the next step, we cluster the BTS using Hartigan's leader clustering algorithm. The advantage of the Hartigan's leader algorithm, unlike clustering algorithms like K-means, we do not need to set the number of clusters at the beginning. We only need to set the radius to cluster the BTS based on their proximity. We set the radius as 1 km. After clustering the BTS, we select the cluster, in which the BTS with the most call days resides and then set the centroid of the cluster as HAP. In order to find the WAP, on the list of possible BTS for WAP, we apply the same steps by applying the Hartigan's leader algorithm and then selecting the centroid of the cluster in which the BTS has the most number of call days in weekday usage. Aside from the hour differences between the

HTP and WTP, to specify the WTP we only look at the calls made or received on the weekdays within the designated work time hours. In Fig. a3, we present the possible WAP, which is represented by the green circle, and HAP, which is represented by the red circle, locations for a MOU living around Siteler, Ankara.


**Fig. a3.** WAP and HAP for a person living around Siteler, Ankara

## 10 Appendix: SOM Clustering

In SOM clustering, we trained the neural network until the average distance between the data points and the node centroids converged to a relatively small distance and then, we get the codes plots of the clusters, an example can be seen in Fig. a4. If we are to analyze some of the clusters, for instance, in Node 1 (enumeration starts from the bottom left and ends at the top right), we observe that MOUs in this node have higher entropy while having close to zero radius of gyration and visited very few numbers of different cities. As their HAP and WAP locations are very close, their daily commute is expected to be small. Even though they visited different numbers of BTS and districts, their travel distance is relatively small and they wondered in a very small area by visiting a lot of places, which are very close to each other. On the other hand, for Node 6, we see that MOUs in this node have low entropy and high radius of gyration indicating that these MOUs are making longer commutes, especially in the work-time periods, while visiting a fewer number of different places. Additionally, larger commutes were also confirmed by the larger distance between HAP and WAP locations. Lastly, we observe MOUs in Node 10 having only higher values of radius of gyration in the home-time period while other features like entropy, radius of gyration in work-time period, number of different cities visited, and WAP-HAP distance are small. Hence, we can deduce that these MOUs are making larger commutes in their home-time period but they are not visiting different places in terms of BTS, districts, or cities and their HAP and WAP locations are very close to each other.

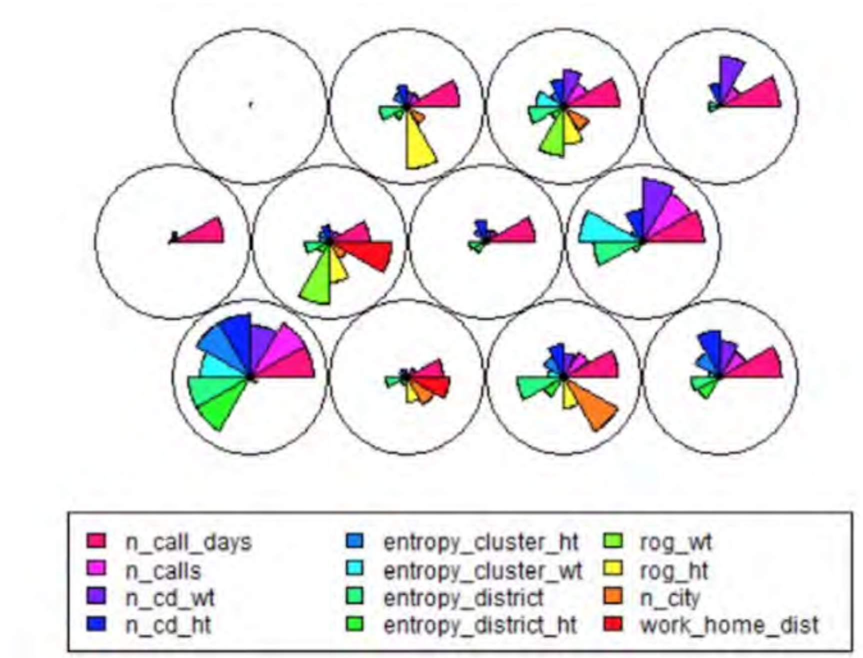


**Fig. a4**. Codes plot of a SOM clustering

Response to Reviewers

We would like to thank reviewers for their constructive feedback and suggestions. We have worked on improving the paper in the light of the provided feedback. Please find below the responses to the reviewer comments (red colored text).

The paper is a well written, comprehensive look at refugee movements, and makes a number of in-depth studies of selected districts. Additional datasets were used, and experts were consulted to interpret seasonal changes in refugee movements. Policy recommendations are given for all themes of the challenge. Interactive tools are made available.

Particular comments in page order:

Page 2: 'In particular, several refugees found jobs as seasonal workers or in small industrial areas ("sanayi siteleri").': This statement needs a reference.

- We added the reference.

Page 3: 'Until the content is approved': This part needs to be rephrased in the final version.

- We removed that part and made everything available online without requiring a password.

Page 5: 'local governorship has barred unauthorized refugees': Was this legislation in place during 2017 as well?

- The official figures in 2017 were low. We came across several news regarding this issue in 2017. Besides, an expert working in Refugees and Migrants Solidarity Association we consulted mentioned about this ban informally. However, we could not reach a related authority or an official document to confirm it. As a result, we removed that part from the paper.

Page 7: Figure 2 can have a legend or a caption addition to indicate what the sizes of the vertices numerically correspond to. Is the population proportional to the area of the circle? Providing 3-4 values, including min and max, could be sufficient. Also useful would be to indicate the sizes of the edges, here the thickness indicates volume, but again, numeric values for min and max would be useful.

- We revised the figure and added related information in the paper.

Page 7: 'missingness': please replace with 'omissions'

- We revised it in the text.

Page 8: 'To be specific, the harvests of Çaykara-Trabzon in July and August are hazelnut and tea': A similar pattern should be observed in Ordu for the hazelnut harvest.

- There is a similar pattern in Ordu but not apparent as in Çaykara. We provided additional results of some districts in the figure but those were chosen according to their ranks in the results (most significant ones). As a result, we did not put Ordu in the Fig.5.

Page 9: 'Due to limited space, we could only have provided the results of six districts in Fig.3.': Since the page restriction is only imposed for fairness' sake w.r.t. awards, the final published report can include more results.

- We provided the results of two more districts, and further details about the proposed algorithm and pre-processing steps in the text.

Page 11: 'Our method makes an assumption that WAP and HAP cannot be the same place.': Why? This seems to be an arbitrary constraint, and too restrictive. 'The detailed steps of the algorithm are provided on our web site': For completeness sake, such additional information should be made available in form of an Appendix, within the paper.

- We added an appendix section where we provided the details of the algorithm, and we removed the statement of "The detailed steps of the algorithm are provided on our web site" from the text. We also revised the algorithm without the aforementioned assumption as to include the locations, which could be marked as both WAP and HAP. After the revision, all the tests were repeated but no significant change was observed in the results other than a few points, where they were updated in the text accordingly. The reason why we had the assumption at the initial development of the algorithm is to ensure whether we can differentiate between the working and living locations correctly. But this is longer needed since SOM clustering helps us understand the nodes, in which WAP and HAP can be identified more safely.

Page 13: 'Kağıthane (a district that is known for its low living standards).': This may be no longer true. There are many luxurious building projects completed in Kağıthane in recent years, and the area is being gentrified: https://www.projepedia.com/konut-projeleri/istanbul/kagithane

- We revised it in the text as follows: "a district known until recently for its low living standards but undergoing a rapid transformation in some parts"

Page 15-16: 'Considering all 31 features...': This paragraph is very difficult to read, as it is full of numbers without interpretation. It is better to put these numbers into a table, and re-write the paragraph in a more intuitive, clear way to indicate the salient points, and relevant patterns.

- We provided the descriptive statistics and the results of the statistics test as two separate tables. We rewrote the paragraph.

Page 17-18: The paper does a very nice job of describing the status of refugees across the country, and illustrating in detail how analysis can be conducted in specific regions. Nonetheless, the policy recommendations do not follow directly from these analyses, and are not empirically supported in a strong way. For instance, is it possible to estimate the percentage of Syrian workforce that is in motion due to seasonal work, and provide a specific map of displacement? Can this be used to inform education related policy decisions?

- In the summary of findings section, we referenced to the related findings of our analyses. Our analyses demonstrate that it is possible to predict population in motion due to seasonal work, which can later be used to inform education related policy decisions if the data is sampled appropriately. However, as the provided data did not have such characteristics[1], we refrain to make generalizations. On the other hand, we believe that this does not degrade our work as we achieved to develop a methodological approach for information extraction from the given data. We are also happy to redo all analyses when an appropriate sample set is prepared.

  [1]The only way to show the location changes of individuals over a year is by using the third dataset. However, due to sampling issues in the dataset, we cannot obtain a correct map for the location changes (we have several missing values on different time periods of individuals and there are no or very few people identified in some regions such as in Trabzon.

Page 18: Please provide a reference for METIP.

- We added a reference as suggested.

Minor comment: The paper is not correctly formatted, please check Springer LNCS guidelines. For instance, references should be numbered inside the text.

- We reformatted the paper.